

The Security Council in Its Entirety: Unveiling 80 Years of Deliberation through the UNSC Meetings and Speeches Dataset

Takuto Sakamoto^{*1,2}, Tomoyuki Matsuoka¹, and Hiroto Ito¹

¹Graduate School of Arts and Sciences, University of Tokyo

²Institute of Digital Observatory, University of Tokyo

February 26, 2025

Abstract

This article presents a new dataset that compiles every available public transcript of the United Nations Security Council (UNSC) from 1946 to 2024 in a machine-readable format. Encompassing over one hundred thousand speeches and more than eighty-seven million words, it offers unprecedented historical depth, capturing both Cold War-era debates and post-1990 expansions in Council activities. The dataset enables analyses of meeting frequencies, resolution adoption rates, evolving rhetorical emphases (e.g., on “sovereignty” versus “human rights”), thematic patterns revealed by topic modeling, and shifts in sentiment across different periods. By spanning nearly eight decades of deliberations, this resource supports diverse research agendas on global security norms, the roles of major powers and smaller states, and the expanding influence of UN officials in shaping Council discourse.

1 Introduction

How have conceptions of international security evolved in the seventy-plus years since the establishment of the United Nations Security Council (UNSC)? What do the Council’s public deliberations—from the heated controversies of the early Cold War to the multifaceted

^{*}Corresponding author. Email: sakamoto@hsp.c.u-tokyo.ac.jp

debates of the post-9/11 era—tell us about shifting understandings of conflict, sovereignty, and human rights? Existing studies have often examined whether specific words or topics were mentioned in meetings (Allen and Bell 2022) or placed on the agenda (Lundgren and Klamberg 2023). More recently, the development of large-scale speech corpora has enabled deeper discourse analysis through quantitative text methods. For instance, Eckhard et al. (2021) used a UNSC speech corpus and speaker-topic network analysis to show that UN bureaucrats shaped and advanced the agenda in debates on Afghanistan. Even so, most existing datasets still provide only partial records of UNSC activities, particularly for the pre-1990 period (e.g., Schönfeld et al. 2019). Although significant text-based analyses have emerged—focusing on specific cases, agenda items, or relatively recent meetings (e.g., Hanania 2020; Sakamoto 2023; Knapp 2024)—comprehensive coverage from 1946 onward has remained elusive.

This article introduces the **UNSC Meetings and Speeches dataset**, which compiles *every available official, public transcript* from January 1946 through December 2024 in a machine-readable format. By transcending the limits of earlier post-Cold War-focused datasets, this collection offers a unique opportunity to observe how major powers, non-permanent members, and UN bureaucrats have articulated—and contested—key global security concepts across historical eras. Combining over 100,000 speeches (over 160,000 if brief procedural remarks are included) and spanning more than 87 million words in total, this dataset enables research on numerous topics: from rhetorical shifts in defining “threats to the peace” (Sakamoto 2023) to patterns of humanitarian discourse that gained prominence in the 1990s (Hanania 2020; Knapp 2024), to the evolving role of international civil servants within the Security Council (Eckhard et al. 2021).

Below, we situate this new resource in the existing text-as-data literature in international relations (IR) and UN studies. After clarifying the data construction methodology, we illustrate the dataset’s scholarly potential via exploratory analyses of meeting frequencies, speech volumes, thematic shifts, and sentiment trends—covering the entire Cold War and post-Cold War periods. We conclude by highlighting how this dataset can unlock new theoretical and empirical lines of inquiry, especially about the Security Council’s transformation from a narrowly focused Cold War instrument into a potentially more inclusive forum, albeit one still marked by power politics and normative disagreements.

2 Literature and Existing Text-Based Approaches

2.1 Textual Data in International Relations

The computational turn in political science (Grimmer and Stewart 2013) has led to a flourishing of text-based studies in IR, enabling large-scale systematic investigation of governments' statements, IO resolutions, and public debates. By “taking text seriously,” IR scholars can examine how states construct shared meanings (Wendt 1995), identify rhetorical strategies in crisis bargaining, and detect normative shifts across time. Many have thus leveraged *public speeches and documents* to locate patterns that might remain invisible to purely qualitative methods.

2.2 UN Speech Datasets and Their Limitations

Specifically within UN research, several text-based studies have advanced the discipline. Knapp (2024) analyzes the development of the norm of civilian protection using an original corpus of UN General Assembly and Security Council resolutions. Hanania (2020) examines long-term normative changes based on a corpus of UN Security Council resolutions from 1946 to 2017. Bethke, Haass, and Niemann (2024) investigate shifts in the concept of “responsibility” in Security Council resolutions from 1946 to 2020. As research utilizing UN textual data expands, projects aimed at constructing large-scale corpora are also increasing:

