http://www.icr.or.kr/apjcr e-ISSN 2733-8096 https://doi.org/10.22925/apjcr.2022.3.1.1

A Corpus-Based Longitudinal Study of Diction in Chinese and British News Reports on *Chang'e Project*

Lu, R., Xie, X., Qi, J., Ali, A. M., & Zhao, J. (2022). A corpus-based longitudinal study of diction in Chinese and British news reports on *Chang'e Project*. Asia Pacific Journal of Corpus Research, 3(1), 1-20.

As a milestone progression in China's space exploration history, *Chang'e Project* has attracted a lot of media attention since its first launching. This study aims to examine and compare the similarities and differences between the Chinese media and the British media in using nouns, verbs, and adjectives to report the *Chang'e Project*. After categorising the documents based on specific project phases, we created two diachronic corpora to explore the linguistic shifts and similarities and differences of diction employed by the Chinese and British media on the *Chang'e Project* ideology. This longitudinal study was performed with Lancsbox and the CLAWS web tagger through critical discourse analysis as the theoretical framework. The findings of the current study showed that the *Chang'e Project* coverage in both media increased on an annual basis, especially after 2019. In contrast to the objectivity and positivity in the Chinese Media, the British Media seemed to be more subjective with more appraisal adjectives in the news reports. Nonetheless, both countries were trying to be objective and formal in choosing nouns and verbs. Ideology-wise, the Chinese news media reports portrayed more positivity on domestic circumstances while the British counterpart was typically more critical. Notably, the study outcomes could catalyse future research on the *Chang'e Project* and facilitate diplomatic policies.

Keywords: Chang'e Project, Corpus-based Study, Diction, Longitudinal, News Discourse

1. Introduction

Moon exploration has always been appealing to scientists and the public around the world, owing to the fact that "the lunar farside is a virgin land and a unique scientific platform on which no humans or robots have ever landed" (Wang & Liu, 2016, p. 1). The *Chang'e Project* or Chinese Lunar Exploration Programme denotes "a global exploration of the resource, energy and special environment on the moon" (Zheng et al., 2008, p. 881). Chinese scientists have demonstrated much interest in lunar exploration since the early 1950s, as it reflects a country's comprehensive national strength (Zheng et al., 2008, p. 882). In 2007, *Chang'e-1* was successfully launched from the Xichang Launch Center. China went on to complete five lunar explorations called the *Chang'e Project* over the next 13 years. The *Chang'e Project* coverage extent in both Chinese and international media has increased yearly. The *Chang'e-5*, which was successfully launched on 24 November, 2020, implied the most intricate mission

within the Chinese space programme. This mission significantly facilitates scientific research on the origin and evolution of the moon as the first unmanned lunar sampling and return mission launched by China. In this vein, worldwide media coverage on *Chang'e-5* has reached a peak over the decades.

News reports denote a novel discourse analysis type that utilises particular news production and editing techniques to reflect and justify social phenomena in adherence to established rules (Hackett, 1984). Communicators' subjectivity and conventions from various aspects (politics, economy, and social culture) deter news from reflecting the truth. The values and ideologies underpinning text objectivity denote news dissemination goals and the actual meaning anticipated by the audience. Van Dijk (2013) strived to examine ideology from three dimensions (cognition, society, and discourse) as the scholar opined that discourse ideology demonstrates individual beliefs. The cognitive dimension primarily emphasised the psychological element of ideology, such as the essence of ideology as opinions and beliefs and its function in representing social sharing. Furthermore, van Dijk (2013) considered news reports as a discourse analysis (DA) form that essentially emphasises the structural analysis of news text, news discourse production, and the analysis of understanding. The DA aims to clearly and systematically describe the language use unit in line with van Dijk (2013). This description highlights two primary perspectives: textual and contextual. The former explains DA at multiple levels while the latter investigates the correlation between structural descriptions and various contextual attributes: people's cognitive comprehension and social and cultural elementsIdeology broadly highlights individuals' overall perspectives and opinions of world events while specifies a one-sided political element that differs from the ruling class ideology (Omotola, 2009). Both implicit and explicit ideology in modern communities, particularly the ideology underlying language, is used to enforce rights. Overt ideology implies constructing the cause and effect of an event while implicit ideology denotes the conditional outcome of re-exposing or shifting the ideology. The term 'implicit' could also establish ideology. Critical linguistics fundamentally outlines explicit and implicit ideology: the discourse representation-underpinning ideology relationship.

Multiple comparative counterparts were performed to examine diction and ideology through a corpus linguistics method despite the presence of space exploration studies. Liu and Zhang (2018) performed a corpus-assisted DA to highlight the health risks of air pollution in China among (i) Chinese and (ii) British and American English newspapers. Resultantly, English newspapers tended to amplify and dramatise the inevitability of health risks through discourse strategies: predicates, nomination, and quantitative speech. The English news in China was inclined to reflect health risk ambiguities through discourse strategies, such as specifying the Chinese background and complicating the causes of health concerns. Deng (2020) performed a corpus-based examination of keywords in news texts associated with the "Belt and Road" Project reported by the Chinese and American English news media. The author created two corpora and employed the corpus tool for a thorough consistency analysis to demonstrate the distinct attitudes of both nations towards the project. The Chinese news reports emphasised mutual benefits between the involved parties while the American counterparts focused on national (American) concerns regarding the development of China that potentially threatened its status and interests.

Yu et al. (2021) predicted that variations in COVID-19 reports potentially reflect ideological differences with explanatory power in COVID-19 development within different countries based on Fairclough's CDA model. Under this premise, the database analysis software was operated on the self-built corpus to analyse the news reports of different COVID-19 phases in China and the UK and highlight both media variances. The China Daily was found to employ more objective and neutral words in describing COVID-19 and confronting the pandemic on a positive note. Contrarily, the Guardian utilised more negative and less restrictive words to highlight the COVID-19 outbreak and report on pandemic control and prevention policies. The British media description of COVID-19 proved more objective and neutral compared to discourses pre- and post-lockdown where more restrictive and socially-conflicting expressions were used. Likewise, a substantial increase in discourses on

experience-sharing and cooperation was identified in the China Daily.

