

ANALYSIS OF UZBEKISTAN’S RELATIONS WITH CHINA, RUSSIA AND SOUTH KOREA: UTILIZING TEXT MINING BASED ON GDELT BIG DATA

Taesang Yoo

Master’s course

Hankuk University of Foreign Studies, South Korea

Jieun Lee

Professor

Department of Central Asian Studies,

Hankuk University of Foreign Studies, South Korea

Annotation. *GDELT (Global Dataset on Events, Language, and Tone), the ‘big’ event data developed in 2013, provides quantifiable data on cooperative and conflictual relations between countries and events related to specific phenomena. Its utility has been recognized and it is actively used in international relations and foreign policy research. This study also utilizes GDELT data to introduce a method for analyzing the relationships between Uzbekistan and China, Russia, and South Korea through keyword correlation indicators derived from “Text Mining”, and to present the actual analysis results. By conducting a keyword correlation analysis based on event data between Uzbekistan and the three analyzed countries, Russia and South Korea, it was possible to identify diplomatic action patterns through the correlation of core keywords such as ‘engage,’ ‘express intent,’ and ‘make’ with keywords indicating specific actions. Uzbekistan’s foreign policy under the Mirziyoyev government has shown a diplomatic action pattern of seeking or participating in negotiations, expressing intentions for negotiation or cooperation, and pursuing meetings for negotiation and cooperation with the analyzed countries, while inviting or visiting counterpart countries as part of these processes. This demonstrates that Uzbekistan has actively worked towards establishing cooperative relations, which were set as the goal or direction of its foreign relations named “Good Neighbor Policy” by President Mirziyoyev. Such keyword correlation indicator analysis facilitates the explanation of cooperative relationships.*

Keywords: *relations, big data, GDELT, CAMEO, text mining, quantitative analysis, keyword correlation analysis, keyword correlation matrix.*

Introduction.

Research on foreign policy and international relations has primarily relied on qualitative research methods. This typically involves selecting a few representative cases, applying relevant theories or concepts to interpret them, and analyzing them from the researcher’s perspective. While this method provides a deep logical understanding of phenomena, it is difficult to consider it a sufficiently objective approach, as the researcher’s subjectivity is involved in every step of the research process, from selecting theories or concepts to interpreting them. To supplement the limitations of qualitative research methods, efforts have been made to secure sufficient quantitative data for empirical verification and to analyze it to enhance objective persuasiveness. Researchers have attempted to accumulate data reflecting international relations from newspaper articles and media reports as quantitative data. As a result, event data sets such as WEIS (World Event/Interaction Survey) and COPDAB (Conflict and Peace Data Bank) were developed to measure inter-state relations based on events reported in the media. Event data started with "human coding" methods and gradually transitioned to "machine coding" methods with the advancement of natural sciences, allowing for handling vast amounts of data. This technological advancement led to the emergence of “big data” that can be utilized for research in foreign policy and international relations.¹

GDELT(Global Data on Events, Location, and Tone), which was introduced in 2013, combines automated systems with human-classified event types to collect and store event data worldwide, attracting significant attention immediately upon its debut. Research utilizing GDELT, which updates data quickly and in real-time compared to other data collection and classification programs designed with human coding methods, has been conducted worldwide. Moreover, studies verifying the usefulness of GDELT for international politics and international relations research have been conducted, concluding that, despite some limitations, GDELT can be effectively utilized for research.

In academic circles of the Anglosphere, Europe, China, Japan, and Korea, research using GDELT data is actively conducted. Nevertheless, in the study of foreign policy and international relations of Central Asian countries, including Uzbekistan, quantitative research attempts are still rare, with qualitative research being overwhelmingly predominant. That is, research has primarily relied on interpretations involving the researcher’s subjectivity rather than attempting to secure objective evidence and empirical verification of phenomena.

¹ 정승철(Jeong Seungcheol). [빅데이터로 살펴본 한미 평화관계: 2017-2020(ROK-US peace relations examined through big data: 2017-2020)]. JPI Policy Forum, 2021, no. 239, pp. 2-3.