- **UN General Debate Corpus (UNGDA) (Jankin et al. 2024).** This dataset compiles annual addresses by heads of state in the General Assembly (UNGA), enabling cross-national comparisons of policy priorities and rhetorical change over decades. Research utilizing the UNGDC has advanced in areas such as the securitization process (Arias 2022), the distinctive characteristics of political leaders' speeches (Baturu and Gray 2024), and support for or challenges to the liberal international order (Kentikelenis and Voeten 2021). However, these once-a-year speeches only partially reflect the routine decision-making processes of the UN system, leaving day-to-day deliberations uncharted.
- **UNGA Speech Corpus (Turco 2024).** Expanding beyond the high-level General Debate, this collection gathers a broader set of statements delivered in the UNGA. It adds important nuance to how states discuss pressing issues (e.g., climate change, development) in more “everyday” settings, yet it does not cover the Cold War period.
- **The UN Security Council Debates (Schönfeld et al. 2019).** A pioneering text corpus targeting the UNSC, these transcripts (1992–2023) shed light on post–Cold War

dynamics, including the roles of key actors such as permanent members and UN officials (Eckhard et al. 2021). Although crucial for systematic analyses of policy deliberations in the UNSC, it lacks coverage of earlier decades. Consequently, it cannot address how Cold War–era negotiations, veto politics, or certain historical conflicts shaped Council discourse.

- **Earlier version of The UNSC Meetings and Speeches (Sakamoto 2023).** An earlier version of The UNSC Meetings and Speeches also offers a UNSC speech corpus, covering the period from 1990 to 2023. Using this corpus, Sakamoto (2023) analyzed how “threat to peace” has been discussed through word embedding. Sakamoto et al. (2024) examined differences between offline and online diplomatic discourse by comparing UNSC speeches with the social media activity of UN permanent representatives. Still, many older transcripts (especially from 1946 to the late 1980s) were outside the scope of these data resources.

2.3 Filling the Gap

The **UNSC Meetings and Speeches dataset** addresses this fundamental gap by including every available public meeting transcript from the Council’s inception (Table 1). This historically deep approach illuminates how different periods shaped (and continue to shape) UNSC norms and practices. For instance, we can now systematically examine how the language of sovereignty, intervention, or collective security was deployed during the Korean War debates of the early 1950s, the Suez crisis of 1956, or the pivotal year of 1989 when the Cold War began to wane. By integrating modern computational techniques with archives dating back to 1946, this dataset provides a longitudinal foundation for IR scholars to evaluate continuity, change, and the interplay between superpowers, regional actors, and the UN bureaucracy itself.

Moreover, from a theoretical perspective, the dataset aligns with *constructivist* frameworks emphasizing the role of discursive practices (Johnstone 2011). It also supports *realist* or institutionalist analyses that investigate whether states’ rhetorical strategies correspond to their material interests or, conversely, how they might adopt normative frames to justify certain positions (Hanania 2020). In short, it opens the door to a wide range of empirical and theoretical projects, building upon the valuable, but temporally limited, corpora introduced in prior research (Schönfeld et al. 2019; Turco 2024; Jankin et al. 2024).

Table 1: Comparison of Major UN Document Corpora

Dataset	Speech Types	Coverage Period	Corpus Size
UNSC Meetings and Speeches Dataset	UNSC speeches	1946–2024	167,642 speeches; ~87M words
The UN Security Council Debates (Schönfeld et al. 2019)	UNSC speeches	1992–2023	106,302 speeches
UN General Debate Corpus (Jankin et al. 2024)	UNGA General Debate	1946–2023	10,760 speeches
UNGA Speech Corpus (Turco 2024)	UNGA speeches	1993–2018	34,058 speeches

Note: Adapted from Schönfeld et al. 2019; Jankin et al. 2024; Turco 2024; Sakamoto 2023.

3 Dataset and Its Construction

The **UNSC Meetings and Speeches dataset**, which is publicly available from the Harvard Dataverse, is constructed from a wealth of archival material provided by the United Nations. Our primary data sources are the UN Official Document System and the Dag Hammarskjöld Library’s online repositories. The process involved several meticulous steps to ensure both breadth and quality:

3.1 Data Acquisition

We initiated the project by scraping the UN Security Council Meetings & Outcomes Tables to retrieve meeting-specific information such as dates, agenda items, document symbols (e.g., “S_PV8942”), and outcome details (e.g., adopted resolutions, voting results). Simultaneously, PDF documents containing the official verbatim records were downloaded via hyperlinks from the UN Official Document System. This initial phase provided the structural backbone for the dataset.

3.2 Text Extraction and Cleaning

PDFs containing the verbatim records of UNSC meetings were processed using ABBYY FineReader, which enabled high-precision extraction of text. Given the diversity of document formats spanning eight decades, our cleaning process employed advanced regular expressions to correct formatting issues, remove non-ASCII characters, and accurately segment texts into

discrete speech contributions. Manual corrections by research assistants further enhanced the quality, especially for older scanned documents where OCR errors were more prevalent.

3.3 Data Structuring

The extracted information was organized into two main files:

- **Meeting records (meetings.tsv):** This file contains comprehensive metadata for each UNSC meeting. Variables include unique record identifiers, date (year, month, day), document names, meeting numbers, agenda items, the president’s name and country, indicators for speech availability, word counts, and links to official documents such as resolutions and presidential statements.
- **Speech records (speeches.tsv):** This file includes detailed records for each speech. Variables encompass unique speech identifiers, linkage to meeting records, speaker names, affiliations, position titles, procedural flags (to distinguish between routine and policy-oriented statements), word counts, full transcript texts, and auxiliary variables indicating actor types (e.g., permanent member, elected member, international organization).

