EXPLORING NETWORK OF COMPARATIVE STRUCTURES AND COMPARATIVE OPINIONS IN NEWSPAPER CUTS

Erdal Ayan, PhD Candidate
Computational Linguistics, Kassel University, Germany
Email: erdal_ayan@yahoo.com

Introduction

- Comparative opinion mining: focus is more on customer or consumers' opinions on certain products.
- There are very few studies on opinion mining presented in newspaper articles.
- Apart from some studies in Chinese, Korean and English, there has not been much research carried out on comparative sentence detection and learning in other languages by using supervised and unsupervised techniques in the field.
- Opinion mining/sentiment analysis is not adequate in defining exact opinions of the people since they only provide judgments and feelings/emotions that people have during their talks about products and services.
- Comparative opinion mining is an alternative that can produce more precise, distinct and concrete information on the topics.
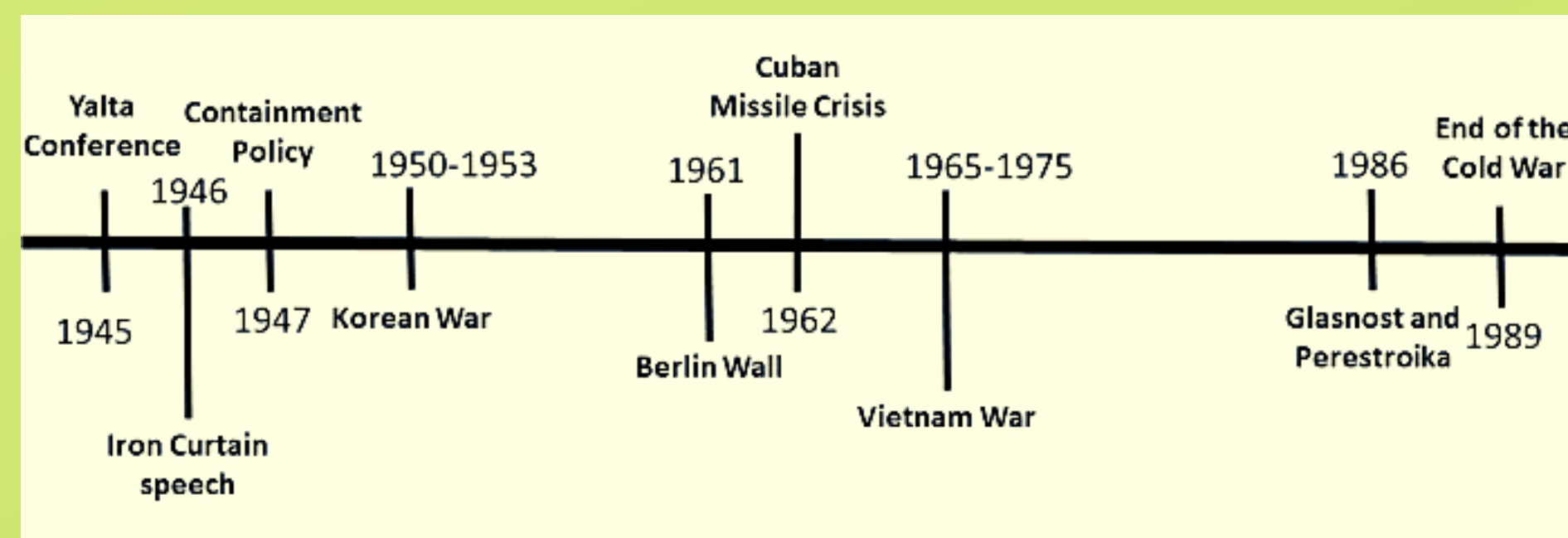
Related Work

