Expanding the UNSC Conflicts Corpus by Incorporating Domain Expert Annotations and LLM Experiments

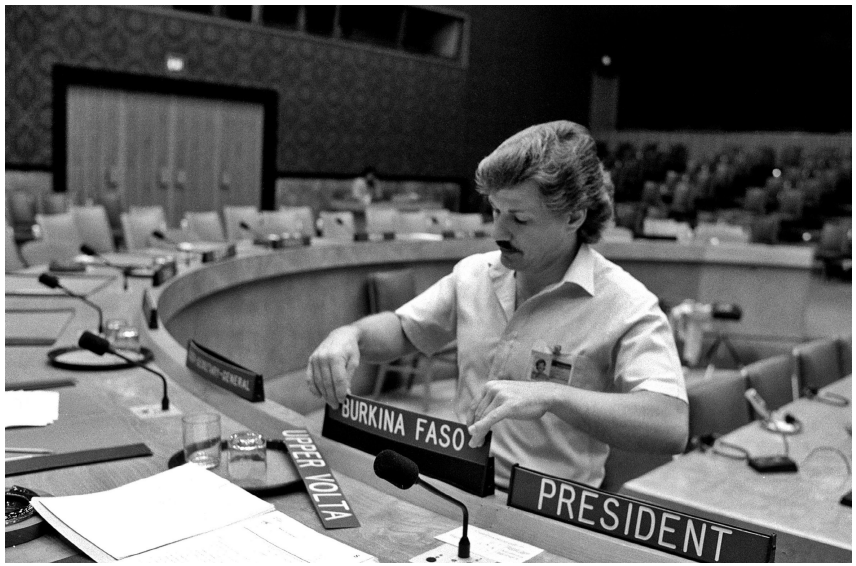
Karolina Zaczynska

Applied Computational Linguistics Discourse Research Lab, University of Potsdam
Institute of Formal and Applied Linguistics, Charles University

July 31, 2025

Introduction

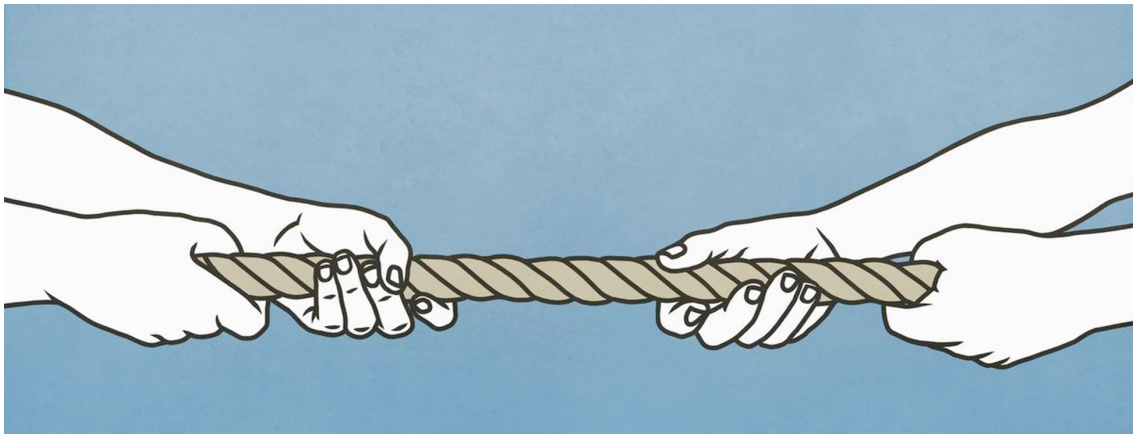
- **UNSC Conflicts Corpus (UNSCon)** analyses verbal disputes in 87 diplomatic speeches given in the UN Security Council
- Presented in Zaczynska, Bourgonje, and Stede 2024
- **Linguistically-motivated** framework for understanding Conflicts



Introduction

Definition of Conflict:

Speaker expresses critique or disagreement with the positions or actions of one or more countries present at the debate.