In addition, we enhanced the dataset by cross-referencing speaker information with the UN Digital Library to ascertain correct affiliations and using external datasets (e.g., Correlates of War) for standardized country codes. Dummy variables indicating actor categories further facilitate quantitative analysis.

3.4 Quality Control

To ensure the dataset’s reliability, we conducted systematic sampling and manual verification of text extractions. Although some discrepancies remain—especially in legacy documents—the overall quality is robust enough to support diverse quantitative and qualitative analyses. The extraction and segmentation code is publicly available on GitHub, ensuring transparency and reproducibility.

4 Illustrations: The Eighty Years of the UNSC

To demonstrate the dataset’s potential, we summarize five exploratory analyses of the entire 1946–2024 timespan. Specifically, we computed descriptive statistics, word-frequency trends, topic structures, and sentiment patterns. These illustrations offer a glimpse of how the dataset can illuminate fundamental shifts in the Council’s discursive landscape.

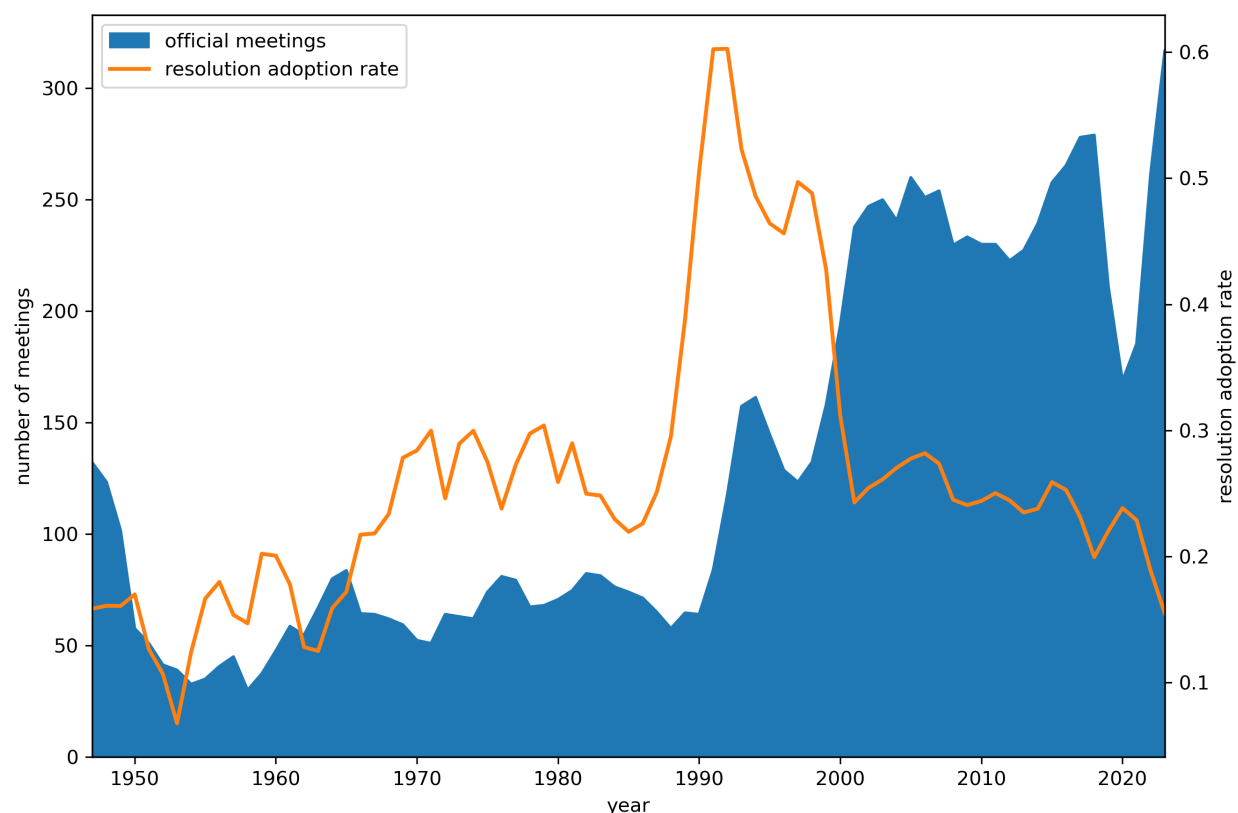


Figure 1: Frequency of Security Council meetings (three-year moving averages) and proportion of sessions producing formal resolutions, 1946–2024.

4.1 Meeting-Level Trends

We first examined the *frequency of Security Council meetings* (three-year moving averages) from 1946 to 2024, alongside the proportion of those sessions that produced formal resolutions (see Figure 1). While the Cold War was marked by fluctuations—sometimes driven by acute crises—overall session numbers increased dramatically around the early 1990s. By the late 1990s and early 2000s, meeting volumes rose to unprecedented levels, only leveling off in recent years. Interestingly, the proportion of meetings yielding resolutions peaked in the 1990s and gradually decreased thereafter, possibly indicating that routine or follow-up sessions (e.g., renewing existing mandates) increasingly occupy the Council’s agenda, reducing the fraction of new binding decisions relative to the total.