- Opinion mining has emerged as a significant field and method focusing on analyzing public opinion. The method is based on automatic detection of polarized opinion(s) (negative, positive or neutral) presented in the textual materials.
- Opinion mining is widely used in business sectors, which helps companies figure out customers' decision-making processes.
- Comparative opinion mining is an improving field of opinion mining among the experts particularly working on text analytics. Comparative forms (e.g. *My car is better than your car.*), which depict comparisons and evaluations on any topic/thing, are mostly explored in comparative opinion mining research.
- Jindal and Liu (2006a) have defined four types of comparatives (i.e. a) non-equal gradable, b) equative, c) superlative, and d) non-gradable).
- The approaches employed in the comparative opinion mining process can be listed as a) Machine Learning and b) Natural Language Processing (NLP) (Varathan, Giachanou and Crestani, 2017).
- Lexical-semantic network analysis is basically a methodology based on visualizing the hierarchical order of lexical items, networks and relationships of topics and words in the textual material. The networks appear via links between topics and words that occur in proximity and create concepts (Choi & Lecy, 2012).
- Lexical-semantic analysis for written languages could be helpful to understand cases activating linguistics processes. Lexical-semantic network analysis as a form of text mining is closely connected to information processing. (Smit & Van Der Graaf, 2012; Soriano, Au, & Banks, 2013).

Purpose & RQs

- Comparative opinion mining methodologies highly regard only comparatives but other forms of comparatives produced by adverbials and conjunctions deserve a closer look.
- Main purpose of the current research is to explore lexical network of comparatives including comparative adverbials and conjunctions, as well as the topics emerged during the Cold War Era.
- The study will contribute to the field by shedding light on unique reactions and opinions referring to agreements and disagreements in ideological perspectives with certain linguistic structures by the authors, which simply create public opinion.
- Visualizing lexical-semantic networks of the comparative forms and structures which were frequently used in both blocks of newspapers will contribute to the field of comparative opinion mining.

Research Questions:

- What kind of comparative forms are used for comparative opinions with respect to certain topics in the press clippings?
- How do comparative opinions, based on certain topics, change over time in western and eastern newspapers during the Cold War Era?
- Are there similarities or differences between western and eastern newspapers in terms of comparative opinions (positive, neutral or negative) of the authors?
- Are there similarities or differences in comparative opinions (positive, neutral or negative) of the authors in terms of their a) gender, b) profession, c) level of education and d) origin?

