

UNSC-NE

A Named Entity Extension to the UN Security Council Debates Corpus

Luis Glaser¹, Ronny Patz² & Manfred Stede¹

¹University of Potsdam, ²Hertie School

Contacts:

luis.glaser@uni-potsdam.de

patz@hertie-school.org

stede@uni-potsdam.de



Our Contribution

We present the *UNSC Debates Corpus Named Entity Linking (NEL)* add-on, an entity-tagged extension to the UN Security Council debates corpus, a large corpus previously published by Schoenfeld et al. (2019).

Motivation

- Growing interest in Natural Language Processing (NLP) in political science research (Baturu et al., 2017; Schoenfeld et al., 2018; Bayram and Ta, 2019)
- We assume that choice of words is an important dimension in political / diplomatic speech. Example: Speakers can focus on states' rights for non-interference or on human rights as deliberate attempts to frame a multilateral debate in favor of interfering (or not) in a conflict.
- We propose Named Entities (NEs) as a means to investigate the dimension of choice of different concepts in political speech.

Main Objectives

1. Create a NE corpus addon that can easily be kept updated
2. Evaluate the quality of the NEL annotations.
3. Empirically evaluate the potential of NE in a substantive experiment
4. Make the corpus addon publically available.

⚡ **NEL is not meant as a replacement for NER but tries to solve a different task!** The annotation quality of DBpedia is significantly worse than spaCy but the problem is different. Instead of tagging a word with a class, NEL links it to its entity.

Type	Count	Measure	spaCy	DBpedia
speeches	82,165	Accuracy	.478	.405
sentences	1,921,352	Precision	.503	.444
entities	2,377,371	Recall	.904	.821
wikidata classes	40,916	F1	.647	.576
distinct entity classes	4,907			
distinct entity superclasses	10,989			

Table 1: Descriptives for UNSC-NE

Table 2: Comparison of annotation quality metrics between spaCy and DBpedia

Key Take-Aways

- NEL addon for the UNSC corpus available at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/OV1FLX>
- As new editions of the UN Security Council Corpus become available, the corpus addon will be kept updated using the build scripts provided on GitHub <https://github.com/glaserL/unsc-ne>.

Named Entity Recognition (NER) compared to NEL



Corpus Structure

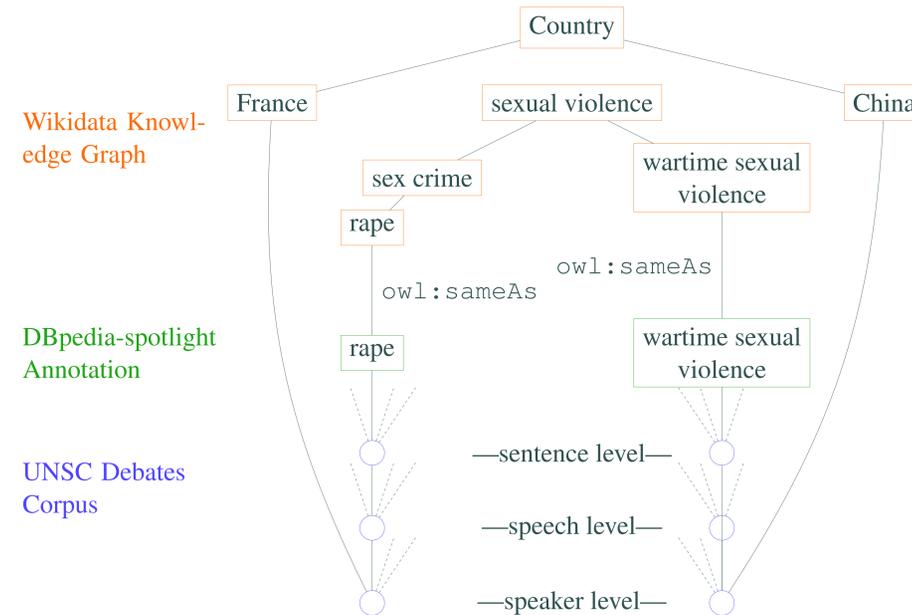


Figure 1: Example of two annotated NEs with their connections in the three annotation layers.

