Perception on global economy through world trade report: A corpus and computation-driven approach

American Journal of Social Sciences and Humanities
Vol. 8, No. 2, 251-262, 2023
e-ISSN: 2520-5382

Zilong Zhong

Research Institute of Foreign Languages, Beijing Foreign Studies University, China.
Email: zhongzilong1106@bfsu.edu.cn

ABSTRACT

This research aimed to investigate the lexical trends and sentiment shifts in the World Trade Report spanning 2018-2020 using corpus and sentiment analysis tools. Data of the *World Trade Report* from 2018 to 2020 was analyzed. We employed the Words tool, Whelk tool, and GraphColl tool within the LancsBox corpus tool to count high-frequency nouns and verbs, scrutinize the distribution characteristics of key words, and assess their semantic collocation features. Furthermore, sentiment analysis was conducted using the VADER algorithm. The results indicated that the 2020 *World Trade Report* prominently featured high-frequency nouns such as policy, innovation, and government, as well as verbs like support, mirroring the challenging global economic climate in that year. Semantic collocation analysis of key words from the 2020 report highlighted the significant challenges COVID-19 posed to global economic stability. Additionally, the sentiment scores from 2019 to 2020 exhibited notable differences, with a consistent decline in average scores annually and a more pronounced negative sentiment in the 2020 report compared to the previous years. Recognizing these linguistic and sentiment trends can aid policymakers and businesses in understanding the nuanced shifts in global economic narratives, especially in response to significant events like the COVID-19 pandemic.

Keywords: *Global economy, LancsBox, Semantic collocation, Sentiment analysis, VADER, World trade report.*

DOI: 10.55284/ajssh.v8i2.977
Citation | Zhong, Z. (2023). Perception on global economy through world trade report: A corpus and computation-driven approach. *American Journal of Social Sciences and Humanities, 8*(2), 251–262.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper..
History: Received: 28 August 2023/ Revised: 2 October 2023/ Accepted: 12 October 2023/ Published: 20 October 2023
Publisher: Online Science Publishing

Highlights of this paper

- Utilizing corpus and sentiment analysis tools, this study uncovers lexical trends and sentiment shifts in the *World Trade Report* from 2018 to 2020.
- The 2020 report shows a dominant presence of words like policy, innovation, and support, reflecting the global economic challenges, with a particular emphasis on the disruptions caused by COVID-19.
- Sentiment scores consistently declined over the years, revealing a more pronounced negative sentiment in 2020, shedding light on the changing global economic narratives in the wake of significant events.

1. INTRODUCTION

With the continuous development and improvement of the discourse system, economic discourse, as an important part of the discourse system, has gradually attracted researchers' attention. *World Trade Report* (WTR), an official document published by World Trade Organization (WTO), is part of the world economic discourse. To date, few studies investigate the discourse of trade reports. The current study attempts to examine the features of lexical usage, semantic collocation and sentiment tendency of WTR in order to expand the research field. In today's world, economic globalization is under intensive development and trade plays a crucial role in the healthy and sustainable development of economic growth. Exploring discourse features in WTR can, on the one hand, help readers grasp main ideas conveyed by the report more easily and, on the other hand, reveal the global economic trend. The present study is based on a quantitative corpus approach to investigate features of lexical usage and semantic collocation and conduct a natural language algorithm to explore sentiment tendency characteristics, ensuring objectivity and persuasiveness of the findings.

A new generation of corpus software, LancsBox 6.0 (Brezina, McEnery, & Wattam, 2015), will be adopted in the present study, which is a new generation of software developed at Lancaster University for analysing language data and corpora. LancsBox is innovative and optimized in many aspects such as information retrieval, data processing and result visualization. As the core functions and unique advantages, it allows users to work with their own data to automatically annotate data for part-of-speech, visualize language data and find out more details about language support. These merits are well typified in the present study. Several previous studies based on LancsBox have been reported. As the project lead of the development team, Vaclav Brezina conducted a series of studies related to discourse analysis using LancsBox and came up with some interesting findings (Brezina, 2016, 2018; Brezina et al., 2015). For instance, Brezina (2018) adopted LancsBox to explore views and attitudes of readers of *the Guardian* and *the Daily Mail* towards Eastern European immigrants. He compares the semantic networks of keyword *immigrant* in both corpora. It revealed that readers of *the Guardian* were neutral or defensive towards Eastern European immigrants, while readers of *the Daily Mail* had a clear opposition and discontent. Additionally, Lillqvist and Anu (2018) investigated discourse patterns of guidelines for Facebook community platform and found that Facebook portrayed itself as a powerful and willing company to help and meet the needs of its users. Germond and Fong (2019) examined the collocation networks and shared collocations between node words *climate change* and *maritime security* via LancsBox and found that two words were related indirectly through *migration* and *displacement*, suggesting that International Maritime Organization (IMO) practitioners need to pay attention to the intrinsic link between climate change and maritime security.

