
Topic Modeling with UN Speeches

Mary Krysllette C. Bunyi
DS3 Project 2 Presentation
8 November 2021

United Nations General Debate Corpus

- Transcripts of speeches given by **heads of state** at the **UN General Assembly** from **almost 200 countries** from **1970 to 2015**
- **UN General Assembly**
 - **Main deliberative, policymaking and representative organ** of the UN
 - Comprises **all Member States**
 - Forum for **multilateral discussion** of the full spectrum of international issues covered by the Charter of the United Nations
 - Official statements on their **government's perspective** on major global political issues

Exploratory Data Analysis



Cleaned data **with** stopwords



Cleaned data **excluding** stopwords

LDA Implementation

- **CountVectorizer**

- LDA relies on counts to generate the distributions
- `ngram_range=(1, 2)`
- `min_df=0.1`
- `max_df=0.75`

- **Document-Term Matrix**

- 7,507 rows
- 1,972 columns

- **3 runs for the whole dataset**

- 10 topics
- 7 topics
- 5 topics

- **15 terms per topic**

Sample Unigrams

ability

accept

acceptable

acceptance

accepted

young

yugoslavia

zimbabwe

zone

zones

Sample Bigrams

african continent

african unity

aligned movement

arab territories

armed forces

western sahara

wish express

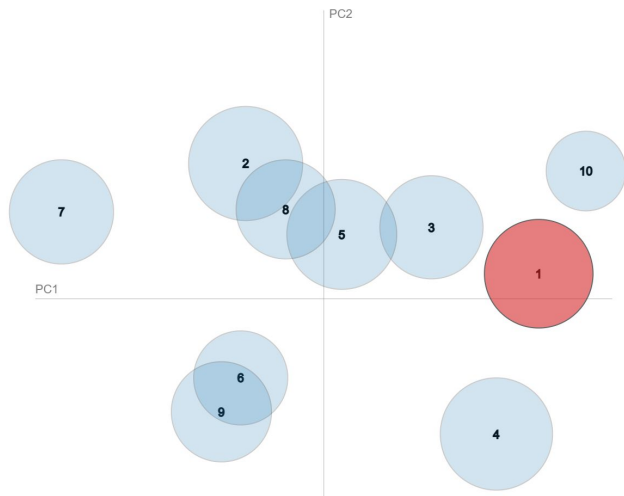
wish success

withdrawal foreign

women children

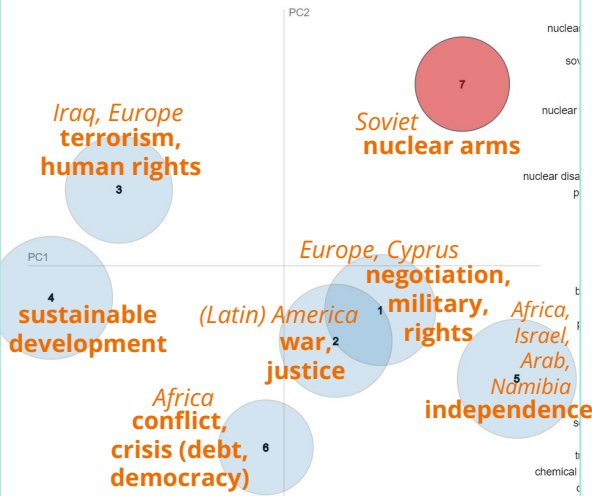
Conceptual Mapping

Intertopic Distance Map (via multidimensional scaling)



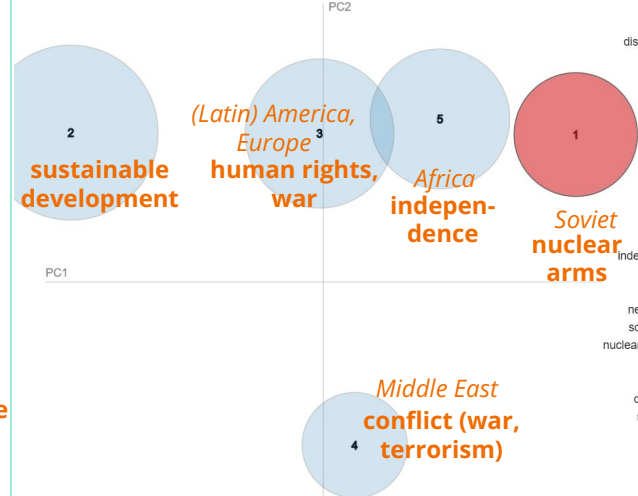
10 topics

Intertopic Distance Map (via multidimensional scaling)