Introduction

Example:

Again, the external anti-Ukranian (sic) and anti-Western propaganda machine is in full swing, inciting suspicion, mistrust and hatred waiting to explode.
(UNSC_2014_SPV.7154, Lithuania)

Goals of this work

- Annotations in orig. UNSCon done by computational linguists
- Test & refine the annotations by incorporating UNSC domain expert (**UDE**) annotations
- Expand the corpus by size
- Expand classification experiments reported in Zaczynska, Bourgonje, and Stede 2024

Background: **Conflict Types** in orig. UNSCon

1. **Direct Negative Evaluation**
   of positions or actions of another country
2. **Indirect Negative Evaluation**
   of positions or actions of another country by criticising a proxy
3. **Challenge**:
   Speaker accuses another one of lying;
4. **Correction**:
   Speaker provides a correction to that allegedly false statement.

- Units of annotation: *Elementary Discourse Units* (EDUs)

Expanding the UNSCon Annotation scheme

Expanding the UNSCon

When creating a corpus find the middle ground between:

- Domain expert's corpus:
  Partially only interpretable by domain experts
- Corpus grounded in the text and lexical markers:
  Domain-independent, robust, interpretable by language models

Variation in Conflict Annotation

Parallel annotations **UDE**[1] and original UNSCon corpus:

- ⇒ Moderate Krippendorff's Alpha 0.30 up to 0.46
- ⇒ Correction often not annotated by UDE (21% of EDUs labelled as No Conflict)
- ⇒ Similar disagreements to those reported for original UNSCon (`IndirectNegE` versus `DirectNegE`)

[1] Political Science and International Relations PhD on language use in UNSC, Chairman of the Data Diplomacy Academy

Key Disagreements: Diplomatic Phrasing

- UDE additionally annotated Conflicts based on **diplomatic rules**

⇒ To keep clear linguistic operationalisation of Conflicts, we chose **not** to include these implicit Conflicts

Example

It is the seventh time that the Security Council has convened to discuss the urgent crisis in Ukraine.
The Council is meeting on Ukraine because it is the job of this body to stand up for peace and to defend those in danger.
(S/PV.7138, United States)

Key Disagreements: Instructive & Emotive Words

- UDE's additionally annotated Conflicts motivated by **instructive or emotive words** as marker

⇒ Update Conflicts including taxonomy by Gruenberg 2009

Example

Russia **must** pull back its forces to their bases and decrease their numbers to agreed levels.
It **must** allow international observers access to Crimea. It **must** demonstrate its respect for the sovereignty and territorial integrity of Ukraine, [...].
(S/PV.7138, Australia)

Extending UNSCon

- Added new Conflict Type Figurative Language, covering:
  - rhetorical questions
  - sarcasm
- Re-annotated the corpus with refined guidelines
- Expanded corpus by size

Expanding UNSCon by Size

- 128 (+41) speeches from 12 (+6) debates, 143.212 (+80.881) tokens
- increasing the #EDUs by around 39%

	#EDUs	
Conflict Type	**UNSCon**	**extended**
Direct NegE	771	1621
Indirect NegE	501	516
Challenge	101	138
Correction	128	214
Sarcasm	-	52
Rhetorical Question	-	120
Conflict	1501	2642
No Conflict	4497	7162
Sum	**5998**	**9804**

Classification Experiments

4-class setup: Model trained on similar domain and task gains the best results.

	RoBERTa^{arg}[2]	RoBERTa$^{policy}_{topics}$[3]	RoBERTa$^{arg}_{relations}$[4]	maj. voting
	4-class setup (FigL / NegE / CC / No Conflict)			
precision	0.34	0.58	**0.62**	0.18
recall	0.34	0.33	**0.42**	0.25
f1 macro	0.33	0.34	**0.47**	0.21
accuracy	0.77	0.76	**0.77**	0.73

Table 1: Classification results 10-fold cv, precision and recall macro-avg. (learning rate: 1e-5, batch size: 32, training epochs: 2, weight decay: 0.01)

[2] https://huggingface.co/chkla/roberta-argument
[3] https://huggingface.co/niksmer/PolicyBERTa-7d
[4] https://huggingface.co/raruidol/ArgumentMining-EN-ARI-AIF-RoBERTa_L

Take-Aways

Main Take-Aways

- Presented an expanded UNSCon corpus refined **iteratively relying both on linguistic and political domain** knowledge
- Present new set of **classification experiments** on Conflict types
- Pilot study with GPT4o to **extract lexical markers** of Conflict (see our paper)

Links:

- Code with experiments & expanded UNSCon:
  https://github.com/linatal/Expanding_UNSCon
- The UP Multilayer UNSC Corpus:
  https://github.com/discourse-lab/UMUC
- Demo with visualisations of the UNSCon:
  https://github.com/linatal/UNSCon-UI

Thank you for your attention!

zaczynska@uni-potsdam.de

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Universität Potsdam

ÚFAL

DAAD Deutscher Akademischer Austauschdienst German Academic Exchange Service

Literature I


Zaczynska, Karolina, Peter Bourgonje, and Manfred Stede (2024). "How Diplomats Dispute: The UN Security Council Conflict Corpus". In: *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Turino. URL: https://aclanthology.org/2024.lrec-main.716/.