The concept of sentiment analysis was originally proposed by Nasukawa and Jeonghee (2003). Sentiment analysis refers to a natural language processing technique that extracts subjective information from texts based on computational algorithms (Chojnicka & Wawer, 2020). Text sentiment analysis mainly includes supervised learning methods, unsupervised learning methods, semi-supervised learning methods and deep learning methods. Among them, supervised learning methods include Naive Bayes, Support Vector Machine, Max Entropy, Decision Tree, Term Frequency-Inverse Document Frequency, etc. Unsupervised learning approaches include sentiment-based lexicon matching, semantic pattern matching and hybrid methods. Semi-supervised learning methods achieve sentiment tendency analysis by using a small number of manually annotated subset of samples to automatically generate high-quality annotations for a large number of the rest unlabelled samples. Deep learning approaches are based on deep learning networks, such as Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) for sentiment classification. Current research in sentiment analysis has been done primarily in a domain-dependent manner, concentrating on particular types of text or language (Zhang, Wang, & Liu, 2018), which cannot be universally applied due to the low accuracy of other topics and text genres. Further, instead of quantifying positive and negative words, it focuses on categorizing the entire text. To address these issues, a more widely applicable method for economic discourse analysis based on a custom dictionary of sentiment is to be carried out in the present study.

The remaining parts of this paper are outlined as follows. Research methodology will be introduced in the next section. Results will be presented and discussed in section three. Conclusion of this research will be described in section four, and limitations as well as future work will be elaborated in the final section.

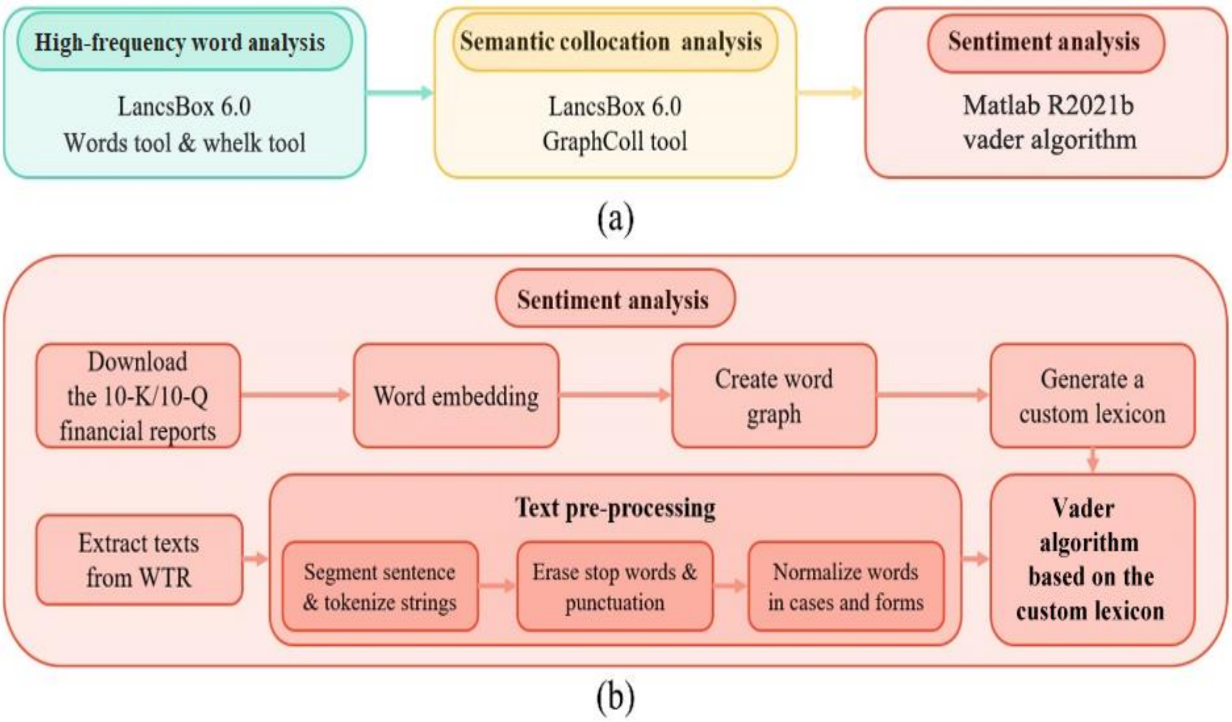

Figure 1. Flow chart of the study (a) and the technical routes to sentiment analysis (b).

Table 1. Composition of the corpus 2018-2020 WTR.

Report	Token	Type
2018 WTR	149041	11520
2019 WTR	133785	10519
2020 WTR	134720	10574

Table 2. Main functions in sentiment analysis.

Procedures	Functions	What functions do
Custom lexicon	Finance reports()	Download the 10-K/10-Q financial reports
	Train word embedding()	Word embedding
	Graph()	Create the word graph
	Polarity scores()	Generate polarity scores
Text extraction	Extract file text()	Extract words from the PDF files
	Readtable()	Read contents in tables
Text pre-processing	Tokenized document()	Tokenize strings
	Erase punctuation()	Erase punctuation
	Remove stop words()	Erase stop words
	Normalize words()	Normalize words
Sentiment analysis	VADER sentiment scores()	Generate sentiment score by VADER algorithm
Visualization	Bar()	Create bar charts
	Swarmchart()	Create swarmscatter charts
	Subgraph()	Display the neighborhood in word graph

2. METHODOLOGY

To clarify the features of lexical usage, semantic collocation and sentiment tendency of WTR, the present study collected 2018-2020 WTR from the WTO website as the analytical corpus Table 1. The overall workflow of the current study is shown in Figure 1a.