The purpose of this study is to attempt an analysis of Uzbekistan’s international relations from a different approach than traditional research methods. As is well known, under President Shavkat Mirziyoyev, Uzbekistan has observed changes in both overall politics and foreign policy. The main policy, named “Good Neighboring Policy”, is to establish cooperative relations with neighboring countries to aid national development.² To achieve this, Uzbekistan has been building a leading position in Central Asia and strengthening cooperation, especially in the economic sector, establishing amicable relations with most countries and enhancing cooperation with South Korea in areas such as the economy and technology. These analyses are the results of qualitative research, where representative cases and literature are interpreted through theories or concepts to derive implications. While qualitative research allows for in-depth understanding and flexible research, it heavily relies on the researcher’s subjectivity, thus lacking objectivity or reliability. To address these issues, efforts to empirically revalidate phenomena established through qualitative research using quantitative methods are increasing. This study also attempts to incorporate quantitative analysis into the study of Uzbekistan’s international relations, which has been relatively understudied in this manner.

The quantitative analysis method chosen for this study involves using GDELT data, which accumulates events occurring between countries as data, and conducting text mining to identify the types of events contained within. The research sequence is as follows. First, the raw data for quantitative analysis of the relationships between Uzbekistan and South Korea, China, and Russia during President Shavkat Mirziyoyev’s government is introduced, along with the text mining analysis method. Next, the extracted keywords and the structure of the five keywords with the highest correlation are analyzed. Through this, the patterns and key elements in the political and diplomatic relations between Uzbekistan and the analyzed countries can be identified. The conclusion will summarize the analysis results and discuss their implications.

Research Methodology: GDELT, Text Mining

GDELT data is a CAMEO(Conflict and Mediation Event Observations) event data set developed by Kalev H. Leetaru and Philip A. Schrodtt, including over 2.5 billion events worldwide from 1979 to the present.³ It collects data from broadcasts, print, and online news articles written in over 100 languages, capturing the context of global

² 이지은(Lee Jieon). [우즈베키스탄 새로운 리더십과 변화(Uzbekistan’s New Leadership and Changes)]. 2019, 『Korean Journal of the Middle East Studies』, no.40(1):1, pp.1-24.

³ Hammond J., Weidmann N.B. [Using machine-coded event data for the micro-level study of political violence]. Research & politics, 2014, no. 1(2), pp. 1-8.

issues, the actors involved, the nature and tone of events, and more.⁴ The CAMEO event classification system, used to identify the nature and types of events, inherited and developed from the WEIS (World Event/Interaction Survey) project conducted from 1966 to 1978. The GDELT data, collected from original sources, is automatically detected and categorized according to the CAMEO event classification system by the TABARI (Textual Analysis by Augmented Replacement Instructions) program, which identifies information about actors, targets, event types, event dates, event locations, and the tone of the event. The CAMEO system broadly classifies events into 20 categories, further subdividing these into 290 specific event types. While primarily a cooperative and conflictual binary classification system, it also includes neutral events, allowing for the identification of cooperative, conflictual, and neutral events.⁵

Due to the ease of event classification provided by the CAMEO system, GDELT data can easily identify the nature of events involving two or more actors during a specific period. Applied to international relations research, GDELT data serves as a valuable quantitative resource for understanding how inter-state relations have evolved over a desired analysis period. The widespread use of GDELT data in various fields of social science research attests to its utility. Representative research examples include the time-series analysis of cooperative and conflictual dynamics in North-South Korean relations using GDELT data in Korea,⁶ constructing political distance networks among Asian countries and analyzing international cooperation patterns in Japan,⁷ and quantitatively analyzing changes in Turkey’s ruling party’s foreign policy from 2015 to 2020 using GDELT data alongside qualitative analysis.⁸ Additionally, studies conducted in China analyzing interactions of cooperation and conflict among the U.S., Russia, and China through quantitative analysis of event data showcase the diverse applications of GDELT data.⁹ Reviewing GDELT data and related research examples revealed that while it is highly feasible for international relations research, there are almost no examples of quantitative analysis, including GDELT data, in studies on Uzbekistan’s international relations. This study addresses this gap by choosing to use GDELT data for a quantitative analysis of Uzbekistan’s international

⁴ Leetaru K., P.A Schrod. [GDELT: Global data on events, location, and tone, 1979-2012]. ISA annual convention, San Francisco, 2013, pp. 1-49.

