# Analyzing EU Politics with Big Data

### Summer Semester 2021
### Online Seminar

**Prof. Sven-Oliver Proksch**
Cologne Center for Comparative Politics (CCCP)
E-mail: so.proksch@uni-koeln.de
Online Office Hours: Tuesday 12:00-13:30 (by online appointment)

## 1. Course Description

Political scientists increasingly rely on big data approaches for discovery and testing theories. This research seminar surveys such methods for systematically extracting information from political text for social scientific purposes and teaches participants how to apply these methods in practical research on European politics. We begin by studying methods to compare documents, the construction and use of dictionaries, and the operationalization of speech tone and complexity. We then turn to methods for placing speeches on continuous dimensions or scales and for measuring the issue attention from texts. The class will be taught in an applied manner using the statistical software R. The lessons will consist of a mixture of theoretical grounding in quantitative content analysis approaches and techniques, with hands-on analysis of real texts using the R package Quanteda.

## 2. Course Objectives

- Learn various state-of-the-art text-as-data approaches

- Learn how to collect textual data from the web with R

- Learn how to use R to analyze textual data and present the results appropriately

- Learn the promises and pitfalls of quantitative text analysis in political science

## 3. Prerequisites

This is an advanced seminar. Students must have completed the MA course in quantitative methods (or equivalent). Basic experience with R is expected. This is not an introductory seminar to R. All methods in this course will be implemented in R, using primarily - but not exclusively - the R package quanteda.

### 4. Online Course Structure

The course is planned to take place online in the following format:

1. **Introductory Session:** The introductory session will take place on April 15 from 14:00 to 15:30 on Zoom. The session will introduce the topics of the course. Course participants will be divided up randomly into several groups. Each group will subsequently receive a group assignment, the results of which will be presented to the class online on May 31.

2. **Video Lectures and R Tutorials:** A video stream with the slides and audio commentary will be made available on ILIAS after the introductory session. Additionally you will receive the set of slides as PDF file. In addition, a walk-through R studio tutorial of the lab session will be available. The videos and R tutorials can be accessed at any time. Together with the required readings, they are necessary to prepare the group project.

3. **Live Session (May 31-June 2):** We will meet live on two and a half days on Zoom. If the university's Corona policy allows, we may be able to meet in person. A decision if the class meets online or in person will be communicated well in advance. On May 31, each group will present the results of the group project with a subsequent Q&A session. On June 1, we will talk about how to use more advanced text models for the study of European politics before starting with presentations by each participant on the individual research designs for the final papers. We will continue with these presentation on June 2.

4. **Online office hours:** You are required to make at least two online office hours appointments with me: one appointment (before May 17) together with your group members to present progress on the group project, and the other one to discuss your final paper proposal (after June 2). I am available for additional online office hours every Tuesday from 12-13.30. Please use the online scheduling form to book an appointment.

### 4. Course Requirements

This module follows the portfolio model of examination, meaning you will complete several assignments. The assignments will take the following form:

- A group project presentation (due May 31)

- A two-page response paper to a research article of your choice (due May 31)

- An individual presentation of your own text analysis project design, including a written 1-page writeup (write up due May 21, presentation on June 1)

- A five-page writeup of final project with visualizations (due July 2)

### 5. Course Material

Readings will be available as electronic articles through Ilias.

## 6. Course Outline

Live Session 1     **Overview**
April 15, 14:00-15:30

No assignment this week other than install R, RStudio, Quanteda, Latex

Readings:

- Brady, Henry. 2019. "The Challenge of Big Data and Data Science". *Annual Review of Political Science* 22: 297–323.
- Jonathan B. Slapin and Sven-Oliver Proksch. 2014. "Words as Data: Content Analysis in Legislative Studies", In S. Martin, T. Saalfeld, and K. Strom (eds), The Oxford Handbook of Legislative Studies. Oxford: Oxford University Press.
- Grimmer Justin and Brandon Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*. 21:267-297.
- Getting started with the Quanteda R package

Video Lecture 1     **Political Speech I: Dictionaries, Sentiment, and Complexity**
Readings:

- Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle, and Stuart N. Soroka. 2019. 'Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches". *Legislative Studies Quarterly* 44 (1), 97-131
- Bischof, Daniel and Roman Senninger. 2018. "Simple Politics for the People? Complexity in Campaign Messages and Political Knowledge". *European Journal of Political Research* 57 (2): 473–495.
- Rauh, C., Bes, B.J. and Schoonvelde, M. 2020. "Undermining, defusing or defending European integration? Assessing public communication of European executives in times of EU politicisation." *European Journal of Political Research.*