Methodology

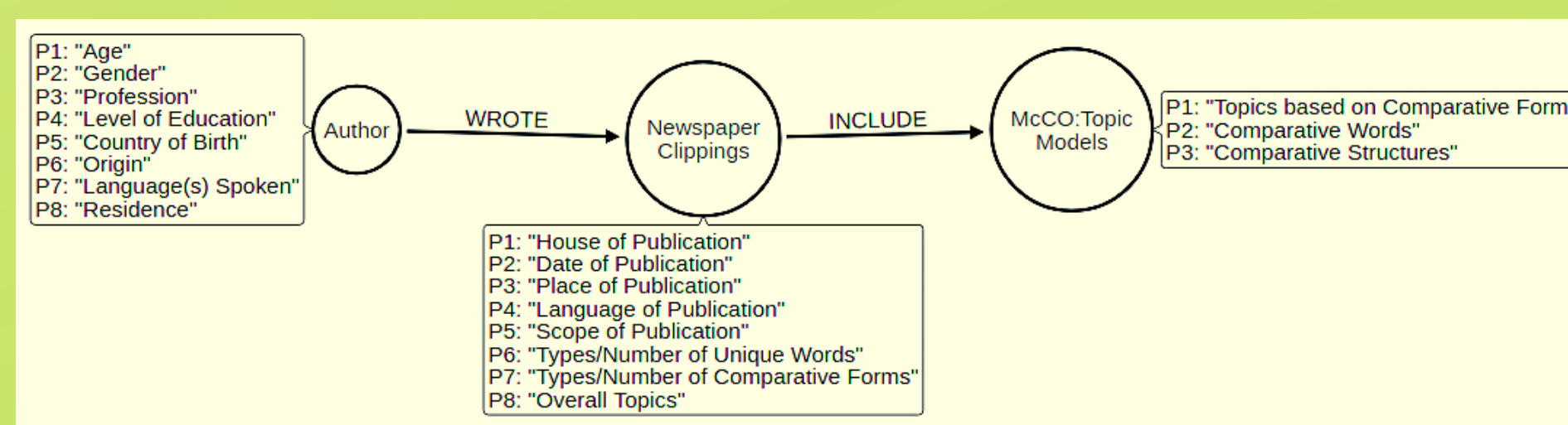### Corpus-based Media Content Analysis

- Press clippings -> collections of particular filtered news articles.
- Press Clippings Archive at Herder Institute -> 5 Million press clippings from 1916 to today.
- Digitization process -> 9750 press-clippings by 1318 different authors.