⁵ 정승철, 2021, op. cit., 3-5.

⁶ Seungwoo YI, Jasmine Jeong. [Time-series Analyses for Inter-Korean Relations with GDELT Data Set]. Gyeonggi Research Institute, 2020, no. 5, ; Seungwoo Yi., Jaehyoung Lim. [Dynamics of Inter-Korean Cooperation and Dispute: 2000~2020]. Dispute resolution studies review, 2021, no. 19(2), pp. 65-96. Etc.

⁷ Sotaro S., Keita O., Fusanori I., Yuichi I. [International cooperation analysis of Asian political distance network constructed using event data]. Frontier in Physics., 2022, no. 10, pp. 1-16.

⁸ Hikmet M., Sadullah Ç. [What does “Big data” Tell? A network Analysis Approach to The Justice and Development Party’s Role Performance in The Middle East Between 2015 and 2020]. Journal of Liberty and International Affairs, no. 9(1), pp. 48-72.

⁹ Yuan L., Song C., Cheng., Shen S., Chen X., Wang Y. [The cooperative and conflictual interactions between the United States, Russia and China: A quantitative analysis of event data]. Journal of Geographical Sciences, no. 30(10), pp. 1702-1720.

relations as an initial step. GDELT data will be used to extract event data for Uzbekistan’s relations with China, Russia, and South Korea, and to derive keywords through text mining based on the CAMEO event classification.

Text mining involves the use of knowledge from linguistics, mathematics, statistics, and computer science to extract meaningful information or concepts from text. The main objective is to derive implications from collections of texts by extracting themes or issues and the keywords that constitute them. Like event data, text mining has been actively utilized in various social science fields due to advancements in data accumulation technology and natural science engineering.

Text mining generally involves the processes of data collection, preprocessing, and keyword extraction. Preprocessing and formatting text for research involves natural language processing (NLP) and morphological analysis, followed by tokenization, which breaks sentences into parts of speech. The refined text then undergoes stemming or lemmatization to extract keywords.¹⁰ To analyze the relationships between countries, the co-occurrence frequency of keywords—how often specific keywords appear together—is measured. For instance, if "cooperation (keyword a)" and "economy (keyword b)" frequently appear together in a large dataset of events between Country A and Country B, these keywords have a high co-occurrence frequency, indicating that "economic cooperation" is a major issue between the two countries. This study uses GDELT data for text mining, follows the data preprocessing steps and keyword extraction process, and derives keywords for Uzbekistan’s relations with China, Russia, and South Korea from 2017 to 2023. The derived keywords are then analyzed for co-occurrence frequency and visualized as a Keywords Correlation Matrix, which helps identify patterns and key elements in the political and diplomatic relations between Uzbekistan and the analyzed countries.

Extraction and Correlation Analysis of Keywords by Country.

