Topic Modeling and Sentiment Analysis of UN Speeches: A Historical and Geographic Analysis

Introduction

Since 1946, officials from all of the world's 193 countries have gathered at the United Nations building in New York for the UN General Assembly (UNGA). Here, representatives of member countries discuss the most pressing issues of the day. These sessions start with the UNGA general debate, which typically gives each member country of the UN the chance to voice their opinions, concerns, and visions for the future of the international order. The speeches given here offer valuable insight into how different nations and regions view diplomacy, international trade, geopolitical crises, and more. For my final project, I propose the following research question(s): *What topics exist in the speeches at the UN debates? What is the sentiment of these speeches? How do sentiment and topics change over time and between countries? What is the relationship between sentiment and world metrics? Can results be validated with real-world events?*

This report will provide an overview of the data used, an analysis of the methodology employed, and results and next steps for future research. It will ultimately find that while most world metrics (save democracy indicators) do not present statistically significant relationships with the sentiment of speeches in a given country and year, speech topics and sentiment speeches can indeed be validated historically and geographically.

Data

For this project, I utilized the UN General Assembly Debate Corpus, containing opening speeches from each companies UN representative, from 1946 to 2023. The web source for this corpus can be found at the footnote below.[1] This corpus contains information at the country-year level—that is, the year the speech was given, the speaker's country, the text of the speech, as well as other metrics such as democracy score, regime types, gender equity scores, and whether or not the country is considered a regional power. Overall, there are 10,760 speeches and observations, with the average length of these speeches being 2,930 words. While the data was retrieved in a tidy form from Kaggle, the full raw corpus is from Slava Jankin, Alexander Baturo, and Niheer Dasandi and their paper: "Words to unite nations: The complete UN General Debate Corpus, 1946-present," from the Journal of Peace Research (forthcoming). To augment this data, I also used the World Bank's World Development Indicators (WDI) API to retrieve country-year data on GDP per capita and Gini coefficient.

The final data is composed of 2 parts: (1) a corpus and document frequency matrix of all UNGA debate speeches from 1946 to 2023, and (2) the original tabular data frame with the final sentiment scores and topic titles merged on to it. When creating the corpus, all speeches were tokenized following several standard procedural steps: converting all words to lowercase; omitting punctuation, numbers and symbols; removing stop words; and stemming. The final

---

[1] Slava Jankin, Alexander Baturo, and Niheer Dasandi. (2024). NLP with UN General Debate Corpus 1946-2023 [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/9304324.
https://www.kaggle.com/datasets/namigabbasov/united-nations-general-debate-corpus-1946-2023?resource=download

document frequency matrix was comprised only of tokens that were in no less than 5 percent of documents and no more than 95 percent of documents.

Methods & Analysis

I. LDA

I employed Latent Dirichlet Allocation (LDA) topic modeling to decipher which topics in UNGA speeches are prevalent, and to see how these change over time and across regions. After observing different perplexity scores for different values of K (2, 5, 7, and 10), a value of K = 10 was decided due to this value achieving the lowest perplexity score. Gibbs was chosen as the inference method for its accuracy. After observing the top words with the highest beta values for each topic, topic titles were assigned. The topic titles chosen, along with some of the tokens with highest betas scores, can be found below:

1. International Development & Economic Inequality: “countri, intern, econom”
2. Africa & the Global South: “africa, african, peopl, south”
3. The Middle East: “Intern, peace, secur, Arab, israel”
4. War, Conflict, & Diplomacy: “world, must, war”
5. Peacebuilding and Disarmament: “intern, world, peace, nuclear”
6. Climate, Sustainability & Development Goals: “global, develop, climat, sustain”
7. Peacekeeping operations: “nation, unit, power, organ, govern, charter”
8. The Cold War: “peace, war, soviet, forc”
9. Island Nations: “develop, nation, island”
10. International Cooperation: “unit, intern, secur, council, cooper”

II. Sentiment Analysis

In addition to topic modeling, I compared sentiment and topic trends to other social, economic, and government metrics by country and year. To retrieve the sentiment for each speech, the Lexicoder Sentiment Dictionary (LSD) was used. LSD is a bag-of-words sentiment dictionary designed for texts that revolve around policy and news (Soroka 2024). While other sentiment dictionaries like VADER may offer more contextual nuance, they are more computationally expensive and work better on shorter texts. Given the length of each speech in this corpus, LSD was considered a more feasible option.