- Regional scope -> Eastern Central Europe, the Soviet Union/Russia, the GDR, the Baltic Sea region, S. East Europe
- Thematic Scope -> Personal Archives, Local Archives, Thematic Archives (History, Politics, etc.)
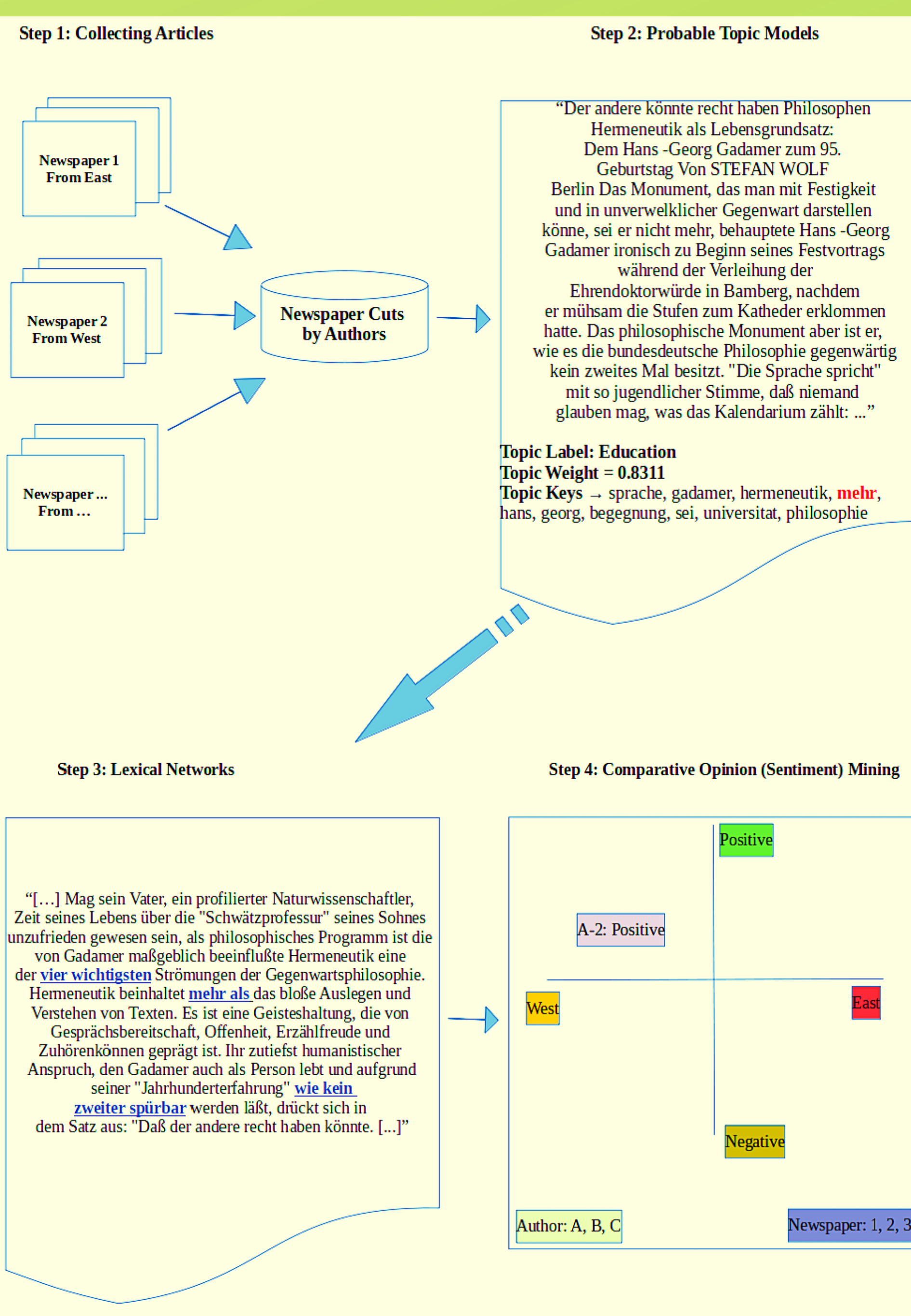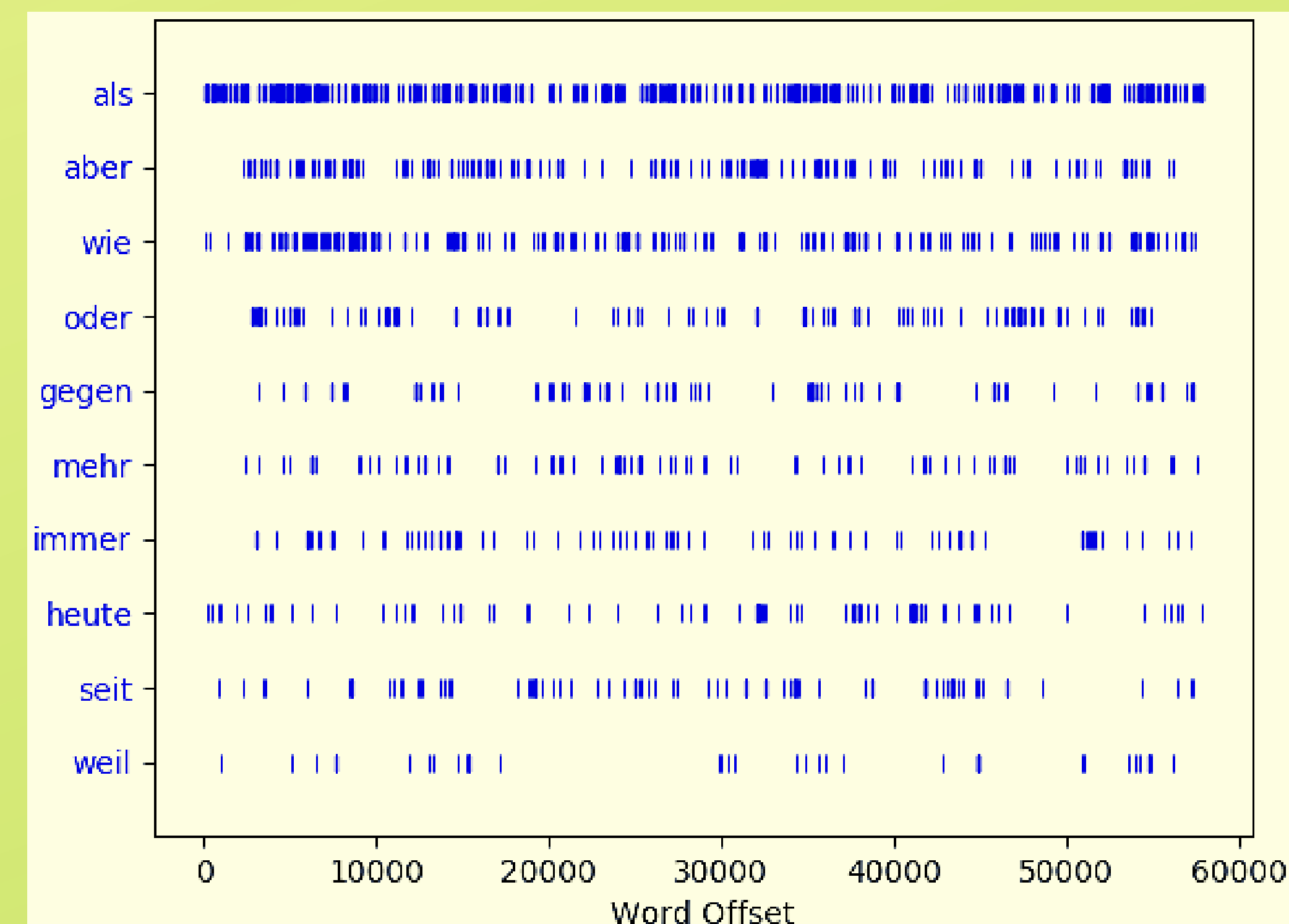