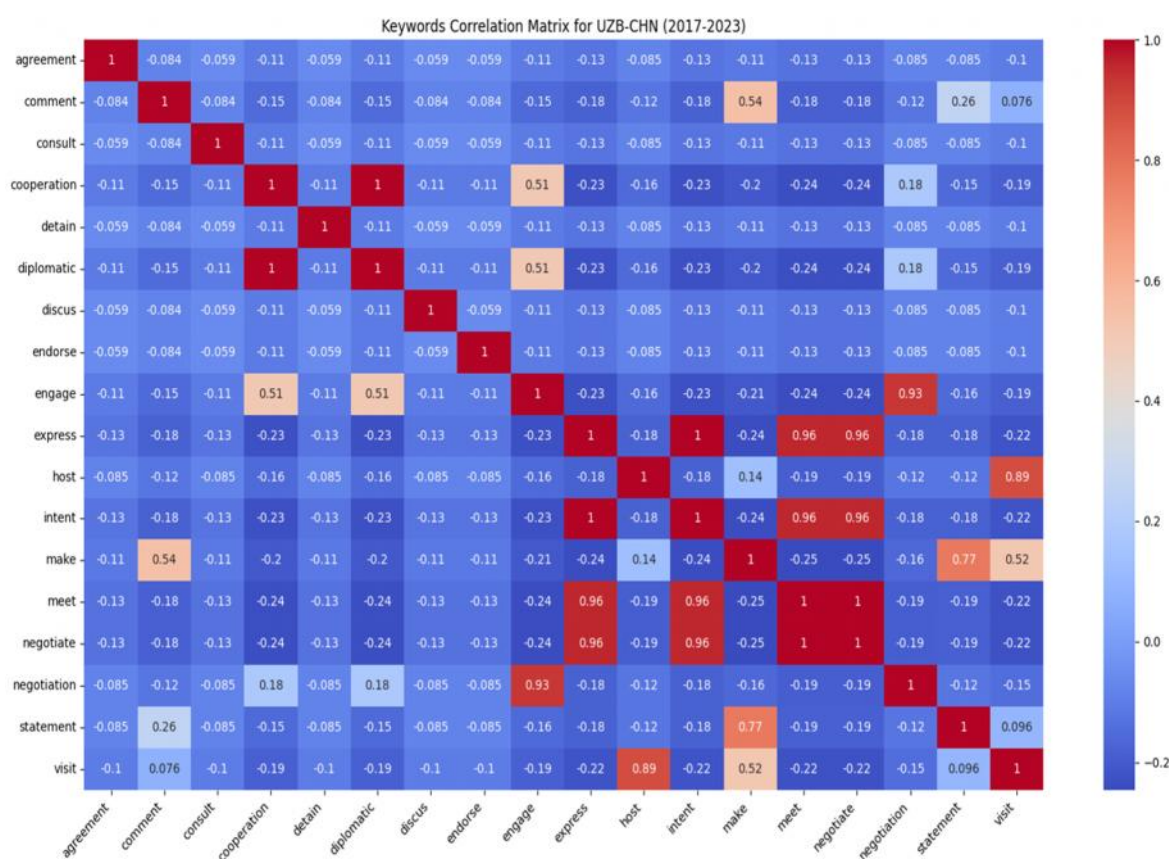
Using the extracted keywords, the keyword correlation indicators for Uzbekistan’s relations with China, Russia, and South Korea during the rule of President Shavkat Mirziyoyev from 2017 to 2023 were visualized as matrices. The x and y axes of the matrix are the keywords derived from the event data between the two countries, and the intersection points show the correlation coefficients calculated using Pearson’s correlation formula. Therefore, the lower-right diagonal of the matrix has a value of 1, and the closer the coefficient is to 1, the higher the correlation between the keywords;

¹⁰ 이한준(Lee Hanjun). [한국언론은 어떻게 신생 대통령 권력을 정당화하는가?(How Korean Press Legitimizes New Presidential Power: Applying Max Weber’s Concept of Legitimacy and Text Mining Analysis)]. M.S. thesis, Hankuk University of Foreign Studies, 2024.

the closer it is to -1, the lower the correlation. Thus, practically, the keywords with positive correlation coefficients provide insight into inter-state relations.