Several articles emphasised the discourse-ideology relationship in terms of diction. As one of the most prominent field experts, Fairclough (2013) demonstrated the essentiality of CDA and how substantial socio-economic changes impact language and power relations. Notwithstanding, studies to investigate such reports from linguistic perspectives or comparative corpora remained lacking. This study (i) gathered relevant reports on the *Chang'e Project* in mainstream China and Britain media, (ii) categorised the documents based on specific project phases, and (iii) explored the linguistic shifts and similarities and differences employed by the Chinese and British media on the *Chang'e Project* ideology. Essentially, the study sought to address the following questions:

- (1) What changes have taken place in reporting the *Chang'e Project* by the Chinese and British media over time?
- (2) What are the similarities and differences between the Chinese and British media in choosing nouns, verbs, and adjectives to report the *Chang'e Project*?
- (3) What ideologies do the mainstream media in both countries reflect in their diction?

2. Data and Methodology

The present study was framed with critical discourse analysis (CDA), an approach to the study of discourse that perceives language as a social practice. Various approaches were identified to analyse and bridge social macro and micro gaps for a unified critical analysis as follows: member-group, action-process, context-social structure, and individual and social cognition. Social power denotes the central focus of most critical works. For example, groups that could relatively control their members' behaviours and thoughts possessed power. The premise of this ability reflects a power base following the attainment of scarce social resources: force, money, status, fame, knowledge, information, culture, or multiple public discourse and communication types (Mayr, 2008).

Ideology encompasses one or more control categories over communication. For example, journalists and news agencies control if and when news is published or reported. The issue of who owns the topic (semantic macrostructure) is crucial in terms of textual control and dialogue structure, (van Dijk, 2013). The local details of (i) vocabulary and syntactic styles, (ii) suggested connotations, (iii) order, and (iv) rhetorical and narrative structures rely on authoritative team members apart from other discourse counterparts despite various contextual and topical discourse controls. Summarily, multiple levels and contextual, textual, and dialogue-oriented contexts could be relatively controlled by powerful speakers and institutions. The same element of power could be abused at the expense of particular recipients, groups, and community members.

Powerful group leaders could control specific information and opinions and the general knowledge, attitudes, and ideologies highlighted in the subjective mental model of specific recipients. Such recipients who are defined as authoritative or trustworthy sources are inclined to accept personal or institutional beliefs, knowledge, and opinions in the context model unless the aspects are inconsistent with scholars', experts', professionals', or media beliefs and experiences (Nesler et al., 1993).

2.1. Research Corpora

The current study involved media reports in China and Britain on China's *Chang'e Project*. News reports from 24 October 2007 to 31 December 2020 by the mainstream media in China and Britain were collected as corpora with 'Chang'e' as the key search term. The two study corpora are presented as follows:

- (1) Chinese news corpus based on the GOC, Global Times, and China Daily with a total of 561 reports and 217, 051 words
 - (https://drive.google.com/file/d/10Q_4Vn1_6n140_qDknQ0MCx2bDvmhHB/view?usp=sharing);
- (2) British news corpus based on *The Sun, The Daily Express, Independent, The Guardian*, and *Daily Mail* with a total of 397 reports and 292, 520 words (https://drive.google.com/file/d/1SO5_6rIQ92jaZ2zi8KUpHrmEQ2_aeyLU/view?usp=sharing).

The news reports gathered from the aforementioned media were categorised into five small corpora following the specific *Chang'e Project* period.

2.2. Research Methods

This study primarily performed corpus analysis based on diachronic study of high-frequency words and concordances. Reports that were not highly pertinent to the *Chang'e Project* were manually omitted at the preliminary study stage. Notably, part of speech (POS) tagging was implemented to annotate the POS of all the words for a sound understanding of word use. The high-frequency nouns, verbs, and adjectives employed in different *Chang'e Project* report periods within China and Britain were longitudinally examined. Similar and distinct words utilised in both media within the five stages were elicited from the aforementioned wordlists for vertical analysis and horizontal comparison. Finally, the similarities and differences of the diction in both nations were compared and contrasted.

The corpus-based method utilises a corpus for hypothetical or theoretical explorations while the corpus-driven approach claims the corpus itself to be the sole source of hypothesis concerning language in line with Tognini-Bonelli (2001). Regardless, the dichotomy was unsupported by McEnery and Hardie (2012) given the absence of a clear cut-off between corpus-based and corpus-driven research. Specifically, "all corpus linguistics can justly be described as corpus-based" (McEnery & Hardie, 2012, p. 6). Three basic tools are extensively-utilised for the corpus-based technique: frequency analysis, collocation analysis, and concordance lines. Frequency analysis is the most rudimentary instrument involving the quantitative corpus-based method components. The tool originated from statistical theory and computer technology. Two frequency types are distinguished: absolute frequency (AF) and relative frequency (RF). The corpus size could impact AF as AF indicates the actual count of specific word occurrences within a corpus.

As another significant instrument underlying the core study method, collocation analysis is deemed necessary to present the definition of collocation and its subsequent expression in corpus linguistics or other language units excluded from the word (McEnery & Hardie, 2012). Collocates are typically generated by statistical association measures in corpus linguistic practices. Collocation denotes the relationship between a node word or 'an item whose total pattern of co-occurrence with other words is under examination' and its collocate or 'any one of the items which appear with the node within a specified span' within a particular span (Sinclair et al., 2004, p. 10).

A corpus-based study was performed in the current study with Lancsbox 5.1.2 and the free CLAWS web tagger based on the corpora above. Lancsbox 5.1.2 denotes a novel language data and corpus analysis software package developed by Lancaster University. The statistical function in Lancsbox is incorporated into common functions that could be adjusted for user needs and language data visualisation. Meanwhile, the CLAWS web tagger is a free instrument that could tag the POS of textual words. A total of ten small wordlists were established with Lancsbox parallel to period divisions while the free CLAWS web tagger was employed to identify the POS of one particular word in the wordlist.

3. Results and Discussion

The study outcomes were presented following the news report diction on the *Chang'e Project* parallel to the two corpora and word frequency lists based on the analytical structure justified in the previous section. Predictably, the similarities and variances could be identified in the diction of Chinese and British media. The corpus-based approach was used to examine the association of word usage with ideology in different nations.