To begin with, high-frequency nouns and verbs in the 2018-2020 WTR were counted, and distribution characteristics of keywords were examined with the help of the Words tool and Whelk tool which are built in LancsBox. Secondly, the GraphColl tool within LancsBox was adopted to generate a semantic collocation network to demonstrate the semantic collocation features of the 2018-2020 WTR. The following thresholding rule was applied: Span selected from five left to five right. Mutual Information score (MI) was chosen for Statistics. Collocation frequency (five or more times) was set for Threshold, and Type was selected as Unit. Finally, sentiment analysis was conducted via MATLAB R2021b. The specific technical route for text sentiment analysis is shown in Figure 1b and the main functions used in the Text Analytics Toolbox that operates within MATLAB R2021b are shown in Table 2.

On the one hand, the 10-K/10-Q financial reports from the first quarter of 2018 to the fourth quarter of 2021 were obtained from the official website of the Securities and Exchange Commission (SEC) as the text samples for a custom lexicon, and the number of downloaded reports reached 2,000 per quarter for sufficient sample texts. The word embeddings were trained on the above text samples, and the similarities between words in the training data were used to create the word graph. The nodes in the word graph correspond to words in the text samples, the connecting lines indicate whether the words are in each other's neighbourhood, and the weights in the word graph correspond to the cosine distance between the word vectors. Words with short distances and high weights from the seed words are identified in the word graph path, and after a full traversal of the word graph, the polarity of all the words in the text samples is calculated starting from the positive/negative seed words see Table . In the polarity calculation loop for each word, the updated score of two nodes is equal to the maximum of the score of the seed word and neighbouring words in the last loop, α_{ij}, and the score of the seed word and search node α_{ik} multiplied by the weight from the word graph edge, ω_{ik}, shown in Equation 1, while the polarity (*POL*) of each word is the sum of all its connected scores, shown in Equation 2 (Velikovich, Sasha, Kerry, & Ryan, 2010),

$$\alpha_{ij} = \max\{\alpha_{ij}, \alpha_{ik} \cdot \omega_{kj}\} \quad (1)$$

$$POL_j = \sum_i \alpha_{ij} \quad (2)$$

β was introduced to calculate the score of each word based on its polarity, which is defined in Equation 3,

$$\beta = \frac{\sum POL^{+}}{\sum POL^{-}} \quad (3)$$

Then Equation 4 presents the score of the j-th word,

$$Score_j = POL_j^{+} - \beta \cdot POL_j^{-} \quad (4)$$

Words were extracted throughout from the text samples as a lexicon for the sentiment analysis based on the postscript VADER (Valence Aware Dictionary and sEntiment Reasoner) algorithm (Rose, Dave, Nick, & Wendy, 2010).

On the other hand, the text was pre-processed for the corpus to be analysed. Based on sentence segmentation and tokenization, potential Influencing factors including variations in case (e.g., *good* and *Good*), variations in word forms (e.g., *promote* and *promotes*), stop words (e.g., *the* and *to*), punctuation and special characters were excluded. The text was then analysed for sentiment tendency using the VADER algorithm based on the custom lexicon.

During the present study, the visualization in sections of high-frequency words analysis and text sentiment analysis were implemented using MATLAB R2021b, and in the part of semantic collocation analysis was performed by LancsBox.

Table 3. Positive/negative seed words.

Positive seed words		Negative seed words	
Achieve	**Advantage**	**Adverse**	**Slowdown**
Better	Creative	Inefficient	Challenge
Efficiency	Improve	Concern	Pandemic
Developed	Greater	Fallout	Barrier
Promote	Improving	Limited	Limitation
Innovation	Innovations	Negative	Tension
Innovative	Opportunities	Failure	Crisis
Boost	Positive	Restriction	Restrict
Success	Strengthen	Restrictive	Argument
Growth	Promotion	Inequality	Gap

3. RESULTS AND DISCUSSION

3.1. High-Frequency Word Analysis

Word frequency reflects basic linguistic features including semantic meaning of discourse (McEnery, Xiao, & Tono, 2006). Due to the variation in the corpus size of the 2018-2020 WTR, word frequency was analysed in unit of per ten thousand words to ensure comparability between the data used in this study. Nouns and verbs in a semantic cluster contain more information, so this study focuses on analysing the similarities and differences between high-frequency nouns and high-frequency verbs in the 2018-2020 WTR. The high-frequency words were extracted through the Words tool of LancsBox, and the invalid data such as *be_v*, *may_v*, *can_v* with no real meaning were eliminated to make a high-frequency noun list Table 4 and a high-frequency verb list Table 5.

Comparing the high-frequency nouns in the 2018-2020 WTR, it can be seen that there is some overlap between the three reports, such as *trade*, *service*, *technology*, *world*, *country* and so on. At the same time, some new words emerge from the 2020 WTR, such as *policy*, *innovation*, *government*. These words do not appear in the high-frequency nouns of the previous two years' reports, but are located in the head of the noun list of the 2020 WTR. Analogously, 2018-2020 WTR also share some high-frequency verbs, such as *have*, *include*, *develop*, *use*, *increase*, etc. In contrast to the 2018-2020 WTR, the word *support*, which ranks eighth in the list of high-frequency verbs of the 2020 WTR, does not appear in the word lists of the previous two years' reports.

Table 1. High-frequency nouns in 2018-2020 WTR.

