
Cross-Lingual Analyses of Parliamentary Debates

— Federico Nanni, Goran Glavaš —
and Simone Paolo Ponzetto

Text Scaling

Current issues:

1. Topic-based analyses (e.g., extract ideological statements)
2. Cross-lingual analyses
3. Evaluation of unsupervised scaling methods

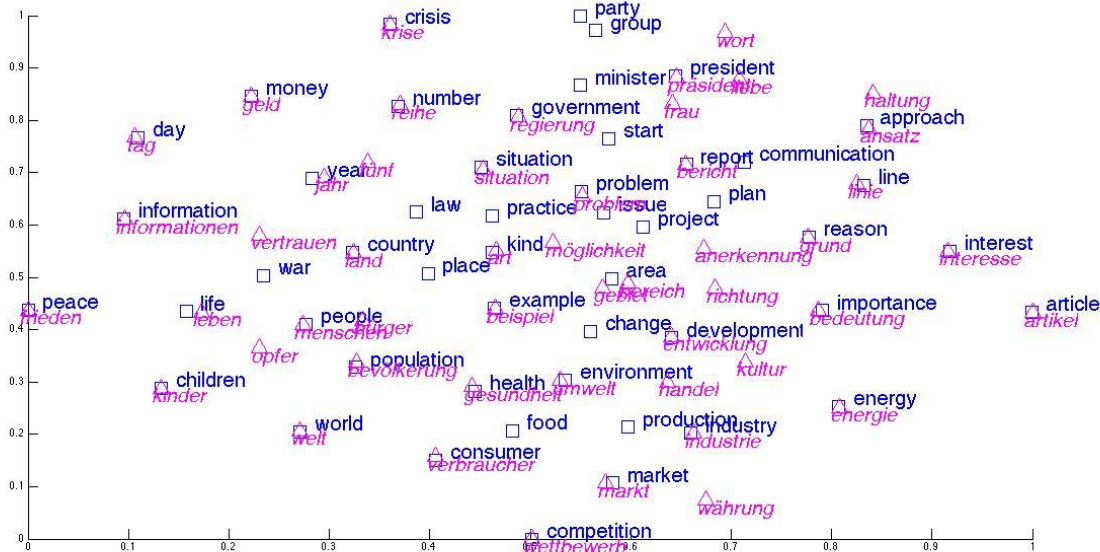
Text Scaling

Current issues:

1. Topic-based analyses (e.g., extract ideological statements)
2. **Cross-lingual analyses**
3. Evaluation of unsupervised scaling methods

Cross-Lingual Mapping of Word Embeddings

Linear translation matrix model to map word embeddings across languages.



Cross-Lingual Scaling (speeches from EP 5th leg)

	Monolingual			Cross-lingual		
	PA	r_P	r_S	PA	r_P	r_S
Random	49.7	-.03	.00	49.7	-.03	.00
Wordfish	55.0	.21	.20	–	–	–
AL-HFLP	61.3	.35	.31	57.3	.20	.25
AGG-HFLP	67.0	.53	.46	63.3	.34	.39

TopFish: Cross-Lingual Classification and Scaling

- 1) Mono and Cross-lingual Classification (based on Conv. Neural Net)
- 2) Mono and Cross-lingual Scaling
- 3) Wordfish (our python implementation)

<https://github.com/codogogo/topfish>

Issues

1. Our tool **does not work** when the texts come from national parliaments

Issues

1. Our tool **does not work** when the texts come from national parliaments
2. We did not really approach this project in a **“interdisciplinary”** way

Issues

1. Our tool **does not work** when the texts come from national parliaments
2. We did not really approach this project in a **“interdisciplinary”** way
3. Our work has been oriented towards making a **NLP** contribution

Larger Issue

Absence of an interdisciplinary venue, relevant for both communities, for publishing works that involve NLP and political science research.

Thanks

Federico Nanni

Data and Web Science Group

University of Mannheim

federico@informatik.uni-mannheim.de