

Sentiment Analysis of China-Related News in *The Star Online* Newspaper

Hong Wu^a

nlilywu@gmail.com

Guangxi University, China

&

Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

Kesumawati A. Bakar^b

kesuma@ukm.edu.my

Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

Azhar Jaludin

azharj@ukm.edu.my

Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

Norsimah Mat Awal

norsimah@ukm.edu.my

Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

ABSTRACT

As China and Malaysia approach their 47th year of diplomatic relationships, cooperation and trust between the two countries have deepened in aspects ranging from politics to economy. Despite this mutual reliance, the relationship between Malaysia and China is not without its conflicts and these conflicts are often manifested in media reports. How China is presented in Malaysia news is a field that has been scarcely explored. As part of the research on media sentiment towards China, this research investigates the general sentiment of China-related news in Malaysian media through sentiment analysis of some selected news coverages. Selecting China-related news in *The Star Online* from 2012 to 2021 as the data for investigation, -the Excel Add-in tool Azure Machine Learning was used to generate polarity of these news reports automatically and corpus tool Wordsmith was used for the analysis of news discourse. A total of 137,475 pieces of news have been collected as the research sample. The finding reveals that: 1) despite the large proportion of news with negative sentiment in China-related news in *The Star Online*, the monthly trend of sentiment shows a slight increase of positiveness over time; 2) an investigation into the keyword lists of three months with highest proportion of negativeness and collocates of the top keywords, however, shows that negative sentiment of the news may be due to a global conflict at that particular time and does not necessarily indicate negative sentiment towards China. A combination of sentiment analysis and corpus approach on the study of China-related news in Malaysian media enriches the study of news discourse from the perspective of corpus linguistics.

Keywords: sentiment analysis; polarity; China; Malaysian English news; *The Star Online*

^a Main author

^b Corresponding author

INTRODUCTION

People's opinion matters. To find out what is in other people's mind has always been an important aspect of our information-gathering behavior (Pang & Lee 2008) because "our beliefs and perceptions of reality, and the choices we make, are, to a considerable degree, conditioned upon how others see and evaluate the world" (Liu, 2012, p. 5). Media play an important role in expressing ideas and influencing people's opinions. They are powerful in that "the influence they exert both on our governments and major institutions as well as their ability to shape our ideas and behaviours" (Bednarek & Caple, 2012, p. 6). News is one popular form of the media content. According to Fowler (1991), news is a product instead of a natural phenomenon emerging straight from 'reality'. "From a broader perspective, it[news] reflects, and in return shapes, the prevailing values of a society in a particular historical context." (ibid., p. 222) The study of news in a certain context may help people understand what kind of ideas or values is constructed by news and how it is constructed. Sentiment analysis, which is also called opinion mining, is "the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" (Liu, 2012, p. 7). Sentiment analysis of news is to uncover "opinions, sentiments or evaluation, etc." reflected in news articles. The aim of this paper is to investigate the general sentiment of China-related news in Malaysian media through sentiment analysis.

China, as a country who owns one-fifth of the world population, has made significant economic achievements in the past 40 years since the reform and opening-up policy. As of 2020, China ranked the second in the world after the United States with a GDP of 14,722,731 million US dollars (World Bank 2020). China's communication and cooperation with other parts of the world has also increased significantly. China has been the largest trading partner of ASEAN for 10 consecutive years, and in 2020, ASEAN jumped to be China's largest trading partner, taking the place of the European Union (Global Times, 2021). Malaysia, as one of the ASEAN countries, also has close connection with China. Malaysia and China established diplomatic relations on May 31st, 1974. China remains Malaysia's largest trading partner for 12 consecutive years (Xinhua, 2021). Although Malaysia is cooperating closely with China in many fields such as trade, education, tourism, etc., during the on-going East Coast Rail Link (ECRL) project, the biggest economic and trade project between China and Malaysia, there have been contested voices from various entities of interest such as local contractors and residents along the new Silk and Spice route in Malaysian news, expressing mixed views towards the project, as well as towards China. Since news plays an important role in shaping people's opinion, the mixed view in Malaysian news makes the sentiment analysis involving China in the Malaysian context an interesting topic. Therefore, Malaysia is taken as the context for the study to investigate sentiment of China-related news in Malaysian media between 2012 and 2021, as the past 10 years has witnessed close cooperation between Malaysia and China. China-related news in *The Star Online*, a Malaysian-based news website, is collected as data for investigation.

LITERATURE REVIEW

Previous work from several fields has contributed to the multidisciplinary nature of this study, namely, news discourse, corpus linguistics and sentiment analysis. In the following section, news discourse, especially studies on China in the media, corpus linguistics and its application in social linguistics and sentiment analysis on news will be illustrated briefly.

NEWS DISCOURSE & CHINA IN THE MEDIA

Approaches to media discourse analysis include “linguistic and sociolinguistic approach, conversation analysis, semiotic analysis, critical linguistics and social semiotic, social-cognitive analysis and cultural-generic analysis” (Fairclough, 1995, p. 20). Studies on news discourse, sometimes as a sub-category of media discourse (van Dijk, 1988), have been carried out based on the approaches mentioned by Fairclough above, separately or combined, such as political texts on the construction of country image (Brookes, 1995), non-protesters’ view from protest scripts (Stamou, 2001), and power relations construed in news (Briones, 2017), to name just a few. In this section, previous research on news discourse and China in the news representation of other countries will be illustrated.

When it comes to the study of news discourse, a majority of such studies are carried out based on news in western countries. For example, Maryam Jahedi (2012) used CDA to investigate the discursive features of news headlines and leads of *The New York Times* of events related to Iran from 1979 to 2009 to find out how the image of Iran was constructed. The researcher found that the ideological representations of Iran were realized through transitivity, thematization and lexicalization, and that the stereotypical image of Iran was the “negative Other”. In addition, Cheng and Lam (2013) compared the changes of perceptions of Hong Kong by western news reports corpora on two 3-year bases. Using corpus-driven critical discourse analysis, they investigated two-word congrams (two words occurring together) and key semantic categories and found dynamic ideological shifts of western news reports towards Hong Kong and China. They found that for western media, focus had shifted to political related issues instead of business environment, and for Chinese media, there was an increase in quoting negative views about the political situation in Hong Kong. Another example is Liu and Jiang (2019), who conducted a synchronic and diachronic corpus-assisted discourse analysis of the press representations of post-colonial Hong Kong through a self-built corpus of *The Times* news reports spanning over 20 years, starting from the handover of Hong Kong to China on 1 July 1997. Using the nodeword “Hong Kong”, the study found that *The Times* still regards Hong Kong, the former colony of Britain — as Britain’s responsibility. The findings from these diachronic studies, with the longest being a 30-year span, can well reflect the shift of representations or focus of news media. There are only a few studies on news discourse which were based on non-western media (Ng, 2008; Asad et al., 2019).

There are even fewer studies on sentiment towards China in the news. The researchers searched WOS and Google Scholar, using “news discourse”, “sentiment” and “other country” as key words in the time frame from 2011 to present and was able to find a number of results focusing on sentiment of the countries/regions being investigated towards other countries (Chan, 2012; Maryam Jahedi, 2012; Sohoni & Mendez, 2012; Alasuutari, 2013; Cheng & Lam, 2013; Bickes, 2014; Kim, 2014; Armony & Velásquez, 2015; Duff, 2015; Bui, 2017; Ospina Estupinan, 2017; Ruhrberg et al., 2018; Liu & Jiang, 2019; Idris, 2019; Haw, 2020) Very few are based on news in countries other than western powers, such as Latin America (Armony & Velásquez, 2015; Ospina Estupinan, 2017), VietNam (Bui, 2017) and Nigeria (Anunne & Yan, 2018). One article on Malaysian media is by Wang and Reagan (2020), who carried out research on media sentiment towards Forest City, a Chinese investment project in Malaysia, by comparing sentiments of news from mainstream and independent media in Malaysia, and found that the sentiment of mainstream media towards the project was more positive, which is consistent with local government’s attitudes.