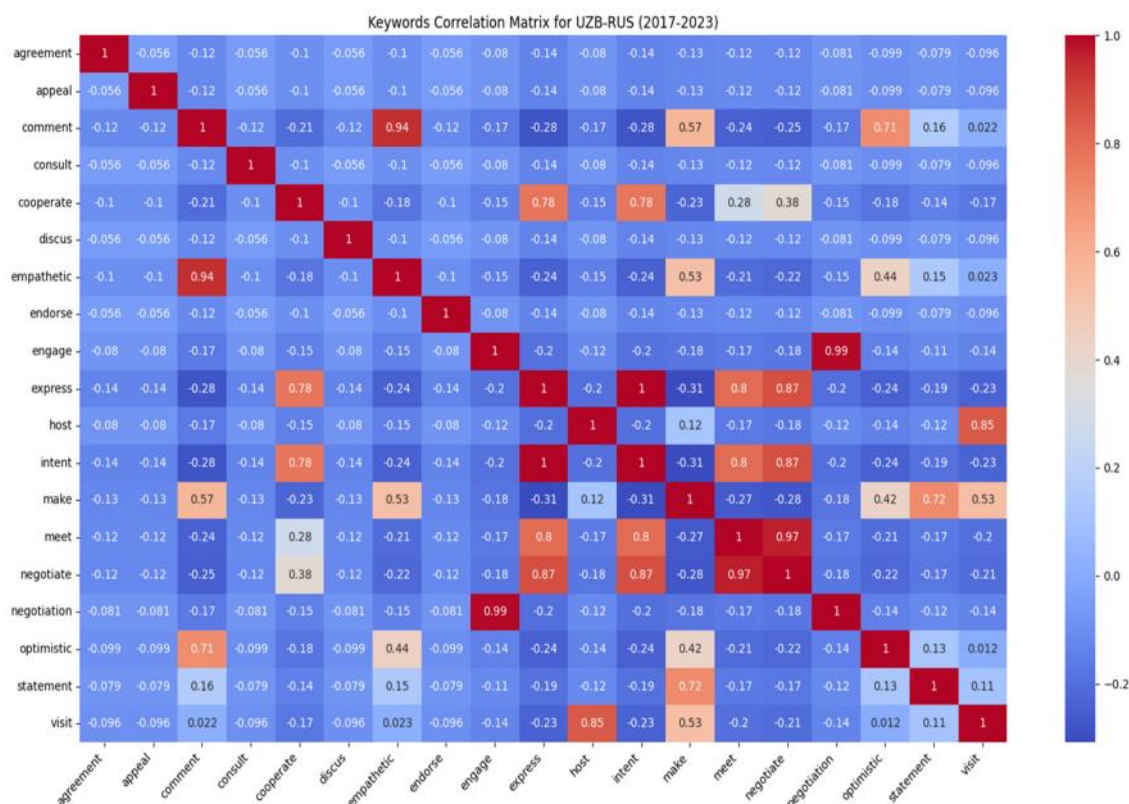
From the event data between Uzbekistan and China from 2017 to 2023, the derived keywords with significant correlation coefficients were "comment," "cooperation," "diplomatic," "engage," "express," "host," "intent," "make," "meet," "negotiate," "negotiation," "statement," and "visit." Keywords such as "diplomatic" and "cooperation," "express" and "intent," "meet" and "negotiate" were treated as the same keywords. Hence, they were analyzed as "diplomatic cooperation," "express intent," and "meet negotiate." The positively correlated keywords included "express intent" and "meet negotiate" (0.96), "engage" and "negotiation" (0.93), "host" and "visit" (0.89), "make" and "statement" (0.77), and "make" and "comment" (0.54). This suggests that from 2017 to 2023, the bilateral relationship prominently featured expressions of intent for meetings and negotiations, actual negotiations, diplomatic cooperation, invitations and visits, and joint statements or comments (see Figure 1).