7 topics

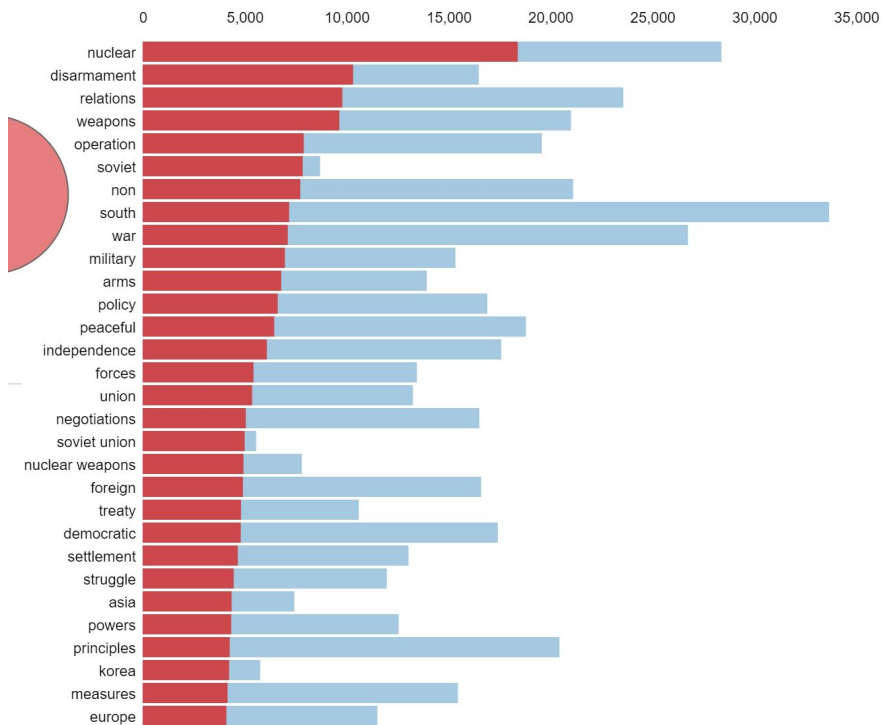
Intertopic Distance Map (via multidimensional scaling)



5 topics

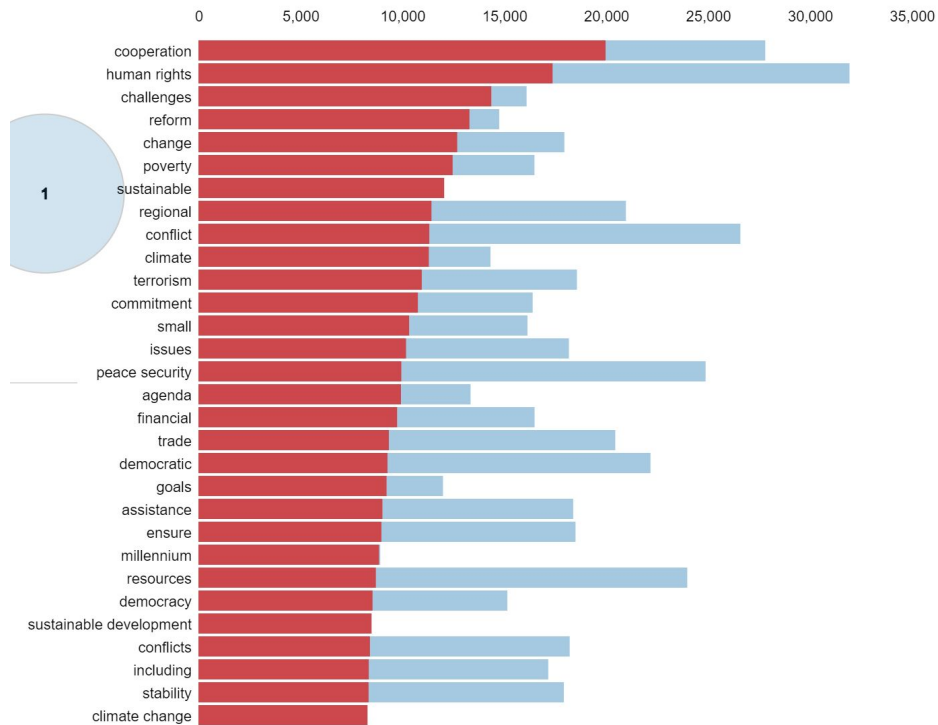
Conceptual Mapping

Top-30 Most Relevant Terms for Topic 1 (15.5% of tokens)