4.2 Speech Volume by Key Actors

We then calculated overall *speech and word counts* disaggregated by major powers, selected non-permanent states, and the UN itself (see Table 2). Notably:

Table 2: Speech and Word Count Statistics by Major Actors

Country/Organization	1946-2024 Speech Count	1946-2024 Word Count	1946-1989 Speech Count	1946-1989 Word Count	1990-2024 Speech Count	1990-2024 Word Count
Russia/Soviet Union	6,972	5,024,205	3,569	2,409,589	3,403	2,614,616
United States of America	5,981	3,943,762	2,682	1,431,707	3,299	2,512,055
United Kingdom	5,214	3,179,338	2,179	1,231,585	3,035	1,947,753
France	4,548	2,761,267	1,577	696,052	2,971	2,065,215
China	4,074	2,149,025	1,105	502,651	2,969	1,646,374
Japan	1,728	994,042	247	125,279	1,481	868,763
Egypt	1,527	1,241,158	833	661,339	694	579,819
Brazil	1,449	907,985	276	124,946	1,173	783,039
India	1,441	1,585,769	610	958,073	831	627,696
United Nations	5,188	6,557,513	903	420,408	4,285	6,137,105
Total (excl. procedural statements)	101,493	82,382,277	30,724	25,389,352	70,769	56,992,925
Total (incl. procedural statements)	167,642	87,678,254	53,254	27,946,504	114,388	59,731,750

Note: Produced by the authors.

- **Russia (including the Soviet era)** leads in the total number of speeches, followed by the United States, then the United Kingdom and France (the other permanent members). China shows a lower volume historically but experiences a sharp post-1990 increase.
- **UN officials** (e.g., Secretary-General, Under-Secretary-General) have collectively spoken more words than any single permanent member in the post-Cold War period. More than 90% of their statements’ total words appeared after 1990, reflecting the shift toward frequent Secretariat briefings, thematic debates (Knapp 2024), and expansions of the UN bureaucracy’s role (Eckhard et al. 2021).
- **Non-permanent members** like Japan and Brazil see their share of total speech volume grow after the 1990s, while certain historically prominent countries like Egypt hold diminished relative footprints.

4.3 Changing Rhetorical Emphasis: “Sovereignty” vs. “Human Rights”

We probed how references to “sovereignty” and “human rights” (per 1,000 words) evolved across the full 1946–2024 range (see Figure 2). The results affirm that “**sovereignty**” dominated Cold War discourse, echoing the era’s norm of non-interference, whereas “**human rights**” surged after 1990. In the last decade, “human rights” usage has stagnated while “sovereignty” has modestly rebounded—suggesting new friction over how far the Council should intrude into domestic matters. Disaggregation by speaker reveals that China is the strongest proponent of “sovereignty,” while the UN Secretariat leads in “human rights” mentions. Among the P5, an apparent rhetorical split has emerged in the 2010s, surpassing even the polarization of certain Cold War periods (see Figure 3).

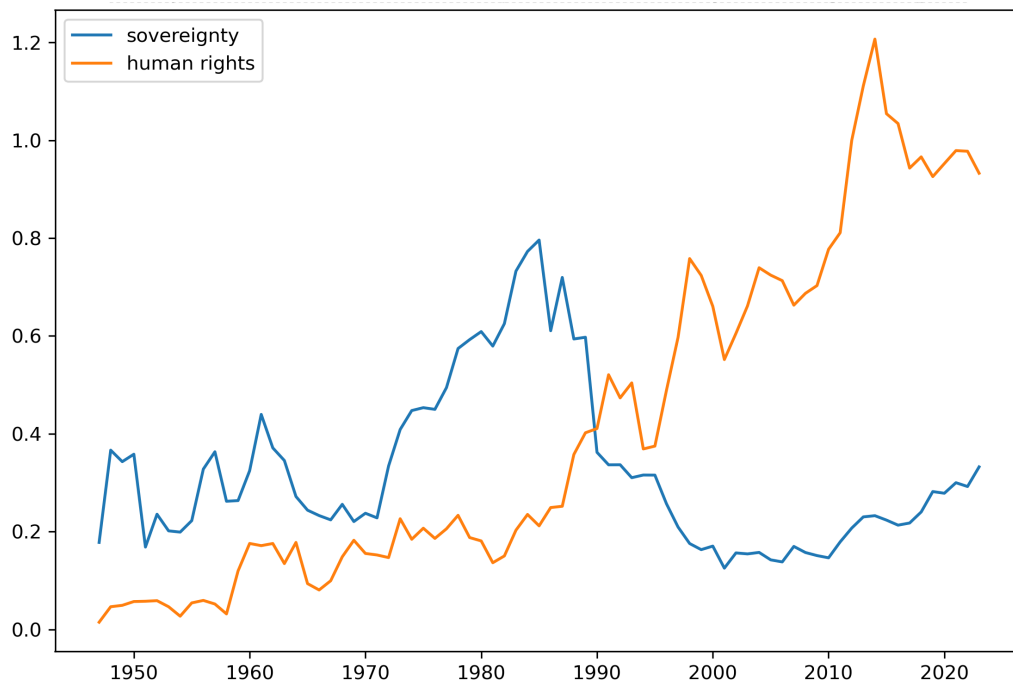


Figure 2: Evolution of references to “sovereignty” and “human rights” (per 1,000 words; 3-year MA) from 1946 to 2024.