The scenario in publications in the Chinese language is almost the same. The researchers searched the core journal database in CNKI, the electronic platform of Chinese knowledge-based

information resources, using “China-related news” and “image” as key words (there is no result for “sentiment/attitude”, so the word “image” is used instead), and found 29 articles from 2011 till now. 15 of which are based on western news media (Zhang, 2018; Zhao & Lu, 2021; Lin, 2022; to name just a few). And the rest is sparsely distributed to some other countries, such as ASEAN countries as a whole (Luo, 2019), South Korea (Yin, 2020), etc. There is only one article about Malaysian news media, the *Nanyang Siang Pau* in Chinese language. Wu (2020) collected China-related news on *Nanyang Siang Pau* between January 1 to June 30, 2019, analysed the total number, themes and news sources of these reports and found that: 1) 50% of these reports were neutral while 15% of them were negative, and 2) 19% of these reports were citing from western news media.

In the studies of news discourse, there are only a few to investigate sentiment of one country towards another country, even fewer on news from other than western media using corpus linguistics. Besides, the western press rarely gives news with positive sentiment coverage to communists and others on the ‘far left’ (Demers, 2002, as cited in Leung & Huang, 2007, p. 676). In view of China’s bilateral efforts and cooperation with other nations, sentiments of news in these countries towards China other than from the western powers, should also be paid attention to. Therefore, a study of China-related news in Malaysian media helps to enrich the studies on news discourse, especially on how China is represented in media of the other countries.

CORPUS LINGUISTICS & SOCIAL LINGUISTICS

A corpus is a large collection of naturally occurring texts (Egbert & Baker, 2020, p. 4), and Corpus Linguistics (CL) is “the study of language based on examples of real life language use” (McEnery & Wilson, 1996, p. 1). The corpus approach has been applied independently or combined with other theories in the studies of almost all linguistic fields, such as lexicography (Sinclair 1991; Hanks 2004), syntactics (Gries, 2005; Lu, 2011), social linguistics (Baker, 2010), language teaching (O’keeffe et al., 2007), etc. In this section, previous studies on how corpus linguistics is combined with research in social linguistics are reviewed.

Social linguistics is the study of “the relationship between language and society . . . the various functions of language in society” (Wardhaugh, 2010, p. 1). Many scholars, such as Romaine (2001) and McEnery et al. (2006), have applied corpus approach in the study of social linguistics. A lot of studies on social linguistics use news articles as the data for investigation. For example, Gabrilatos and Baker (2008) explored the discursive construction of refugees and asylum seekers in the UK press through a corpus-based approach. They collected 175,139 articles from daily tabloids and broadsheets from January 1996 to October 2005. Through keyness, collocates and concordance analysis as well as critical discourse analysis, they discovered that a moral panic around refugees and asylum seekers had become the dominant discourse in the UK press. Cavasso and Taboada (2021) also carried out a corpus analysis of online news comments. With a corpus of comments posted in response to a variety of opinion pieces on the website of the Canadian English-language newspaper *The Globe and Mail* between 2012 and 2016, they used the Appraisal framework for discourse analysis and found that highly opinionated language is usually categorized as Judgement and Appreciation rather than as Affect. Nor Fariza Mohd Nor et al. (2021) used a corpus-driven approach to explore the adjective collocates for “mental health” in *New Straits Times* and *The Star Online* and found that the portrayal of mental health in the two newspapers in 2020 was positive as these news reports focused on initiatives taken by the government and NGOs to handle mental health situations. Of course, there are also other studies

using data other than news discourse. A study of *halal* and *haram* concepts in Malaysian parliamentary discourse by Norsimah Mat Awal et al. (2020) revealed the trend and semantic flexibility of the terms over 60 years. With data covering Parliament 1 (1959-1964) to Parliament 13 (2013-2018) in Malaysian Hansard Corpus (MHC), they looked at the frequency and context of these two words, and found different usage patterns of the two words, which also displayed variations of meaning over time. What a large-scale corpus-based approach has been successful in achieving is the identification of semantic and sociolinguistic variation and patterns of use while reducing researcher's bias in the analysis and interpretation of findings.

In conclusion, corpus approach has been widely applied in studies of social linguistics. The use of corpus can aid sociolinguistics "in providing large amounts of existing data, along with computational tools and procedures which allow common (and rare) language patterns and frequencies to be identified quickly and accurately and compared across different populations" (Baker, 2010, p. 9). As a large quantity of news data is involved in this study, a corpus of news was built and corpus tool WordSmith was used for keyword and concordance analysis to investigate in depth sentiment reflected in news.

SENTIMENT ANALYSIS ON NEWS DISCOURSE

Sentiment analysis is originally a domain in computational linguistics. The term "sentiment" has several synonyms in different fields of studies, such as evaluation in news discourse (Bednarek, 2006), attitude/stance in academic English (Biber et al., 1999), appraisal in Systemic Functional Linguistics (Martin & White, 2005), etc. Sentiment in this study is used as the general term for attitude/stance/viewpoint/feelings reflected in news reports, which is in line with news discourse. Since 1990s, there have been some preliminary works on subjectivity and sentiment adjectives, but research on sentiment analysis mainly started in the early 2000 and became a popular research area from then on (Liu 2012). Scholars and data scientists (Pang & Lee, 2008; Liu, 2012) have carried out various research in this field. Now sentiment analysis is widely used in almost all domains such as economics (Scaffidi et al., 2007; Zhai et al., 2011), political science (Antonakaki et al., 2017; Ansari, 2020), and social sciences (Godbole et al., 2007; Balahur et al., 2013) because people's opinions have a vital influence on people's behavior.

Most work on sentiment analysis has been carried out on highly subjective texts such as product reviews, tweets, and blogs because "the target is clearly defined and unique across the text" (Balahur & Steinberger, 2009, p. 1). Later, researchers began to apply sentiment analysis on news articles and social media posts, the analysis of which is more complicated as there are more entities involved in the texts. Godbole et al. (2007) conducted sentiment analysis on news and blogs by assigning positive/negative scores to each entity in the corpus and found that controversial politicians in the U.S. were often talked about with different sentiments in news and blogs, which reflected the different stand between mainstream news agencies and citizen journalists like bloggers. Similarly, by analyzing more than 1,500 quotes from news articles, Balahur et al. (2013 & 2016) attempted to "separate positive or negative opinion from good or bad news" and pointed out that sentiment analysis of news is different and more difficult than that of movie or product reviews because news articles cover more content and different points of view. Besides sentiment analysis of news in English, analysis of news in other language has also been carried out. Bučar et al. (2018) carried out a comprehensive sentiment analysis of Slovene news texts by manually annotating 10,000 pieces of news for polarity in 3 levels of passage, paragraph and sentence respectively and found that positive and negative sentiments were often more explicit at the

beginning of the news reports but more neutral at the end of the reports. The studies above have shown that sentiment analysis of news requires a combination of in-depth analysis and different algorithm because of the complexity of the genre (Edgerly & Vraga, 2020) and its different norms and practices (Cook, 2006).

Previous studies on sentiment analysis of news have proven the feasibility of such approach in discovering sentiment in the news. However, most of such research is carried out in the field of computation, whose main focus is to improve algorithm and accuracy in sentiment analysis. Combining sentiment analysis and corpus approach on news discourse will be an innovative move in linguistic studies, to see how computational linguistics can aid discourse analysis.

