The UNSC Meetings and Speeches

This dataset contains detailed information on the official public meetings of the United Nations Security Council (UNSC) covering the period from 1946 to 2024. The dataset consists of two major parts: meeting records and speech records. The former compiles relevant pieces of information (date, agenda items, the monthly president's name and country, policy outputs, etc.) on each of the Council meetings (more than 9,800 in total) held consecutively during this period. The latter comprises the transcripts of more than 160,000 public statements delivered by delegates and officials representing various countries and organizations at these meetings, along with related details on each statement (the speaker's name and affiliation, procedural or policy statement, etc.). The speech transcripts were converted to machine-readable format (plain text). The dataset is under constant update and expansion.

Content

The dataset consists of the following text files (encoding: UTF-8):

- *meetings.tsv*: contains tab-separated tabular data wherein each row gives information on the corresponding UNSC official meeting, including its date, agenda, president, and outcome description.
- *speeches.tsv*: contains tab-separated tabular data wherein each row gives information on the corresponding speech, including the speaker's name and affiliation, and its transcripts in plain-text format.

Description of variables

Meeting records (meetings.tsv)

- *record_id* (string): Unique identifier for the meeting
- *year* (integer): the year in which the meeting took place.
- *month* (integer): the month in which the meeting took place.
- *day* (integer): the day on which the meeting took place.
- *doc_name* (string): The name of the meeting, as indicated by the UN Document Symbol of its Verbatim record (e.g., `"S_PV8942"`).
- *meeting_num* (integer): Numeric identifier of the meeting (e.g., `8942`).
- *closed* (boolean): True if the meeting was closed to the public; False otherwise.
- *agenda* (string): The official agenda item(s) adopted for the meeting. Multiple items are separated by semicolons (;).

- *agenda_category* (string): The main category of the agenda item as classified in the *UN's Repertoire of the Practice of the Security Council*. Examples include "Africa", "Middle East", "Europe", "Thematic Issues", etc.
- *agenda_sub_category* (string): The sub-category of the agenda item as classified in the UN's *Repertoire of the Practice of the Security Council*. For example, "Libya", "Syria", "Children and Armed Conflict", "Protection of Civilians in Armed Conflict", etc.
- *pres_name* (string): the surname of the president presiding over the meeting.
- *pres_country* (string): the country that the president represented.
- *speeches* (boolean): True if the corresponding speech records are available; False otherwise.
- *word_count* (integer): the total number of words uttered during the meeting (if available).
- *outcome* (string): short description of the meeting outcome obtained from the Dag Hammarskjöld Library, including voting outcome(s) if available.
- *record* (string): the UN document symbol of the verbatim record of the meeting.
- *record_url* (string): the location of the PDF of the meeting's record in the UN Official Document System.
- *RES* (string): the UN document symbol of the Resolution adopted during the meeting (if available).
- *RES_url* (string): the location of the PDF of the adopted Resolution in the UN Official Document System.
- *PRST* (string): the UN document symbol of the Presidential Statement adopted during the meeting (if available).
- *PRST_url* (string): the location of the PDF of the adopted Presidential Statement in the UN Official Document System.

Speech Records (speeches.tsv)

- *speech_id* (string): Unique identifier for each speech.
- *record_id* (string): Unique identifier for the meeting to which the speech belongs.
- *doc_name* (string): The name of the meeting, as indicated by the UN Document Symbol of its Verbatim record (e.g., `"S_PV8942"`).
- *meeting_num* (integer): Numeric identifier of the meeting (e.g., `8942`).
- *year* (integer): The year in which the statement was made.
- *month* (integer): The month in which the statement was made.
- *day* (integer): The day on which the statement was made.
- *agenda* (string): The official agenda item(s) adopted for the meeting. Multiple items are separated by semicolons (;).
- *agenda_category* (string): The main category of the agenda item as classified in the *UN's*

Repertoire of the Practice of the Security Council. Examples include "Africa", "Middle East", "Europe", "Thematic Issues", etc.

