What Is This All About?

This publication is part of the DYSEN Project which stands for: Dynamic Sentiment Analysis as Emotional Compass for the Digital Media Landscape.

Research Question: How do print media report about Viennese Politicians?

Aim of this project: Develop a tool that can detect change of emotional polarization of politicians in Austria.

Problem

Currently there is no dictionary based on Austrian-German in the domain of news media and politics.

To resolve this research gap the Austrian Language Polarity in Newspapers (ALPIN) sentiment dictionary is introduced.

Data Sources

Viennese Politicians

- Online archive of Vienna (OLAV) of the Vienna City and State Archives
- Members of the Vienna City Council
- Vienna City State
- Vienna Urban Parliament
- Vienna State Parliament
- action between the 13th and 20th parliamentary term
- 497 politicians

Standard Posts (STP)

- German-Austrian print media
- Preprocessed and linguistically annotated (Bernauer et al., 2017)

Our Data

- Print media relates to Vienna between 1996 and 2017
- 14,849 and 1271 articles (The New York Times)
- Test snippets of around 40 articles around the politicians’ terms were retrieved

- Fleiss-Kappa: 0.295 (fair inter-annotator agreement)

- (70 annotators after excluding the 14 bad ones)

Methods: AMC & STP

- 3) Department of European and Comparative Literature and Language Studies
- tanja.wissik@oeaw.ac.at

Austriacisms

- Based on: 
  - Austrian Language Polarity in Newspapers (DYSEN)
  - Vienna City and State Archives
  - Politician archive of Vienna (POLAR) of the 13th and 20th parliamentary term

Austriacism Webpage based on the ALPIN dictionary. Link: https://dysen-tool.acdh-dev.oeaw.ac.at/

Methodology

- Data sources
  - Data annotation
  - Data processing
  - Evaluation

- Data annotation

- AMCo: AFCı System

- German-Austrian print media
- Preprocessed and linguistically annotated (Bernauer et al., 2017)

- Data processing

- Output: 3046 labeled text snippets including Viennese politicians

Post Processing

- Relabeling of annotation (1/4 of annotated data)
- F1-score: 0.78
- $F_\beta$: 0.68

Methods: Austriacisms

- Crowd Sourcing: Austriacism
  - Standard Posts (STP)

- Data annotation

- AMC

- STP

- Data processing

- Evaluation

- ALPIN sentiment dictionary (Austrian Language Polarity In Newspapers)

Web Application

- Post Processing

- Comparison of words which occur in both dictionaries (amc+derStandard vs austriacisms)

- Output: 4417 tuples

Results

- Precision: 0.78
- $F_\beta$: 0.68
- Recall: 0.79

Future Work

- Mitigate the potential bias due to labelling
  - Improvement of the text extraction by using Aspect-based sentiment analysis
  - Investing more money to label a bigger dataset
  - Expanding the scope of the project to all politicians and media in Austria

Discussion

- Difficult to label news media (mainly "neutral"
  - Limited text length
  - No external dataset for evaluation
  - Potential bias during labelling e.g. words like "Flüchtling" negatively annotated

Reference

- Thomas Koltô, Katharina Sekanina, Bettina J. M. Kern, Julia Neidhardt, Tanja Wissak, Andreas Baumann

Contact

1) Faculty of Informatics, TU Wien
2) Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences
3) Department of European and Comparative Literature and Language Studies and Department of English and American Studies, University of Vienna

Data

https://doi.org/10.5281/zenodo.5857150