METHODOLOGY

The aim of the study is to use a mixed method approach, namely sentiment analysis, corpus approach and discourse analysis, to investigate sentiment towards China in *The Star Online* between 2012 and 2021. As part of the research, this paper focuses on the general trend of sentiment of China-related news in the past 10 years and the possible reason(s) behind such sentiment. China-related news retrieved from *The Star Online* is used as the data for investigation.

DATA

Based on convenience sampling (Neuman, 2014, p. 248), whose “primary criteria for selecting cases are that they are easy to reach, convenient, or readily available”, online news was chosen because of its easy accessibility. *The Star Online*, the digital companion to the newspaper *The Star*, was launched on June 23rd, 1995, as Malaysia’s first news website. In 2014, the website was recognized as one of the best in Asia by the World Association of Newspapers and News Publishers (WAN-IFRA) (*The Star Online*, 2021). According to the Reuters Institute Digital News Report (2021), *The Star* is the first largest weekly reach English newspaper in print, while *The Star Online* is the second largest weekly reach English news website, after *Malaysiakini*. In the meantime, *The Star* enjoys the second highest brand trust score, after Astro Awani. The Star Online has a million strong following on both Twitter and Facebook (*The Star Online*, 2021).

Using “China” as a search word, news articles mentioning China, either in news title or in the main body from the time frame of January 2012 to December 2021, were collected from *The Star Online*. The time frame was set as such because the cooperation between China and Malaysia in the past 10 years is more frequent and comprehensive. Instead of focusing on only one type of news, the study collected various types, including hard news and opinion pieces, covering world’s economy, politics, culture etc., hoping to get a comprehensive view of the sentiment of China-related news in *The Star Online*. Either news by local writers or reprinted from other news agencies was included because when readers browse the news website, they will focus more on the content instead of on the writer or source. Altogether, 137,475 pieces of China-related news with a total of more than 65 million words were saved in 120 files by each calendar month.

TABLE 1. News Articles from 2012 to 2021

Month Year	1	2	3	4	5	6	7	8	9	10	11	12	Year Total
2012	1168	1100	1106	1271	1221	1163	1457	1677	1316	1213	1069	908	14669
2013	413	805	972	947	424	778	925	894	939	1063	1046	884	10090
2014	1111	914	1456	1161	1246	923	940	999	1180	1015	1017	893	12855
2015	981	883	1003	1097	996	1131	1290	1313	1302	1341	1270	1161	13768
2016	1356	1048	1243	1118	1144	974	1124	1284	1137	1101	1200	1099	13828
2017	1256	1079	1355	1175	1427	1163	1272	1265	1169	1191	1223	1025	14600
2018	1140	896	1161	1227	1120	1089	1240	1466	1152	1206	1151	1072	13920
2019	1350	1019	1222	1197	1274	1236	1249	1427	994	970	1219	1292	14449
2020	1955	2026	1498	1176	1053	1141	1428	915	1390	1775	1473	1552	17382
2021	1693	926	1142	999	881	946	966	963	842	900	792	864	11914
Total Number of News Articles													137,475

TOOLS FOR ANALYSIS

As the aim of this paper is to examine the general trend of sentiment of China-related news, the analysis is divided into two phases. In the first phase, Azure Machine Learning was used as the analytic tool to obtain the sentiment of these news reports, and in the second phase, corpus tool Wordsmith (Scott, 2020) was used for discourse analysis. The following part introduces how data was processed.

Published research on sentiment analysis can be divided into classifier-based (Duda et al., 2001; Theodoridis & Koutroumbas, 2008; Nasa & Suman, 2012) and lexicon-based approaches (Esuli & Sebastiani, 2006; Baccianella et al., 2010). The classifier-based approach involves machine learning, in which the computer has to learn a classifier (such as Naïve Bayes or Support Vector Machines etc.) before analyzing sentiment automatically. The lexicon-based approach is to assign words with sentiment scores and the sentiment of a sentence or an article will be the sum of the sentiment scores of the words in it. It is generally agreed that classifier-based approach outperforms the lexicon-based one (Salah, 2014). Microsoft Azure's Text Sentiment Analysis is a web-based machine learning classification algorithm used to generate a sentiment value between 0 and 1 (Andersson, Dryden & Variawa, 2018; Peslak, 2018). It is available in Azure Machine Learning (hereinafter referred to as AML), the add-in tool in Microsoft Excel. As AML already has an extensive body of trained data and both AML platform and python programming have been confirmed to be used for sentiment analysis on various data (Singh & Sharma, 2021), we used the sample data set provided by AML for analysis. As the news reports were saved in Microsoft Excel, AML was applied directly, and numeric scores between 0 and 1 were generated for each article and polarity was calculated automatically. News reports with polarity values closer to 1 are labeled as positive sentiment, news reports with scores closer to 0 are negative, while news reports with scores around 0.5 are taken as neutral sentiment.

title	time	tweet_text	author	Sentiment	Score
1kg bird's nest lim	Tuesday, 17 Jan 201	PETALING JAYA: TOURISTS to IMalaysia are now only allowed to take home a maximum 1kg of bird's nest. "This is to ensure that no smuggling of this expensive commodity takes place," Agriculture and Agro-based Industry Minister Datuk Seri Noh Omar.He also announced that local companies would require three certificates to export edible bird's nest products to China.The certificates are from the Veterinary Services Department, the Health Ministry and the Malaysian Communications and Multimedia Commission (MCMC). "After a company receives all three certificates, it will be allowed to use the 1Malaysia Best' logo and can export to China," Noh told reporters after launching the first authentic edible bird's nest trade and depository centre by PT Swift Marketing Sdn Bhd in	NURHIDAY#	negative	0.105

FIGURE 1. A Sample of Sentiment Analysis Result

Figure 1 shows an example of sentiment analysis result for one piece of news. In the figure it can be seen that the overall score for this piece of news is 0.105, which is closer to 0. As a result, the sentiment of this piece of news is labeled as negative by AML automatically.

The polarity of China-related news reports from 2012 to 2021 was obtained using AML (as shown in Figure 1). Each spreadsheet contained news reports for the whole month. As the number of news articles for each month ranged from several hundred to more than 1,000, the percentage of news with negative, neutral and positive sentiment was calculated to compare results of different months. The percentage is presented in Table 2. As can be seen in Table 2, in Feb. 2012, there are 1,100 pieces of news mentioning China. 88.27% (971 pieces) of news is labelled as negative, 9.45% (104 pieces) is labelled as positive and only 2.27% (24 pieces) as neutral.

TABLE 2. Percentage of Sentiment of Feb. 2012

Polarity	No. of Articles	Percentage
Negative	971	88.27%
Neutral	25	2.27%
Positive	104	9.45%
Grand Total	1100	100%

With percentages of news with negative, positive and neutral sentiment of all the 120 months keyed in Excel, a chart showing the sentiment in China-related news over 120 months is generated by Excel. As is shown in Figure 2 below, the blue columns represent the percentages of news with negative sentiment, the columns in red represent the percentages of news with neutral sentiment while the green line represents the percentages of news with positive sentiment. Overall, the results show that the sentiment for most of the China-related news in *The Star Online* is negative.