3.1. Diachronic Changes in the Number of News Reports on Chang'e Project

Table 1. Data on Five Periods

Periods	No. of Years/Months	Dates	Main Coverage
Period 1	3 years	24 October, 2007 – 30 September, 2010	Chang'e-1
Period 2	3 years	1 October, 2010 - December 1, 2013	Chang'e-2
Period 3	5 years	December 2, 2013 - December 7, 2018	Chang'e-3
Period 4	2 years	December 8, 2018 - November 23, 2020	Chang'e-4, Chang'e-5
Period 5	1 month	November 24, 2020 - December 31, 2020	Chang'e-5

This longitudinal study on *Chang'e Project* news reports encompassed five periods. The division of the five periods was primarily based on the launching time of each *Chang'e Project*. The first period ran from 2007 to 2010 when *Chang'e-1* was mainly reported by the media. Period 2 spanned three years from 2010 to 2013, and period 3 lasted for five years due to the longer interval between the launch time of Chang'e-3 and Chang'e-4. Period 4 began in 2018 and ended on 23 November 2020, when Chang'e-5 was launched on the next day. Although the period divisions generally paralleled the timeframe of every *Chang'e Project* launched, each period did not solely involve one project. For example, the first period ranging between October 2007 and September 2010 primarily concerned the *Chang'e-1 Project*. Nevertheless, news on *Chang'e-2*, *Chang'e-3*, *Chang'e-4*, and even *Chang'e-5* were occasionally reported during Period 1. Period 5, which ranged between 24 November 2020 and 31 December 2020, was the shortest of the five periods as this study was performed in late 2020.

The original corpora of Chinese and British news reports encompassed all the reports with the keyword 'Chang'e'. Manual processing was duly performed as some reports were not significantly associated with this study, excluding reports that (i) proved irrelevant to the *Chang'e Project* with titles that did not explicitly include *Chang'e-1* to *Chang'e-5* and (i) contained similar information were omitted from the study.

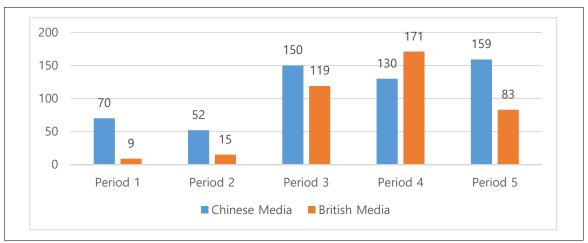


Figure 1. Number of News Reports on Chang'e Project in Each Period

Figure 1 demonstrates alterations in the number of news reports on the *Chang'e Project* between 2007 and 2020. Vertically, the number indicated the China and Britain media coverage while the different *Chang'e Project* periods were categorised horizontally. In this vein, the Chinese media placed more emphasis on the *Chang'e Project* compared to the British counterpart during the initial moon exploration period in China. The overall number for Periods 1 and 2 in the British media (24) significantly contrasted with the overall number of the same periods in the Chinese counterpart (122).

Period 3 reflected a drastic increase in the number of news reports within the Chinese and British media. While the number of Chinese media coverage on the *Chang'e Project* in Period 3 was twice more than that in Period 1, the British media coverage was approximately thirteen times more since 2013. The number of news reports in both China and Britain grew steadily post-2018. The British media demonstrated the most extensive media coverage in Period 4 while the Chinese counterpart reflected a decline in the same period following the Chinese corpus size cut in line with Figure 2.

The overall number of media coverage in both nations was not low, specifically in China, although the final period only involved news reports of approximately one month. Conversely, the number of media coverage between 24 November 2020 and 31 December 2020 proved higher than the other four periods in China due to their citizens' enthusiasm for the successful launching of *Chang'e-5*. The general trend in the number of *Chang'e Project* news reports within the five periods reflected a steady rise. A sharp increase was indicated in the number of news reports within the third period. The highest number of media coverage appeared in both nations within the fourth and fifth periods.

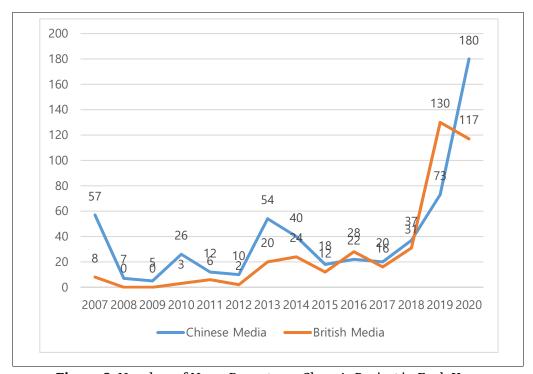


Figure 2. Number of News Reports on Chang'e Project in Each Year

Figure 2 clearly demonstrates the diachronic shift of news reports by the Chinese and British media between 2007 and 2020. The blue graph line represents China while the orange counterpart indicates Britain. Specifically, both media have witnessed fluctuations over the past thirteen years. The peaks of every graph line are approximately interconnected with each *Chang'e Project's* launch time. For example, the Chinese media graph line in 2007, 2010, 2013, 2018, and 2020 indicated five peaks with each representing the year *Chang'e-1* to *Chang'e-5* were launched.

The number of news reports from the British media remained level between 2007 and 2017, drastically increased in 2018, and attained a peak in 2019 (see Figure 2). Meanwhile, the highest number of Chinese media reports appeared in 2020 when *Chang'e-5* was launched. A drastic increase

was identified in the number of *Chang'e Project* news reports in both media over thirteen years. Turning points appeared in 2019 and 2020 when the news reports primarily involved *Chang'e-5* and China as a substantial global power in scientific research development, specifically aerospace exploration, despite signs of levelling off.

3.2. Analysis of High-frequency Words in Chinese and British Media

Notably, word frequency lists parallel to the POS tagging by the CLAWS WWW tagger through tagset C7 is presented in Table 2 below. The commonly-listed English POS includes nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions, interjections, numerals, articles, or determiners. This study aimed to analyse the diction of three POS (nouns, verbs, and adjectives) as they denote content words with semantic values. The lexical category was preferred following a high percentage in the corpora given that most of the words encompassed multiple POS in the wordlist. For example, the word 'launch' was utilised as a noun and a verb 155 times and 27 times, respectively, in the Period 1 Chinese news report corpus. The word was deemed to be a noun in the corpus. Other words with multiple POS were similarly evaluated. Such words that reflected a considerably even percentage in the corpora were excluded from the wordlist.

