

Social Knowledge in Controversial Discourses

Challenges in Annotating and Automating Topoi

Carina Kiemes, Marcus Müller



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Research setting

Our project is part of the DFG research group *Controversial Discourses – Language History as Contemporary History since 1990*. The group consists of **five subprojects** and **two associated subprojects**:

Participation & Equality
Noah Bubenhofer /
Davide Ventre (Zürich)

Humans & Technology
Constanze Spieß /
Lesley-Ann Kern (Marburg)

Individual & Society
Nina Janich /
Patrick Johnson (Darmstadt)

Freedom & Security
Martin Wengeler/
Swantje Köhler (Trier)

Culture & Identity
Martin Wengeler/
Nicolai Jacobs (Trier)

Diversity & Unity
Martin Wengeler /
Erik Schröder (Trier)

Methodology & Reflexion
Marcus Müller /
Carina Kiemes (Darmstadt)

The research group aims to write the **history of German political discourse since 1990** and to develop **discourse history as collaborative research**. Our subproject *Methodology & Reflexion* builds on the qualitative annotations of the other subprojects. These are expert annotations as part of dissertation projects. We conduct automation experiments to build taggers for topoi of mid-level abstraction (e.g., utility topoi, danger topoi). **Our methodological goal is to improve quantitative methods of discourse analysis with insights from qualitative research.**

Topos Analysis

A topos is [...] a **social pattern of thought for the formation of rhetorical-enthymematic argumentations** whose conclusiveness – the conclusion – follows from premises that are recognised opinions. (Römer 2017: 104)

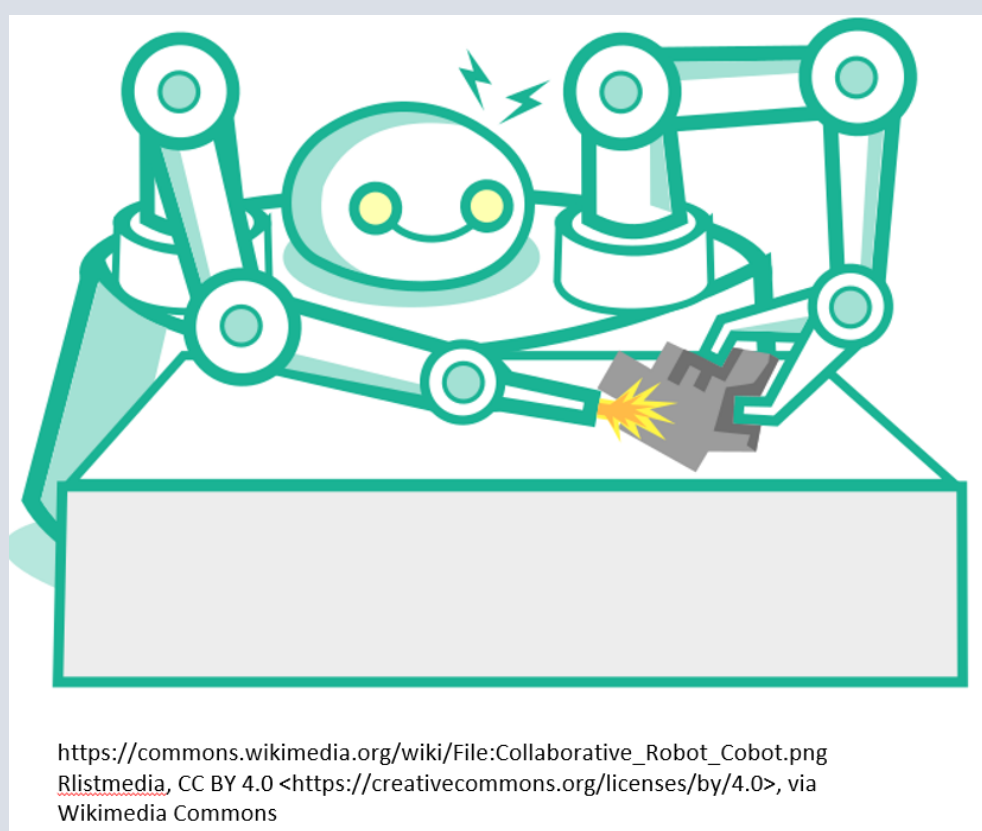
Example of the topos of utility: *If an action provides benefit, it should be realised.*