Cold War Timeline
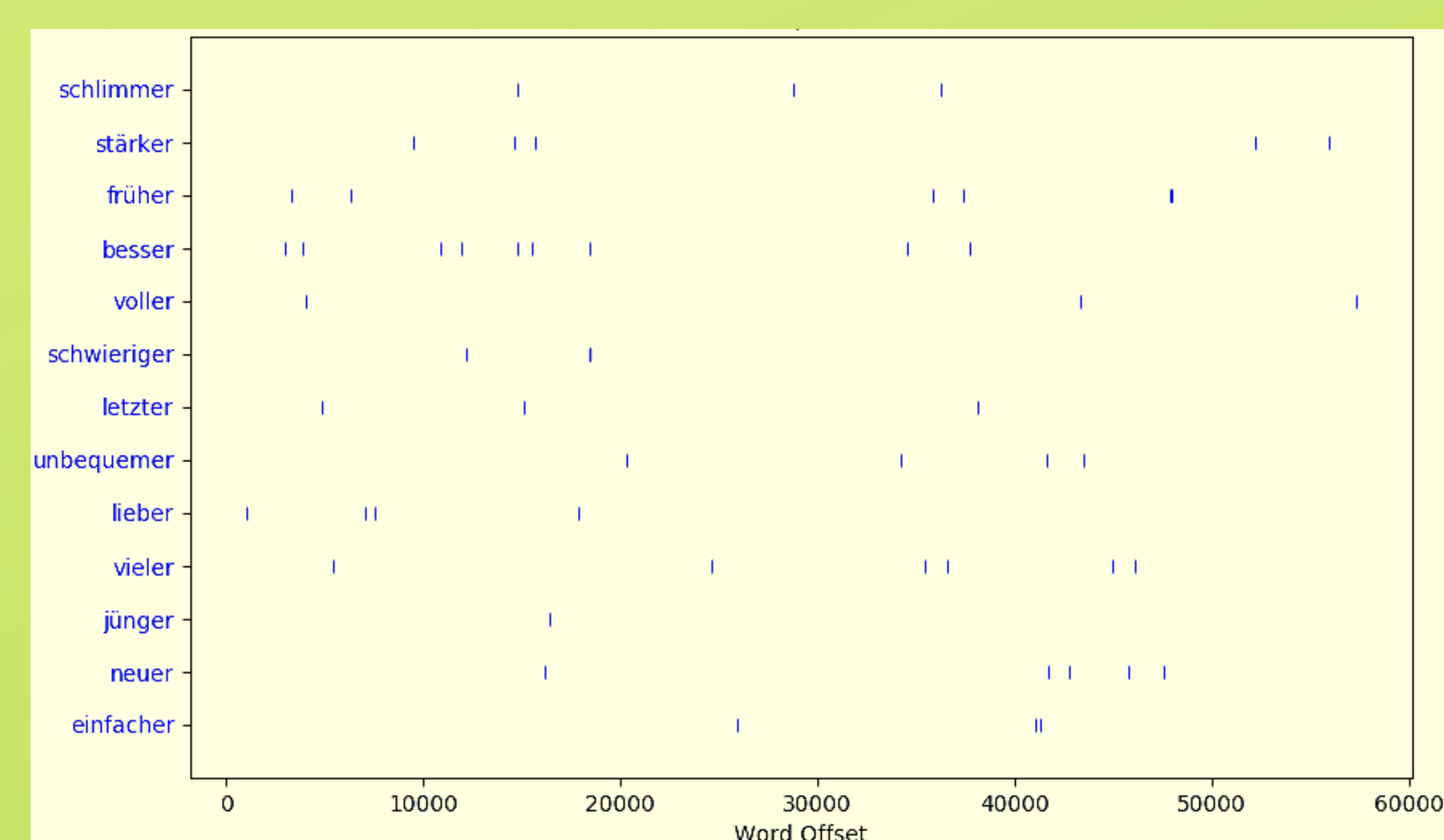

Data Model

Data Processing Model

Tools for Data Collection & Analysis

Natural Language Analysis with Python NLTK

neo4j

Preliminary Findings