Table 2. High-frequency	Nouns in Chinese News Reports
-------------------------	-------------------------------

Period 1	Period 2	Period 3	Period 4	Period 5
Moon	Moon	Moon	Space	Space
Satellite	China	China	China	China
Space	Chang'e-2	Probe	Moon	Moon
Chang'e	Probe	Space	Probe	Chang'e
Probe	Space	Chang'e-3	Rover	Samples
Orbit	Orbit	Mission	Chang'e	Mission
Launch	Launch	Launch	Side	Probe
China	Mission	Earth	Exploration	Orbit
Center	Chang'e-3	Technology	Mission	Capsule
Orbiter	Satellite	Lander	Earth	Spacecraft

Table 3. High-frequency Nouns in British News Reports

Period 1	Period 2	Period 3	Period 4	Period 5
Space	Space	Space	Moon	Space
Moon	China	Moon	Space	Moon
China	Moon	China	Side	China
Probe	Launch	Mission	China	Samples
Programme	Probe	Rover	Mission	Missions
Launch	Mission	Earth	Earth	Earth
Chang'e	Station	Surface	Probe	Probe
Mission	Programme	Station	Surface	Surface
Technology	Surface	Probe	Rover	Spacecraft
Years	Missions	Side	Chang'e-4	Chang'e-5

A noun functions as "a word (other than a pronoun) used to identify any of a class of people, places, or things, or to name a particular of these" (Stevenson, 2010, p. 979). Table 2 outlines a list of high-frequency nouns in Chinese and British news reports within the five aforementioned periods. The wordlist includes nouns that encompass content meaning with high analytical values. The high-frequency nouns within both Chinese and British news reports proved highly similar regarding moon exploration and the China *Chang'e Project* in Tables 2 and 3. The top five high-frequency nouns in the Chinese media are presented as follows: 'moon', 'space', 'China', 'probe', and 'mission'. The top five high-frequency nouns out of the ten nouns in British media are presented as follows: 'space', 'moon', 'China', 'mission', 'Earth', and 'probe'. This similarity

of high-frequency nouns in both media justified the objectivity of the news report, which has been widely accepted as a "basic principle of news reporting" (Akpan et al., 2012, p. 713). As Stovall (2005, as cited in Akpan et al., 2012) pointed out, "the expectation of news organization on their journalists is that they will gather and handle information objectively" (p. 712).

Table 4. High-frequency Verbs in Chinese News Reports

Period 1	Period 2	Period 3	Period 4	Period 5
Said	Said	Said	Said	Said
Expected	Launched	Launched	Launched	Launched
Transfer	Expected	Land	Made	Return
Completed	Test	Return	Work	Made
Launched	Conduct	Made	Bring	Collected
Working	Taken	Sent	Make	Work
Named	Made	Expected	Taken	Completed
Sent	Scheduled	Entered	Shows	Expected
Entered	Entered	Carry	Expected	Landed
Carried	Carry	Landed	Conduct	Conduct

Table 5. High-frequency Verbs in British News Reports

Period 1	Period 2	Period 3	Period 4	Period 5
Said	Said	Said	Said	Said
Launched	Launched	Plans	Launched	Return
Put	Plans	Launched	Landed	Launched
Plans	Docking	Land	Land	Landed
Named	Making	Send	Made	Land
Given	Build	Sent	Taken	Bring
Return	Consent	Landed	Study	Collect
Announced	Read	Return	Released	Crewed
Claims	Send	Carry	Named	Carrying
Made	Followed	Crewed	Pictured	Collected

A verb conveys an action, occurrence, or state of being. The wordlist in Tables 4 and 5 present high-frequency verbs in Chinese and British news reports, respectively. Additionally, the verbs encompass high semantic and analytical values. Several words, such as 'updated' and 'mailing' were excluded from the wordlist despite demonstrating a relatively high frequency as they were utilised in issue dates or at the end of news reports. The annotated corpora were accessed to confirm its POS if the lexical word category proved hard to identify.

The highest-frequency verb within both Chinese and British media over the five periods was 'said', an indication of the typical linguistic features of news reports (see Tables 4 and 5). In other words, news report language was more formalised with reported speech for paraphrasing purposes. Other high-frequency verbs, such as 'launched', 'plans', 'entered', 'completed', 'carry', 'collected', and 'return' also indicated linguistic formality and objectivity in news reports. Very few words generally reflected opinions and emotions as news reports primarily aimed to inform the public of current events, which echoes the opinion that "journalists value fairness in their quest for the truth" (Canella, 2021, p. 2). It is, therefore, an ideal practice for news reports to attain objectivity from every aspect.

Table 6. High-frequency Adjectives in Chinese News Reports

Period 1	Period 2	Period 3	Period 4	Period 5
Lunar	Lunar	Lunar	Lunar	Lunar
Chinese	Chinese	Chinese	Far	National
Long	National	Scientific	National	Chinese
Scientific	Chief	National	Chinese	Long
Chief	Scientific	Soviet	Scientific	Scientific

Orbital	Unmanned	Long	New	Manned
National	Long	New	Long	New
Manned	New	Soft	Solar	Early
International	Full	Successful	International	International
Circumlunar	Manned	International	Soft	Robotic

Table 7. High-frequency Adjectives in British News Reports

Period 1	Period 2	Period 3	Period 4	Period 5
Lunar	Lunar	Lunar	Lunar	Lunar
Chinese	Chinese	Chinese	Chinese	Chinese
New	Manned	Far	Dark	National
Long	Unmanned	Long	National	Soviet
National	International	Manned	Solar	Solar
Official	Future	National	New	Long
Fake	Military	Official	Near	International
Asian	National	New	Military	Latest
Patriotic	Huge	Soviet	Successful	Ambitious
Manned	Ambitious	Dark	International	Successful

An adjective is one of the primary POS in English despite its historical classification with nouns (Trask, 1993). In linguistics, an adjective modifies a noun or noun phrase and describes its referent with a semantic component to alter the information provided by a noun. Adjectives may be semantically inclined in news report language based on the context. several adjectives reflected semantic preferences, such as 'successful', 'new', 'patriotic', 'ambitious', 'fake', 'military', and 'dark' based on the high-frequency adjective list in Tables 6 and 7. The British media seemed to be more subjective in this respect by employing more adjectives for evaluating functions compared to the objective description of Chinese news reports. Only two adjectives, 'successful' and 'new', may have appraisal value in the Chinese media while all the aforementioned adjectives in British media demonstrated semantic preferences. This corroborates the ideas that "news objectivity, which is the basic principle of news reporting, is not really taken into consideration in most online contents of the news stories" (Akpan et al., 2012, p. 713).

Most of the adjectives employed in Chinese and British news reports were associated with the China's *Chang'e Project* as news report language remained objective rather than subjective. For example, 'lunar', 'Chinese', 'manned', 'unmanned', 'long', 'national', 'international', and 'scientific' were primarily utilised to describe the China moon probe without opinion preferences.

