BGTS Workshop

Content Analysis in the Political Sciences

Dr. Christian Rauh
Senior researcher, WZB Berlin Social Science Center
www.christian-rauh.eu
rauh@wzb.eu

Sessions 1-3
Monday, 27 May 2019, 9:30 a.m. – 5 p.m.
Hertie School of Governance, Room 2.30

Sessions 4-6
Tuesday, 28 May 2017, 9:30 a.m. – 5 p.m.
Hertie School of Governance, Room 2.30

A. Workshop outline

Text is the most prevalent and easy-to-access type of information about various political phenomena in time and space. Sources as diverse as speeches, position papers, diaries, media reports, or blogs and tweets bear witness of the priorities, positions, demands, or even values and norms of political actors. However, texts often come in comparatively unstructured form and – depending on the research question - require different levels of interpretation in order to extract the systematic information that political scientists require.

The workshop thus addresses PhD students considering a content analysis. It aims to support them in reading studies using content analyses and should help in choosing and implementing adequate and feasible methods for their own research questions. Based on seminal literature and my experiences in conducting and supervising respective projects we will first identify key *methodological* and *pragmatic* challenges. Then we will focus on the strength and weaknesses of *individual human-coding and (semi-)automated approaches to content analysis.*

While the first two sessions offer general discussions of the methodological fundamentals of content analyses, the later sessions present the intuition, exemplary applications and software tools of specific individual approaches to extract systematic information from texts.

On this basis, we will then commonly discuss how these methods relate to your research projects. Each participant is asked to prepare a 7-10 mins presentation on his/her project roughly structured around the following for points:

(1) What is your research question (and what are possible answers)?
(2) Why may content analysis help you to answer this question?
(3) What are the relevant documents/texts and where can you get them from?
(4) How would you extract the information you need from these texts?

Note that these presentations do not have to be highly polished – the may range from

preliminary results, initial drafts of a research design, to very initial ideas or options you might have with regard to the four structuring questions. This will initially help individual participants by gathering ideas and suggestions for their specific projects. But this way we will also commonly train how to evaluate different methods in the light of concrete research questions.

B. Course organization and literature

(O): obligatory literature,

(P): optional literature (also available for student presentations)

Course materials can be accessed via:
https://www.dropbox.com/sh/27t0r434dwfod0k/AACH9PYT9kraG1WJQT9WTJZ1a

1. Content analysis: History, scope and workflow

May 27; 9:30-10:15 am, Hertie School, Room 2.30

This teaching session provides a short historical overview of applied content analysis to then derive a working definition of this class of data collection methods. On this basis we will also discuss the typical workflow of a CA project.

Literature

(O) Krippendorff, Klaus (2004) Content Analysis: An Introduction to Its Methodology. London: Sage Publications, Inc.: Chapters 1-3

(O) Neuendorf, Kimberly (2001) *The Content Analysis Guidebook*. SAGE Publications, Inc.: Chapters 1-2

(P) Macdonald, Keith (2001) 'Using Documents', in: Gilbert, Nigel (ed.) *Researching Social Life.* 2nd ed: SAGE Publications Ltd.: pp. 194-210

2. Key methodological challenges: Unitizing, sampling, reliability and validity

May 27; 10:30-12:00, Hertie School, Room 2.30

This teaching session introduces the four key major methodological challenges of content analyses and briefly discusses pragmatic approaches to handle them.

Literature

(O) Krippendorff, Klaus (2004) Content Analysis: An Introduction to Its Methodology. London: Sage Publications, Inc.: Chapters 5, 6, 11 & 13

(O) Neuendorf, Kimberly (2001): Chapters 4, 6, 7

3. Quantitative Content Analysis: Human coding

May 27; 13:00-16:30 (with variable breaks in between), Hertie School, Room 2.30

Based on the preceding methodological discussions, this session will zoom in on human coding of qualitative text content. To this end we will briefly review the approaches in three seminal political science projects in this regard to then discuss pragmatic, computer-assisted approaches to coding documents. This will be followed by a discussion of PhD project research designs resorting to human-coded content analysis and/or presentations of exemplary literature.

Literature:

(O) Baumgartner, F., Jones, B. and MacLeod, M. (1998) 'Lessons from the trenches: quality, reliability, and usability in a new data source', *The Political Methodologist* 8(2): 1–10.

(O) Koopmans, R. & Statham, P. (1999) Political Claims Analysis: Integrating Protest Event and Political Discourse Approaches. Mobilization: An International Quarterly, 4, 203-221.

(O) Volkens, Andrea, Judith Bara and Ian Budge (2009) 'Data Quality in Content Analysis. The Case of the Comparative Manifestos Project', *Historical Social Research* 34(1): 234-51.

(P) De Wilde, Pieter (2013) 'Representative claims analysis: theory meets method', Journal of European Public Policy: 278-294.

(P) Kriesi, H., Grande, E., Lachat, R., Dolezal, M., Bornschier, S. & Frey, T. (2006) Globalization and the Transformation of the National Political Space: Six European Countries Compared. European Journal of Political Research, 45, 921-956.

Presentations:

Farag, Mahmoud	Mass elite cleavages in Tunisia (coding Tunisian party manifestos)
Benson, Robert	Effect of anti-establishment parties in Europe on political discourse (political speeches, media discourses)
Rodriguez Olivari, Denisse	Fight against corruption in Latin America along the Lava Jato case (newspapers, magazine, and digital media outlets)
Zhang, Yaning	Strategic toleration of non-compliance by international organisations (WTO/EU policy documents and communications)

4 (Semi-)Automated analyses of large text corpora: Promises and Pitfalls

May 27; 9:30-10:00, Hertie School, Room 2.30

This brief teaching session provides a general introduction to semi-automated approaches meant to retrieve systematic information from larger text corpora. It highlights advantages with regard to efficiency and reliability but also emphasizes how the underlying simplified representations of language may affect the validity of these methods with a view to particular research questions.