Once LSD sentiment scores were calculated for each speech, scores and topic instances were compared with indicators like democracy scores, gender equality, regime type, GDP per capita, and Gini coefficient. Trends and changes over time in relation to specific regions and historical events were later analyzed. For instance, after the collapse of the Soviet Union in 1991, did Russia's speeches contain more negative sentiment? Did the fall of the Berlin Wall lead to more positive speeches from German UN reps?

III. Hypothesis testing

For my main multivariate regression, I wanted to observe the effect of metrics like GDP per capita, Gini coefficient, democracy, gender equality score, and regime type on sentiment. However, given the clustered structure of the data (i.e. the country-year schema), simple bivariate and multivariate regressions would be insufficient to see the larger picture. I therefore opted to choose a two-way fixed effects model, with the effects themselves being at the country and year levels.

Results

Sentiment and topic distribution

Of all topic categories, "International Cooperation" had the most speeches in the corpus at 2,002, with "The Cold War" having the least at 472. "The Cold War" also had the lowest mean net sentiment score at 19.6, while "Peacebuilding & Disarmament" had the highest at 71.9. Regarding the distribution of sentiment scores as a whole, the mean sentiment score was 52.9 and the median score was 51. The minimum score was -216 and the maximum was 412. The standard deviation of sentiment scores was 42.35. The year with the highest mean net sentiment score was 1955, at 93, while 2001 had the lowest at 23.5.

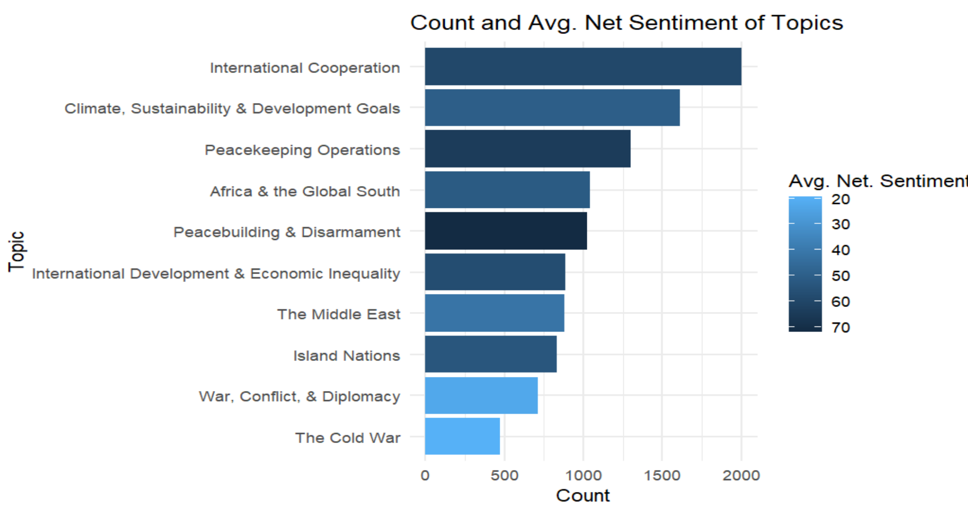


Figure 1: Count and mean net sentiment by topic

Metrics and effects on sentiment

The results from the primary two-way fixed effects multivariate regression (that is, how various metrics affect sentiment while controlling for country and year effects) can be found below:

Two-Way Fixed Effects Multivariate Regression

	Dependent variable: Net Sentiment (LSD)
GDP per Capita (2015 US)	-0.0002
	(0.0004)
Gini Index	0.561
	(0.375)
Non-Western Country (Binary)	10.672*
	(5.889)
VDEM Gender Equality	-52.484
	(32.020)
Democracy (Binary)	12.758**
	(5.414)
Observations	1,424
R2	0.009
Adjusted R2	-0.118
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 2: Multivariate two-way fixed effects model

$$LSD\ Net\ Sentiment_{\mathrm{it}} = \beta 0 + \beta_1 \mathrm{GDP}_{\mathrm{it}} + \beta_2 \mathrm{Gini}_{\mathrm{it}} + \beta_3 \mathrm{Democracy}_{\mathrm{it}} + \beta_4 NonWestern_{\mathrm{it}} + \beta_5 GenderEquality_{\mathrm{it}} + \alpha i + \gamma t + \epsilon it$$

Interestingly, the only statistically significant coefficient is that of the binary democracy variable (1 if democratic, 0 if non-democratic). Results suggest that democracies give speeches with a mean sentiment score 12.7 points higher than those of non-democracies, while controlling for GDP per capita, Gini index, whether or not the country was in the “West,” and gender equality scores. Generally, though, the p-value is only slightly less than 0.05, and the adjusted R-squared is quite small.

Geographic and historical trends

One of the primary motivations for this project was to determine if the outcomes from topic modeling and sentiment analysis could be validated through historical and geographic analysis. Using a dashboard created in R Shiny, I was able to visualize how certain sentiment, topic, and metric trends evolved over time and between countries to determine how well-validated these outcomes were.

I. Temporal sentiment trends by topic

The first tab of the dashboard illustrates a line graph showing the number of speeches made year by topic, along with the mean sentiment for that topic in that year. Some of these speech topics exhibit trends that line up well with historical events. For instance, the "Peacebuilding & Disarmament" topic (which generally deals with nuclear weapons and détente) has two large spikes in the mid 1970s and late 1980s. These roughly coincide with the SALT I, SALT II, and START talks between the United States and Soviet Union, aimed at mitigating the tensions resulting from the Cold War and arms race (US Department of State, Office of the Historian, 2025).

In addition, regarding the "Climate, Sustainability, & Development Goals" topic, there is a large increase from almost zero speeches in 2000 to 150 in 2021. Over this roughly 20-year span, the threat of climate change became more and more well-known and well-documented, suggested that this topic is relatively well-validated through historic analysis as well.