Corpus Format

```

1 {"type": "node", "id": "1", "labels": ["Sentence"]}
2 {"type": "node", "id": "2", "labels": ["DBConcept"], "properties":
  ↳ {"uri": "dbpedia.org/Wartime_sexual_violence"}}
3 {"type": "node", "id": "3", "labels": ["WDConcept"], "properties": {"label": "wartime
  ↳ sexual violence", "uri": "wikidata.org/Q2408310"}}
4 {"id": "4", "type": "relationship", "label": "CONTAINS", "start": {"id": "1",
  ↳ "labels": ["Sentence"], "end": {"id": "2", "labels": ["DBConcept"]}}
5 {"id": "5", "type": "relationship", "label": "owl_sameAs", "start":
  ↳ {"id": "2", "labels": ["DBConcept"], "end": {"id": "3", "labels": ["WDConcept"]}}

```

Figure 2: Data sample from UNSC-NE corpus addon

Exemplary Experiment: Women, Peace and Security (WPS) in the UNSC

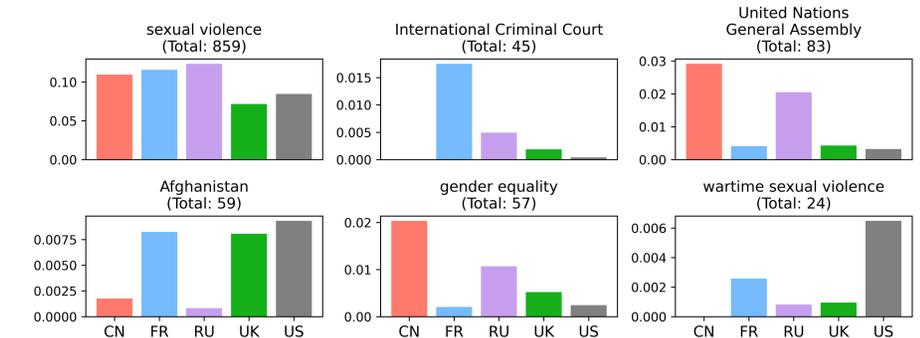


Figure 3: Distribution of an extract of the top 25 entities used most frequently by the P5 in their speeches during WPS meetings

The UNSC-NE addon can replicate insights found in recent research but also reveals new insights

- China and Russia want to limit the scope of WPS in the UNSC debates (True and Wiener, 2019) and thus mention UNGA more (Figure 3).
- Mainly western countries mention WPS during UNSC debates on Afghanistan (Schoenfeld et al., 2018). Conversely, they point to Afghanistan in debates on WPS more frequently (Figure 3).
- There is a difference between “sexual violence”, “wartime sexual violence”, and “rape” in international law (Macfarlane, 2021), and the UNSC-NE demonstrates how different states refer to these (Figure 4). E.g.: China does not refer to “rape” but to “sexual assault”.

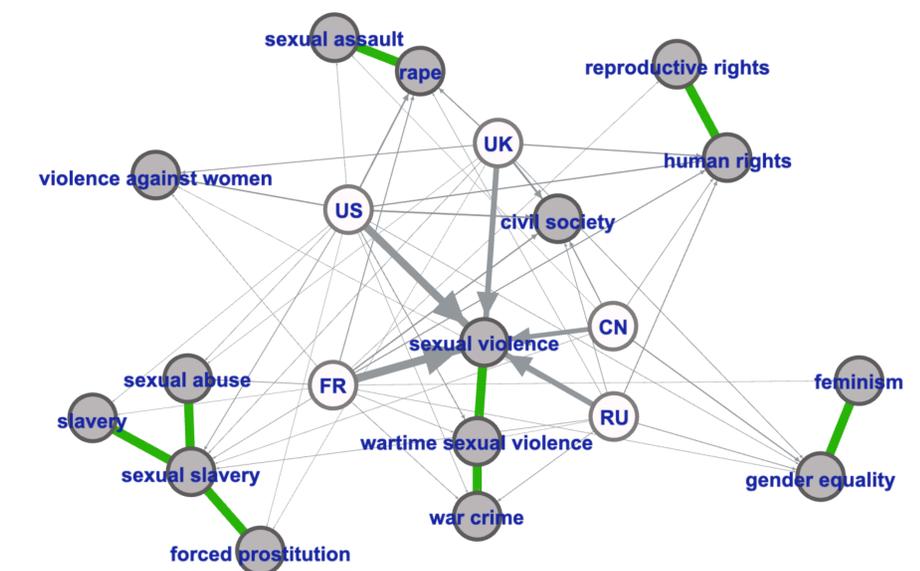


Figure 4: Extract of top 25 entities in WPS debates. Green lines between neighbors in the knowledge graph. The width of grey edges refers to the frequency the P5 talk about the NE.