Literature:

(O) Grimmer, Justin, and Brandon Stewart (2013) 'Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts', *Political Analysis* 21(3): 267-297.

(P) Gentzkow, Matthew, Brian T. Kelly, Matt Taddy (2019) 'Text as data', *Journal of Economic Literature*: Forthcoming

(P) Lemke, Matthias, and Gregor Wiedemann (2016) *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*. Wiesbaden: Springer VS.

5 Corpus construction, corpus exploration, dictionary-based analysis

May 27; 10:15-13:00, Hertie School, Room 2.30

This session discusses how to find and especially how to construct a corpus (for automated) content analysis. On this basis we will see that rather simple corpus explorations can already provide interesting and relevant insights that may guide further manual or automated analysis. We then move on to discuss dictionary-based approaches of automatically classifying individual texts.

(O) Rauh, Christian (2018) 'Validating a sentiment dictionary for German political language', *Journal of Information Technology & Politics* 15(4): 319-343.

(O) Rauh, Christian, and Pieter De Wilde (2018) 'The Opposition Deficit in EU Accountability: Evidence from over 20 years of plenary debate in four member states', *European Journal of Political Research* 57(1): 194–216.

(P) Young, Lori, and Stuart Soroka (2012) 'Affective News: The Automated Coding of Sentiment in Political Texts', *Political Communication* 29(2): 205-231.

(P) Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle, and Stuart Soroka (2018) 'Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches', *Legislative Studies Quarterly*(Online first).

(P) Rauh, Christian, and Sebastian Bödeker (2013) The international trade regime in the

public sphere, 1986-2012: Evaluating the social legitimacy of global governance with semi-automated text mining approaches. *European Workshop on International Studies (EWIS)*. Tartu, Estonia.

(P) Rauh, Christian, Pieter De Wilde, and Jan Schwalbach (2017) 'The ParlSpeech data set: Annotated full-text vectors of 3.9 million plenary speeches in the key legislative chambers of seven European states' *Harvard Dataverse:* http://dx.doi.org/10.7910/DVN/E4RSP9

(P) Baturo, Alexander, Niheer Dasandi, and Slava J. Mikhaylov (2017) 'Understanding state preferences with text as data: Introducing the UN General Debate corpus', *Research & Politics* 4(2): 2053168017712821.

(P) Merz, Nicolas, Sven Regel, and Jirka Lewandowski (2016) 'The Manifesto Corpus: A new resource for research on political parties and quantitative text analysis', *Research & Politics* 3(2)

Presentations:

Schürmann, Benjamin	Use of populism and crisis framing in political communication (Facebook and Twitter, populism dictionary and manual coding)
Fuß, Julia	Prospects for cooperation and conflict among international institutions (IO treaties, statutes, and cooperation agreements to identify cross-Io references and classification of relationship)
Gelesne, Claire	Links between the welfare state and asylum in the French press (media and political speeches)
de Moraes Azenha, Ana Luisa	Assessing the effects of crowdlaw initiatives (online platforms and reports, press articles, legislation, and interview content)

5 Advanced text analysis methods

May 27; 13:30-16:30 (with variable breaks in between), Hertie School, Room 2.30

In this session we will initially learn about the intuitions of supervised and unsupervised machine learning algorithms that either classify (topic models) or scale texts along latent dimensions (Wordscores, Wordfish, Wordshoal). Before rounding off, we will finally briefly note other text-based measures that tap into concepts of political relevance such as text complexity, text similarity or higher-level approaches to grammatical and semantic parsing.

(O) Klüver, Heike (2009) 'Measuring Interest Group Influence Using Quantitative Text Analysis', *European Union Politics* 10(4): 535-549.

(O) Blei, David (2012) 'Probabilistic topic models', *Commun. ACM* 55(4): 77-84.

(O) Lewis-Kraus, Gideon. 2016. The Great A.I. Awakening. *The New York Times Magazine*, 14.12.2016.

(P) 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.

(P) Slapin, Jonathan, and Sven-Oliver Proksch (2008) 'A Scaling Model for Estimating Time-Series Party Positions from Texts', *American Journal of Political Science* 52(3): 705-722.

(P) Lauderdale, Benjamin E., and Alexander Herzog (2016) 'Measuring Political Positions from Legislative Speech', *Political Analysis* 24(3): 374-394.

(P) Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, 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.

(P) Greene, Derek, and James P. Cross (2017) 'Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach', *Political Analysis* 25(1): 77-94.

(P) Allee, Todd, and Andrew Lugg (2016) 'Who wrote the rules for the Trans-Pacific Partnership?', *Research & Politics* 3(3).

(P) Cross, James P, and Henrik Hermansson (2017) 'Legislative amendments and informal politics in the European Union: A text reuse approach', *European Union Politics* 18(4): 581-602.

(P) 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.

(P) Rauh, Christian, Bart Joachim Bes, and Martijn Schoonvelde (2019) 'Undermining,

defusing, or defending European integration? Assessing public communication of European executives in times of EU politicization', *European Journal of Political Research*: Forthcoming.

Presentations:

Schöfer, Till   EU approaches to rising China (content of text around references to China in European Parliament transcripts)

Mones, Haytham Atef   Algorithmic incentives and public value creation (FB and Twitter communication/impact of public institutions)

Schopmans, Hendrik   Emerging Technologies and the Pre-Emptive Development of Global Norms (Normative vs ontological contestation in UN negotiations/debate)

Raiser, Killian   Effectiveness of the Paris Climate Change agreement (mapping secondary literature)