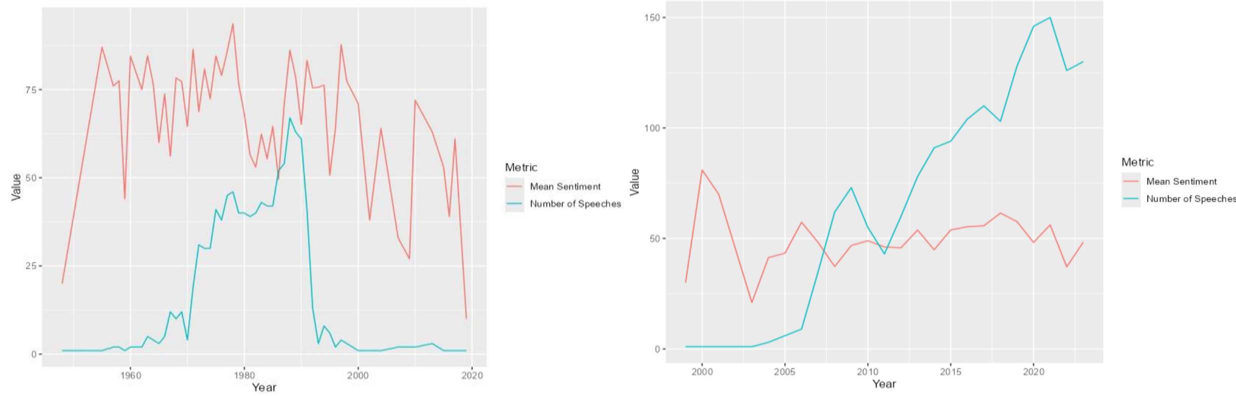


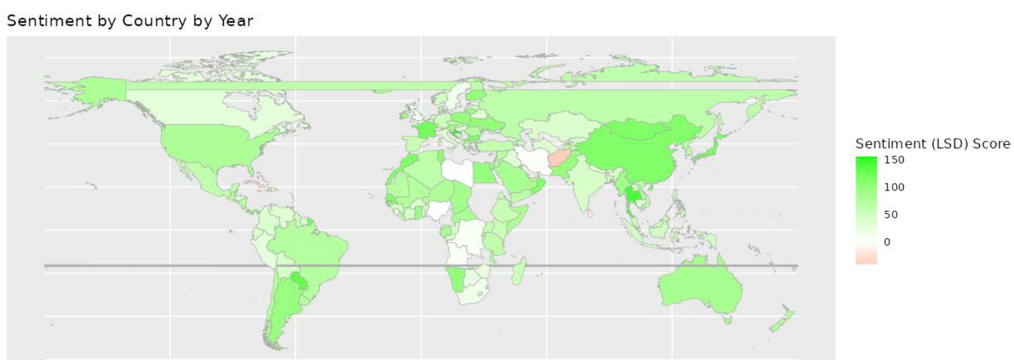
Figure 3: Number of topics and mean sentiment over time for "Peacebuilding and Disarmament" (left) and "Climate, Sustainability & Development Goals

II. Speeches for different countries by topic

The second tab of the dashboard shows the total number of speeches per category given by countries over the course of the UNGA's existence. This provided a relatively simple way to validate findings at the country level—countries associated with certain regions or issues should have a higher number of relative speeches for the categories in question. For example, Bahrain, Saudi Arabia, and many other countries in the Middle East and North Africa have by the most speeches that fall into the "Middle East" topic. Likewise, island nations like Fiji and The Bahamas have a majority of their speeches falling under "Island Nations," and countries in Sub-Saharan Africa like Kenya, Tanzania, and the Democratic Republic of the Congo have a majority of their speeches within the "Africa and the Global South" topic.

III. Sentiment by country and year

The third tab in the dashboard illustrates choropleths of mean sentiment by country by and year. This made validation by both history and geography relatively straightforward. For example, in 2000, most countries exhibited highly positive sentiment scores. This may have been due to the creation and adoption of the UN Millenium Development Goals. However, one year later in 2001, a drastic visible downward trend in sentiment occurred likely spurred by the September 11 attacks and the beginning of the war on terror. This clear decrease in worldwide net sentiment lends support speech sentiment being verifiable through real-world events.



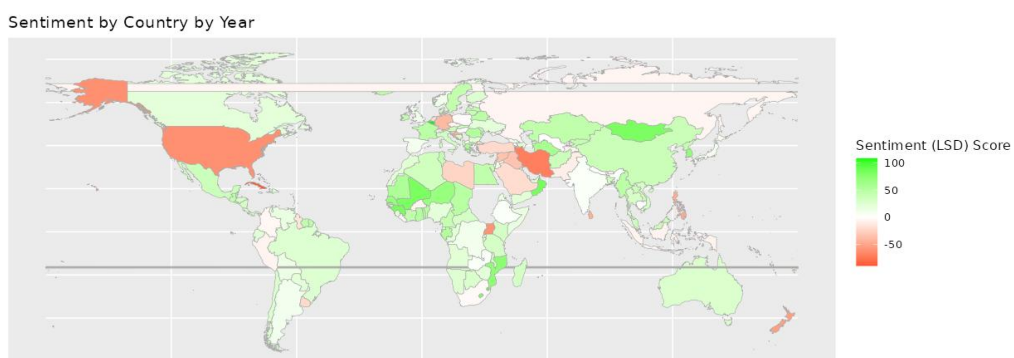


Figure 4: World net sentiment, 2000 (top) and 2001 (bottom)