Dispersion of Selected Conjunctions

Dispersion of Selected Comparatives

Concordance View of a Conjunction, "aber"

Lexical Networks & Selected Topics
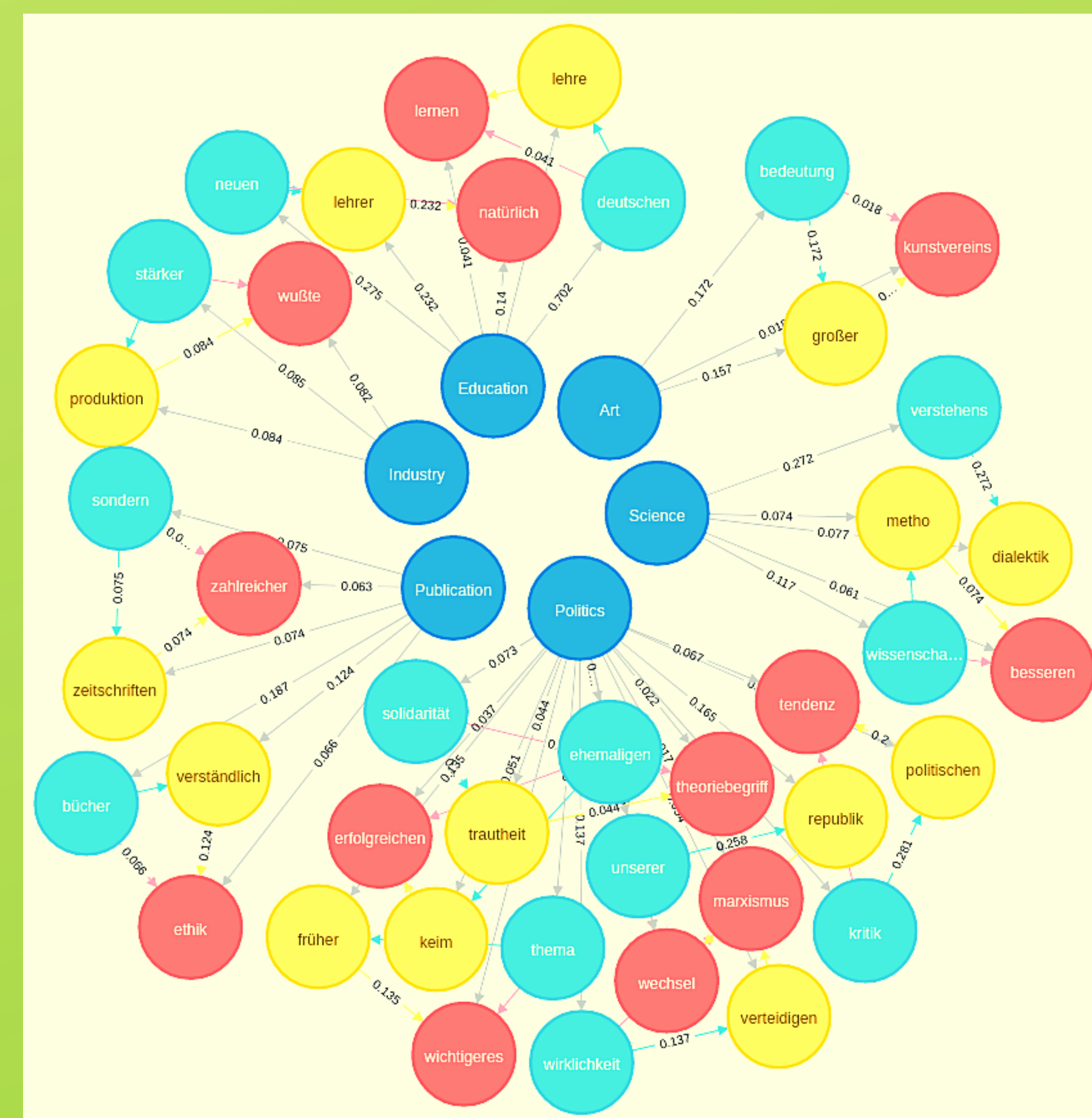
Selected References

