



# Analysis of the political discourse on the example of the GermaParl corpus

## GermaParl

- Corpus of Parliamentary Protocols in the German Bundestag
- covers 72 years of debates
- 1949-2021
- PolMine Project (<http://polmine.github.io>)

## Motivation

### Political Sciences:

- Understanding Political Discourse
- Tracking Political Trends
- Political Behavior
- Policy Analysis and Development

### Computer Science:

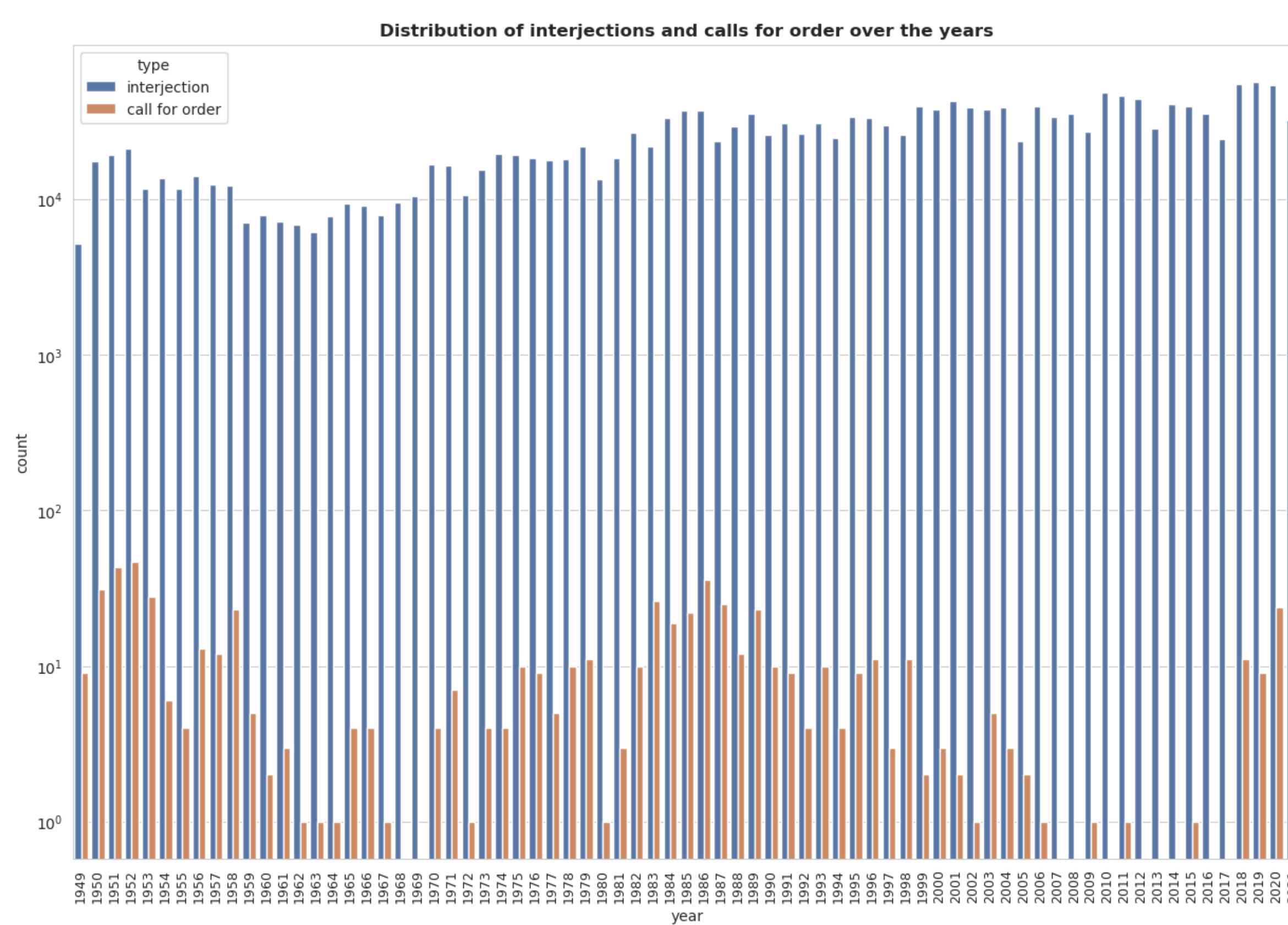