Overall, the study data revealed the news report language in both nations to be typically objective and formal in summarising the use of nouns, verbs, and adjectives within Chinese and British news reports. Several words contained appraisal meaning regardless of the POS (noun, verb, or adjective). Several adjectives in Chinese and British media encompassed evaluating functions albeit with high semantic preferences to add more information on noun modification. The British media appeared more subjective with high appraisal adjectives in news reports (albeit with several negative adjectives) compared to the Chinese media's objectivity and positivity. This accords with earlier observations of Guanah and Obi (2020) that affiliations of media would influence their objectivity and government-owned media tend to have bias in news reporting in pursuing their own interests.

3.3. Analysis of same High-frequency Words in Chinese and British Media

Table 8. Same High-frequency Words from Each Period in Chinese Media

Number of Periods	Same Words	Periods
Same words appear in 5 periods	N: Moon, Space, Probe, China V: Said, Expected, Launched Adj: Lunar, Chinese, Long, Scientific, National	Period 1, 2, 3, 4, 5

Same words appear in 4 periods	Adj: International	Period 1, 3, 4, 5
	N: Mission; V: Made; Adj: New	Period 2, 3, 4, 5
Same words appear in 3 periods	V: Launch, Entered	Period 1, 2, 3
	N: Orbit; Adj: Manned	Period 1, 2, 5
	N: Chang'e	Period 1, 4, 5
	V: Conduct	Period 2, 4, 5
Same words appear in 2 periods	N: Satellite; Adj: Chief	Period 1, 2
	V: Sent	Period 1, 3
	V: Completed	Period 1, 5
	N: Chang'e-3; V: Carry	Period 2, 3
	V: Taken	Period 2, 4
	N: Earth; Adj: Soft	Period 3, 4
	V: Return, Landed	Period 3, 5
	V: Work	Period 4, 5

Table 9. Same High-frequency Words from Each Period in British Media

Number of Periods	Same Words	Periods
Same words appear in 5 periods	N: Space, Moon, China, Probe; Period 1, 2, 3, 4, V: Said, Launched; Adj: Lunar, Chinese, National	
Same words appear in 4 periods	N: Mission N: Surface	Period 1, 2, 3, 4 Period 2, 3, 4, 5
Same words appear in 3 periods	N: Plans; V: Manned Adj: New V: Return; Adj: Long Adj: International N: Earth; V: Land, Landed	Period 1, 2, 3 Period 1, 3, 4 Period 1, 3, 5 Period 2, 4, 5 Period 3, 4, 5
Same words appear in 2 periods	N: Programme; V: Launch Adj: Official V: Named, Made N: Station; V: Send Adj: Military N: Missions; Adj: Ambitious N: Rover, Side; Adj: Dark V: Crewed; Adj: Soviet Adj: Solar, Successful	Period 1, 2 Period 1, 3 Period 1, 4 Period 2, 3 Period 2, 4 Period 2, 5 Period 3, 4 Period 3, 5 Period 4, 5

Tables 8 and 9 present the same high-frequency words utilised in every period within Chinese and British news reports. Most of the same words in Chinese media were significantly related to the entire *Chang'e Project* process: 'moon', 'China', 'space', 'probe', 'launched', and 'lunar'. The previously-mentioned word 'said' reflected the linguistic features of news reports among the same words in five periods. Meanwhile, the word 'expected' indicated the Chinese media to highly anticipate and prioritise this project. The same words appearing in four other periods also demonstrated the substantial impact of the project through the word 'international'. The word 'Chang'e' in Periods 1, 4, and 5 revealed the Chinese media to pay more attention to the *Chang'e Project* compared to the specific number of projects. The 'Chang'e-3' appeared in both wordlists within Periods 2 and 3, thus highlighting the Chinese media focus on the *Chang'e-3 Project* despite being in its infancy.

Most of the same words employed within the five periods were connected with the entire *Chang'e Project* process and its global impact in Chinese and British media. Regardless, the words 'military' in Periods 2 and 4 and 'ambitious' in Periods 2 and 5 indicated the fear experienced by the British media as to whether (i) the scientific development in China would be globally beneficial or (ii) China would undertake military measures to attain technological goals following the launch of *Chang'e-2*. The word 'Soviet' also highlighted the position of China in moon exploration as the fourth country to land on the moon after the United States, Soviet Union, and Japan.

In this vein, the same high-frequency words employed in both Chinese and British news reports were inextricably linked to the entire *Chang'e Project* process, such as 'launching', 'landing', and 'returning'. Both media demonstrated variances in adjective usage. All the adjectives employed in Chinese media were positive while the British counterpart employed multiple positive and negative adjectives. Such negativity reflected a 'Chinese threat' that could adversely depict the development of China. A list of the same high-frequency words from the five stages in Chinese and British media could be derived by integrating Tables 8 and 9 (see Table 10).

Table 10. The Same High-frequency Words in the Five Periods of Chinese and British Media

POS	Total Numbers
Noun	Space, Moon, China, Probe
Verb	Said, Launched
Adjective	Lunar, Chinese, National

Bar charts were developed following the normalised frequency of specific words to demonstrate word comparisons that were divided into different POS.

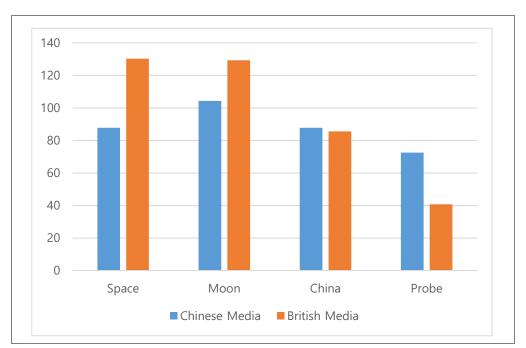


Figure 3. Same High-frequency Nouns in the Five Periods of Chinese and British Media

Given that the Chinese media frequently utilised 'China' and 'probe' while the British counterpart often employed 'space' and 'moon' (see Figure 3), the Chinese media emphasised the *Chang'e Project* itself rather than merely covering the event.

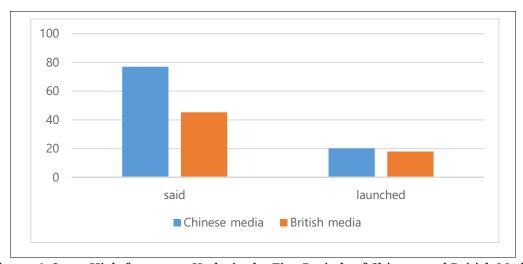


Figure 4. Same High-frequency Verbs in the Five Periods of Chinese and British Media

The Chinese media indicated a high frequency of shared verbs in Figure 4. The Chinese media discourse paralleled the authenticity of news reports as the use of 'said' implied the Chinese media preference to quote organisational or individual speeches. For example, the word 'launched' denoted that China placed more focus on the lunar exploration process compared to Britain.

