Universität Münster

# Text as Data

## Session 12: Wrap-Up

Mirko Wegemann

22 July 2026

# Schedule of today's session

- Review of learning outcomes
- Deep learning with embeddings
- Evaluation Criteria for term paper

# Evaluation of your learning outcomes

- If there is one thing you learned from the seminar, which is it?
- If there is one method you'd like to know more about, which is it?
- If there is one thing you'd change in the course structure, which is it?

# R-Crash Course

We started with a two-week crash course on `R`. Our focus was primarily on...

- Objects
- Import of data
- Data preparation via the `tidyverse`

# Concept clarification I

We learned that quantitative text analysis is a form of content analysis that discovers relationships and regularities between different text fragments.

"Computational text analysis (also called Quantitative Text Analysis, Automated Content Analysis, Text Mining, Text as Data etc.) draws on techniques developed in natural language processing and machine learning to analyse textual documents." (Chun Ting-Ho 2021)

# Data collection

You have learned to collect data.

- Access existing data sources and import into `R`
- Scraping static web pages via `rvest`
- Scraping dynamic web pages using `RSelenium`
- API access through API wrapper or `jsonlite`

# Data preparation I

We went through the conventional data preparation pipeline of *bags-of-words* approaches:

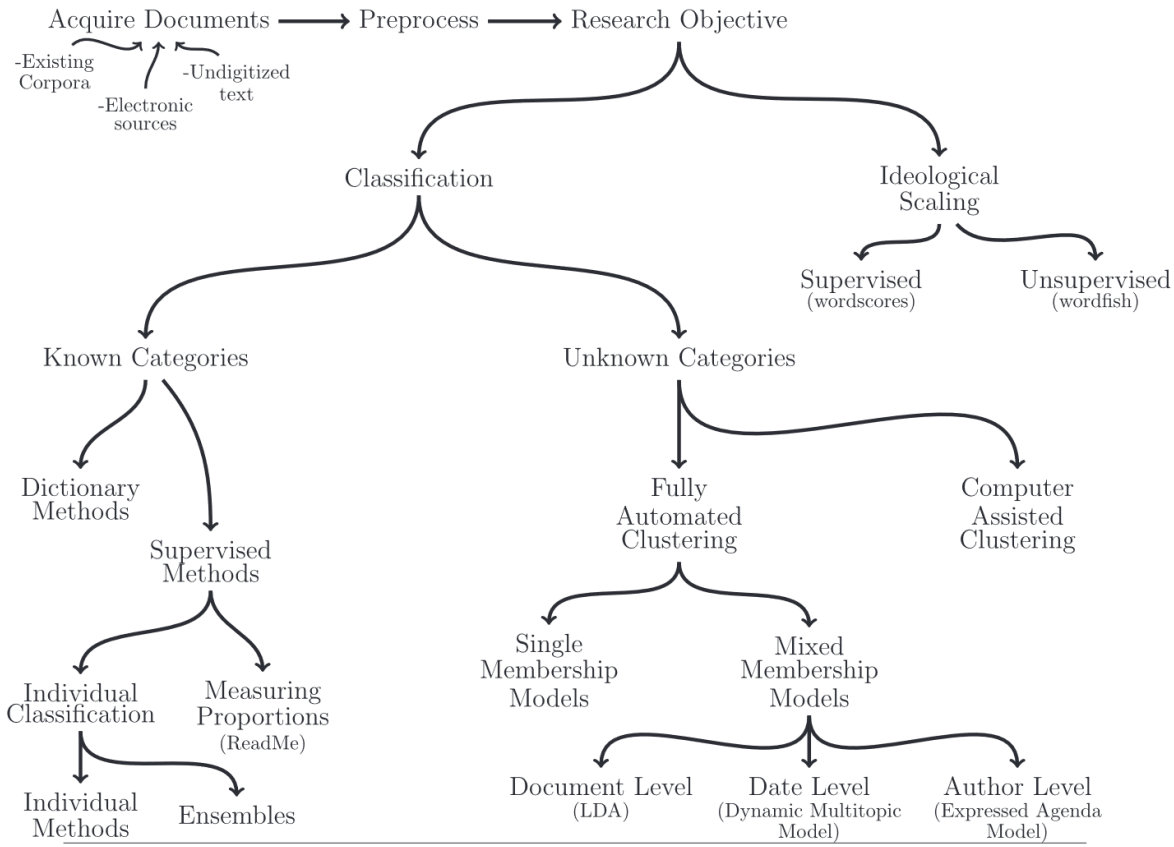
- Import data
- Filtering the data if necessary
- Conversion to a *corpus* object
- ... further conversion to a *tokens* object
- ... until the transformation into a **d**ocument-**f**requency-**m**atrix