Balahur, A., & Steinberger, R. (2009). Rethinking Sentiment Analysis in the News: from Theory to Practice and back. In *Proceedings of the 1st Workshop on Opinion Mining and Sentiment Analysis* (p. 12). Retrieved from https://pdfs.semanticscholar.org/533e/1e815d569820d85d093e00e5b32611d9332a.pdf

Choi, C., & Lecy, J. D. (2012). A Semantic Network Analysis of Changes in North Korea's Economic Policy. *Governance*, 25(4), 589–616. https://doi.org/10.1111/j.1468-0491.2012.01597.x

Gaur, R. C., Chand, M., Gaur, K., & Yadav, A. S. (2013). Online Newspaper clippings & News Services for Libraries: Experiences in Indian Libraries. 16.

Jindal, N., & Liu, B. (2006a). Identifying Comparative Sentences in Text Documents. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 244–251). New York, NY, USA: ACM. https://doi.org/10.1145/1148170.1148215

Lipinsky, J. (2013). Presseausschnittarchiv online: Zwischen Open Access und Urheberrechtsschranken. *Herder-Aktuell*, 36(Januar-Juni 2013), 18.

Lipinsky, J. (2014a). Baltica in der Forschungsbibliothek. *Herder-Aktuell*, 39(Juli-Dezember 2014), 18.

Lipinsky, J. (2014b). *Die Zeitungsausschnittsammlung – digitaler Zugriff statt Blättern in zerfallendem Papier*. Oral Presentation presented at the Deutscher Bibliothekartag, Bremen. Retrieved from https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/year/2014/docId/1447

Pollak, S., Coesemans, R., Daelemans, W., & Lavrač, N. (2011). Detecting contrast patterns in newspaper articles by combining discourse analysis and text mining. *Pragmatics*, 21(4), 647–683. https://doi.org/10.1075/prag.21.4.07pol

Smit, E., & Van Der Graaf, M. (2012). Journal article mining: the scholarly publishers' perspective. *Learned Publishing*, 25(1), 35–46. https://doi.org/10.1087/20120106

Soriano, J., Au, T., & Banks, D. (2013). Text mining in computational advertising. *Statistical Analysis and Data Mining*, 6(4), 273–285. https://doi.org/10.1002/sam.11197

Varathan, K. D., Giachanou, A., & Crestani, F. (2017). Comparative opinion mining: A review. *Journal of the Association for Information Science and Technology*, 68(4), 811–829. https://doi.org/10.1002/asi.23716