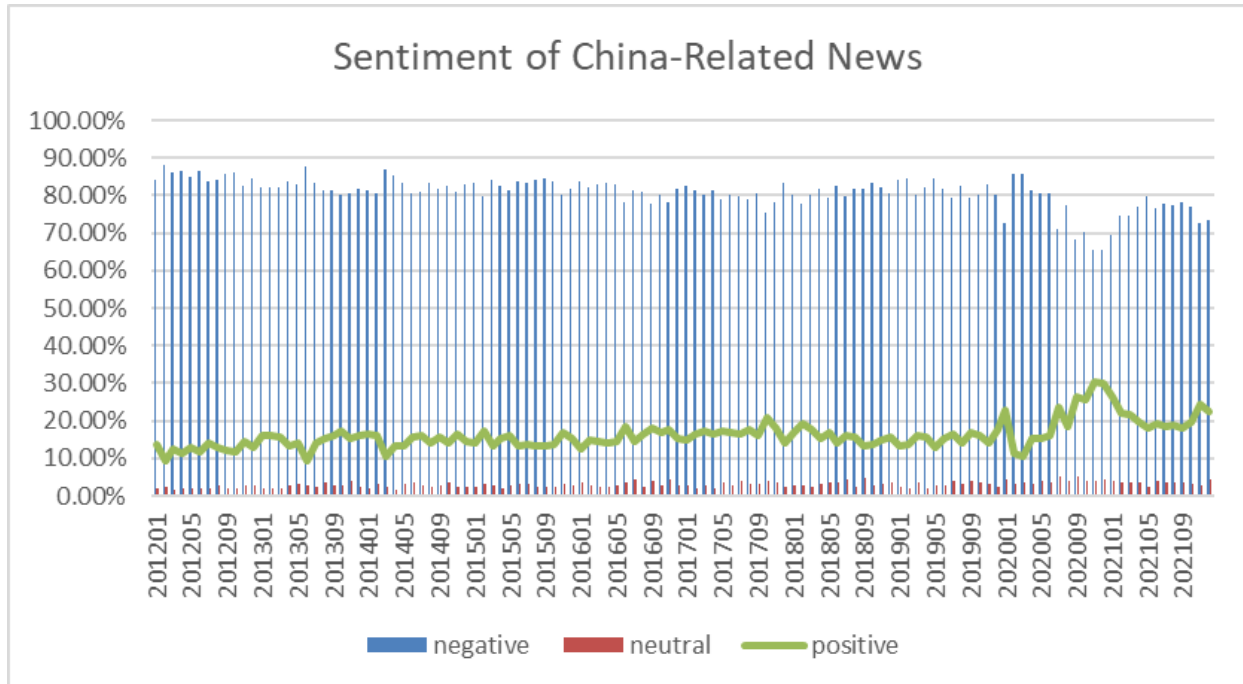


FIGURE 2. Sentiment of China-Related News (in percentage)

AML can only show the polarity and the general sentiment trend reflected in the news. The cause behind such sentiment and trend, however, remains unknown to us. Therefore, in the second phase of the study, Wordsmith 8.0 (henceforth Wordsmith) was used to examine the reason behind such sentiment and trend. The main functions of Wordsmith such as wordlist, keyword, and concord, allow researchers to explore the data further in terms of the unusual frequent words (whose frequency is unusually high as compared to a reference corpus), significant collocates and their concordances. Here, the functions of word list and keyword help to uncover the unusual frequent words in months with the highest proportion of news with negative sentiment (as compared to the wordlist of a reference corpus, the corpus of all the news in this case), and the function of Concord helps to investigate the top collocates (words that co-occur within a 5-word span) of those keywords. As Wordsmith mainly deals with plain text files, the news content of each month was saved in one plain text file, and a corpus of *The Star Online* on China-related news (hereinafter referred to as SOC) of 120 files was created. Then, the text files of the months with highest percentage of news with negative sentiment were used as the target files, and SOC serves as the reference corpus, and the default setting of Log Likelihood was used for calculating the “keyness” of words in the target files. The value of “keyness” of every word is calculated by comparing its frequency in the month with that in SOC through log-likelihood tests. In this way, keywords with higher keyness were then generated in comparison with the words in SOC, and the unusual frequent words in the target file could be identified. Next, collocates of the most frequent keywords were then generated, and context of these collocates was examined to discover their usage and sentiment. In short, keywords, collocates and concordance analysis were used together to investigate the cause behind the high proportion of negativeness in the data.

FINDINGS AND DISCUSSION

As can be seen in Figure 2, the percentages of news with negative sentiment vary greatly even in the same year. Hence, statistics of the month with the highest percentage of news with negative and positive sentiment in each year from 2012 to 2021 are presented in Table 3 and Table 4 respectively.

TABLE 3. Months in Each Year with Highest Percentage of News with Negative Sentiment (percentage of grand total)

	201202	201306	201403	201508	201601	201712	201810	201902	202003	202105
negative	88.27%	87.66%	87.08%	84.23%	83.92%	83.43%	83.50%	84.49%	85.78%	79.68%
neutral	2.27%	2.83%	2.47%	2.44%	3.69%	2.53%	2.65%	1.96%	3.67%	2.38%
positive	9.45%	9.51%	10.45%	13.33%	12.39%	14.04%	13.85%	13.54%	10.55%	17.93%

TABLE 4. Months in Each Year with Highest Percentage of News with Positive Sentiment (percentage of grand total)

	201211	201310	201411	201502	201606	201710	201802	201912	202011	202101
negative	82.69%	80.15%	80.92%	79.73%	78.03%	75.40%	77.99%	80.26%	65.51%	69.34%
neutral	2.62%	2.73%	2.56%	3.06%	3.59%	3.86%	2.79%	2.24%	4.01%	4.08%
positive	14.69%	17.12%	16.52%	17.21%	18.38%	20.74%	19.22%	17.49%	30.48%	26.58%

As can be seen in Table 3 that February of 2012 has the highest percentage of news with negative sentiment, not only in the year 2012, but also among the 120 months in the ten-year span; May 2021, however, has the least proportion of news with negative sentiment in 2021. Table 4 displays that November 2020 has the most proportion of news with positive sentiment (30.48%) of all the 120 months, while in 2012, the month with most proportion of news with positive sentiment is November (14.69%). Both Table 3 and Table 4 have shown a trend of decreasing proportion of news with negative sentiment and increasing percentage of news with positive sentiment. To get the general trend of sentiment, the average percentage of polarity of each year is shown in Table 5 and Figure 3 is a clustered column-line chart of the average percentage of China-related news showing the general trend of sentiment in 10 years.

TABLE 5. Average Percentage of Each Year

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
negative	85.26%	82.42%	82.58%	82.72%	81.14%	80.10%	80.99%	81.84%	75.42%	75.72%
neutral	2.18%	2.73%	2.74%	2.66%	3.21%	3.01%	3.26%	3.00%	4.08%	3.51%
positive	12.56%	14.85%	14.68%	14.62%	15.65%	16.89%	15.75%	15.16%	20.49%	20.77%

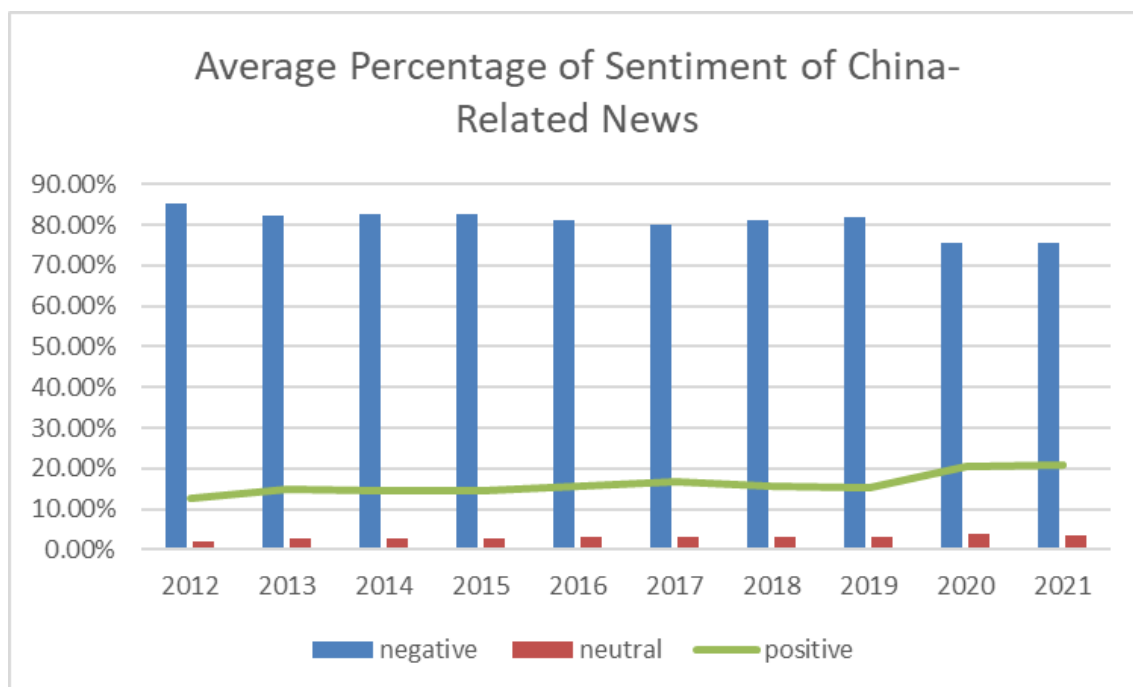


FIGURE 3. Average Sentiment of China-Related News (in percentage)