# Data preparation II

In the process, we often made some changes to the data:

- Lowercasing
- Remove features without semantic value (mostly punctuation, numbers, stop words, symbols)
- Removing Particularly Common/Rare Features
- n-grams
- Stemming or lemmatization

# Data analysis I




Text Analysis Methods

# Data analysis II

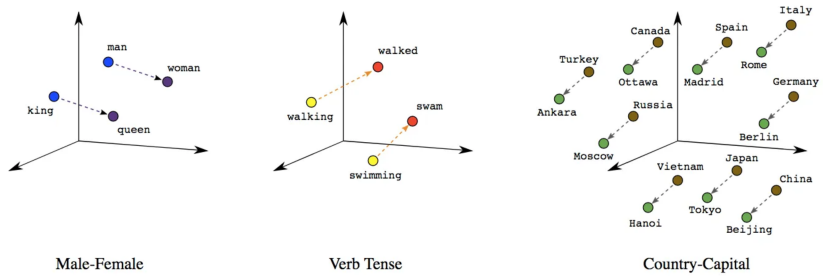
We have dealt with the following analysis methods:

- Dictionary Analysis
- Unsupervised topic models
- Supervised classification

# Data analysis III

In addition, we learned different representation structures:

- bags-of-words
- embeddings




Male-Female Verb Tense Country-Capital

# The evaluation of the project work I

## Formal Criteria

As a reminder:

- Submission of the term paper by **30 September 2026**
- Submission of a .pdf file and an R script. The analysis must be replicable. If a data set is required for this, submit it.
- Length of paper: 4,000 words excl. abstract and bibliography)
- Correct citation, transparency in AI use
- Have all tables and figures been labeled and referenced?
- All other formal criteria, in particular the Word template to be used, can be found here

# Content criteria I

- You can choose between an extensive newspaper report and a policy brief
- Focus in particular on these criteria:
  - Develop a timely research question
  - Embed it briefly in the literature
  - Short theory section that outlines main concepts and theoretical expectations
  - At the heart of the paper: the method part
  - Description of data source and explanation of operationalization of theoretical concepts
  - Method selection and explanation of implementation
  - Weighing up the advantages and disadvantages of the method you chose
  - Validation of the method (classification metrics, possibilities according to Quinn et al. (2010), examples)

# Content criteria II

- Correct description of the results
- Feedback of the results to the research question
- Addressing the strengths and limitations of the present work

# Structural criteria

- Do the individual parts of the work flow well into each other?
- Have repetitions been avoided?

# Literature

Powner, L. C. (2015). *A political science student's practical guide*. Sage/CQ Pres

- Introduction to Academic Writing in Political Science
  Overview of all sub-steps of an empirical work
- The first 3 chapters (research question, theory and literature review) are particularly helpful

On the writing style in academic articles:
Schimel, J. (2012, January). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded*. Oxford University Press, USA

# Outlook

- Those who are writing a term paper (@Jonah): Arrange a consultation
- Feel free to attend the seminars of our department next semester:
  - 'Einführung in die Politikwissenschaft' by Prof. Daniel Bischof
  - 'Marching on the Streets, but not Changing Minds? How Protests Mobilize, Strategize, and Succeed' by Elena Leuschner
  - 'Gleiches Spiel, andere Parteien: Parteipolitik in vergleichender Perspektive' by myself (block seminar)

# References I


Powner, L. C. (2015). *A political science student's practical guide*. Sage/CQ Pres.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010).How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science*, *54*(1), 209–228. https://doi.org/10.1111/j.1540-5907.2009.00427.x

Schimel, J. (2012, January). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded*. Oxford University Press, USA.