Figure 1. 2017~2023 Keyword Correlation Matrix for UZB-CHN



From the event data between Uzbekistan and Russia from 2017 to 2023, the derived keywords with significant correlation coefficients were "comment," "cooperate," "empathetic," "engage," "express," "host," "intent," "make," "meet," "negotiate," "negotiation," "optimistic," and "statement." "Express" and "intent" were treated as the same keyword, analyzed as "express intent." Keywords with significant positive correlations included "engage" and "negotiation" (0.99), "meet" and "negotiate" (0.97), "empathetic" and "comment" (0.94), "express intent" and "negotiate" (0.87), "host" and "visit" (0.85), "express intent" and "meet" (0.80), and "express intent" and "cooperate" (0.78). This indicates that the relationship featured a significant number of negotiation-related events, with meetings playing a major role. The goals appeared to be cooperation, with expressions of intent for negotiations and cooperation, actual negotiations and cooperation, and joint declarations or comments that included empathetic or optimistic elements (see Figure 2).

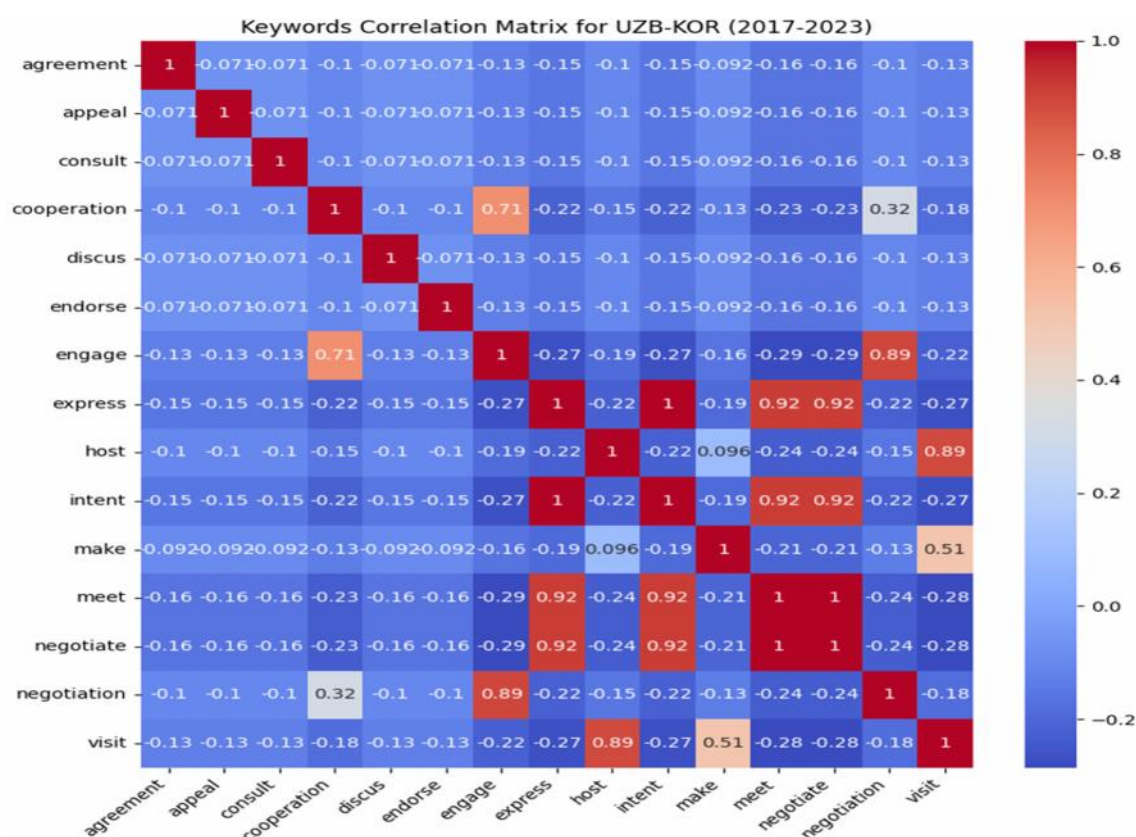
Figure 2. 2017~2023 Keyword Correlation Matrix for UZB-RUS



From the event data between Uzbekistan and South Korea from 2017 to 2023, the derived keywords with significant correlation coefficients were "cooperation," "engage," "express," "host," "intent," "make," "meet," "negotiate," "negotiation," and