4.4 Topic Structure via BERTopic

Using a BERTopic model—a topic modeling technique based on BERT sentence embeddings (Grootendorst 2022), we clustered over one million paragraphs (excluding short procedural remarks) to detect distinct thematic areas (see Supplementary Information). While about one-quarter of the paragraphs coalesce into a “general Council business” category, numerous specialized topics emerged, notably:

- **Israel–Palestine:** By far the largest country-specific conflict topic, encompassing over 10% of paragraphs. Its frequency aligns with key conflict moments, with 2023–2024 data near historical highs.
- **Women, Peace, and Security (WPS):** Gains prominence after the late 1990s, mirroring the Council’s official WPS agenda. Related topics such as “protection of civilians” and “children and armed conflict” likewise intensify in the same period, paralleling well-known Council innovations like Resolution 1325.
- **Country-specific agenda items:** e.g., “Syria,” “Afghanistan,” “Democratic Repub-

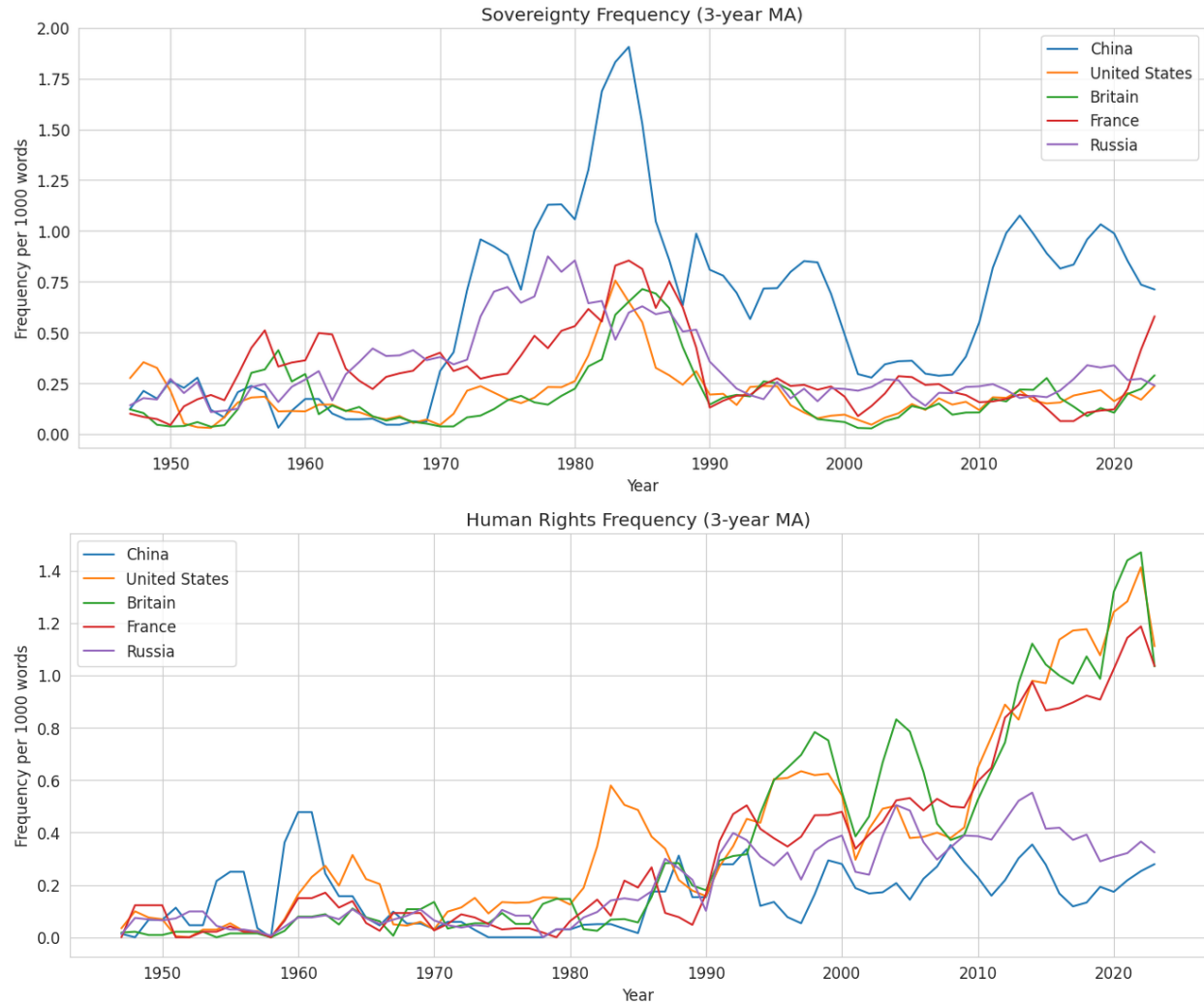


Figure 3: Rhetorical trends of “sovereignty” and “human rights” among P5 members over time.

lic of the Congo,” and “Sudan/Darfur.” Each set of paragraphs experiences time-bound surges, reflecting conflict escalation or new Council resolutions.

Overall, these thematic clusters confirm the Council’s evolution from a primarily geopolitical forum (focusing on major-power conflicts or large-scale crises) to a more varied institutional body hosting repeated thematic debates on humanitarian crises, WPS, and other global security issues (see Figure 4).