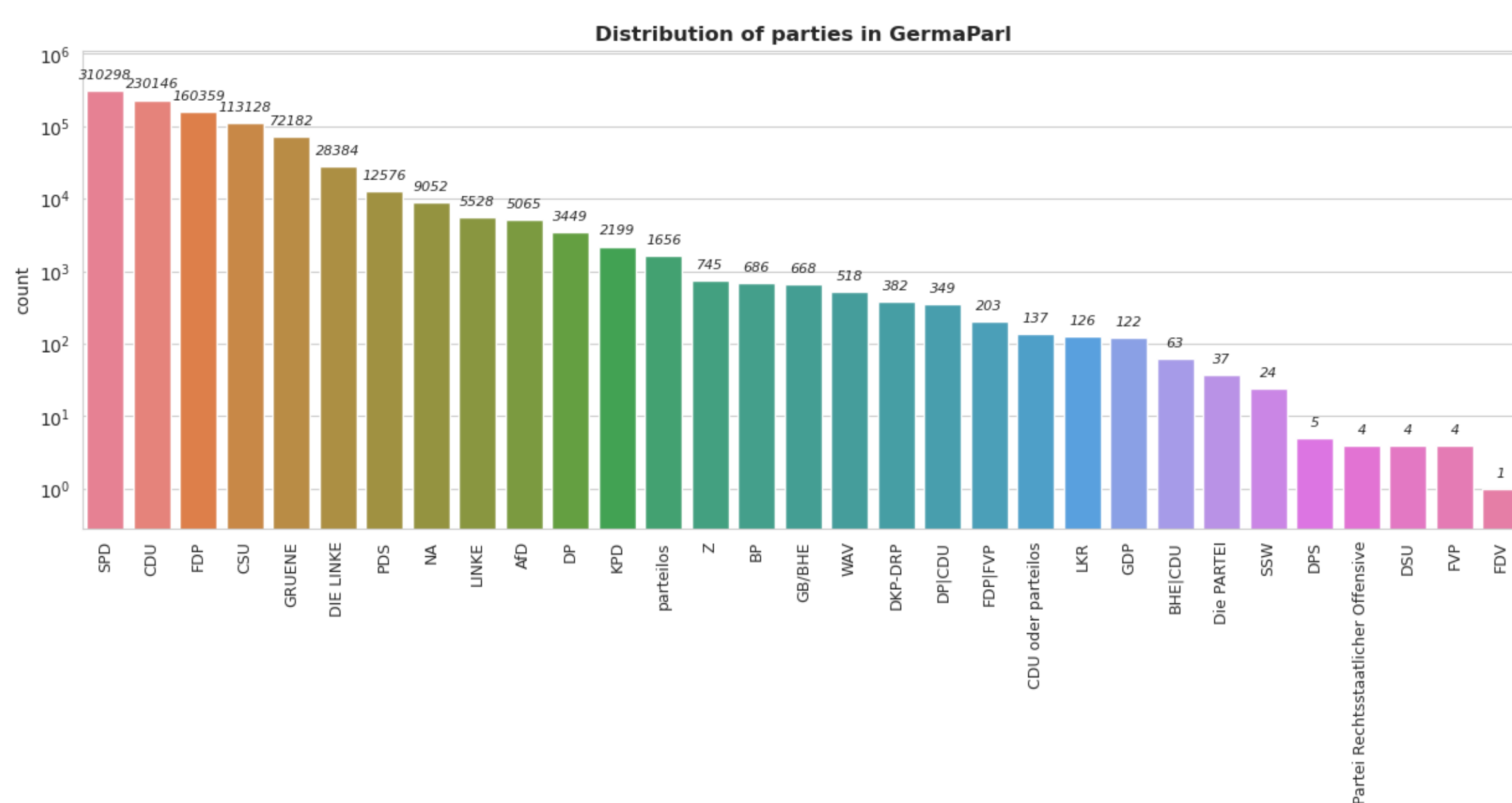
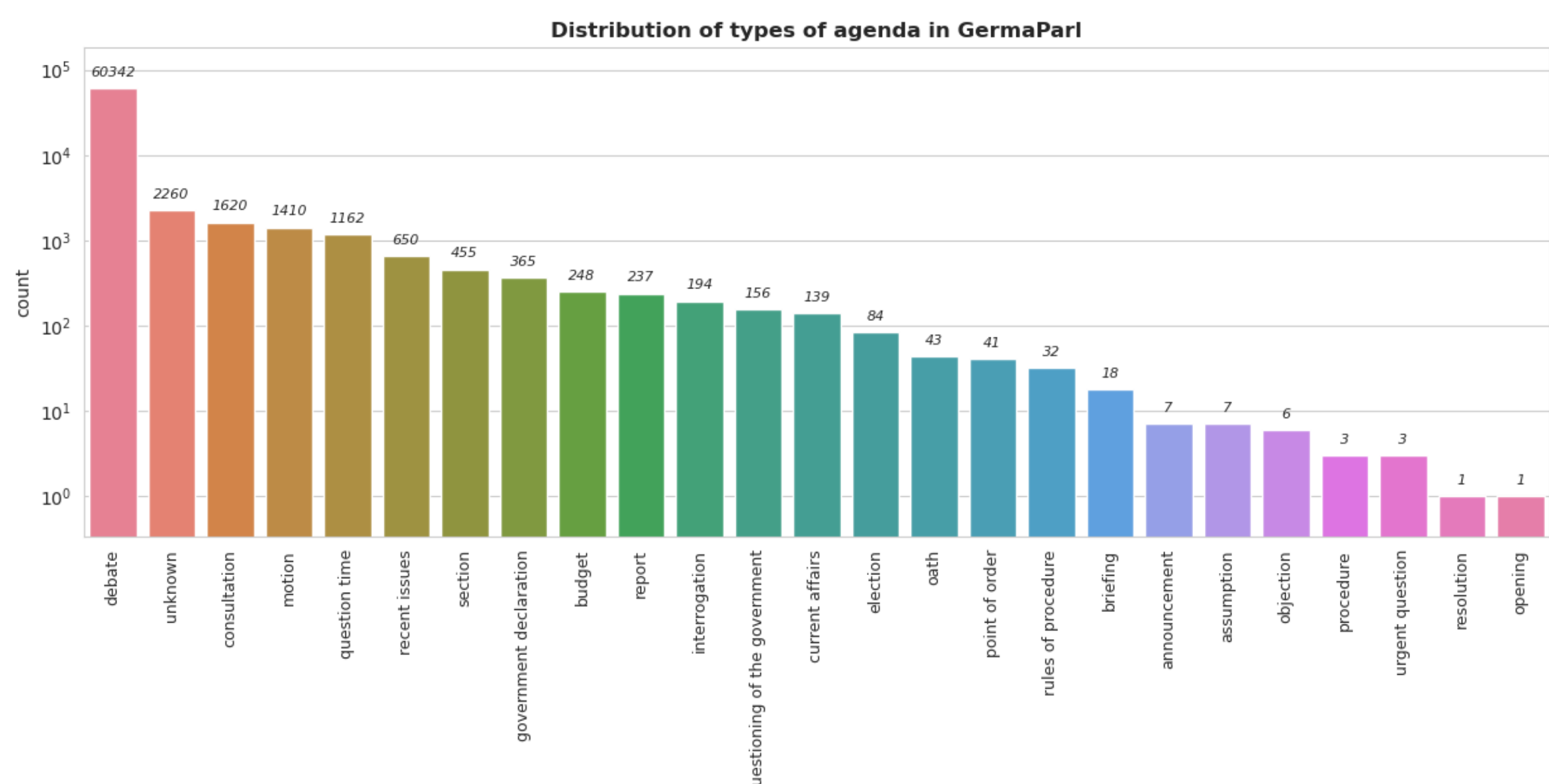
- Scalability
- Accuracy and Reliability
- Interdisciplinary Integration
- Transparency and Replicability
- New Insights