WTR (2018)		WTR (2019)		WTR (2020)	
Nouns	**Frequency**	**Nouns**	**Frequency**	**Nouns**	**Frequency**
Trade_n	128.22	Service_n	224.39	Policy_n	109.12
Service_n	62.06	Trade_n	169.68	Innovation_n	96.57
Technology_n	60.99	Economy_n	50.83	Trade_n	60.05
World_n	39.05	Sector_n	46.27	Government_n	49.96
Country_n	36.97	World_n	37.90	Economy_n	49.06
Cost_n	31.67	Country_n	36.63	Technology_n	45.58
WTO_n	24.49	Cost_n	34.01	Country_n	41.05
Product_n	23.62	Cent_n	32.59	World_n	32.21
Information_n	21.81	WTO_n	30.72	Development_n	31.77
Cent_n	21.81	Policy_n	25.26	Market_n	30.73
Economy_n	21.27	Services_n	25.11	Service_n	28.43
Report_n	21.14	Figure_n	25.04	Firm_n	27.98
Datum_n	20.93	Market_n	23.17	WTO_n	26.72
Development_n	20.40	Firm_n	21.30	Datum_n	25.01
Market_n	20.26	Development_n	21.00	Investment_n	23.08
Protection_n	20.20	Share_n	20.85	Sector_n	22.27
Good_n	20.20	Report_n	20.63	Industry_n	21.82
Provision_n	20.06	Technology_n	20.48	Report_n	21.30
Internet_n	19.99	Growth_n	19.28	Measure_n	20.64
Policy_n	19.26	Change_n	18.91	Effect_n	20.56

Table 5. High-frequency verbs in 2018-2020 WTR.

WTR (2018)		WTR (2019)		WTR (2020)	
Verbs	**Frequency**	**Verbs**	**Frequency**	**Verbs**	**Frequency**
Have_v	62.20	Have_v	54.79	Have_v	64.06
Include_v	23.42	Develop_v	31.09	Develop_v	25.09
Develop_v	21.14	Trade_v	31.02	Use_v	20.78
Use_v	18.45	Increase_v	19.66	Increase_v	17.22
Increase_v	16.57	Include_v	16.89	Include_v	16.48
Make_v	15.10	Provide_v	12.33	Promote_v	16.48
Provide_v	14.96	Use_v	12.03	Provide_v	15.29
Relate_v	11.27	See_v	10.32	Support_v	11.58
Reduce_v	11.14	Find_v	9.87	See_v	10.84
See_v	11.07	Base_v	9.87	Make_v	9.65
Take_v	9.66	Reduce_v	9.87	Find_v	9.28
Trade_v	9.53	Grow_v	9.79	Show_v	9.06
Require_v	9.39	Make_v	9.49	Lead_v	8.54
Transform_v	9.19	Affect_v	9.42	Help_v	8.46
Base_v	8.25	Show_v	8.60	Address_v	8.02
Lead_v	8.11	Become_v	7.40	Allow_v	7.27
Help_v	7.78	Lead_v	7.18	Create_v	7.13
Address_v	7.51	Export_v	7.10	Become_v	7.05
Find_v	7.38	Allow_v	6.43	Require_v	6.98
Affect_v	7.25	Take_v	6.43	Take_v	6.75

Combined with the high-frequency nouns (*policy*, *innovation*, *government*) and high-frequency verbs (*support*), it can be assumed that the world economic environment in 2020 was tougher than 2018 and 2019. To cope with the sluggish economic trend and build up a favourable environment for economic development, policies and announcement had been issued and innovative economic development models had been advocated by governments. To further visualize the differences in the distribution of the emerging words in the 2018-2020 WTR, the study takes the high-frequency nouns *policy** and *government** and the high-frequency verb *support** as examples, and retrieves their distribution in the three reports with the help of LancsBox's Whelk tool, based on which a heat map of word frequency distribution is drawn Figure 2. Different colour blocks in the graph represent different frequency of occurrence, and the colour bar from bottom to top indicates the lowest to the highest frequency of occurrence. It can be seen that the three words appear much more frequently in the 2020 WTR than in the previous two years. Combined with the social context of COVID-19 that began in late 2019, it is broadly assumed that the epidemic led to a crisis in global economic markets. The distribution of the keywords *COVID-19** and *crisis** in the 2018-2020 WTR was also shown in Figure 2. Results validate the speculation to some extent.