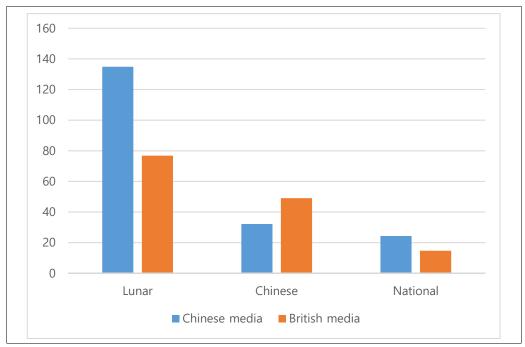


Figure 5. Same High-frequency Adjectives in the Five Periods of Chinese and British Media

The Chinese media tended to utilise 'lunar' than the British counterpart (see Figure 5), the Chinese media preferred to employ the adjective 'lunar' while the British counterpart was inclined to utilise the noun 'moon' in describing the Lunar Exploration Project (see Figure 3). The Chinese media tended to utilise the word 'national' in place of 'Chinese' as *Chang'e* is a Chinese project., The Chinese media proved more realistic and relevant to the *Chang'e Project* while the British counterpart demonstrated a more general coverage based on the same high-frequency words in Chinese and British media within the five periods. These results show that both Chinese and British media tend to be objective in most cases, as objectivity in the news report is "one of the pillars of the journalism profession" (Guanah & Obi, 2020, p. 277).

3.4. Analysis of Concordances in British and Chinese Corpora

A concordance denotes "a collection of the occurrences of a word-form, each in its own textual environment" (Sinclair, 1991, p. 32) and indicates how a word is contextually employed. Concordancer tools effectively explore a corpus. Four generations of concordances were identified following McEnery and Hardie (2012). Lancsbox, a new-generation corpus analysis instrument developed for powerful searches, was utilised to perceive how the word 'Chang'e' co-occurred with other counterparts following the current research focus on the *Chang'e Project*. Meanwhile, Key Word in Context (KWIC) and GraphColl could be employed to demonstrate and measure concordances in corpora analysis.

3.4.1. The KWIC of 'Chang'e' in British and Chinese Corpora

The KWIC enables the perception of how the current study keyword, 'Chang'e', is utilised in British and Chinese media.

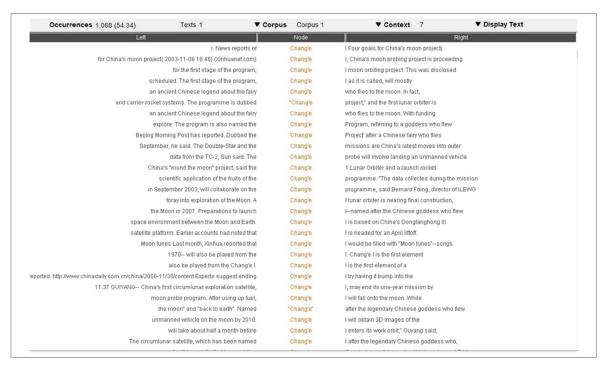


Figure 6. The KWIC of 'Chang'e' in Chinese News Corpus



Figure 7. The KWIC of 'Chang'e' in British News Corpus

The word 'Chang'e' occurred 1088 times in the Chinese news corpus and 521 times in the British counterpart. The normalised frequencies of 'Chang'e' in both corpora were computed to precisely reflect the relative frequencies in each corpus (see Figure 8).

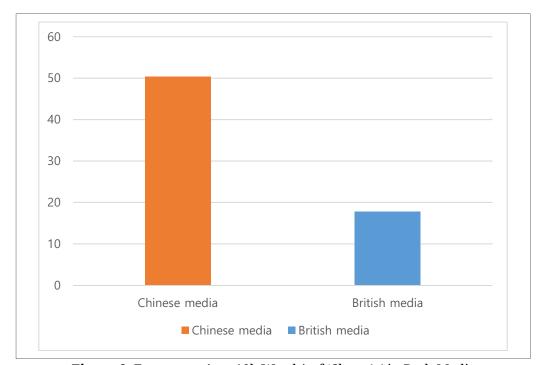


Figure 8. Frequency (per 10k Words) of 'Chang'e' in Both Media

Figure 8 demonstrates the occurrences per 10k tokens of the word 'Chang'e' in Chinese and British media. Specifically, the Chinese media were more prepared to employ this word in pertinent reports. For example, the Chinese media paid more attention to the *Chang'e Project* than the British counterpart. Figures 7 and 8 indicate that the most commonly-collocated words in both Chinese and British news corpora for 'Chang'e' were articles and numbers. The most frequently-occurred words on the left of the word 'Chang'e' were 'the', 'of', and 'by', whereas its collocations on the right were numbers

resembling '1, 2, 3, 4, and 5' to (i) demonstrate the *Chang'e Project* phases and words resembling 'mission' and 'project' and (ii) portray its nature.

The word 'Chang'e' was commonly integrated with definite or indefinite articles as a noun. The word was also extensively complemented with numbers as the project involved a moon exploration project launched at specific time slots. The KWIC is ideal to perceive word usage for multiple examples of word usage in authentic settings. The high number of collocates require much manual work although KWIC is an optimal medium to perform corpus-driven research. GraphColl was more directly and clearly employed in the following part to demonstrate 'Chang'e' collocates as it was challenging to explicitly observe the collocation network.

3.4.2. GraphColl of 'Chang'e' in British and Chinese Corpora

GraphColl in LancxBox is a unique means of explicitly and straightforwardly presenting concordance with graphs to perceive word collocations by displaying its collocation network. The following diagram in Figure 9 appears dense with a threshold of 5, thus implying 'Chang'e' to have a high frequency of collocates. The statistical threshold value was set to 20 with search results to eliminate the clutter (see Figure 9).