Gruenberg, Justin (Jan. 1, 2009). "An Analysis of United Nations Security Council Resolutions: Are All Countries Treated Equally?" In: *Case Western Reserve Journal of International Law* 41.2, p. 513. ISSN: 0008-7254. URL: https://scholarlycommons.law.case.edu/jil/vol41/iss2/12.

Literature II


Krippendorff, Klaus (1995). "On the Reliability of Unitizing Continuous Data". In: *Sociological Methodology* 25, pp. 47–76. ISSN: 00811750, 14679531. URL: http://www.jstor.org/stable/271061 (visited on 06/20/2025).

Appendix

Sarcasm and Rhetorical Questions

Example Sarcasm

Some colleagues today have achieved **high levels of rhetoric**. I must mention that the Ukrainian colleague nevertheless went far beyond anything permissible. (S/PV.7138_spch020, Russia)

Example Rhetorical Question

A few days ago, a Ukrainian helicopter was downed by a rocket-propelled grenade [...].
That certainly does not sound like the implementation of Geneva agreement by the separatists and their state sponsors? (S/PV.7165_spch016, Lithuania)

Experiments: Include Linguistic Markers

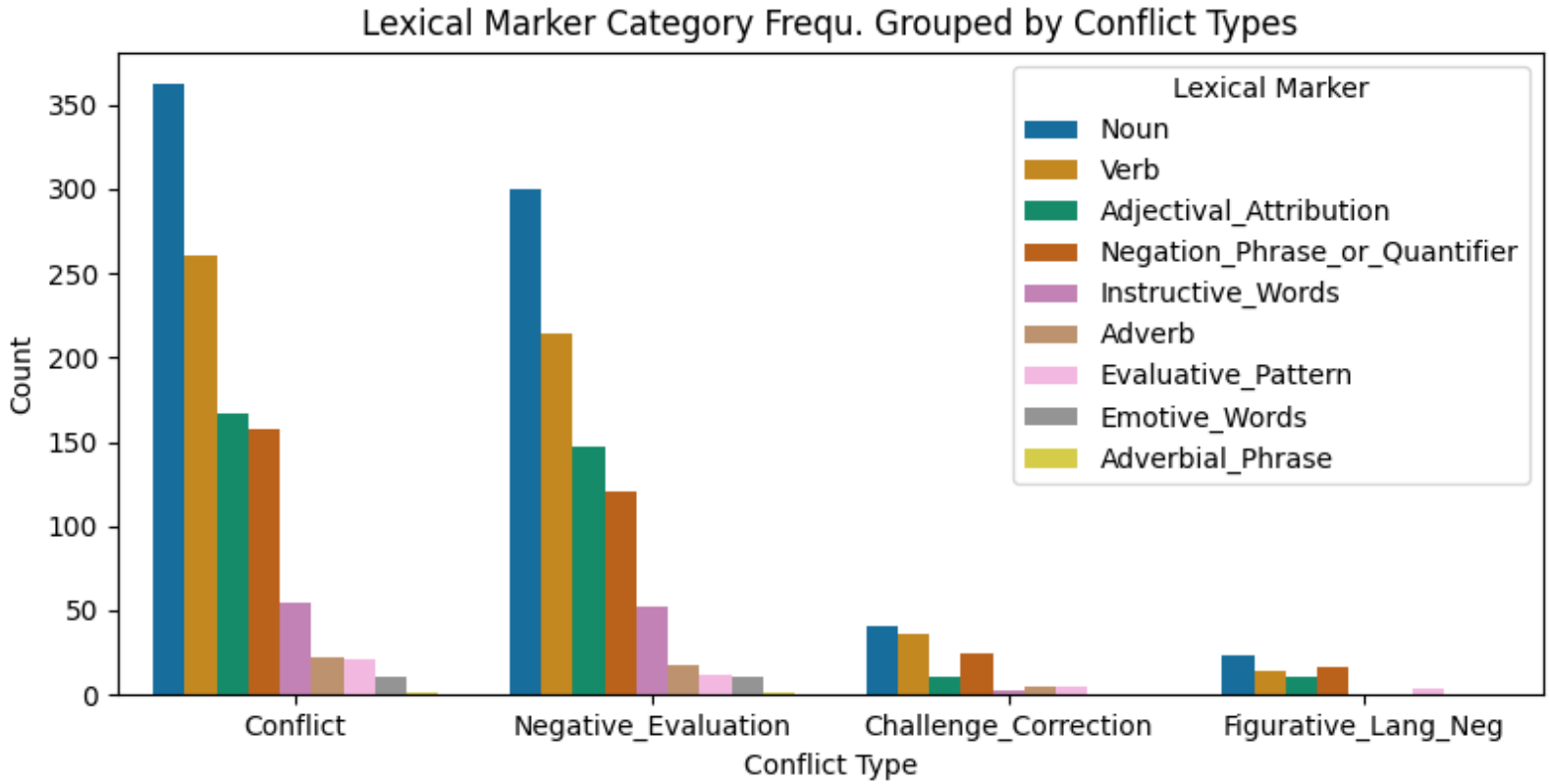


Figure 1: Frequency of Lexical Markers found by GPT4o, Categories per Conflict Type.

Distribution of Conflict Types

- Although we bound our Conflicts by several rules (explicit lexical marker, targets): frequently **directly expressed critique**
- The frequency for CC is higher for Ukraine than for WPS

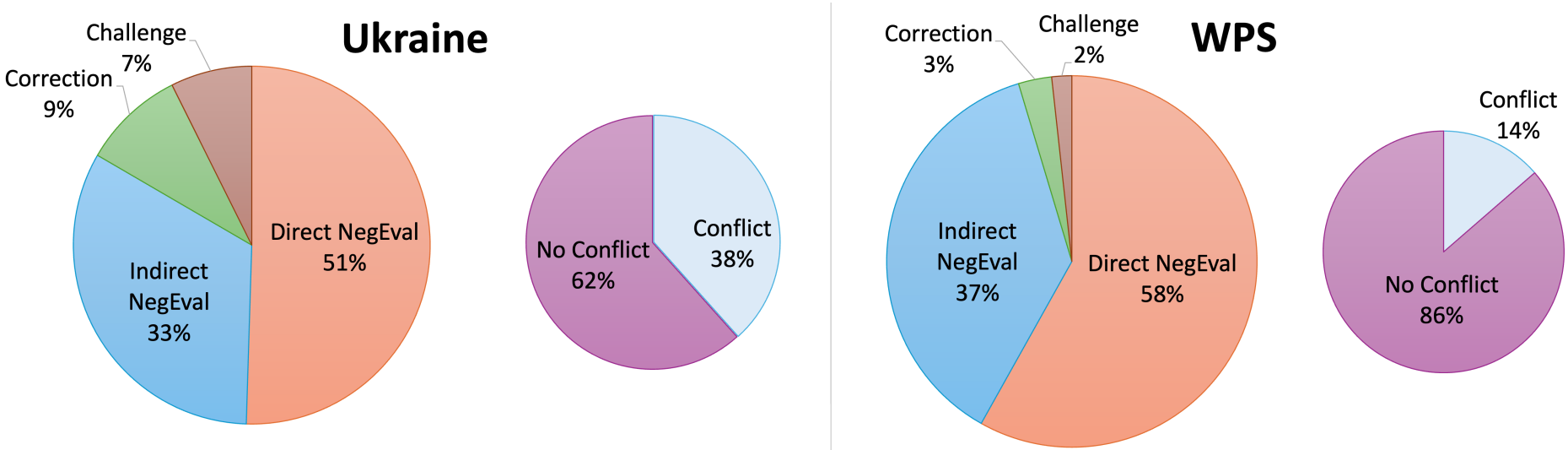


Figure 2: Piechart with distribution of Conflict Types per EDU (#EDUs Ukraine: 3.465, WPS: 1.261)

