

# WPS Subcorpus: Reported Speech in the UNSC debates

Meryem KARALIOĞLU

## 1 Description

The WPS subcorpus is based on the speeches of the United Nations Security Council debates (Schoenfeld et al., 2019). The jsonl file is made up of sentences containing an instance of reported speech. Every instance has a unique identifier and is annotated with the cue, content and source span which can also be uniquely identified, respectively. The unique identifier is a combination of the spv name of the UNSC speech, the speech number and the component of the annotations. The annotations contain the file name of the original file in the UNSC corpus. The sentence index in the original file is provided, as well as the position of the starting token of the sentence, cue, content and source span.

- **file\_name**
  - The file name of the original UNSC corpus without the file extension
    - \* UNSC\_2007\_SPV.5663Resumption1\_spch042
- **id**
  - The unique identifier of the instance and the annotations contained in it
    - \* 2007\_5663Resumption1\_042\_11,  
2007\_5663Resumption1\_042\_11\_cue
- **Sentence index**
  - The index of the sentence in the speech file
- **Sentence**
  - The sentence the instance of reported speech is contained in
- **Cue**
  - The cue word annotation of the instance
- **Content**
  - The content annotation of the instance
- **Source**
  - The source annotation of the instance
- **Position**
  - The byte position of the first token of the annotation in the speech file

## References

Mirco Schoenfeld, Steffen Eckhard, Ronny Patz, Hilde van Meegdenburg, and Antonio Pires. The UN Security Council Debates, 2019. URL <https://doi.org/10.7910/DVN/KGVSYH>.