Topic 1: *nuclear arms*

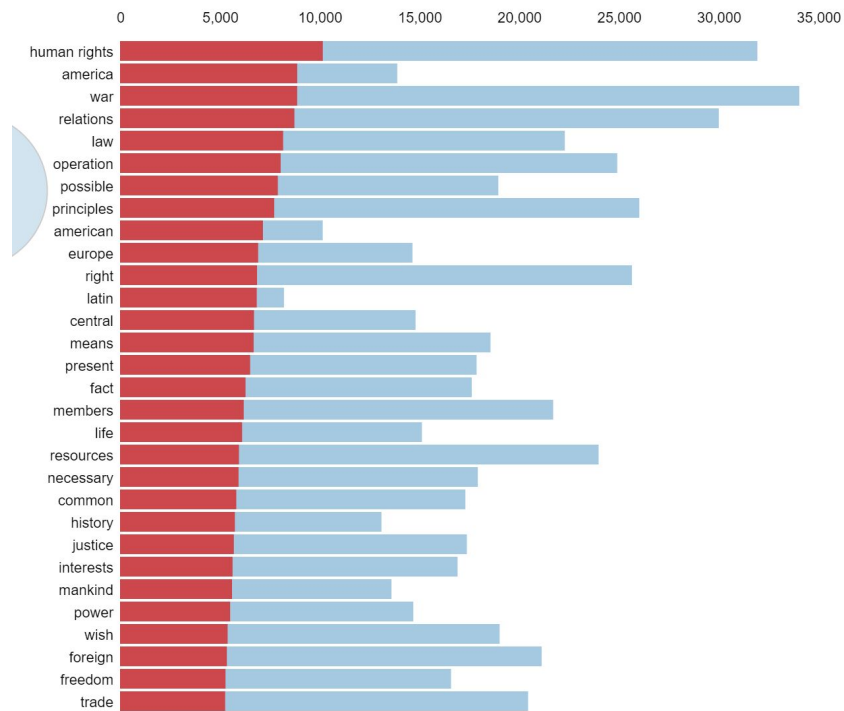
Top-30 Most Relevant Terms for Topic 2 (31% of tokens)



Topic 2: *sustainable development*

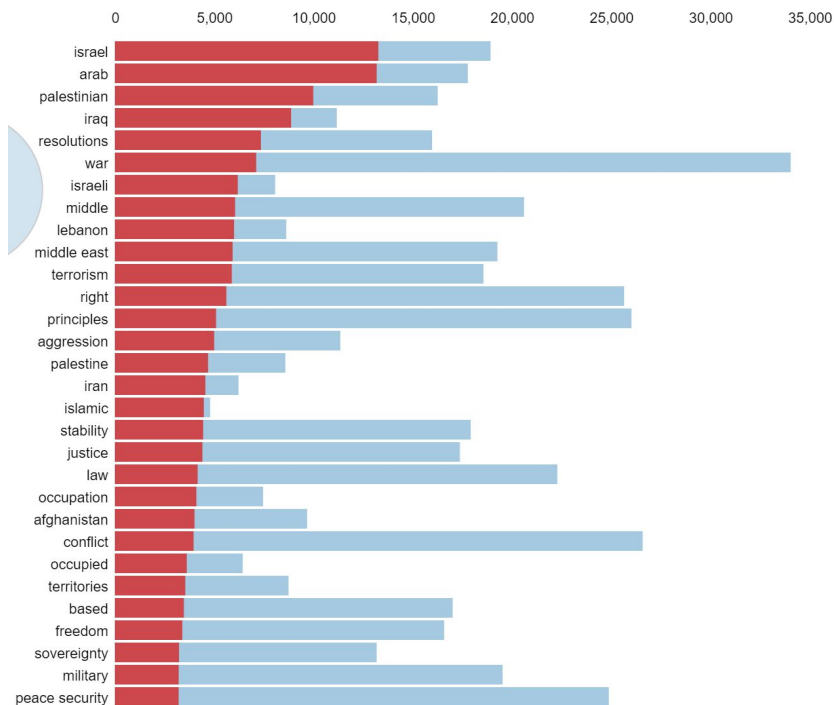
Conceptual Mapping

Top-30 Most Relevant Terms for Topic 3 (22.5% of tokens)