4.5 Sentiment Shifts over Time

Finally, in a complementary analysis, we employed a zero-shot sentiment classification using a lightweight large language model (LLaMa 3.2) to categorize over 1.2 million paragraphs from

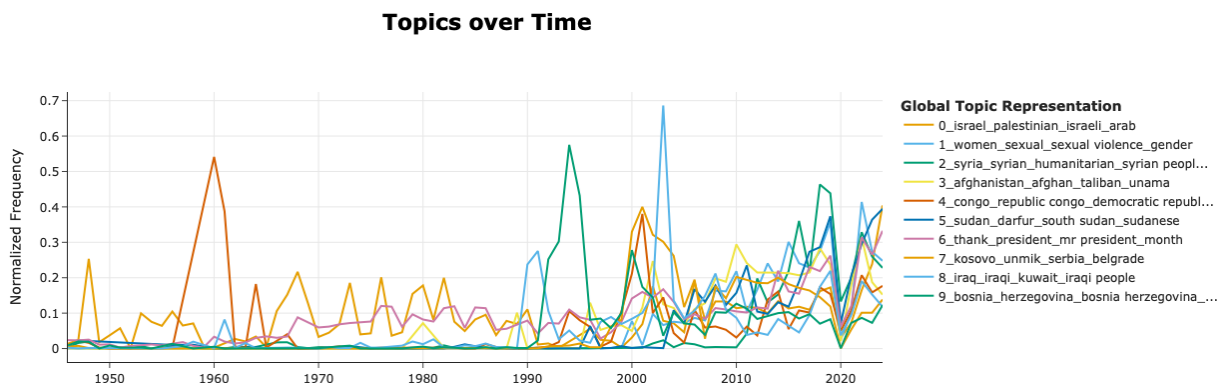


Figure 4: Evolution of the ten most frequent thematic clusters detected using BERTopic (Grootendorst 2022).

the dataset into positive, neutral, or negative tones. Aggregating with three-year moving averages reveals (see Figure 5):

- **Cold War:** A gradual rise of negative and positive sentiments relative to neutral, reflecting intensifying rhetorical clashes at certain points (e.g., the Cuban Missile Crisis era) but also instances of détente or occasional collaboration.
- **1990s Upswing:** After the Cold War, negativity dips while positivity soars, capturing the more cooperative atmosphere of the early 1990s.
- **Mid-2010s Tensions:** Negative sentiment ticks upward again in the past decade, while positive sentiment wanes. This might imply a return to more confrontational rhetoric, consistent with accounts of revived major-power tensions on issues like Syria or Ukraine (Eckhard et al. 2021).

Disaggregating sentiment by speaker group yields further insights (see Figure 6): Russia’s statements lean more negative on average, while China’s are comparatively more positive, echoing patterns identified in other sentiment-based research (Sakamoto et al. 2024). These patterns illustrate how the dataset can capture the tonal contours of Council debates across entire historical epochs.

5 Conclusion

We have introduced the **UNSC Meetings and Speeches dataset**, which, for the first time, collects every publicly available transcript of Security Council meetings from 1946