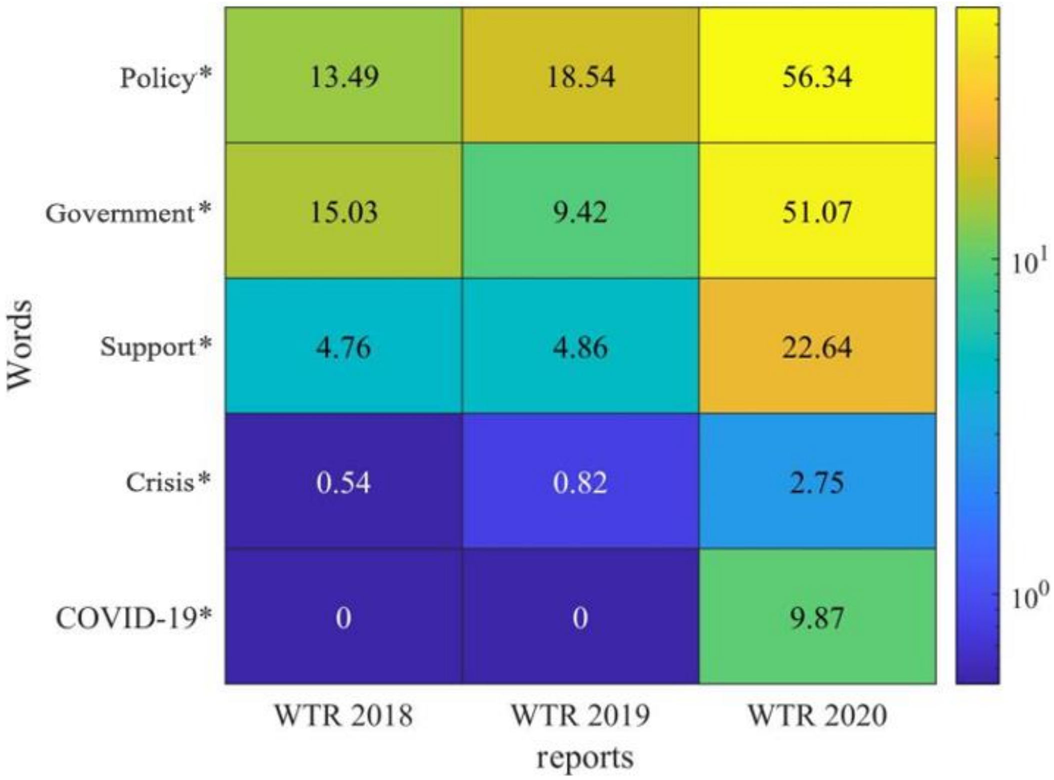

Figure 2. Heat map of the distribution of keywords in 2018-2020 WTR.

Note: The asterisks (*) after the words indicate the use of wildcard characters in a search query. When used in this context, the search would retrieve all variations or derivations of the root word.

3.2. Semantic Collocation Analysis

Collocation networks can reveal the relatedness of keywords and their semantic meaning in a discourse (Baker, 2006). The GraphColl tool built within LancsBox was conducted to generate a collocation network graph for the keyword *crisis** in the 2020 WTR, as shown in Figure 3a. The lines between node words in the graph indicate their relatedness. The length of the line indicates the collocation strength of the node words, and the closer the distance between them, the stronger the collocation strength. The darker the colour of the node, the higher the frequency of co-occurrence between the node words. In descending order of MI, the collocations with MI above 7 are *2008-09*, *current*, *during*, *financial*, *since* and *covid-19*. It can be initially concluded that COVID-19, like the global economic crisis of 2008-2009, poses a great threat to global economic and financial development.

Further mapping of the semantic collocation network for the 2020 WTR keyword *COVID-19** was carried out in Figure 3b. In descending order of MI, the collocations with MI over 7.5 are *fight*, *pandemic*, *vaccine*, *notifications*, *vaccines*, *spread*, *against* and *crisis*. These collocations disclose the serious impact of COVID-19 on the stability of global economy. Combined with the growth rates of the world economy and world trade during the period 2018-2020 published by the International Monetary Fund (IMF) Figure 4, it is clear that the growth rates of the world economy and trade have continued to decline over the three years, with even negative growth occurring in 2020. The two main economic indicators verify the previous judgment.

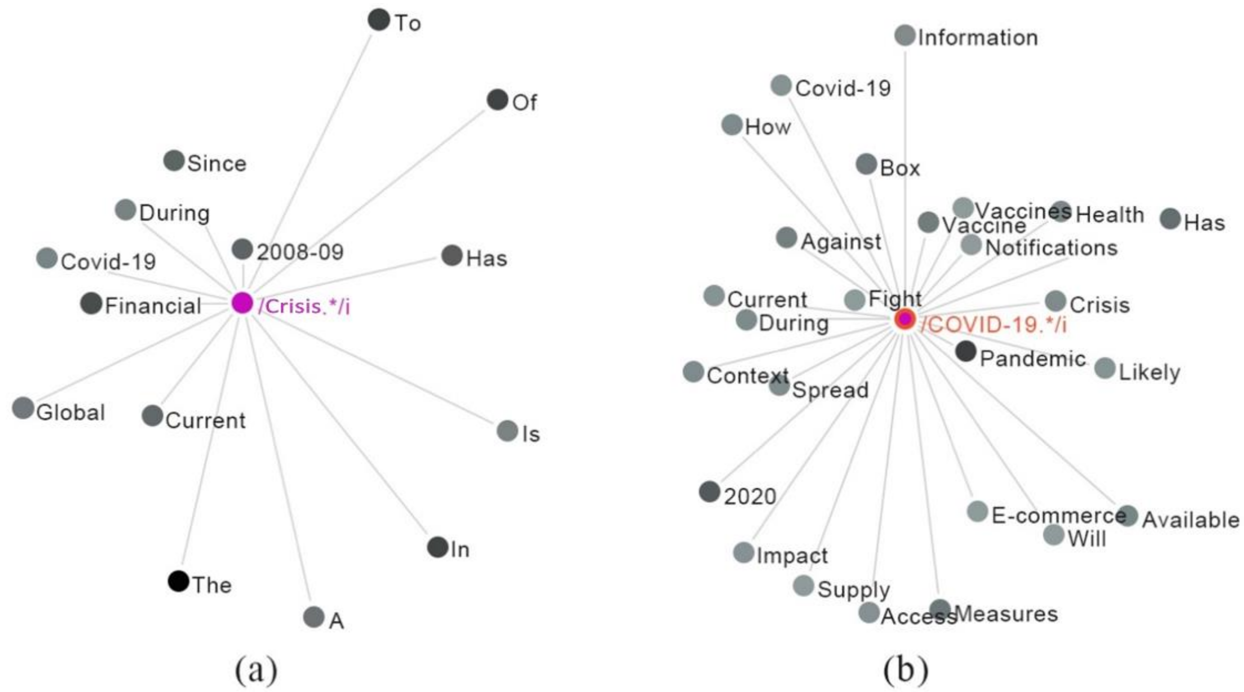

Figure 3. Semantic collocation network of crisis* (a) and COVID-19* (b) in 2018-2020 WTR.