"visit." "Express" and "intent," and "meet" and "negotiate" were treated as the same keywords, analyzed as "express intent" and "meet negotiate." Keywords with significant positive correlations included "express intent" and "meet negotiate" (0.92), "engage" and "negotiation" (0.90), "host" and "visit" (0.89), "engage" and "cooperation" (0.71), "make" and "visit" (0.51), and "negotiation" and "cooperation" (0.32). This suggests that the relationship primarily featured expressions of intent for meetings or negotiations, sought or participated in negotiations, invited or visited each other, and aimed for or participated in cooperation (see Figure 3).

Figure 3. 2017~2023 Keyword Correlation Matrix for UZB-KOR



Based on the analysis results, the key keyword correlation structures for Uzbekistan's relations with the three countries were summarized. The results showed that keywords such as "engage" and "negotiate," "express intent" and "meet negotiate," "host," and "make" and "visit" were the main correlation indicators. Moreover, the keyword correlations between Uzbekistan and China, and Uzbekistan and Russia included rhetorical relations indicated by keywords like "make" and "comment," "make" and "statement," which were not included in the relations with South Korea. Detailed results are shown in Table 1.

As indicated by the analysis results, the main keyword correlation structures derived for Uzbekistan’s relations with the analyzed countries over the entire period show similar patterns. It is important to consider that the significance of keyword extraction through text mining and the analysis of correlation indicators lies in identifying the main elements that exist and operate in political and diplomatic relations between countries through keywords. In other words, consistent patterns or trends in inter-state relations can be discovered, which are useful for evaluating the impact of long-term relationship changes or changes in foreign policy during specific periods on international relations.

Table 1. Top 5 Keyword Correlation Between UZB-CHN, UZB-RUS, UZB-KOR

UZB-CHN	UZB-RUS	UZB-KOR
express intent / meet negotiate	engage / negotiation	express intent / meet negotiate
engage / negotiation	meet / negotiate	engage / negotiation
host / visit	express intent / negotiate	host / visit
make / statement	host / visit	make / visit
make / comment	express intent / meet	negotiation / cooperation

In light of these considerations, the analysis results suggest that the political and diplomatic relations between Uzbekistan and the analyzed countries are structured around a few key keywords that lead to other keywords. Keywords such as "engage," which means "seeking or actual action," and "express intent," meaning "expression of intention," along with "make," originally meaning "to create" but closer to "to do" when combined with nouns, and "host," meaning "to invite or host," are central, followed by "negotiation," meaning "negotiation," "cooperation," meaning "cooperation" or "diplomatic cooperation," "meet negotiate," meaning "meet or negotiate," "visit," meaning "visit," "statement," meaning "statement," and "comment," meaning "comment." Thus, Uzbekistan has primarily focused on seeking or participating in negotiations, (diplomatic) cooperation, and meetings with the six countries, expressing these intentions, and inviting or visiting counterparts. Additionally, it has consistently made statements or comments about counterpart countries, mostly in a positive tone as seen in the previous sections of period-specific and country-specific analyses.

Conclusion.

Summarizing the analysis results, the key keyword correlation structures derived from the event data between Uzbekistan and the three analyzed countries center around keywords such as "engage," "express intent," and "make." These keywords represent "seeking or actual participation," "expression of intention," and "to do," respectively, and are combined with keywords like "negotiation" or "cooperation," "diplomatic cooperation," "meet," and "visit" to represent diplomatic actions. The patterns identified in the event occurrence types between Uzbekistan and the analyzed countries through the main keyword correlations can be summarized as seeking or actual participation in negotiations, expressions of intent for negotiations or cooperation, pursuing meetings for negotiations and cooperation, and inviting or visiting as methods. This indicates that the main elements of diplomatic relations between Uzbekistan and the analyzed countries are negotiations, cooperation, and expressions of intent. The analysis results show that Uzbekistan has actively worked towards establishing cooperative relationships, which were set as the goal or direction of its foreign relations under President Mirziyoyev's administration.

