Textual Data Analytics in EU Contexts

Module: (POL60701) Digital Transformations in Europe - Seminar 2
Lecturer: Dr. Burak Giray

Fall 2024

E-mail: giray.burak@tum.de
Office Hours: By appointment
Office: H.405

Web: burakgiray.com/teaching
Day and Time: Th 13:15–14:45
Course Format: *Onsite*

Seminar Description

This seminar is designed to introduce students to quantitative textual analysis methods using R. Following a brief introduction to R, RStudio, basic R operations, and relevant packages, the seminar progresses to text mining techniques, including reading text into R, tokenizing textual data, and implementing lemmatization and stemming. Students will also learn to perform co-occurrence and sentiment analyses. In the second half of the course, they will apply these R skills to real-world European datasets, with a focus on how text analysis can be used in EU-related contexts such as politics, law, and social media. Key scholarly readings employing these methods will be discussed throughout the course. By the end of the seminar, students will be able to apply text analysis in the EU context and develop a research project on EU politics, presenting a theoretical argument supported by empirical evidence.

Seminar Style

Each session is based on required readings, which participants must review prior to the session. Consequently, all attendees are expected to grasp the key points from the required readings and contribute comments or critiques during group discussions. The reading list indicates which readings are eligible for presentation. Recommended readings are optional. Most sessions will also feature a lab component where participants will learn and practice that week's skills using pre-prepared R scripts. Students are encouraged to run the code in RStudio on their own laptops during the sessions. These scripts will be shared in advance of each session.

Examination of the module

The module grade is based on a written examination in the form of a seminar paper (13 – 15 pages). Through their seminar paper, students will document to which extent they have gained a

foundational understanding in the area of digital transformations in civil conflicts and peace operations. The seminar paper is accompanied by a presentation (10-15 min) to test the communicative competence of presenting scientific topics to an audience. The presentation must be supplemented by a short-written handout.

- **Presentation (20%):** Each session contains a presentation part of 12-15 minutes per student. For each presentation, a handout of max. two pages is prepared. The presentation (1) summarizes the core statements of the supplementary texts briefly, (2) clarifies theoretical connections using empirical examples chosen by the students themselves, and (3) places the topic in the larger seminar context. In addition, (4) media supporting the content are used (e.g. blackboard, presentation slides, etc.) The presentation serves as content preparation for the seminar paper.

- **Seminar Paper (80%):** Written paper of 13-15 pages (without bibliography and title page) based on a theoretical question in the context of the seminar topic. The seminar paper explores the compulsory texts and the supplementary reading for a session in more depth. In addition, further secondary and primary sources should be consulted to answer the chosen question. Topic, question and approach can be discussed and determined together with the lecturers during the lecture period. If you would like to take advantage of this, please prepare a draft of your seminar paper (topic/session, possible question, theory, case selection) of no more than one page and send it by e-mail to the lecturer at least three days before the office hours. The seminar paper must be submitted via Moodle by **March 15, 2025** (midnight) at the latest.

Session Overview

Session	Session Date	Session Title
1	17.10.2024	Course Overview & Introduction to R
2	24.10.2024	Types of Data
3	31.10.2024	Descriptive Statistics and Visualization
4	07.11.2024	Reading Text into R and Preprocessing
5	14.11.2024	Rule-based Approaches and Co-occurrence Analysis
6	21.11.2024	Sentiment Analysis using Tidytext
7	28.11.2024	Practicing with YouTube Data
8	12.12.2024	YouTube in EU Politics
9	19.12.2024	Social Media Data (Twitter)
10	09.01.2024	Political Communication on Twitter
11	16.01.2025	Text Mining in Political Documents
12	23.01.2025	EU Trade Agreements, Policy Papers, and Party Manifestos
13	30.01.2025	Final Presentations
14	06.01.2025	Wrap-up Session

Installing R & RStudio

R: R is the programming language we'll use to import, edit, and analyze data. You'll have to specify your operation system to download the right version:

- R for Windows — `https://cran.r-project.org/bin/windows/base/`

- R for Mac — `https://cran.r-project.org/bin/macosx/`

R Studio: R Studio is a desktop application with a graphical interface that facilitates programming with R. The newest version of R Studio (1.4.1717) can be downloaded via this — `https://posit.co/download/rstudio-desktop/`

1 Course Sessions and Readings

1.1 Textual Analysis in R

Session 1: Course Overview & Introduction to R	
Required Readings	Course Syllabus
	K. Benoit (2020). Text as Data: An Overview. Handbook of Research Methods in Political Science and International Relations. Ed. by L. Curini and R. Franzese. Thousand Oaks: Sage: 461–497.
Lab Session	Installing RStudio and necessary packages, and performing basic R operations (e.g., creating variables).
Recommended Reading(s)	Grimmer and B. M. Stewart (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis 21 (3): 267–297.

