Quantitative Text Analysis

Winter Semester 2017/2018

Wednesday, 10:00 - 13:00 (Seminarraum 1 0.12, Weyertal, 121)
Please note that sessions 5 and 6 will take place on different days.

Prof. Sven-Oliver Proksch
Cologne Center for Comparative Politics (CCCP)
E-mail: so.proksch@uni-koeln.de
Office Hours: Wednesday 13:30-14:30 (by appointment)
Office: Universitätsstrasse 91

---

1. Course Description

Political scientists increasingly rely on large amounts of political text to extract quantitative information in order to test political science theories. This research seminar surveys quantitative methods for systematically extracting information from text for social scientific purposes and teaches participants how to apply these methods in practical research. We begin by describing the field of quantitative content analysis and by studying how political texts can be collected in an automated fashion. Subsequently, we focus on methods to compare documents, the construction and use of dictionaries, and crowd-coding methods. We then turn to methods for placing documents on continuous dimensions or scales. This overview is completed by a discussion how quantitative text analysis methods can (or cannot) accurately reflect theoretical data-generating processes political scientists have in mind when designing research projects. Finally, we will how topic models can be integrated into political science research and how social media data (such as Twitter data) can be collected and analyzed. Due to the nature of the course, 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 content analytic and statistical software.

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 at least one prior MA/PhD-level course in quantitative methods. 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 WISO faculty's 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 (see below) 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: Quantitative Text Analysis in Political Science**
October 11

Readings:

- 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 **Working with Texts: Automated Data Collection**
October 18

Readings:

- Munzert et al. 2015. *Automated Data Collection with R.* Wiley. Selected chapters.

Week 3 **Comparing Texts, Using Dictionaries, and Relying on the Crowd**
October 25

Readings:

- Laver, Michael, and John Garry. 2000. "Estimating policy positions from political texts." *American Journal of Political Science* 619-634.
- Young, Lori, and Stuart Soroka. 2012. "Affective news: The automated coding of sentiment in political texts." *Political Communication* 29.2: 205-231.
- Spirling, Arthur. 2016. "Democratization and Linguistic Complexity: The Effect of Franchise Extension on Parliamentary Discourse, 1832–1915." *The Journal of Politics* 78.1: 120-136.
- Benoit, Ken, Conway D, Lauderdale B, Laver M, Mikhaylov S. 2016. "Crowd-sourced text analysis: reproducible and agile production of political data." *American Political Science Review* 110(2):278–95

Week 4 **Estimating Positions from Texts: Scaling Methods**
November 8

Readings:

- Laver, Michael, Kenneth Benoit, and John Garry. "Extracting policy positions from political texts using words as data." American Political Science Review 97.2 (2003): 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.
- Lowe, Will, Kenneth R. Benoit, Slava Mikhaylov, and Michael Laver. 2011. "Scaling policy preferences from coded political texts." *Legislative Studies Quarterly* 36(1): 123–155.
- Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring political positions from legislative speech." *Political Analysis* 24.3: 374-394.
- Working with the manifestoR package

Week 5 **The Data-Generating Process: Bringing Political Science Theory Back into the Picture**
November 15

Readings:

- 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.
- Proksch, Sven-Oliver, and Jonathan B. Slapin. 2012. "Institutional foundations of legislative speech." *American Journal of Political Science* 56.3: 520-537
- Herzog, Alexander, and Kenneth Benoit. 2015. "The most unkindest cuts: speaker selection and expressed government dissent during economic crisis." *The Journal of Politics* 77.4: 1157-1175.
- Proksch, Sven-Oliver, Will Lowe, and Stuart Soroka. N.d. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Parliamentary Speeches." Working Paper

Day 6 **Topic Models: Workshop with Molly Roberts (UC, San Diego)**
December 8 (Friday Session)

Readings:

- Roberts, Stewart, Tingley, Lucas, Leder-Luis, Gadarian, Albertson, and Rand. 2014. "Structural topic models for open-ended survey responses." *American Journal of Political Science.*
- Structural Topic Model Vignette
- Other readings TBA

Day 7 **Collecting and Analyzing Social Media Data: Workshop with Pablo Barbera (LSE)**
December 19 (Tuesday session)

Readings:

- TBA

Day 8 **Student Presentations**
December 20