Several interesting phenomena can be observed from the tables and chart above:

- 1) Most of the China-related news reports in *The Star Online* displayed either positive or negative sentiment. The amount of news with negative sentiment takes up the majority while news with neutral sentiment accounts for no more than 5% in each month.
- 2) Sentiment of China-related news in *The Star Online* is generally negative. In the 10 years span, February 2012 has the most reports with negative sentiment — about 88.27% of the reports in that month are negative, followed by 87.66% in June 2013 and 87.08% in March 2014. However, a slightly downward trend of the proportion of news with negative sentiment can be detected, decreasing from an average percentage of 85.26% of news with negative sentiment in 2012 to around 75.5% in 2020 and 2021.
- 3) As the proportion of negative China-related news decreases, the proportion of news with positive sentiment increases from average percentage of 12.56% in 2012 to about 20.5% in 2020 and 2021. November 2020 has the most positive reports; 30.48% of the reports are positive, doubled the proportion compared to 14.69% of February 2012.

Knowing the polarity and general trend of sentiment is not enough; the cause behind such sentiment is more important in understanding why such patterns occur, especially why news with negative sentiment makes up the majority of news articles. To investigate the source of negativity, we did a close analysis of months with the highest percentage of news with negative sentiment i.e., February 2012, June 2013, and March 2014. Using Wordsmith, keywords for these three months were generated and concord was used to generate collocates of the most frequent keywords. Analyzing the keywords, their collocates and concordances may help us to understand the reasons behind the high negativity in these three months.

DISCUSSION

KEYWORDS, COLLOCATES AND CONCORDANCE ANALYSIS OF FEB. 2012

Table 6 is a Keyword List for February 2012, which ranks the highest in terms of percentage of news with negative sentiment of all 120 months.

TABLE 6. Top Ten Keyword by Keyness in Feb. 2012

Rank	Keyword	Keyness (Log Likelihood)	Frequency
1	ASSAD	8,893.30	2,719
2	SYRIA	8,492.93	2,976
3	SYRIAN	7,670.38	2,405
4	HOMS	6,165.71	1,252
5	ARAB	6,112.11	1,726
6	OPPOSITION	3,518.92	1,427
7	RESOLUTION	3,285.49	1,083
8	N*	2,273.37	1,370
9	IRAN	2,089.13	1,702
10	AMRO	2,084.06	386

*Note: N as in U.N., since the U.S. is often mentioned in the news, so only N stands out as a unusually frequent word.

In February 2012, the top keywords were *assad*, *syria*, *syrian*, *homs* (a Syrian city), *arab*, *opposition*, *resolution*, *iran* and *amro* (as in Baba Amro, a district in Homs) which are related to Syria (see Table 6). In fact, the word *syria* appeared 282 times in news titles of the month, which means out of the 1,100 news articles of February 2012 collected, more than one-fourth of the news were about Syria. This is related to the Syrian Crisis that began with anti-government protests against the Assad-led Syrian government in March 2011. As Syria played an important role in the Middle East, the Arab League laid out a reform plan in late January 2012 aiming to form a new Syrian government, strongly supported by the U.N. Security Council. China took a stance on political settlement in solving the crisis instead of taking any action that could further militarize the crisis. Both China and Russia, who had strong economic and military ties with Syria, vetoed the U.N. Security Council resolution on February 4, 2012, because the resolution was not a settlement widely accepted by all parties in Syria. The veto was frequently mentioned in news reports on Syria Crisis, explaining why in February 2012, at least one-fourth of China-related news was about Syria.

TABLE 7. Top 5 Noun and Adjective Collocates of *syria* by Frequency in Feb. 2012

Rank	Noun	Frequency	MI	Adjective	Frequency	MI
1	opposition	136	12.32	arab	84	4.12
2	resolution	136	5.06	military	72	7.24
3	meeting	108	5.90	humanitarian	72	5.66
4	arms	84	12.95	western	62	7.44
5	sanctions	84	14.44	syrian	60	6.44

Table 7 is a list of top 5 noun and adjective collocates of the most frequent keyword *Syria*. Among these 10 words, *opposition*, *arms*, *sanctions*, and *military* conveyed negative meaning.

Example (1) and (2) are sentences taken from news in February 2012 containing the collocates *opposition* and *resolution*.

(1) *The setback in diplomatic efforts to defuse the revolt peacefully came after world leaders and Syrian **opposition** activists accused Assad’s forces of killing hundreds of people in a bombardment of the city of Homs, the bloodiest night in 11 months of upheaval in the pivotal Arab country. (Russia, China veto U.N. resolution telling Assad to quit, 04 Feb 2012)*

(2) *Syria called the League’s **resolution** “a flagrant departure from the group’s charter and a hostile act that targets Syria’s security and stability.” (Ties sour over war denial, 27 Feb 2012)*

We may see that although *resolution* alone does not bear any negative meaning by definition, its occurrence with *flagrant* and *hostile* in Example (2) creates a negative tone or prosody, where the negative connotation of the context affects the polarity of the word. The negative sentiment can thus be attributed to its collocations. Additionally, frequently used phrases such as “*Syria’s (political) opposition / to arm the opposition movement in Syria*”, “*arms sales to Syria*”, “*sanctions on (against) Syria*” and “*about/against/for/to military intervention in Syria*”, resulted in a much more pronounced negative sentiment than that of the other months. But this does not show sentiments towards China, only negative meaning of the issues, in this case, the sentiment is towards Syria.

KEYWORDS, COLLOCATES AND CONCORDANCE ANALYSIS OF JUN. 2013

Table 8 lists the top 10 keywords for June 2013, which is with the second highest percentage of news with negative sentiment

TABLE 8. Top Ten Keyword by Keyness in Jun. 2013

Rank	Keyword	Keyness (Log Likelihood)	Frequency
1	SNOWDEN	7,916.08	1,216
2	RELEASE	1,243.74	357
3	NSA	1,015.54	198
4	KUALA	990.93	830
5	LUMPUR	951.94	823
6	ECUADOR	799.38	156
7	RESEARVES	775.07	198
8	SURVEILLANCE	703.55	247
9	WIKILEAKS	657.40	115
10	OBAMA	606.24	411

Table 8 displays the top 10 keywords in June 2013 news, which include *snowden*, *release*, *nsa*, *surveillance* and *wikileaks*, etc. The words led to a former American intelligence contractor. *Edward Snowden*. In June 2013, Snowden revealed the existence of PRISM, a secret wide-ranging surveillance program used by the National Security Agency (NSA) of the United States for collecting data from emails, live chats, and search histories, as well as Britain’s electronic eavesdropping agency GCHQ’s gathering information via PRISM. He also disclosed that US government had been hacking computers in mainland China and Hong Kong for years. On June

23rd, with the assistance of Wikileaks, Snowden left Hong Kong for Ecuador via Russia, but was detained there as his passport was rescinded by the U.S. government.

