



Universität zu Köln

# Analyzing EU Politics with Big Data

Summer Semester 2020

---

Wednesday, 10:00 - 13:30 (IBW, Raum 3.40)

**Prof. Sven-Oliver Proksch**

Cologne Center for Comparative Politics (CCCP)

E-mail: [so.proksch@uni-koeln.de](mailto:so.proksch@uni-koeln.de)

Office Hours: Tuesday 12:30-13:30 (by online appointment)

Office: IBW, room 3.45

---

## 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 politics in the European Union. 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. Due to the nature of the course with lecture and lab session, this seminar will take place in 3-hour (double) sessions. 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*. Students will write several data analysis projects and a larger final paper applying the methods to their own datasets.

## 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. Course Requirements

This module follows the portfolio model of examination, meaning you will complete several weekly assignments. The homework assignments will take the form of applications, replications, or extensions of the analysis in class and the lab. In the last assignment you will present your own text analysis, using a technique presented in the course. The assignments will be handed out at the end of each class and you have (usually) one week to complete them. Assignments must be written in R Markdown and uploaded to the course website prior to the next class meeting (unless noted otherwise). Late assignment will not be accepted. You are not allowed to work on homework assignments together. Details will be communicated at the beginning of the course.

### 5. Course Material

Readings will be available as electronic articles through Ilias.

### 6. Course Outline

---

#### Week 1 **Overview**

April 22

Readings:

- Monroe, B. L., Pan, J., Roberts, M. E., Sen, M., and Sinclair, B. 2015. “No! Formal theory, causal inference, and big data are not contradictory trends in political science.” *PS: Political Science and Politics*, 48(01), 71-74.
  - 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](#)
-

Week 2 **Political Speech I: Dictionaries, Sentiment, and Complexity**

April 29

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.
  - 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)
- 

Week 3 **Political Speech II: Ideology, Policy Positions, and Polarization**

May 6

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.
-

Week 4 **EU Policy-Making and Negotiations**

May 13

Readings:

- 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.
- Proksch, Sven-Oliver, Christopher Wratil, and Jens Wäckerle. 2019. “Testing the Validity of Automatic Speech Recognition for Political Text Analysis.” *Political Analysis*, 27(3), 339-359
- Klüver, Heike. 2009. “Measuring interest group influence using quantitative text analysis.” *European Union Politics* 10.4: 535-549.
- Bunea, Adriana, and Raimondas Ibenskas. 2015. “Quantitative text analysis and the study of EU lobbying and interest groups.” *European Union Politics* 16.3: 429-455.

---

Week 5 **Populism and Euroscepticism**

May 20

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*

Day 6 **Social Media**

June 10

Readings:

- Nulty, P., Theocharis, Y., Popa, S. A., Parnet, O., and Benoit, K. 2016. “Social media and political communication in the 2014 elections to the European Parliament.” *Electoral studies*, 44, 429-444.
- Theocharis, Y. , Barberá, P. , Fazekas, Z. , Popa, S. A. and Parnet, O. 2016. “A Bad Workman Blames His Tweets: The Consequences of Citizens’ Uncivil Twitter Use When Interacting With Party Candidates.” *Journal of Communication*, 66: 1007-1031.
- Silva, Bruno and Sven-Oliver Proksch. Tweeting Europe. Working Paper.

---

Day 7 **Student Presentations**

June 24

---