Keyword correlation indicator analysis using event data provides significant insights into political and diplomatic relations between countries from events. However, it is not a perfect method for analyzing the impact of changes in foreign policy on international relations. This is due to several limitations in the CAMEO event classification system used for text mining, keyword extraction, and correlation analysis. Keyword extraction and correlation analysis using GDELT event data revealed low keyword correlations for conflicting relationships and infrequent events, exposing the limitations of the analysis. This phenomenon occurs because, during keyword extraction from event types composed of multiple words, only the core word of the event type is retained, while other words are treated as stop-words to prevent errors in calculating keyword correlations. This requires meticulous observation and examination by the researcher, indicating that completely relying on specific quantitative data and its analysis methods has limitations.

Despite these limitations, keyword correlation analysis using event types has significant implications. It allows for identifying patterns and key elements in the political and diplomatic relations between Uzbekistan and the analyzed countries. The remaining challenge is to complement the limitations identified in this analysis. The identified limitations suggest the need for diverse quantitative analysis methods and a multifaceted approach in the study of international relations. Therefore, further research incorporating richer and more diverse quantitative data and methods, including keyword correlation analysis through text mining, is anticipated to yield significant insights.

“This work was supported by Hankuk University of Foreign Studies Research Fund (Of 2023)”

References

1. 이지은(Lee Jieon). [우즈베키스탄 새로운 리더십과 변화(Uzbekistan's New Leadership and Changes)]. 2019, 『Korean Journal of the Middle East Studies』, no.40(1):1, pp.1-24.
2. 이한준(Lee Hanjun). [한국언론은 어떻게 신생 대통령 권력을 정당화하는가?(How Korean Press Legitimizes New Presidential Power: Applying Max Weber's Concept of Legitimacy and Text Mining Analysis)]. M.S. thesis, Hankuk University of Foreign Studies, 2024.
3. 정승철(Jeong Seungcheol). [빅데이터로 살펴본 한미 평화관계: 2017-2020(ROK-US peace relations examined through big data: 2017-2020)]. 2021, JPI Peace Forum, no. 239, pp. 1-17.
4. Hammond Jesse and Weidmann Nils B. [Using machine-coded event data for the micro-level study of political violence]. 2014, 『Research & Politics』, no. 1(2), pp. 1-8.
5. Hikmet Mengüslan and Çellik Sadullah. [What does “Big Data” Tell? A network Analysis Approach to The Justice and Development Party's Role Performance In The Middle East Between 2015 and 2020], 2023, 『Journal of Liberty and International Affairs』, no. 9(1), pp. 48-72.
6. Leetaru, K., Phipps A. Schrodt. [GDELT: Global data on events, location, and tone, 1979-2012], 2013, ISA annual convention no. 2(4), pp. 1-49.
7. Linhua, Y., Changqing, S., Chanhxiu, C., Shi, S., Xiaoqiang, C., Yuanhui Wang. [The cooperative and conflictual interactions between the United States, Russia, and China: A quantitative analysis of event data]. 2020, 『Journal of Geographical Sciences』, no. 30(10), pp. 1702-1720.
8. Sada Sotaro, Oikawa Keita, Iwasaki Fusanori, Ikeda Yuichi. [International cooperation analysis of Asian political distance network constructed using event data], 2022, 『Frontiers in Physics』, no. 10, pp. 1-16.
9. Yi Seongwoo and Jeong Jasmine. [Time-series Analyses for Inter-Korean Relations with GDELT Data Set]. Gyeonggi Research Institute, 2020, no. 5
10. Yi Seongwoo and Lim Jaehyoung. [Dynamics of Inter-Korean Cooperation and Dispute: 2000~2020]. Dispute resolution studies review, 2021, no. 19(2), pp. 65-96.
11. The GDELT Project, <https://www.gdeltproject.org/>. (accessed 24.06.24.2024)