Table 9 below lists top 5 noun and adjective collocates for *snowden*.

TABLE 9. Top 5 Noun and Adjective Collocates of *snowden* by Frequency in Jun. 2013

Rank	Noun	Frequency	MI	Adjective	Frequency	MI
1	Contractor	62	7.36	former	35	10.56
2	Asylum	59	7.45	russian	34	6.66
3	Decision	35	13.73	legal	24	10.86
4	Government	30	11.51	new	13	3.18
5	Agency	29	6.68	political	12	10.6

Except for *asylum*, the collocates in Table 9 do not convey negative meaning or connotation by definition. The words *contractor*, *government*, *agency* and *former* were all used as noun modifiers for *Snowden* in the news. Example (3) and (4) are sentences taken from news in June 2013 containing the collocates *former*, *agency* and *new*.

(3) *Snowden, the former National Security Agency contractor accused of leaking secrets of U. S. electronic surveillance, already had spent nearly two weeks in hiding. (On a smooth journey, 20 Jun 2013)*

(4) *Hong Kong's South China Morning Post newspaper earlier quoted Snowden offering new details about U. S. surveillance activities, including accusations of U. S. hacking of Chinese mobile phone firms and targeting of China's Tsinghua University. (U.S. spy fugitive Snowden seeks asylum in Ecuador, 24 Jun 2013)*

Though these collocates themselves are neutral or positive by definition, we may see that both examples were used in contexts with a strong negative valence, as the events related to Snowden involved spying activities, leaking secrets and political asylum. This helps to explain why China-related news reports in June were mostly negative in sentiment.

KEYWORDS, COLLOCATES AND CONCORDANCE ANALYSIS OF MAR. 2014

Table 10 is the Keyword List for March 2014, which ranks the third highest in percentage of news with negative sentiment among all.

TABLE 10. Top Ten Keyword by Keyness in Mar. 2014

Rank	Keyword	Keyness (Log Likelihood)	Frequency
1	PLANE	3,958.07	1,045
2	SEARCH	3,643.34	1,095
3	UKRAINE	2,152.84	636
4	MISSING	1,892.13	451
5	FLIGHT	1,861.65	749
6	AIRCRAFT	1,768.46	774
7	AIRLINES	1,464.16	521
8	PASSENGERS	1,413.60	567
9	RADAR	1,253.81	333
10	OCEAN	972.54	330

In Table 10, we could see that all the top ten keywords in March 2014, such as *plane*, *search*, *missing*, *flight*, *aircraft*, *airlines*, *passengers*, *radar* and *ocean* etc. pointed to a tragic aviation accident on March 8, 2014. Both *missing* and *mh* appeared 269 times in news titles of the month, which means out of the 1,456 news articles of March 2014 collected, nearly one-fifth of the news were about the missing MH370 plane. At about 00:42 Malaysian time on March 8, 2014, a Malaysian Airline aircraft, MH370, left Kuala Lumpur International Airport for Beijing. There were 12 crew members and 227 passengers on board, and about three-fourths of the passengers were Chinese. The flight was expected to arrive at Beijing Capital International Airport at 6:30 that day, however, it was last seen on Malaysian military radar at 2:14 over the Strait of Malacca and had disappeared ever since. The missing MH370 attracted the international community's attention, and follow-up search efforts were widely covered by the media. Table 11 below lists the top 5 noun and adjective collocates of the most frequent keyword *search*.

TABLE 11. Top 5 Noun and Adjective Collocates of *search* by Frequency in June 2013

Rank	Noun	Frequency	MI	Adjective	Frequency	MI
1	rescue	158	19.89	missing	85	15.36
2	area	150	15.23	malaysian	38	17.45
3	operation	109	19.67	southern	36	13.68
4	operations	90	20.81	international	37	13.06
5	plane	61	22.84	new	35	12.36

Except for the word *missing*, most of these 10 words are neutral (such as *operation*, *plane* and *international*) or even positive in meaning (such as *rescue* and *new*). However, when analyzing the concordances, we found that the examples were used in contexts with a strong negative valence. Examples (5) and (6) are sentences containing collocates for *rescue*, *missing*, *new* and *area* taken from news in March 2014.

(5) A Vietnamese **rescue** official denied that the signal of the **missing** Malaysian plane has been detected. (from *Missing MAS flight: Vietnam official denies detecting Malaysian missing plane's signal*, 08 Mar 2014)

(6) A **new** search **area** failed to yield an immediate breakthrough in the hunt for ill-fated Flight MH370 on Sunday, as Australia appointed its former military chief to help coordinate the operation in the Indian Ocean. (from *MH370 crash: Search draws blank as Australia brings in ex-military chief*, 30 Mar 2014)

We may see that both examples conveyed negative sentiment despite those neutral or positive collocates, as the event itself was a tragedy and the search and rescue work were not making much progress. Another keyword *ukraine* in Table 10 is about the Ukraine Crisis, specifically the conflict between Ukraine and Russia in Crimea as reflected in news titles like “China says supports international financial aid for Ukraine” and “China's Xi urges political solution to Ukraine crisis”. Either in the topic of missing MH370 or Ukraine Crisis, news reports on both events were highly likely to be negative as both crises were not something people expect to see. This raises the question: What is it about news with negative sentiment that are prioritised in news content, as demonstrated in the trend of negative sentiment in *The Star Online* news? Interestingly, a number of studies on our news consumption habits (Stafford 2014) have revealed the human tendency to prioritise negative over positive news content (Soroka 2015). Stories with

a negative tone — conflicts, crisis, corruption, crimes — are repeatedly chosen over neutral or positive ones. Correspondingly, the value of news with negative sentiment in terms of readership and sales far exceeded the positive ones. The fact that most of the China-related news articles frequently revealed negative sentiment further support the above findings. In addition, since a portion of news with negative sentiment is not closely related to China, the slight upward trend may indicate the closer cooperation between China and Malaysia.

CONCLUSION

The result of this study indicated that from 2012 to 2021, China-related news in *The Star Online* were mostly negative, but a slightly upward trend of positive sentiment could also be observed. Such result verifies the feasibility of applying sentiment analysis on news texts. The top ten keywords, their collocates and context in the three months with the highest proportion of news with negative sentiment helped to reveal the reasons behind the negative sentiment as the findings have shown, the result of automatic sentiment analysis is more about whether the news is good or bad. When bad things happen, news reports of that period tend to use words strongly associated with negative sentiment. However, sentiment toward a certain entity in the news e.g., China, is not necessarily negative. Therefore, even though China-related news in February 2012, June 2013 and March 2014 has the highest proportion of news with negative sentiment, the negativity was more about the bad news and not necessarily negative sentiment towards China. Combining sentiment analysis with corpus approach on the study of China-related news in Malaysian media enriches the study of news discourse from the perspective of corpus linguistics perspective.

This study focuses on the trend of media sentiment in China-related news in *The Star Online*. However, analyzing the slight upward trend of positive sentiment, especially the sudden increase in news with positive sentiment in November 2020, making it the highest in proportion of all the 120 months, is beyond the scope of this paper. What caused the steady increase of the proportion of news with positive sentiment and what was the sentiment towards China are questions that remain to be answered. Currently, automated sentiment analysis is not able to accurately identify media sentiment towards China in *The Star Online*. In the follow-up study, a more sophisticated linguistic analysis, such as discourse analysis using Appraisal Theory, is needed for further study to examine sentiment towards China in the news.

ACKNOWLEDGEMENTS

This study was funded by Ministry of Education of China under Research Grant of Humanities and Social Sciences (Project No. 18XJC740010).