Session 2: Types of Data	
Required Readings	Winter, B. (2019). Statistics for linguists: An introduction using R. Routledge (Chapter 1)
Lab Session	Importing/exporting various data formats (e.g., CSV, Excel), basic data manipulation using tidyverse, and understanding different data types (numeric, character, factor)

Session 3: Descriptive Statistics and Visualization	
Required Readings	Winter, B. (2019). Statistics for linguists: An introduction using R. Routledge (Chapter 3)
Lab Session	Calculating basic descriptive statistics (mean, median, mode, standard deviation), aggregating data, Creating visualizations using ggplot2 (e.g., histograms, box plots)
Recommended Reading(s)	Hehman, E., & Xie, S. Y. (2021). Doing better data visualization. Advances in Methods and Practices in Psychological Science, 4(4), 25152459211045334.

Session 4: Reading Text into R and Preprocessing	
Required Readings	Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text analysis in R. Communication methods and measures, 11(4), 245-265.
Lab Session	Tokenization of text data, generating frequency distributions, creating word clouds, and removing stopwords and applying lemmatization/stemming techniques
Recommended Reading(s)	Silge, J. (2017). Text mining with R: A tidy approach. O'Reilly Media, Inc. (Chapter 1: The tidy text format)

Session 5: Rule-based Approaches and Co-occurrence Analysis	
Required Readings	Korger, A., & Baumeister, J. (2021, September). Rule-based Semantic Relation Extraction in Regulatory Documents. In LWDA (pp. 26-37).
Lab Session	Conducting keyword-in-context analysis, and analyzing co-occurrences around specific keywords

Session 6: Sentiment Analysis Using Tidytext	
Required Readings	Misuraca, M., Forciniti, A., Scepi, G., & Spano, M. (2020). Sentiment Analysis for Education with R: packages, methods and practical applications. arXiv preprint arXiv:2005.12840.
Lab Session	Calculating positive and negative sentiments, utilizing sentiment lexicons for text analysis, and visualizing sentiment analysis results
Recommended Reading(s)	Saini, S., Punhani, R., Bathla, R., & Shukla, V. K. (2019, April). Sentiment analysis on twitter data using R. (pp. 68-72). IEEE.

1.2   Potential Data Sources in the EU Context

Session 7: Practicing with YouTube Data	
Required Readings	Breuer, J., Kohne, J., & Mohseni, M. R. (2023). Using YouTube data for social science research. In Research Handbook on Digital Sociology (pp. 258-277). Edward Elgar Publishing.
Lab Session	Extracting, cleaning, and transforming YouTube data (views, likes, comments, etc.) from YouTube API
Recommended Reading(s)	Ackland R., Gertzel B., & Borquez F. (2022). Introduction to vosonSML (YouTube section): `https://cran.r-project.org/web/packages/vosonSML/vignettes/Intro-to-vosonSML.html`

Session 8: YouTube in EU Politics	
Required Readings	Czymara, C. S., Dochow-Sondershaus, S., Drouhot, L. G., Simsek, M., & Spörlein, C. (2023). Catalyst of hate? Ethnic insulting on YouTube in the aftermath of terror attacks in France, Germany and the United Kingdom 2014–2017. Journal of Ethnic and Migration Studies, 49(2), 535-553. **(Presentation)**
	Uba, K., & Jansson, J. (2021). Political campaigns on YouTube: trade unions' mobilisation in Europe. New technology, work and employment, 36(2), 240-260. **(Presentation)**
Recommended Reading(s)	Heuer, H., Hoch, H., Breiter, A., & Theocharis, Y. (2021). Auditing the biases enacted by YouTube for political topics in Germany. In Proceedings of Mensch und Computer 2021 (pp. 456-468).

Session 9: Social Media Data (Twitter)	
Required Readings	Chen, K., Duan, Z., & Yang, S. (2022). Twitter as research data: Tools, costs, skill sets, and lessons learned. Politics and the Life Sciences, 41(1), 114-130.
Lab Session	Preparing raw Twitter data by cleaning tweets (removing hashtags, mentions, URLs, stopwords, and special characters) and tokenizing them for deeper keyword and sentiment analysis
Recommended Reading(s)	Jungherr, A. (2023). Digital campaigning: How digital media change the work of parties and campaign organizations and impact elections. In Research Handbook on Digital Sociology (pp. 446-462). Edward Elgar Publishing.