- A topos
- is “habitually and collectively spread”,
 - “can be used both in favour of and against the positions in question”,
 - is repeated by the “speaking individuals with their interests and intentions”, who however ‘simultaneously modify it with every linguistic act’, and
 - can be “linguistically/symbolically realised in various ways”. (Wengeler 2007: 167 f.)

Social Knowledge in Discourse Analysis:
“Discourse in Foucault’s sense is therefore always also and above all **the area in which social knowledge is characterised and directed as social knowledge.**” (Busse 2013: 149)

By automatically transferring topoi annotated in one subproject to others, we learn about the **topic specificity and discourse function of social knowledge.**

Challenges in Annotating and Automating Topoi



Since topoi are social thought patterns, they are usually not fully verbalised and remain implicit.

Topoi can be indicated by certain metaphors, motifs or idioms, e.g., ‘designer baby’
BUT: **Every linguistic expression can realise a topos under certain conditions of use.** (Wengeler 2003: 197)

→ The implicitness of topoi and their different realisation patterns lead to **topos analysis being described as an interpretative act based on subjective understanding** (Römer 2017: 127)

Context-specific topoi according to Wengeler are **specific to certain discourses and therefore require discourse knowledge.** This knowledge is unevenly distributed in the subprojects.

Initial Insights

Lessons learnt from collaborative annotation:

- **segmentation:**
sentence to paragraph level (BIO scheme)
- **interpretation depth:**
Annotators must agree on how much background knowledge to use when annotating → cases of doubt need to be discussed intensively and the results should be included in the guidelines.
- **co-text and discourse knowledge are relevant**

Pre-test – experimental set-up:

- Use of a **BERT model** (Devlin et al. 2019) for German (deepset/gbert-base)
- **Training the model with annotated data from the Humans & Technology subproject** (single annotation)
- **Balancing the data** by ensuring that there are 345 sentences with topoi annotation and 1397 sentences without topoi annotation, achieving a **1:4 ratio**
- **Splitting the data in a Training set (80 %) and test set (20 %)**
- **Aim: binary classification decision for each sentence as to whether a topos is present or not**

Approach 1: Sentence-based classification

Stats:	accuracy	83.67
	f1	61.90

This means that there were

- **18 incorrect classifications in the test set**, of which
- 9 weren’t recognised as topoi and
- 9 were incorrectly classified as topoi.

Approach 2: Sentence classification with co-textual information including 5 sentences before and after the target sentence

Stats:	accuracy	79.16
	f1	8.77

This means that there were

- **104 incorrect classifications in the test set**, of which
- 95 (out of 100 instances) weren’t recognised as topoi and
- 9 were incorrectly classified as topoi.

Discussion & Further Steps

When evaluating incorrect classifications from the first approach, the question arose as to whether the co-text is relevant for automation.

Example: *God has given us the gift of using medicine to help people. And pre-implantation genetic diagnosis is about helping people in serious need.* [topoi of utility]

In the example, the topos becomes apparent in the combination of both sentences. However, the model performs significantly worse with co-text window of five sentences in each direction. The high accuracy is only due to the predominant proportion of sentences in which there are no topos annotations.

- Further steps:**
- Evaluate model in which **function words are filtered from the co-text**
 - Create a model that is provided with **connectors typical for argumentation**
 - Model certain topoi, e.g., the danger topoi. There, additional **annotation of relevant phrases and words that led to the annotation decision.** These annotations are then added to the model as a manual attention mechanism.