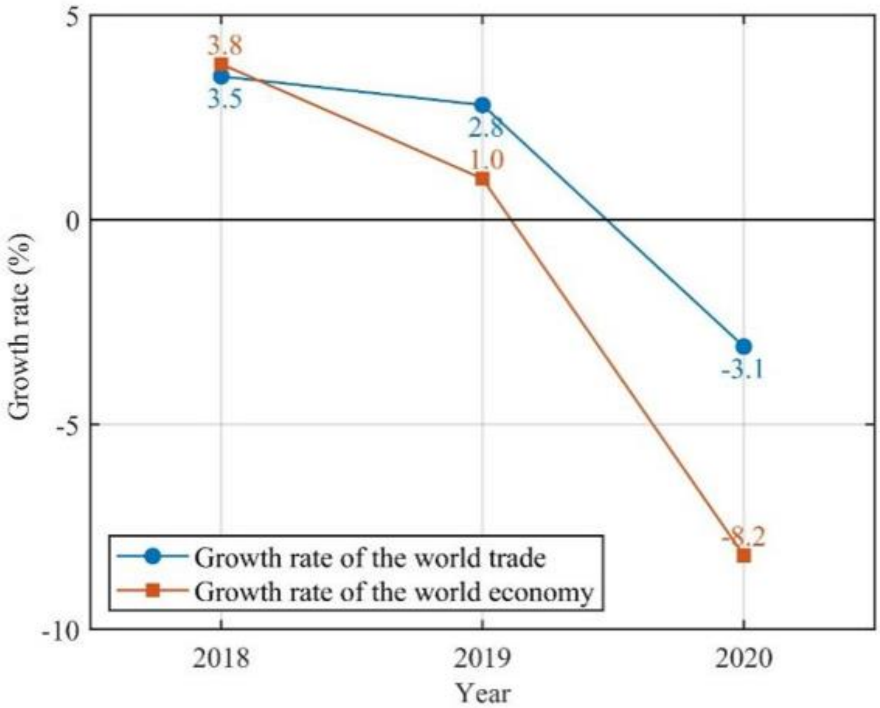

Figure 4. World economic growth rate and world trade growth rate (2018-2020).

3.3. Sentiment Analysis

In the current study, a self-built word bank containing 833 words was created based on the 10-K/10-Q financial reports for 2018-2021. The top ten words with the highest (positive) and lowest (negative) scores are shown in Table 6. The neighbourhoods of the highest scoring word *efficiency* and the lowest scoring word *limitation* in the word graph are shown in Figure 5.

Discourse usually shows corresponding sentiment tendencies due to the personal stance and preference of the text creator. In order to investigate the differences in sentiment tendencies in the 2018-2020 WTR, the Foreword by the WTO Director General or Foreword by Deputy Directors General and Executive summary sections of the 2018-2020 WTR were selected to measure sentiment scores. The former is the subjective statement of the WTO Director General, which has a more pronounced sentiment tendency than the objective description. The latter is a brief description of the WTR, which brings together the central ideas of the report and is placed at the beginning of the report to help the reader quickly grasp its main points.

Table 6. The top 10 positive/negative words in the custom lexicon.

Positive seed words		Negative seed words	
Words	**Sentiment score**	**Words**	**Sentiment score**
Efficiency	0.9971	Limitation	-1.0000
Improve	0.9802	Restriction	-0.9115
Innovation	0.9410	Adverse	-0.8277
Innovative	0.9287	Barrier	-0.8099
Creative	0.7228	Failure	-0.7717
Success	0.7128	Crisis	-0.7670
Strengthen	0.7100	Restrictive	-0.7369
Promote	0.7017	Slowdown	-0.7244
Promotion	0.6834	Pandemic	-0.7231
Boost	0.6814	Gap	-0.6844