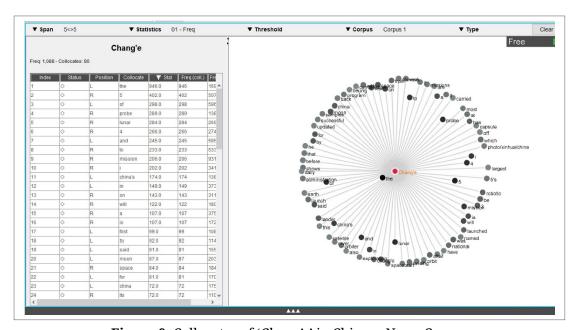


Figure 9. Collocates of 'Chang'e' in Chinese News Corpus

In Figure 9, the left side indicated the 'Chang'e' frequency and collocates. The search word occurred 1088 times in the Chinese news corpus with 80 collocates and a span value of 5. The first 10 collocates on the left side of 'Chang'e' were 'the', 'of', 'and', 'China's', 'in', 'first', 'by', 'said', 'moon', and 'for' while the collocates on its right in the Chinese news corpus were '5', 'probe', 'lunar', '4', 'to', 'mission', 'I', 'on', 'will', and 'a'. The right side indicated the diagram of the node word 'Chang'e' collocates in Figure 10. Words that came close to the diagram centre reflected a high frequency of its node word collocation. The collocation network was explicitly displayed in Figure 10.

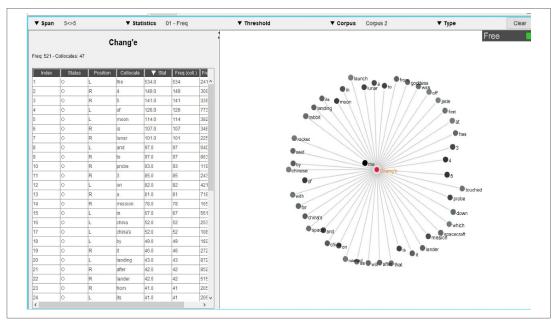


Figure 10. Collocates of 'Chang'e' in British News Corpus

The 'Chang'e' collocates in the British counterparts seemed sparse due to its frequency in the British news corpus (521) compared to the GraphColl of 'Chang'e' in the Chinese news corpus: approximately twice lesser than the Chinese counterpart. The entire 'Chang'e' collocates in the British news corpus were 47. The first two most frequently-occurred words on the left side of 'Chang'e' were precisely the same as that of the Chinese counterpart: 'the' and 'of'. The remaining eight words from the top 10 collocates on the left side of the node word are as follows: 'moon', 'and', 'on', 'in', 'China', 'China's', 'by', and 'landing'. The top 10 words that frequently occurred together with it on the right of 'Chang'e' were '4', '5', 'is', 'lunar', 'to', 'probe', '3', 'a', 'mission', and 'it'.

Table 11. Top 10 Collocates of 'Chang'e' in British and Chinese News Corpora

-	G	-
	British News Corpus	Chinese News Corpus
Left	the	the
	of	of
	moon	and
	and	China's
	on	in
	in	first
	China	by
	China's	said
	by	moon
	landing	for
	4	5
	5	probe
Right	is	lunar
	lunar	4
	to	to
	probe	mission
	3	i
	a	on
	mission	will
	it	a

Based on Table 11, the 'Chang'e' collocates in the British and Chinese news corpora were highly similar on the left and right. Seven words occurred in both news corpora either on the left or right of

'Chang'e'. The same collocates were 'the', 'of', 'moon', 'and', 'in', 'China's', and 'by' on the left of the 'Chang'e' node word and '5', '4', 'probe', 'to', 'lunar', 'mission', and 'a' on its right side. As such, the *Chang'e-5* and *Chang'e-4 Projects* garnered more attention compared to other project periods.

Content words involving 'moon', 'probe', 'lunar', 'mission', and 'China's' implied the *Chang'e Project* name and nature while functional words involving 'the', 'of', 'to', 'by', 'a,' and 'in' were employed to express the correlation between 'Chang'e'-oriented notions. Summarily, the top 10 'Chang'e' collocates in both the Chinese and British news corpora were mostly similar. The corpora either reflected the correlation of 'Chang'e' to other words or demonstrated the project properties and operation.

Taken together, the analysis of concordances in British and Chinese media showed that similarities outweighed differences in the collocates of the word 'Chang'e'. First, the *Chang'e Project* was associated with 'moon' or 'lunar' explorations which aimed at the China *Chang'e* 'probe' to land on the moon. Second, *Chang'e-4* and 5 garnered both Chinese and British media attention among all the five project periods.

4. Conclusions

This research set out to carry out a corpus-based longitudinal study of diction in Chinese and British news reports on China's *Chang'e Project*. In this investigation, we examined how word usage was associated with different country ideologies by combining CDA and corpus linguistics methods. Resultantly, sorting, induction, and analysis of real corpus revealed a gradual increase in the amount of coverage on the *Chang'e Project* in both Chinese and British media, specifically post-2019. Both nations strived to be objective and formal in noun and verb selection although several evaluative words were identified through adjectives. More adjectives reflecting praise or criticism were found in the British media compared to positive reviews in the Chinese counterpart. Notably, the use of adjectives was gradually turning positive. Furthermore, investigation into concordances of 'Chang'e' in both corpora revealed similarities of Chinese and British media in using the word.

However, the Chinese media reflected more closeness to the entire *Chang'e Project* process in terms of the high-frequency words, whereas the British media employed more positive and negative evaluative vocabulary from the viewpoint of novel keywords among high-frequency words. The utilisation demonstrated that the British media took a critical attitude compared to the praising of the Chinese media. Both the Chinese and British media placed much emphasis on the project nature and operation from the 'Chang'e' collocation analysis. The *Chang'e-4* and *Chang'e-5* gained more popularity among the Chinese and British media. Lastly, both country ideologies could be inferred through the aforementioned words. The Chinese news media were prone to 'report the good but not the bad' on domestic circumstances and seldom publicised events that degraded the nation or government. Although the British news media were typically more critical with implications of the global threats posed by China, their stance reflected a pro-China attitude.

Since the present study only collected data from two countries in the world, it was not possible to compare and contrast the diction in other countries' news reports on China's *Chang'e Project*. Further research in the comparison of *Chang'e Project* corpora across different countries would be of great help in examining the diction and ideology reflected in the news discourse. Another weakness of this study was the paucity of in-depth analysis of diction on sentence levels. Further investigation into sentence-based examples from the corpora is strongly recommended. Notwithstanding these limitations, this study pioneered the use of space exploration as research subject to analyse the diction-ideology relationship with a corpus method and contribute to the corpus linguistics field. The reporting time for each *Chang'e Project* period was relatively concentrated although the corpus was diachronically gathered. As such, the ideology of some periods might be exaggerated or hidden following particular historical occurrences. Overall, this study could enable readers to comprehend

different news media attitudes towards the same project and facilitate critical thinking when reading different news reports on the same subject.