Piechart UNSCon

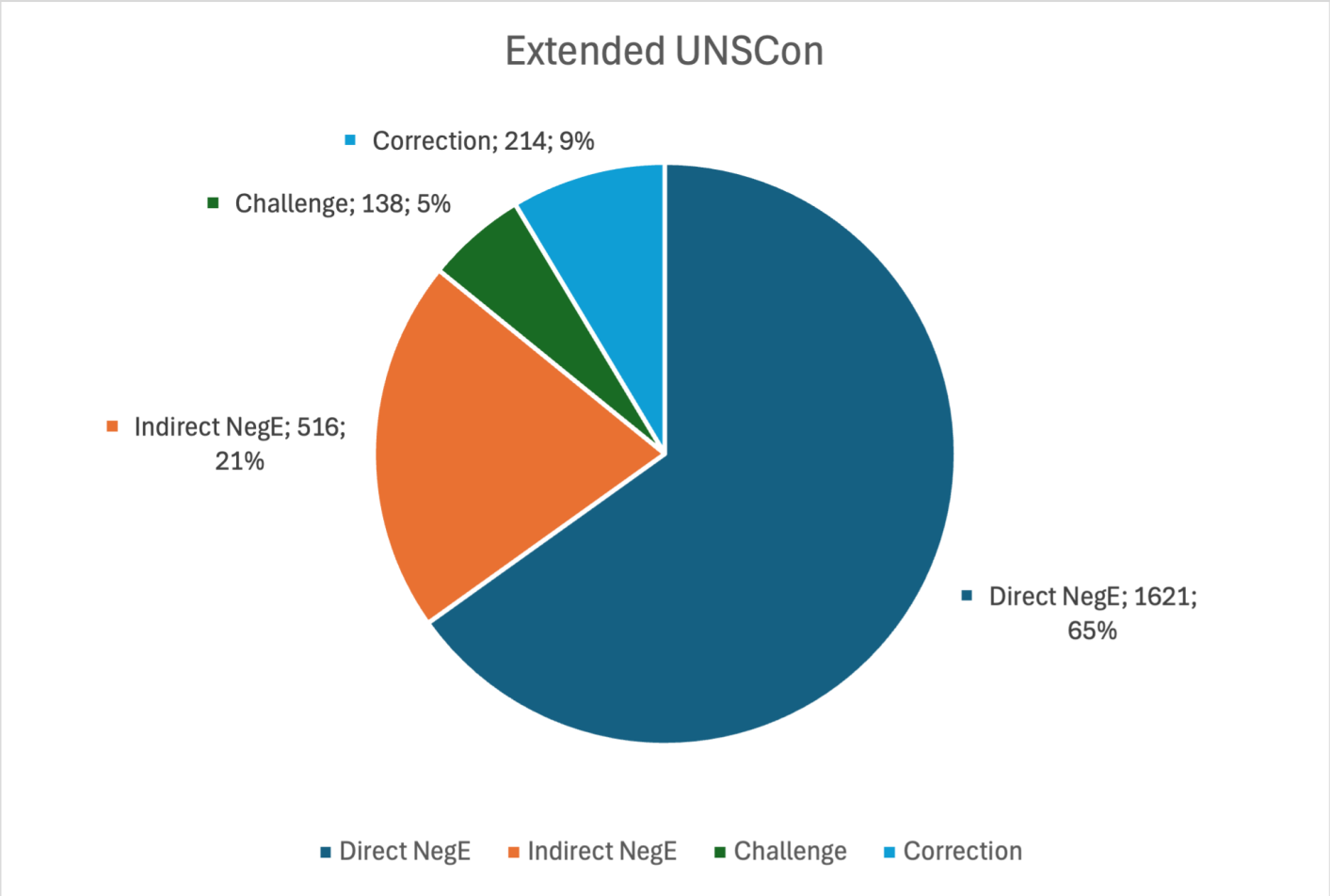


Figure 3: Distribution Conflict Types in extended corpus

.

Evaluation: Inter Annotator Agreement

- Used **Krippendorff's Alpha unitising** to compute IAA (Krippendorff 1995)
- Based on a distance function for comparing units with identical, overlapping, or disjoint boundaries
- Suitable in cases with partial overlap of annotations
- Suitable in cases with arbitrary number of annotators

$$\alpha = 1 - \frac{D_o}{D_e}$$

Where D_o is the disagreement observed and D_e is the disagreement expected by chance.