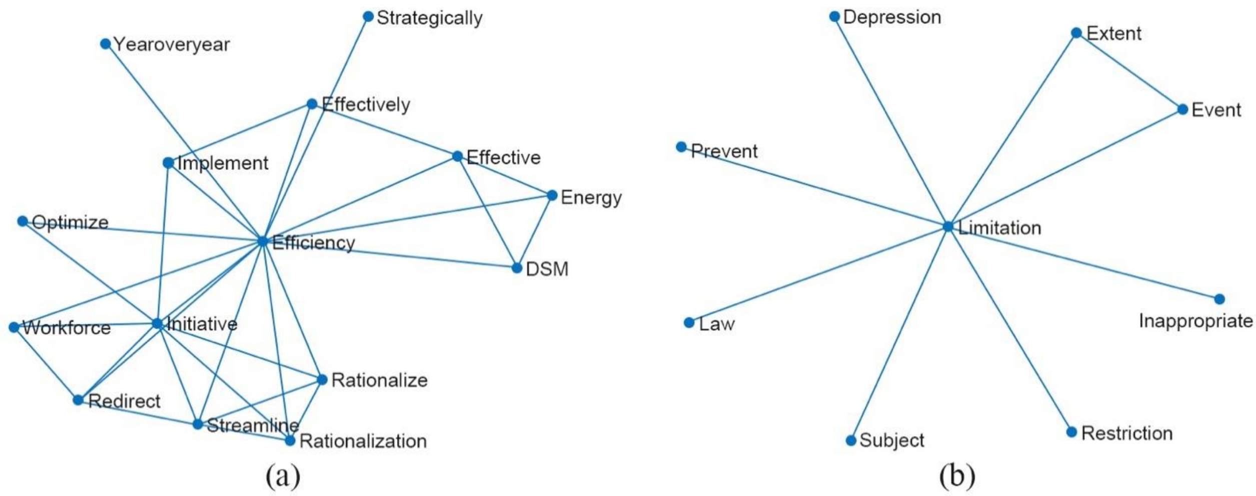

Figure 5. The neighbors of efficiency (a) and limitation (b) in the word graph.

The 2018-2020 WTR sentiment score measure was implemented by using the VADER algorithm embedded in MATLAB R2021b. Sentences were judged to be positive when the score was higher than 0.05, negative when the score was less than -0.05, and neutral when the score was between -0.05 and 0.05. Results showed that positive sentences in the 2018-2020 WTR decreased year by year, accounting for 22.40%, 19.46%, and 14.73% respectively. The average score of sentiment tendency reduced year by year, with -0.0139, -0.0223, and -0.0298 respectively Figure 6. Further analysis of the positive sentiment scores for the 2018-2020 WTR was conducted using R version 3.5.3. normality was tested firstly and the results were not all greater than 0.1 ($p1 = 0.0000046699$, $p2 = 6.6986\times10^{-8}$, $p3 = 0.0957$), indicating that the samples were not all normally distributed. Therefore, a Kruskal-Wallis H-test was performed on the data and the results revealed that the distribution of positive sentiment scores differed among the groups ($p = 0.0041 < 0.05$). A two-way comparison was then performed and the results presented a significant difference in the distribution of positive sentiment scores for the 2019-2020 WTR

(2019/2020: $p = 0.0024 < 0.05$). To give a more visible representation of the distribution of sentiment scores, a swarmscatter plot of the sentiment scores for the 2018-2020 WTR is illustrated in Figure 7. Different coloured scatter clusters in the graph represent different sentiment tendencies. The area covered by the scatter clusters indicates the density of the distribution, with a larger area in a given region indicating a larger number of sentiment scores in that region. It can be seen that the area covered by scatter clusters of positive sentiment tendencies in the 2018-2020 WTR is decreasing year by year. Combined with Figure 4 and Figure 6, it is reasonable to speculate that the global economic situation is becoming increasingly negative during the period 2018-2020.

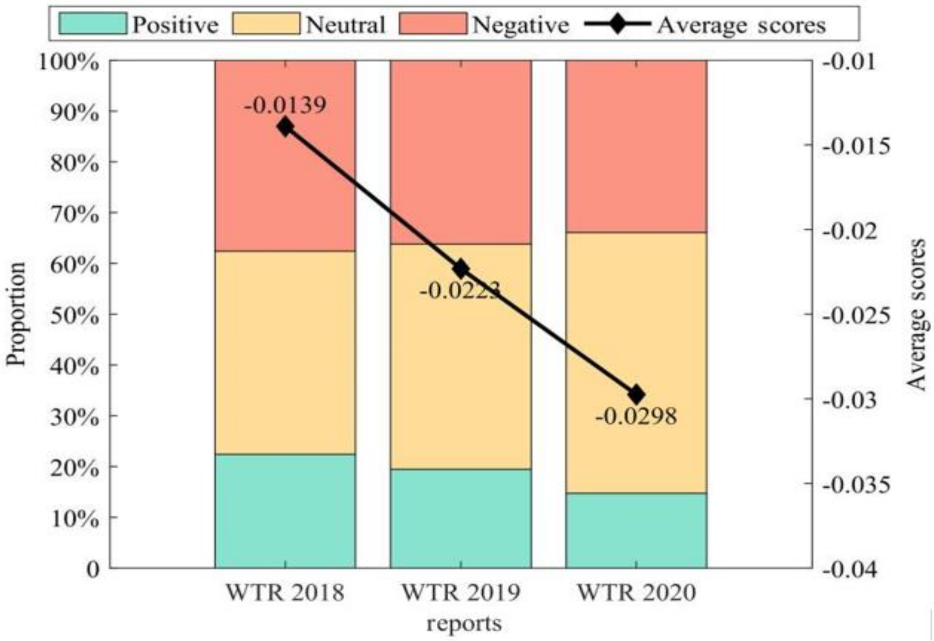

Figure 6. Sentiment tendency proportion and average scores.