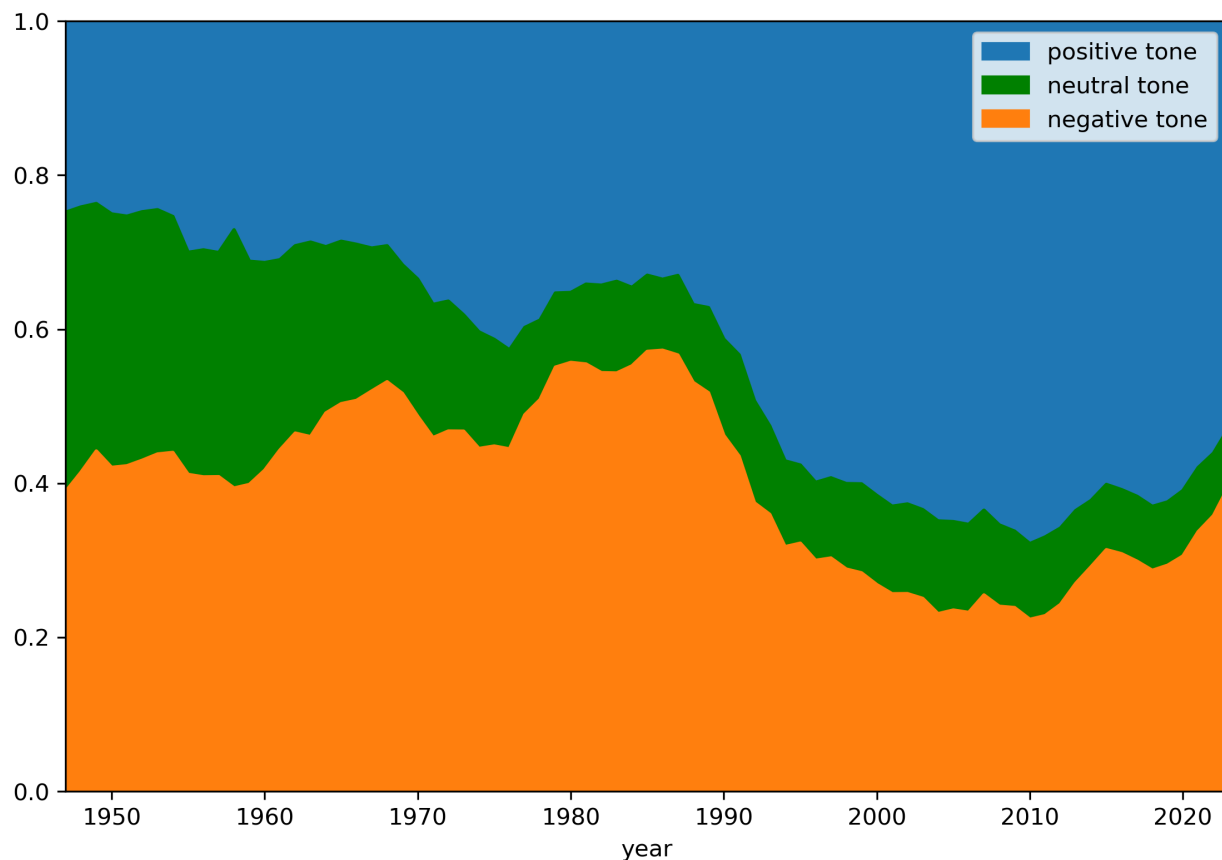


Figure 5: Proportion of positive, neutral, and negative tones over time, based on sentiment classification using LLaMa 3.2.

up to 2024 in a unified, machine-readable format. This project addresses a longstanding need for holistic, historical coverage of the Council’s day-to-day deliberations, bridging a gap identified by many IR scholars who wished to analyze older materials (Hanania 2020; Sakamoto 2023). By combining tens of thousands of speeches totaling more than 87 million words, the dataset offers a robust resource for studying how discursive practices among the P5, non-permanent members, and UN officials have shaped global security norms.

Our illustrative analyses underscore several takeaways. First, the Council’s meeting frequency and resolution adoption patterns exhibit striking shifts between the Cold War and post-Cold War eras, with the 1990s emerging as a high point for formalized Council outputs. Second, the composition of speakers has diversified as UN officials’ statements have grown and certain non-permanent states have become more vocal, demonstrating that Council debates today often feature administrative briefs and extensive thematic discussions (Eckhard et al. 2021). Third, clear rhetorical trends—such as a relative decline in “sovereignty” talk in the 1990s, followed by its more recent resurgence—underscore how language in the Council

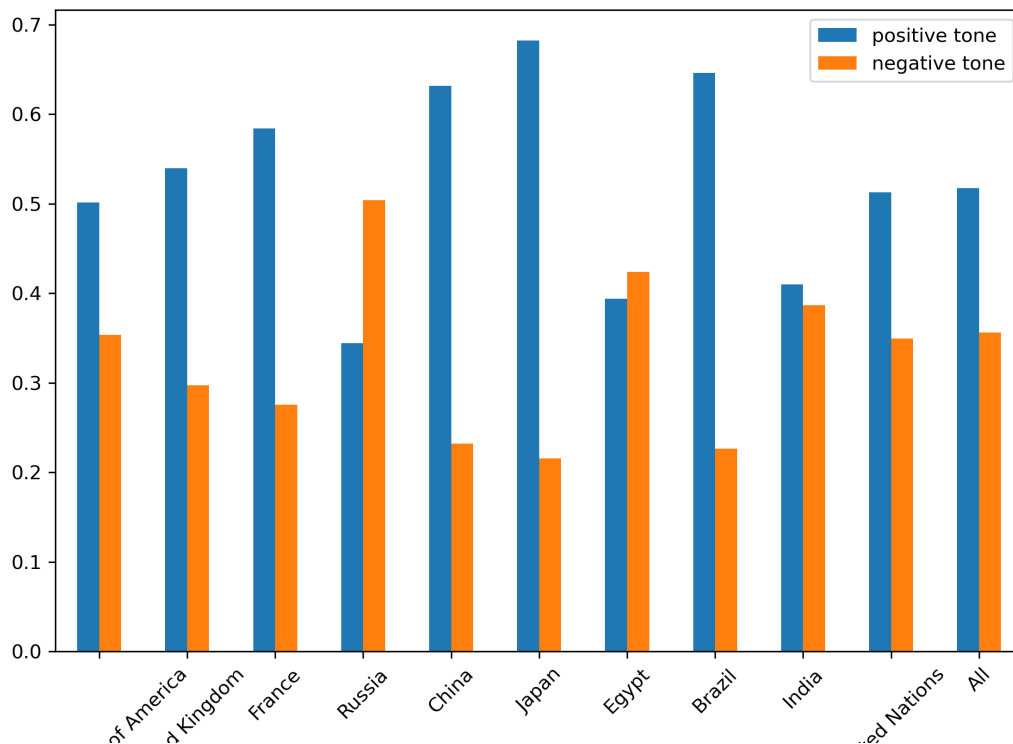


Figure 6: Cross-national comparison of sentiment proportions, based on sentiment classification using LLaMa 3.2.

tracks broader shifts in global power politics and norm contention (Knapp 2024). Finally, sentiment analysis suggests recurring cycles of conflict and cooperation, with post–Cold War optimism giving way to renewed negativity in the 2010s.