## Challenges

- **Clustering**
  - multiple topics in one speech are possible
  - language
  - context
- **Classification**
  - ideally, we should know topics, to do classification
- **Keyword extraction**
  - language
  - stop words specific to the political discourse
- **NER**
  - entities specific to the political discourse
- **Entity Linking**
  - small number of German models for Entity Linking
  - none found for the political texts

Topic	Count	Name
0	6142	0_fr_herr_herren_damen
1	3467	1_fr_dm_millionen_ber
2	3380	2_fr_bundesregierung_ber_deutschen
3	2954	3_damen_herren_herr_geschftsordnung
4	2825	4_abgeordneter_herr_redezeit_wort
5	2151	5_tagesordnung_punkt_beratung_nr
6	2132	6_gegenprobe_hand_erheben_angenommen
7	1726	7_abgeordneter_herr_wort_abgeordnete
8	1379	8_wort_aussprache_berichterstatter_minuten
9	1216	9_wort_abgeordnete_abgeordneter_herr
10	762	10_bundesminister_wort_herr_finanzen
11	478	11_dr_abgeordneten_krankheit_entschuldigt
12	355	12_beamtin_beamte_fr_gesetz
13	347	13_wort_abgeordnete_dr_reisnarr
14	220	14_richter_abgeordnete_wort_merkatz
15	196	15_abgeordnete_wort_bucarius_dr
16	188	16_loritz_abgeordneter_wort_herr
17	187	17_abgelaufen_redezeit_abgeordneter_abgeordnete

```
[{"minister": "tatbestande", "score": 0.5526}, {"minister": "staatschliessen", "score": 0.5223}, {"minister": "unrechtm\u00e4\u00dfige", "score": 0.5381}, {"minister": "herr", "score": 0.5853}, {"minister": "minister", "score": 0.5831}]
```



## Aims

1. Develop methods for comprehensive analysis of political texts in German
2. Compare different methods for clustering and keyword extraction for political discourse
3. Compare the performance of different models and approaches for clustering, keyword extraction and classification including LLMs

Beim Benzin kommt nat\u00fcrlich die \u00d6kosteuer hinzu. Um bei dem Beispiel zu bleiben: Was folgt denn letztendlich daraus? Im Fr\u00fchjahr 2000, Herr Wieseh\u00fcgel, haben fast alle Gewerkschaften, alle Tarifpartner f\u00fcr zwei Jahre Lohnabschl\u00fcsse get\u00e4tigt, wobei sie von einer Inflationsrate von um 1 Prozent ausgingen.

INTERJECTION: (Klaus Wieseh\u00fcgel [SPD]: Sie reizen mich, den Betrieb hier aufzuhalten!)

Im n\u00e4chsten Fr\u00fchjahr haben wir wieder Lohnverhandlungen. Ich frage mich: Wie weit soll sich denn diese Lohn-Preis-Spirale noch drehen, wenn keine wirksamen Entlastungen \u00fcber die Steuer bei den Leuten ankommen? Sie erw\u00e4hnen nun immer, dass die Entlastung einge treten w\u00e4re, und meinen damit nat\u00fcrlich, dass die Rentenversicherungsbeitr\u00e4ge um 1,2 Prozentpunkte gesenkt wurden. In diesen Tagen hat die AOK Baden-W\u00fcrttemberg ihre Beitr\u00e4ge um 0,7 Prozentpunkte erh\u00f6ht.

INTERJECTION: (Zuruf von der CDU/CSU: Wahnsinn!)

F\u00fcr die Menschen ist es v\u00f6llig egal, in welche Tasche das Geld geht. Sie stellen nur fest, dass immer weniger von ihrem Bruttolohn \u00fcbrig bleibt. Das ist Faktum. Wenn jetzt diese Erh\u00f6hung bei der Krankenkasse kommt, dann werden wir am Ende feststellen, dass die ganze Operation der Einf\u00fchrung der \u00d6kosteuer zur Senkung der Lohnzusatzkosten eine Chim\u00e4re gewesen ist, weil n\u00e4mlich am Ende dieser Operation genauso hohe Lohnzusatzkosten stehen wie vor dieser Operation.

INTERJECTION: (Beifall bei der CDU/CSU sowie bei Abgeordneten der F.D.P.)

Allerdings werden Sie den Menschen dann insgesamt rund 37 Milliarden DM, einschlie\u00dflich Mehrwertsteuern, durch die \u00d6kosteuer aus der Tasche gezogen haben. Deshalb ist diese Steuer das, was wir von Anfang an gesagt haben: Sie ist \u00f6kologisch unbrauchbar und dient nur dazu, die Menschen abzukassieren.

INTERJECTION: (Beifall bei der CDU/CSU sowie bei Abgeordneten der F.D.P.)

Operation

## Method

### Clustering:

- Testing of traditional methods for clustering (e.g., K-means, hierarchical clustering, DBSCAN)
- Clustering Performance: silhouette score, Davies-Bouldin index, and cluster purity to evaluate clustering results

### Keyword extraction:

- Testing of traditional techniques for keywords extraction (e.g., TF-IDF, RAKE, TextRank)
- Keywords Relevance: precision, recall, and F1-score?

Analysis of the strengths and weaknesses of each method.

Analysis of which methods perform best under various conditions (e.g., different text lengths, topics)

Comparison of traditional methods with models like GPT or Llama