Acknowledgments

This study is part of a research program *School-level Scientific Research Project of Tianshui Normal University*. Research No. CXJ2021-13.

References

- Akpan, C. S., Ifeanyi, A. L., Martin, O. P., Alexander, O. C., & Uchenna, A. (2012). Rethinking objectivity in news reporting in the digital age. *Journal of Alternative Perspectives in the Social Sciences*, 4(4), 711-729.
- Canella, G. (2021). Journalistic power: Constructing the "Truth" and the economics of objectivity. *Journalism Practice*, DOI: 10.1080/17512786.2021.1914708.
- Deng, F. (2020). Computerized corpus keyword approaches to evaluation: A case study of evaluative attitudes of "the Belt and Road" reports in mainstream media of China and America, *International Journal of Electrical Engineering & Education*, DOI: 10.1177/0020720920923303.
- Fairclough, N. (2013). Language and power 2000. In Candlin, C. N. (Ed.), *Language and Power* (2nd ed.) (pp. 203-218). London: Routledge.
- Guanah, J. S., & Obi, I. (2020). Newspaper reportage and objectivity: An examination of newspaper owners' affiliation to political parties. *Edo Journal of Arts Management and Social Science*, 2(1), 275-293.
- Hackett, R. A. (1984). Decline of a paradigm? Bias and objectivity in news media studies. *Critical Studies in Mass Communication*, 1(3), 229-259.
- Liu, M., & Zhang, Y. (2018). Discursive constructions of scientific (Un) certainty about the health risks of China's air pollution: A corpus-assisted discourse study. *Language & Communication*, 60, 1-10.
- Mayr, A. (2008). Introduction: Power, discourse and institutions. In Mayr, A. (Ed.), *Language and Power:* An Introduction to Institutional Discourse (pp. 1-25). London: Continuum.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Nesler, M. S., Aguinis, H., Quigley, B. M., & Tedeschi, J. T. (1993). The effect of credibility on perceived power. *Journal of Applied Social Psychology*, *23*(17), 1407-1425.
- Omotola, J. S. (2009). Nigerian parties and political ideology. *Journal of Alternative Perspectives in the Social Sciences*, 1(3), 612-634.
- Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford: Oxford University Press.
- Sinclair, J. (2004). Current issues in corpus linguistics. In J. Sinclair & Carter, R. (Eds.), *Trust the Text: Language, Corpus and Discourse* (pp. 185-193). London: Routledge.
- Stevenson, A. (Ed.). (2010). Oxford Dictionary of English. Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001). Corpus Linguistics at Work. Amsterdam: John Benjamins.
- Trask, R. L. (1993). A Dictionary of Grammatical Terms in Linguistics. London: Routledge.
- van Dijk, T. A. (2013). *News Analysis: Case Studies of International and National News in the Press.* New York: Routledge.
- Wang, Q., & Liu, J. (2016). A Chang'e-4 mission concept and vision of future Chinese lunar exploration activities. *Acta Astronautica*, 127, 678-683.
- Yu, H., Lu, H., & Hu, J. (2021). A corpus-based critical discourse analysis of news reports on the COVID-

19 pandemic in China and the UK. *International Journal of English Linguistics*, 11(2), 36-45. Zheng, Y., Ouyang, Z., Li, C., Liu, J., & Zou, Y. (2008). China's lunar exploration program: Present and future. *Planetary and Space Science*, 56(7), 881-886.

THE AUTHORS

Rong Lu is a PhD candidate at Cross Department, Faculty of Modern Languages and Communication, Universiti Putra Malaysia. She is also working as a Lecturer in Tianshui Normal University, China. She got her master's degree in linguistics and applied linguistics in Shanghai Maritime University, China. She was also appointed as a visiting scholar in Wuhan University, China from 2015 to 2016. Her research areas include consecutive interpreting, tourism translation, and corpus linguistics.

Xue Xie is a MA student at the Faculty of Modern Languages and Communication in Universiti Putra Malaysia, Malaysia. Her research interest lies in corpus linguistics and translation studies.

Jiashuang Qi is a MA student at the Faculty of Modern Languages and Communication in Universiti Putra Malaysia, Malaysia. Her research interest lies in TESOL.

Afida Mohamad Ali is an Associate Professor at the Faculty of Modern Languages and Communication in Universiti Putra Malaysia, Malaysia. Her research areas include Corpus Linguistics, English/Languages for Specific Purposes/Language for the Professions, Applied Linguistics, Text Linguistics, Written Discourse/Genre Analysis.

Jie Zhao is a Lecturer in the Department of Foreign Languages at Lanzhou Jiaotong University, Gansu Province, China. She received her master's degree in Foreign Linguistics and Applied Linguistics from Lanzhou Jiaotong University. Her research interests lie in the fields of applied linguistics, discourse analysis and language for specific purpose

THE AUTHORS' ADDRESSES

First Author

Rong Lu

1. Lecturer

School of Foreign Languages

Tianshui Normal University

South Xihe Road, Qinzhou District, Tianshui City, Gansu Province, 741000, CHINA

2. PhD Student

Faculty of Modern Languages and Communication

Universiti Putra Malaysia

Serdang, 43400 Seri Kembangan, Selangor, MALAYSIA

E-mail: GS59350@student.upm.edu.my

Corresponding Author

Xue Xie

MA student

Faculty of Modern Languages and Communication

Universiti Putra Malaysia

Serdang, 43400 Seri Kembangan, Selangor, MALAYSIA

E-mail: gs56165@student.upm.edu.my

Co-authors

Jiashuang Qi

MA student

Faculty of Modern Languages and Communication

Universiti Putra Malaysia

Serdang, 43400 Seri Kembangan, Selangor, MALAYSIA

E-mail: gs56329@student.upm.edu.my

Afida Mohamad Ali

Associate Professor Faculty of Modern Languages and Communication Universiti Putra Malaysia Serdang, 43400 Seri Kembangan, Selangor, MALAYSIA E-mail: afida@upm.edu.my

Jie Zhao

Lecturer
Department of Foreign Languages
Lanzhou Jiaotong University
NO.188, Bao Shihua Road, Anning District, Lanzhou City, Gansu Province, CHINA
E-mail: gs59588@student.upm.edu.my

Received: 30 May 2022

Received in Revised Form: 1 August 2022

Accepted: 16 August 2022