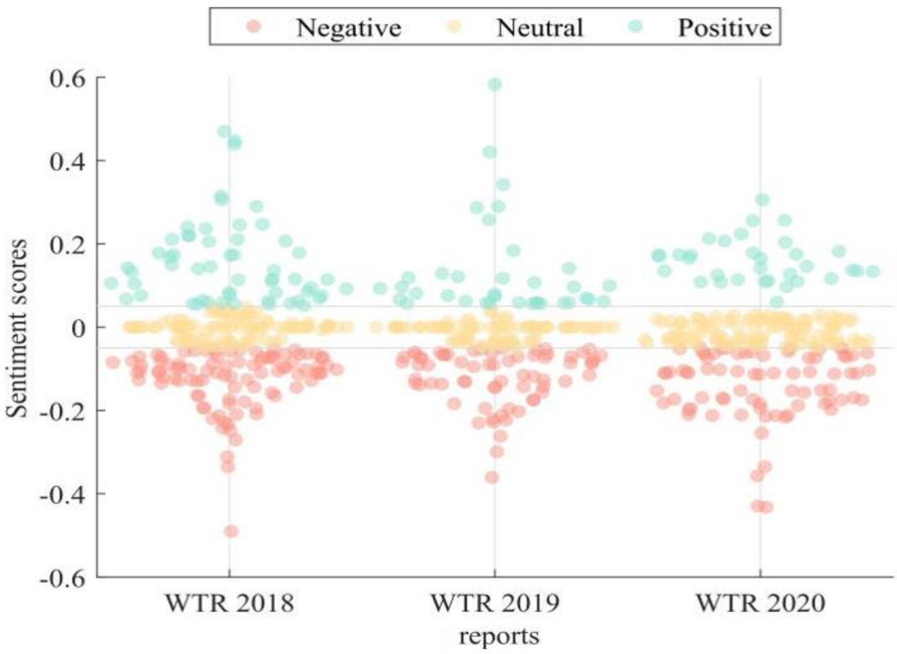

Figure 7. Swarmscatter plot of sentiment scores.

4. CONCLUSION

From the longitudinal perspective, 2018-2020 WTR was collected in this study. The Words tool, Whelk tool and GraphColl tool of the corpus tool LancsBox were used to count high-frequency words, and examine distribution characteristics of high-frequency words as well as semantic collocation features. Sentiment analysis was conducted with the help of VADER. High-frequency Word Analysis revealed some emerging high-frequency nouns and verbs, reflecting the seriousness of the global economic environment in 2020. Semantic collocation analysis indicated that COVID-19 had brought huge challenges to the stability of global economy. Sentiment analysis found that there were significant differences in sentiment scores in the 2019-2020 WTR, with the average sentiment score decreasing year by year during the period, and the proportion of negative sentiment in the 2020 WTR being greater than that in 2018-2019 WTR.

5. LIMITATIONS AND FUTURE WORK

As with any method, corpus and computation-based approach has its limitations, which could be tackled in future work. In terms of data collection, the current study only selected three years' WTR as the corpus for the analysis. Therefore, in the following study, if time permits, the scope of the self-built corpus should be expanded to investigate diachronic change of economy through economic discourse more comprehensively. From the perspective of sentiment classification, one limitation of the VADER algorithm for sentiment analysis is located in its underlying mechanism. As the word-based classifier, the implicit sentiment transformation rooted in texts may not be recognized in sentences with conjunctions, which are regarded as the stop words in the text pre-processing. The rapid development of artificial intelligence brings new possibilities to sentiment analysis. The implicit sentiment requiring further inference based on that readers' experience may be extracted via the developing neural networks. Besides, sentiment analysis based on the social comments on the economic situation may unveil the true living environment and expectations from the mass.

REFERENCES

Baker, P. (2006). *Using corpora in discourse analysis*. London & New York: Continuum.

Brezina, V. (2016). Collocation networks in Paul Baker & Jesse Egbert (eds.). Triangulating methodological ap-proaches in corpus linguistic research. In (pp. 90-107). New York: Routledge.

Brezina, V. (2018). Collocation graphs and networks: Selected applications in pascual cantos-Gómez & Moisés Al-mela-Sánchez (eds.) Lexical collocation analysis. In (pp. 59-82). Cham: Springer.

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*(2), 139-173. https://doi.org/10.1075/ijcl.20.2.01bre

Chojnicka, I., & Wawer, A. (2020). Social language in autism spectrum disorder: A computational analysis of sentiment and linguistic abstraction. *PLoS One*, *15*(3), e0229985. https://doi.org/10.1371/journal.pone.0229985

Germond, B. Y., & Fong, W. H. (2019). Climate change and maritime security narrative: The case of the international maritime organization. *Journal of Environmental Studies and Sciences*, *9*(1), 1-12. https://doi.org/10.1007/s13412-018-0509-2

Lillqvist, E., & Anu, A. H. (2018). Discourse of enticement: How facebook solicits users. *Critical Approaches to Dis-course Analysis across Disciplines*, *10*(1), 63-80.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.

Nasukawa, T., & Jeonghee, Y. (2003). *Sentiment analysis: Capturing favorability using natural language processing*. Paper presented at the 2nd International Conference on Knowledge Capture Association for Computing Machinery, 23-25 October.

Rose, S., Dave, E., Nick, C., & Wendy, C. (2010). Automatic keyword extraction from individual docu-ments in Michael W. Berry

& Jacob Kogan (eds.). Text mining: Applications and theory. In (pp. 1-20). New York: Wiley.

Velikovich, L., Sasha, B.-G., Kerry, H., & Ryan, T. M. (2010). *The viability of web-derived po-larity lexicons*. Paper presented at the Paper presented at the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2-4 June.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.