REFERENCES

- Alasuutari, P., Qadir, A. & Creutz, K. (2013). The domestication of foreign news: news stories related to the 2011 Egyptian revolution in British, Finnish and Pakistani newspapers. *Media, Culture & Society*. 35(6), 692-707. <https://doi.org/10.1177/0163443713491299>
- Andersson, E., Dryden, C. & Variawa, C. (2018). Applying Machine Learning to Student Feedback Through Clustering and Sentiment Analysis. Proceedings of the Canadian Engineering Education Association (CEEA) Conference, 3-6 June, Vancouver BC. <https://doi.org/10.24908/pceea.v0i0.13059>

- Ansari, M. Z., Aziz, M. B., Siddiqui, M. O., Mehra, H. & Singh, K. P. (2020). Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science*. (167), 1821-1828. <https://doi.org/10.1016/j.procs.2020.03.201>
- Antonakaki, D., Spiliotopoulos, D. V., Samaras, C, Pratikakis, P, Ioannidis, S and Fragopoulou, P. (2017). Social media analysis during political turbulence. *PLoS ONE*. 12(10), e0186836. <https://doi.org/10.1371/journal.pone.0186836>
- Anunne, U. K. & Yan, Lifeng (2018). China in Foreign Media: Assessing China's Image in Online Editions of Nigeria's Leading Newspapers in 2017. *European Scientific Journal*, ESJ, 14(35), 165-188. <https://doi.org/10.19044/esj.2018.v14n35p165>
- Armony, A. C. & Velásquez, N. (2015). Anti-Chinese Sentiment in Latin America: An Analysis of Online Discourse. *Journal of Chinese Political Science*. 20, 319–346. <https://doi.org/10.1007/s11366-015-9365-z>
- Asad, S., Mohd Noor, S. N. F. & Jaes, L. (2019). Transitivity analysis of election coverage in online newspapers of Malaysia & Pakistan: A study with critical discourse analysis & systematic functional linguistics' perspective. *Amazonia Investiga*. 8(21), 168–176.
- Baccianella, S., Esuli, A. and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA).
- Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburg: Edinburg University Press.
- Balahur, A. et al. (2013). Sentiment analysis in the news. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010)*. 19-21 May, pp. 2216–2220. <https://arxiv.org/pdf/1309.6202> [Oct 12th, 2019]
- Balahur, A., & Steinberger, R. (2009). Rethinking Sentiment Analysis in the News: from Theory to Practice and back. *Proceeding of WOMSA*. 9, 1-12.
- Bednarek, M. (2006). *Evaluation in Media Discourse – Analysis of a Newspaper Corpus*. London & New York: Continuum.
- Bednarek, M., & Caple, H. (2014). Why Do News Values Matter? Towards a New Methodological Framework for Analysing News Discourse in Critical Discourse Analysis and Beyond. *Discourse & Society*. 25(2), 135-158.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Bickes, H., Otten, T. & Weymann, L. C. (2014). The financial crisis in the German and English press: Metaphorical structures in the media coverage on Greece, Spain and Italy. *Discourse & Society*. 25(4): 424-445. doi: 10.1177/0957926514536956
- Bučar, J., Žnidaršič, M. & Povh, J. (2018). Annotated news corpora and a lexicon for sentiment analysis in Slovene. *Lang Resources & Evaluation*. 52, 895–919. <https://doi.org/10.1007/s10579-018-9413-3>
- Bui, N. T. (2017). Managing anti-China nationalism in Vietnam: evidence from the media during the 2014 oil rig crisis. *The Pacific Review*. 30(2), 169-187. <https://doi.org/10.1080/09512748.2016.1201132>
- Brookes, H. J. (1995). 'Suit, Tie and a Touch of Juju'—The Ideological Construction of Africa: A Critical Discourse Analysis of News on Africa in the British Press. *Discourse Society*. 6(4), 461-494. <https://doi.org/10.1177/0957926595006004002>
- Briones, R. R. Y. (2017). Using Fairclough's CDA Framework on News Articles. *Beyond Words* 5(1), 83-97.

- Cavasso, L. & Taboada, M. (2021). A corpus analysis of online news comments using the Appraisal framework. *Journal of Corpora and Discourse Studies*. 4, 1-38. <https://doi.org/10.18573/jcads.61>
- Chan, M. (2012). The discursive reproduction of ideologies and national identities in the Chinese and Japanese English-language press. *Discourse and Communication*. 6(4), 361-378. <https://doi.org/10.1177/1750481312457496>
- Cheng, W. & Lam, P. W. Y. (2013). Western perceptions of Hong Kong ten years on: A corpus-driven critical discourse study. *Applied Linguistics*. 34(2), 173-190. <https://doi.org/10.1093/applin/ams038>
- Cook, T. E. (2006). The News Media as a Political Institution: Looking Backward and Looking Forward. *Political Communication*. 23 (2), 159-171.
- Duda, R., Hart, P. and Stork, D. (2001). *Pattern Classification*. 2nd edition. New York: John Wiley&Son.
- Duff, P. A., Anderson, T., Doherty, L., & Wang, R. (2015). Representations of Chinese Language Learning in Contemporary English-language News Media: Hope, Hype, and Fear. *Global Chinese*. 1(1), 139-168. <https://doi.org/10.1515/glochi-2015-1006>
- Ederly, S. & Vraga, E. K. (2020). News-Ness as an Audience Concept for the Hybrid Media Environment. *Journalism & Mass Communication Quarterly*. 97 (2), 416-434.
- Esuli, A. and Sebastiani. F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. In Proceedings from the International Conference on Language Resources and Evaluation (LREC).
- Fairclough, N. (1995). *Media Discourse*. London: Edward Arnold.
- Fowler, R. (1991). *Language in the News Discourse and Ideology in the Press*. Oxon: Routledge.
- Gabrielatos, C. & Baker, P. (2008). Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics*. 36(1), 5-38. <https://doi.org/10.1177/0075424207311247>
- Global Times. (2021). ASEAN becomes China's largest trading partner in 2020, with 7% growth. Retrieved January 15th, 2021 from <https://www.globaltimes.cn/page/202101/1212785.shtml>
- Godbole, N., Srinivasaiah, M. & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. ICWSM 2007 - International Conference on Weblogs and Social Media. <https://www.icwsm.org/papers/3--Godbole-Srinivasaiah-Skienna.pdf>
- Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*. 34(4), 365-399.
- Hanks, P. (2004). Corpus Pattern Analysis. In Euralex Proceedings. (1), 87-98. Lorient: Université de Bretagne-Sud.
- Haw, A. L. (2020). Duelling Discourses: A Rhetorical Device for Challenging Anti-Asylum Sentiment in Western Australia. *Journal of Australian Studies*. 44(3), 303-317. <https://doi.org/10.1080/14443058.2020.1737178>
- Idris, A. (2019). Implicit stances taken by ABC and BBC towards Indonesia presidential election : CDA Perspectives. *Journal of English Language Studies*. 4(2), 197-211.
- Kim, K. H. (2014). Examining US news media discourses about North Korea: A corpus-based critical discourse analysis. *Discourse and Society*. 25(2), 221-244. <https://doi.org/10.1177/0957926513516043>