For IR theory, the data facilitate new inquiries into the interplay of discursive practices and international outcomes. Constructivists may investigate how certain security frames gain traction; rationalists may ask whether rhetorical shifts precede or follow negotiation successes. Researchers of IO bureaucracies can measure how the Secretariat’s voice, once muted, has come to shape the Council agenda. Additionally, the historically deep coverage allows scholars to track the formation and evolution of norms such as human rights protection, the responsibility to protect, or women’s participation in peace processes (Hanania 2020; Sakamoto 2023).

While our dataset offers unprecedented insights, several challenges remain. The quality of text extraction for older, scanned documents requires ongoing refinement, and the current classification of agenda items could be enhanced by incorporating additional contextual information (e.g., regional or event-specific classifications). Future work will focus on

leveraging emerging technologies, such as more advanced large language models, to further automate text correction and enrich metadata.

By publicly releasing the dataset and documentation, we hope to empower future research on questions that have remained largely unexplored due to limited archival accessibility and incomplete digital resources. This new corpus should facilitate a range of methodological approaches: advanced text-as-data analytics, in-depth qualitative readings, or mixed-method studies that integrate textual data with archival sources on key turning points in UNSC history. Ultimately, the **UNSC Meetings and Speeches dataset** unveils how nearly eight decades of Council debates reflect—and sometimes reshape—fundamental concepts of global peace, security, and governance.

Acknowledgments

The authors would like to express their sincere gratitude for the valuable research assistance provided by Momoko Araki, Kyoungwan Jang, Yuka Matsubara, Mi Kyoung Park, Takuto Tohkairin, and Ryouun Ukita. We also appreciate the support provided by the Institute of Digital Observatory, the University of Tokyo. This research was financially supported by the Japan Society for the Promotion of Science (JSPS) under JSPS KAKENHI Grant Number 22H00815.

References

- Allen, Susan Hannah, and Sam R. Bell.** 2022. “The United Nations Security Council and Human Rights: Who Ends Up in the Spotlight?.” *Journal of Global Security Studies* 7 (4): ogac013.
- Arias, Sabrina B.** 2022. “Who Securitizes? Climate Change Discourse in the United Nations.” *International Studies Quarterly* 66 (2): sqac020.
- Baturo, Alexander, and Julia Gray.** 2024. “Leaders in the United Nations General Assembly: Revitalization or Politicization?.” *The Review of International Organizations* 19 (4): 721–752.
- Bethke, Felix S., Felix Haass, and Holger Niemann.** 2024. “The Language of Responsibility in the United Nations Security Council, 1946–2020.” *International Studies Quarterly* 68 (2): sqae025.
- Eckhard, Steffen, Ronny Patz, Mirco Schönfeld, and Hilde van Meegdenburg.** 2021. “International Bureaucrats in the UN Security Council Debates: A Speaker-Topic Network Analysis.” *Journal of European Public Policy* 1–20.

- Grimmer, Justin, and Brandon M. Stewart.** 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21 (3): 267–297.
- Grootendorst, Maarten.** 2022. “BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure.” *arXiv* 2203.05794.
- Hanania, Richard.** 2020. “The Humanitarian Turn at the UNSC: Explaining the Development of International Norms through Machine Learning Algorithms.” *Journal of Peace Research* 58 (4): 655–670.
- Jankin, Slava, Alexander Baturo, and Niheer Dasandi.** 2024. “Words to Unite Nations: The Complete United Nations General Debate Corpus, 1946–Present.” *Journal of Peace Research*: 00223433241275335.
- Johnstone, Ian.** 2011. *The Power of Deliberation: International Law, Politics and Organizations*. Oxford: Oxford University Press.
- Kentikelenis, Alexander, and Erik Voeten.** 2021. “Legitimacy Challenges to the Liberal World Order: Evidence from United Nations Speeches, 1970–2018.” *The Review of International Organizations* 16 (4): 721–754.
- Knapp, Andrea.** 2024. “Speaking of Civilians: Automated Text Analysis of the United Nations’ Framing of Complex Humanitarian Emergencies.” *International Interactions* 50 (2): 300–331.
- Lundgren, Magnus, and Mark Klamberg.** 2023. “Selective Attention: The United Nations Security Council and Armed Conflict.” *British Journal of Political Science* 53 (3): 958–979.
- Sakamoto, Takuto.** 2023. “Threat Conceptions in Global Security Discourse: Analyzing the Speech Records of the United Nations Security Council, 1990–2019.” *International Studies Quarterly* 67 (3): sqad067.
- Sakamoto, Takuto, Momoko Araki, Hiroto Ito, and Tomoyuki Matsuoka.** 2024. “How Different Are Offline and Online Diplomacy? A Comparative Analysis of Public Statements and SNS Posts by Delegates to the United Nations.” *Frontiers in Big Data* 7: 1–11.
- Schönfeld, Mirco, Steffen Eckhard, Ronny Patz, and Hilde van Meegdenburg.** 2019. “The UN Security Council Debates 1995–2017.” *arXiv* 1906.10969: 1–13.
- Turco, Linnea R.** 2024. “Speaking Volumes: Introducing the UNGA Speech Corpus.” *International Studies Quarterly* 68 (1): sqae001.
- Wendt, Alexander.** 1995. “Constructing International Politics.” *International Security* 20 (1): 71–81.