Possible articles for response paper (choose one):

- Slapin, Jonathan B. and Justin H. Kirkland. 2019. "The Sound of Rebellion: Voting Dissent and Legislative Speech in the UK House of Commons". *Legislative Studies Quarterly* .
- Benoit, Kenneth, Kevin Munger, and Arthur Spirling. 2019. "Measuring and Explaining Political Sophistication Through Textual Complexity". *American Journal of Political Science* 63 (2): 491–508.
- Schoonvelde, Martijn, Anna Brosius, Gijs Schumacher, and Bert N. Bakker. 2019. "Liberals Lecture, Conservatives Communicate: Analyzing Complexity and Ideology in 381,609 Political Speeches". *PLoS One* 14 (2)
- McDonnell, D., and Ondelli, S. 2020. "The Language of Right-Wing Populist Leaders: Not So Simple." *Perspectives on Politics*, 1-14.

---

Video Lecture 2 **Political Speech II: Ideology, Policy Positions, and Polarization**

Readings:

- Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97(2): 311-331.
- Jonathan B. Slapin and Sven-Oliver Proksch. 2008. "A Scaling Model For Estimating Time-Series Policy Positions from Texts", *American Journal of Political Science* 52(3), 705-722.
- Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring political positions from legislative speech." *Political Analysis* 24(3): 374-394.
- Peterson, Andrew, and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis*, 26(1), 120-128.

Possible articles for response paper (choose one):

- Herzog, Alexander and Kenneth Benoit. 2015. "The Most Unkindest Cuts: Speaker Selection and Expressed Goverment Dissent During Economic Crisis". *The Journal of Politics* 77 (4): 1157–1175.
- Greene, Zachary and Matthias Haber. 2016. "Leadership Competition and Disagreement at Party National Congresses". *British Journal of Political Science* 46 (3): 611–632.
- Klüver, Heike. 2009. "Measuring interest group influence using quantitative text analysis." *European Union Politics* 10.4: 535-549.

Video Lecture 3  **Supervised classification: Populism and Euroscepticism**

Readings:

- Hawkins, Kirk, and Bruno Castanho Silva. 2018. "Textual Analysis: Big data approaches", In: *The Ideational Approach to Populism: Concept, Theory, and Analysis.* Routledge.
- Rooduijn, Matthijs and Teun Pauwels. 2011. "Measuring Populism: Comparing Two Methods of Content Analysis." *West European Politics* 34 (6): 1272–1283.
- Nanni, Federico et al. 2018. "Findings from the Hackathon on Understanding Euroscepticism Through the Lens of Textual Data", *Proceedings of the LREC 2018 Workshop "ParlaCLARIN : LREC2018 workshop on creating and using parliamentary corpora", 7 May 2018*

Possible articles for response paper (choose one):

- Bernhard, Laurent, and Hanspeter Kriesi. 2019. "Populism in election times: a comparative analysis of 11 countries in Western Europe." *West European Politics* 42.6: 1188-1208.
- Castanho Silva, Bruno, and Sven-Oliver Proksch. 2021. "Politicians Unleashed? Political Communication on Twitter and in Parliament in Western Europe." *Political Science Research and Methods*
- Decadri, Silvia, and Constantine Boussalis. 2020. "Populism, party membership, and language complexity in the Italian chamber of deputies." *Journal of Elections, Public Opinion and Parties* 30.4: 484-503.

Live Session 2  **Group Presentations**
May 31, 9:00-17:00
Each group will present its research project in class.

Live Session 3 **Advanced Text Models and Individual Research Proposal Presentation**
June 1, 9:00-17:00

Readings:

- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson LederLuis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses". American Journal of Political Science 58 (4): 1064–1082.
- Cross, James P., and H. Hermansson. 2017. "Legislative amendments and informal politics in the European Union: A text reuse approach." *European Union Politics*, 18(4), 581–602.
- Rodriguez, Pedro and Arthur Spirling. 2020. "Word Embeddings: What works, what doesn't, and how to tell the difference for applied research" Working Paper

Live Session 4 **Continuation: Presentation of Individual Research Proposals**
June 2, 9:00-12:00