Session 10: Political Communication on Twitter	
Required Readings	Silva, B. C., & Proksch, S. O. (2022). Politicians unleashed? Political communication on Twitter and in parliament in Western Europe. Political science research and methods, 10(4), 776-792. **(Presentation)**
	Usherwood, S., & Wright, K. (2017). Sticks and stones: Comparing Twitter campaigning strategies in the EU referendum. The British Journal of Politics and International Relations, 19(2), 371-388. **(Presentation)**
Recommended Reading(s)	Bauer, P. C., Ecker, A., Imre, M., Landesvatter, C., & Malich, S. (2023). Who tweets, and how freely? Evidence from an elite survey among German politicians. Research & Politics, 10(1), 20531680221144237.
	Jungherr, A. (2023). Digital campaigning: How digital media change the work of parties and campaign organizations and impact elections. In Research Handbook on Digital Sociology (pp. 446-462). Edward Elgar Publishing.
	Rocca, R., Lawall, K., Tsakiris, M., & Cram, L. (2024). Communicating Europe: a computational analysis of the evolution of the European Commission's communication on Twitter. Journal of Computational Social Science, 1-52.

Session 11: Text Mining in Political Documents	
Required Readings	Gyódi, K., Nawaro, Ł., Paliński, M., & Wilamowski, M. (2023). Informing policy with text mining: technological change and social challenges. Quality & Quantity, 57(1), 933-954.
Lab Session	Extracting and processing text from policy documents (e.g., trade agreements, progress reports) in PDF format

Session 12: EU Trade Agreements, Policy Papers, and Party Manifestos	
Required Readings	Allee, T., Elsig, M., & Lugg, A. (2017). Is the European Union Trade Deal with Canada new or recycled? A text-as-data approach. Global Policy, 8(2), 246-252. **(Presentation)**
	Mi, X. (2023). Gavras, K., Mader, M., & Schoen, H. (2022). Convergence of European security and defense preferences? A quantitative text analysis of strategy papers, 1994–2018. European Union Politics, 23(4), 662-679. **(Presentation)**
	Pauwels, T. (2011). Measuring populism: A quantitative text analysis of party literature in Belgium. Journal of Elections, Public Opinion and Parties, 21(1), 97-119. **(Presentation)**
Recommended Reading(s)	Strategic cultures between the EU member states: convergence or divergence?. European Security, 32(4), 558-582.
	Scarrow, S. E., Wright, J. M., & Gauja, A. (2023). Party statutes and party institutionalization. Party Politics, 29(2), 217-228.

Session 13: Final Presentations	
Structure	10-minute presentation, supplemented by a short- written handout that outlines the research question, theoretical expectations, and the methodology.

Session 14: Wrap-up Session	
Structure	The final session includes a brief review of the previous sessions, a discussion of any remaining questions, and feedback from students on the course structure.

Research Paper Guidelines

Formal

Length	13-15 pages, excluding references and title page
Font-size	Standard (Times New Roman 12)
Line spacing	1.5
Submission	Upload via Moodle: LASTNAME_ModuleNumber.pdf
Citation style	Any common social science citation style is accepted, e.g. Harvard in-text citation – Author (Year: pages). Important is that the used citation style is uniform and consistent throughout the paper. (Hint: Use a reference management program, e.g. EndNote)

Structure

Title page (1 page)	- Title of the paper (as precise and pertinent as possible, indicating the topic and approach, not be the same as the research question) - Name, student number, degree/program, email for contact, date, seminar name and reference, lecturer
Introduction (2-3 pages)	- Clear and precise presentation of the research problem and question, as well as its relevance - Development of a research question (formulated as a direct question that is answerable)
Main part (10-12 pages)	- Concise presentation of the current literature on the topic - Argumentation for the selection of your theoretical approach - Formulation of theoretical propositions, eventually hypotheses - Presentation of the selected methods and scientific approach - Justification for case selection and empirical focus - Empirical research and evidence of one or more cases that can support (or falsify) the theoretical propositions - Evaluation, analysis and interpretation of the empirical results/evidence collected on the basis of the theoretical propositions
Conclusion (2-3 pages)	- Clear answer to the research question - Summary of results and conclusions from the theoretical analysis
References	Complete list of sources used (primary and secondary literature)
Appendix	Signed declaration of independence