- *agenda_sub_category* (string): The sub-category of the agenda item as classified in the UN's *Repertoire of the Practice of the Security Council*. For example, "Libya", "Syria", "Children and Armed Conflict", "Protection of Civilians in Armed Conflict", etc.
- *order* (integer): The sequential order of the speech within the meeting.
- *speaker* (string): The surname of the speaker who delivered the statement.
- *affiliation* (string): The country or organization that the speaker represented.
- *position* (string): The position title of the speaker, sourced primarily from meeting records or the UN Digital Library.
- *president* (boolean): `True` if the speaker was serving as the Council President at the time of the statement; `False` otherwise.
- *secretary_general* (boolean): `True` if the speaker was the UN Secretary-General at the time of the statement; `False` otherwise.
- *procedural* (boolean): `True` if the statement was a procedural statement by the President (opening and closing remarks, vote announcements, ceremonial greetings, etc.), which was not made in his/her capacity as a national representative; `False` otherwise.
- *count* (integer): The number of words in the speech.
- *speech* (string): The full recorded text of the statement.
- *cow_ccode* (integer): The Correlates of War country code, if applicable.
- *permanent_member* (boolean): `True` if the speaker's affiliation is that of a permanent member state of the council; `False` otherwise.
- *elected_member* (boolean): `True` if the speaker's affiliation is that of an elected member state of the council; `False` otherwise.
- *state* (boolean): `True` if the affiliation is that of a country; `False` otherwise.
- *igo* (boolean): `True` if the affiliation is that of an intergovernmental organization; `False` otherwise.
- *un_org* (boolean): `True` if the affiliation is that of a United Nations organization; `False` otherwise.
- *ngo* (boolean): `True` if the affiliation is that of a non-governmental organization; `False` otherwise.

Usage

Import either or both of the component tables into appropriate data objects using whatever coding environment you prefer (e.g., Python, R). Use relevant parts of the imported data for further quantitative operation and analysis.

To illustrate, the following steps describe the possible procedures taken for applying word-embedding to the speech records to investigate how a certain notion such as “threat” has been conceived among the members of the UNSC.

1. Export relevant parts (e.g., the statements made by the U.S. delegates during the 1990s) of the *speech* column into a single plain-text file in which each line corresponds to a single speech.
2. Execute some word-embedding program, for example, GloVe (Global Vectors for Word Representation), using the exported file as input. Follow the instructions for the program concerned.
3. Obtain words and phrases semantically associated with a focus word, such as “threat,” using a similarity measure (e.g., cosine similarity) for the derived vector representations. For Python users, the *gensim.models.KeyedVectors* library provides useful functions for these operations.

Data source and generation

- Data source: the main source for this dataset is the electronic version (PDF) of the official verbatim records of the UNSC meetings, which are available from the UN Official Document System. These documents were downloaded via hyperlinks embedded in the summary tables provided on the website of the Dag Hammarskjöld Library. Other relevant information, such as the meeting date and outcome, was also acquired from these tables. We also consulted the speech information compiled in the UN Digital Library particularly for locating the affiliations (as well as the position titles if available) of officials of the UN and other organs.
- Source code: the Python code used for constructing this dataset is available from GitHub.

Author

- Takuto Sakamoto: Professor, the Graduate School of Arts and Sciences, the University of Tokyo
- Tomoyuki Matsuoka: Academic Assistant, the Graduate School of Arts and Sciences, the University of Tokyo

Contributors

The contributions of the following persons are highly appreciated.

- Mihoko Iijima: former graduate student, the University of Tokyo
- Yuri Goto: former graduate student, the University of Tokyo
- Nozomu Takaguchi: former graduate student, the University of Tokyo
- Li Qingqian: former research student, the University of Tokyo

- Liu Jinwu: former graduate student, the University of Tokyo
- Ryoun Ukita: graduate student, the London School of Economics and Political Science (LSE)
- Yuka Matsubara: graduate student, the University of Tokyo
- Kotaro Otsuka: graduate student, the London School of Economics and Political Science (LSE)
- Nao Hashimoto: graduate student, the University of Tokyo
- Kazuhiko Kimura: former undergraduate student, the University of Tokyo
- Momoko Araki: graduate student, the University of Tokyo
- Aika Inooka: graduate student, the University of Tokyo
- Shoichiro Hashimoto: student, Shibuya Kyoiku Gakuen Makuhari Senior High School
- Hiroto Ito: graduate student, the University of Tokyo
- Takuto Tohkairin: graduate student, the University of Tokyo
- Kyounghwan Jang: graduate student, the University of Tokyo
- Mi Kyoung Park: graduate student, the University of Tokyo

Versions

- Version 1.0 (March 15, 2023)
- Version 2.0 (May 15, 2024)
- Version 3.0 (February 26, 2025)
- Version 4.0 (February 27, 2025)
- Version 5.0 (March 31, 2026)