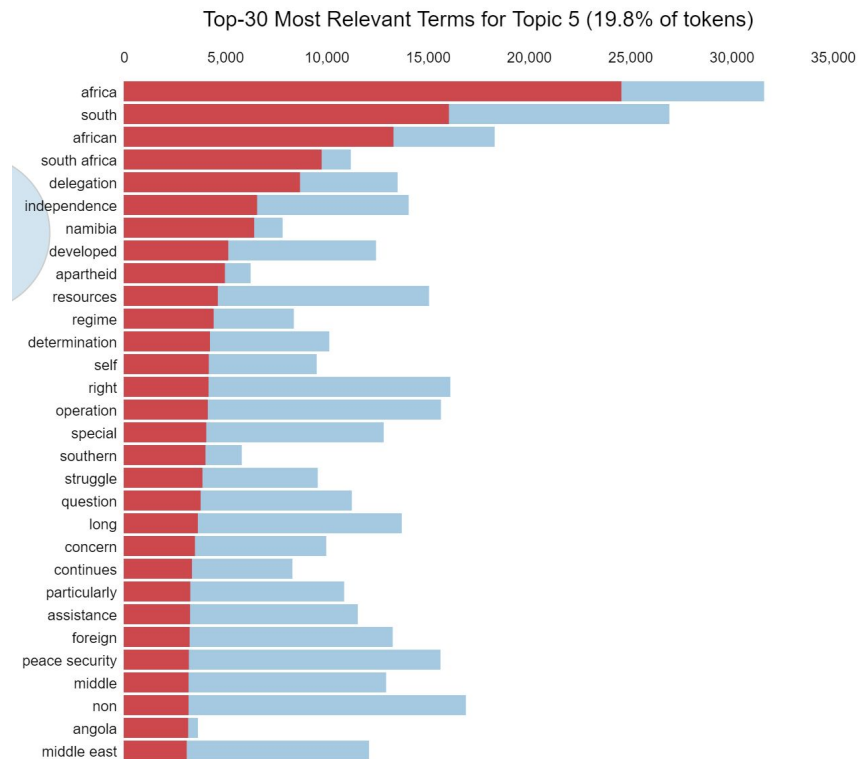
Topic 3: *justice (human rights, war)*

Top-30 Most Relevant Terms for Topic 4 (11.2% of tokens)



Topic 4: *conflict (war, terrorism)*

Conceptual Mapping



Topic 5: *independence*

Topic Trends over Time

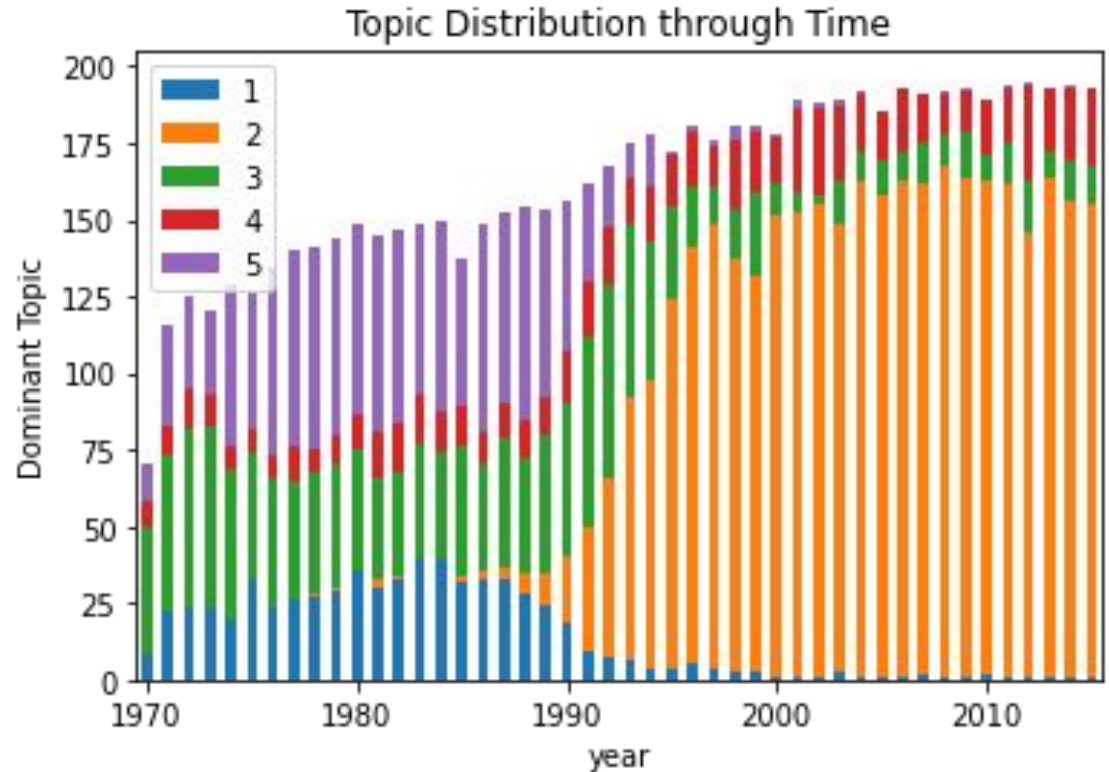
Topic 1: *nuclear arms*

Topic 2: *sustainable development*

Topic 3: *justice (human rights, war)*

Topic 4: *conflict (war, terrorism)*

Topic 5: *independence*



Policy Relevance of Topic Modeling

- Directs attention to salient issues
- Quickly and efficiently parse through large volumes of data (compared to subjective human reading)
- Capture new patterns that are not immediately perceptible
- Trade-offs
 - Abstracts away nuances
 - Susceptibility to noise
 - Interpretability
- Other uses in public policy
 - Gauge of public opinion on various issues
 - Anti-money laundering: identify typologies of financial crime

Thank you for your time!

Any questions?