- Leung, C. C. M. & Huang, Y. (2007). The paradox of journalistic representation of the other: The case of SARS coverage on China and Vietnam by western-led English-language media in five countries. *Journalism*. 8(6), 675-697.
- Lin Sixian. (2022). On the Framing Strategy of China-Related News in Western Mainstream Media. *Contemporary International Relations*. 1, 53-60.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. (Synthesis Lectures on Human Language Technologies). Vermont: Morgan & Claypool Publishers.
- Liu, M. & Jiang, C. (2019). Constant fear, but lingering nostalgia: British press representations of post-colonial Hong Kong 20 years on. *Discourse & Communication*. 13(6), 630–646. <https://doi.org/10.1177/1750481319868852>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*. 45(1), 36-62.
- Luo Yi. (2019). Image of China as Media Representation of Others. *Media*. 7, 79-82.
- Maryam Jahedi (2012). The Discursive Construction of Iran from 1979 to 2009 in the New York Times. Thesis for Doctor of Philosophy. Faculty of Modern Languages and Communication. Universiti Putra Malaysia.
- McEnery, T. and Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. London: Routledge.
- Nasa, C. & Suman. (2012). Evaluation of different classification techniques for web data. *International Journal of Computer Applications*. 52(9), 34–40.
- Neuman, W. L. (2014). *Social Research Methods: Qualitative and Quantitative Approaches* (Seventh Edition). Essex: Pearson.
- Ng Siew Hua. (2008). A Critical Discourse Analysis of Representations of Bilateral Issues Concerning Malaysia and Singapore in Mainstream Newspaper Editorials. Thesis for Doctor of Philosophy. Universiti Putra Malaysia.
- Nor Fariza Mohd Nor, Novelia Bernice Jeffree & Hilwa Abdullah@Mohd Nor. (2021). Health is Wealth: A Corpus-driven Analysis of the Portrayal of Mental Health in Malaysian English Online Newspapers. *Journal of Language Studies*. 21(2), 46-71. <https://doi.org/10.17576/gema-2021-2102-03>
- Norsimah Mat Awal, Kesumawati A. Bakar, Anis Nadiyah Che Abdul Rahman, & Imran Ho Abdullah (2021). Representasi Halal dan Haram dalam Wacana Parlimen Malaysia. *GEMA Online® Journal of Language Studies*. 21(4), 186-207.
- O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Ospina Estupinan, J. D. (2017). The coverage of China in the Latin American Press: Media framing study. *Cogent Arts & Humanities*. 4, 1287319, <https://doi.org/10.1080/23311983.2017.1287319>
- Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(2), 1–135. <https://doi.org/10.1561/1500000001>
- Peslak, A. R. (2018). Facebook Fanatics: A Linguistic and Sentiment Analysis of the Most “Fanned” Facebook Pages. *Journal of Information Systems Applied Research (JISAR)*. 11(1), 23-33.
- Reuters Institute (2021). Digital News Report 2021. Retrieved September 20, 2021 from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/202106/Digital_News_Report_2021_FINAL.pdf

- Romaine, S. (2001). 'A corpus- based view of gender in British and American English.' In M. Hellinger and H. Bußmann (Eds.), *Gender across Languages*. Vol. 1 (pp. 153–175). Amsterdam and Philadelphia: John Benjamins.
- Ruhrberg, S., Kirstein, G., Habermann, T., Nikolic, J. & Stock, W. (2018). #ISIS—A Comparative Analysis of Country-Specific Sentiment on Twitter. *Open Journal of Social Sciences*. 6, 142-158. <https://doi.org/10.4236/jss.2018.66014>.
- Salah, Z. (2014). Machine learning and sentiment analysis approaches for the analysis of Parliamentary debates. Doctoral dissertation. University of Liverpool.
- Scaffidi, C., Kevin B., Chang, E. Felker, M. Ng, H. & Jin, C. (2007). Red Opal: product-feature scoring from reviews. in Proceedings of Twelfth ACM Conference on Electronic Commerce (EC-2007).
- Scott, M. (2020). WordSmith Tools version 8, Stroud: Lexical Analysis Software.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Singh, R. & Sharma, P. (2021). Sentiment Analysis using Microsoft Azure Machine Learning and Python. *International Journal of Engineering Research & Technology (IJERT)*. 10 (11), 241-244. <https://doi.org/10.17577/IJERTV10IS110099>
- Sohoni, D. & Mendez, J. B. (2012). Defining immigrant newcomers in new destinations: symbolic boundaries in Williamsburg, Virginia. *Ethnicities & Racial Studies*. 37(3), 496-516. <https://doi.org/10.1080/01419870.2012.716521>
- Soroka, S. (2015). Why do we pay more attention to news with negative sentiment than to news with positive sentiment? The London School of Economics. Retrieved May 25, 2019 from <https://blogs.lse.ac.uk/politicsandpolicy/why-is-there-no-good-news/>
- Stafford, T. (2014). Psychology: Why bad news dominates the headlines. BBC Future. Retrieved July 29, 2019 from <https://www.bbc.com/future/article/20140728-why-is-all-the-news-bad>
- Stamou, A. G. (2001). The Representation of Non-Protesters in a Student and Teacher Protest: A Critical Discourse Analysis of News Reporting in a Greek Newspaper. *Discourse & Society*. 12(5), pp. 653-680. <https://doi.org/10.1177/0957926501012005005>
- Theodoridis, S. & Koutroumbas, K. (2008). *Pattern Recognition*. 4th edition. 2008. Burlington/San Diego/London: Academic Press.
- The Star Online. (2021). Retrieved September 15, 2021 from <https://www.thestar.com.my/AboutUs>
- Wang, Y. & Reagan, J. (2020). Media Sentiment Towards Chinese Investments in Malaysia: An Examination of the Forest City Project. *Asian Journal for Public Opinion Research*. ISSN:2288-6168 (Online), 8(3), 197-221. <https://doi.org/10.15206/ajpor.2020.8.3.197>
- Wardhaugh, R. (2010). *An Introduction to Sociolinguistics*. Sixth edition. Oxford: Blackwell.
- World Bank (2020). GDP Ranking. Retrieved June 26, 2022 from <https://databank.worldbank.org/data/download/GDP.pdf>
- Wu Junjing (2020). Construction of China's National Image in Foreign Chinese Newspaper. *Chinese Editors Journal*. 124, 109-114.
- Xinhua. (2021). China becomes Malaysia's largest trading partner for 12 consecutive year as bilateral trade grow in 2020. Retrieved February 21, 2021 from http://www.china.org.cn/world/Off_the_Wire/2021-01/30/content_77172119.htm
- Yin Yue (2020). China's National Image Constructed by Korean Daily and Korean National Daily. *Journal of Yanbian University (Social Sciences)*. 53(3), 20-28 & 139-140.

- Zhai, Z. W., Liu, X., Xu, H. & Jia, P. F. (2011). Clustering Product Features for Opinion Mining. in Proceedings of ACM International Conference on Web Search and Data Mining (WSDM-2011).
- Zhang Yijiang (2018). Construction of China's National Image in Mainstream Newspaper in Spain. *Youth Journalist*. 6, 93-94.
- Zhao Yonghua & Lu Junyu. (2021). Metaphors in News Discourse and the Selective Construction of National Image. *Contemporary Communication*. 6, 17-22.

ABOUT THE AUTHORS

Hong Wu is an English teacher from Guangxi University, Nanning, China. Currently, she is a PhD candidate in linguistics at the Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia. Her research interests include corpus linguistics, discourse analysis, and translation studies.

Kesumawati A. Bakar (Ph.D.) is a senior lecturer at the Center for Research in Language Studies and Linguistics, Universiti Kebangsaan Malaysia. Her research interests include multimodal discourse analysis, corpus linguistics, gender and identity studies.

Azhar Jaludin (Ph.D.) is a Senior Lecturer at the Center for Language and Linguistic Research, Faculty of Social Sciences and Humanities at Universiti Kebangsaan Malaysia. His research, publication and teaching interests include general linguistics, corpus linguistics, computational linguistics, and natural language processing.

Norsimah Mat Awal (Ph.D.) was an associate professor at the Centre for Research in Language Studies and Linguistics, Faculty of Social Sciences and Humanities, Univeristi Kebangsaan Malaysia. Her areas of interest include Translation, Pragmatics and Semantics.