Discussion

Generally, speeches fell well into their assigned topic model categories. However, allocation was not perfect– for example, there were some countries that gave speeches classified as “Island Nation” speeches that were not in fact island nations. In addition, despite any indicators to the contrary, the world metrics multivariate regression did not provide many insights into the relationships between UN speech sentiment and metrics like GDP, Gini, democracy, western country status, and gender equality. With the exception of democracy, most world metrics were not indicative of UN speech sentiment. Despite these imperfections, many aspects of sentiment and topics were indeed validated by temporal trends and geographic analysis. This was promising, and one of the primary goals of this project.

Given more time and computational power, a higher K (which would yield more topics during LDA), would allow me to consolidate further for more accurate results. Setting a higher value of K (say, 15 or 20) may yield higher gammas and more accurate categorization. These categories could then be further consolidated. Another step I could take is using a different topic modeling method, such as structural topic modeling (STM), which involves encoding documents based on their metadata and characteristics. This may offer a different perspective regarding the latent topics found in these speeches. Further text analysis would also be insightful. For example, TF-IDF weighting and n-gram analysis would be a simple yet informative addition. This could involve sub-setting the corpus to observe n-grams of only a certain length and above a certain TF-IDF score, and compare those to the betas of tokens found for each category. Doing so would provide further validation, as well as more insight as to what words and phrases are common in specific topics, countries, and/or time periods. Finally, a different method of sentiment analysis would be a welcome addition. For instance, the NRC sentiment dictionary looks at the presence of multiple emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) rather than just indicating “positive” or “negative” sentiment

(Mohammed, 2022). This would give a more nuanced and detailed distribution of emotion, making analysis and conclusions more insightful.

Overall, this project contributes to the growing literature that utilizes and observes natural language processing techniques in a social science context with the employment of topic modeling and sentiment analysis.

References

Abbasov, Namig. (2024). *NLP with UN General Debate Corpus 1946-2023.* www.Kaggle.com. https://www.kaggle.com/datasets/namigabbasov/united-nations-general-debate-corpus-1946-2023?resource=download

Council on Foreign Relations. (2024 September). *What is the UN General Assembly and what does it do? Public Broadcasting Station*. https://www.pbs.org/newshour/world/what-is-the-un-general-assembly-and-what-does-it-do

Heiss, Andrew. (2021). *A guide to working with country-year panel data and Bayesian multilevel models*. www.andrewheiss.com. DOI: 10.59350/t19jz-ds665. https://www.andrewheiss.com/blog/2021/12/01/multilevel-models-panel-data-guide/

Jankin, Slava, et. al. (2024). *NLP with UN General Debate Corpus 1946-2023* [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/9304324. https://www.kaggle.com/datasets/namigabbasov/united-nations-general-debate-corpus-1946-2023?resource=download

Mohammed, Seif. (2022). NRC Word-Emotion Association Lexicon. National Research Council Canada. https://saifmohammad.com/. https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Soroka, Stuart. (2025). *Lexicoder Sentiment Dictionary.* www.snsoroka.com. https://www.snsoroka.com/data-lexicoder

United Nations. (2025). *Main Bodies.* www.un.org. https://www.un.org/en/about-us/main-bodies

United States Department of State, Office of the Historian. (2025). *Strategic Arms Limitations Talks/Treaty (SALT) I and II*. U.S. Department of State. www.State.gov. https://history.state.gov/milestones/1969-1976/salt

World Bank Group. (2025). *World Development Indicators*. https://databank.worldbank.org/source/world-development-indicators