

BTS Workshop

Content Analysis in the Social Sciences

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

Sessions 1-2

Monday, 8 May 2017, 9 a.m. – 5 p.m.
WZB, Room D 112/13

Sessions 3-5

Tuesday, 9 May 2017, 9 a.m. – 5 p.m.
WZB, Room B 002/03

A. Workshop outline

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

The workshop thus aims to support PhD students considering a content analysis in choosing and implementing adequate and feasible methods for their research questions. Based on seminal literature and my experiences in conducting and supervising respective projects (mainly in the political sciences) we will first identify key *methodological* and *pragmatic* challenges to then discuss the students' research design ideas in this light.

The first two sessions offer general introductions to the fundamentals of content analyses. Depending on student demand, sessions three to five then zoom in to specific qualitative, but especially quantitative (human-coding), and (semi-)automated approaches to data collection. Based on short introductions to these methods, these latter sessions will primarily revolve around the discussion of your corresponding research projects.

Along this line I expect a thorough study of the obligatory *literature* before the workshop, an active *participation* during the workshop and one *presentation* per participant. To adapt the workshop to your needs, I would appreciate a *short summary of your project plans* (including working title, research question, and envisaged usage of content analysis) until April 24. Students not willing to discuss an own project may also present one of the texts marked with a (P) below.

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/pqxr94psquhudk5/AAB886JlcJQTZUIm8nRACPOKa>

1. Content analysis: History, scope and workflow

May 8; 9-10:30 am, WZB Room D 112/13

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

2. Methodological challenges: Unitizing, sampling, reliability and validity

May 8; 10:45-12:30, WZB Room D 112/13

This teaching session introduces the four major methodological challenges of content analyses and 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. Qualitative Content Analysis and expert interviews: Guided interpretation

May 8; 13:30-16:30 (with variable breaks in between), WZB Room D 112/13

This session assesses in how far the methodological cornerstones developed earlier hold in more qualitatively oriented research designs. We will also briefly touch upon expert interviews as one particular type of 'content' to be analyzed. Then four PhD candidates will outline their research projects to discuss applicable CA research designs in the forum.

Literature:

- (O) Mayring, P. (2000) Qualitative Content Analysis. Forum: Qualitative Social Research. (<http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385>)
- (P) Bevir, Mark and Rhodes, R.A.W. (2002) 'Interpretive Theory', in: Marsh, David and Stoker, Gerry (eds.) *Theory and Methods in Political Science*. 2nd Edition, Basingstoke: Palgrave MacMillan: pp. 131-152
- (P) Gamson, W. A. & Modigliani, A. (1989) Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. *American Journal of Sociology*, 95, 1-37.
- (P) Aberbach, Joel D., and Bert A. Rockman (2002) 'Conducting and Coding Elite Interviews', *Political Science & Politics* 35(4): 673-676.
- (P) Goldstein, Kenneth (2002) 'Getting in the Door: Sampling and Completing Elite Interviews', *PS: Political Science and Politics* 35(4): 669-672.
- (P) Leech, Beth L. (2002) 'Asking Questions: Techniques for Semistructured Interviews', *PS: Political Science and Politics* 35(4): 665-668.

Presentations:

- | | |
|------------------------|---|
| Seongcheol Kim | <i>How do how populist discourses operate in the Visegrád countries?</i> |
| Ilyas Saliba: | <i>The influence of diffusing protest on perceptions amongst authoritarian regime elites during critical junctures</i> |
| Ibrahim Khatib: | <i>Attitudes of Palestinians in Israel towards the Arab Spring: Between Identity, Conflict, and Values</i> |
| Sarah Hinz | <i>Normative contestation-Ethiopia's justifications for violations of civil and political human rights in the interplay with transnational actors</i> |

4. Quantitative Content Analysis: Human coding

May 9; 9:30-12:30 (with breaks in between), WZB Room B002/003

Based on the preceding methodological discussions, this session will zoom in on human coding of qualitative content. To this end we will briefly review the three most prominent political science approaches in this regard to then discuss pragmatic, computer-assisted approaches to coding documents. This will be followed by another round of presentation and discussion of PhD project research designs.

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:

- | | |
|---------------------------|---|
| Nora Ratzmann | <i>De Wilde, Pieter (2013) 'Representative claims analysis: theory meets method', Journal of European Public Policy: 278-294.</i> |
| Enora Palaric | <i>Policy transfer as a driver of convergence between public administrations in the European Union?</i> |
| Ann-Sophie Gast | <i>Re-Building a Region – Explaining the Development of Regionalism in Eurasia</i> |
| Franziska Loschert | <i>When Populism meets Economic Interests: member states' preference formation regarding EU asylum and immigration policies</i> |

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

May 9; 13:30-16:30 (with variable breaks in between), WZB Room B002/003

This final session introduces more recent, semi-automated approaches to retrieving systematic information from large text corpora. It highlights advantages with regard to efficiency and reliability but also emphasizes key assumptions that may affect the validity of these methods with a view to particular research questions. Three students will finally present key applications of automated text analysis.

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.
- (O) Blei, David (2012) 'Probabilistic topic models', *Commun. ACM* 55(4): 77-84.
- (P) Rauh, C. (2015) 'Communicating supranational governance? The salience of EU affairs in the German Bundestag, 1991-2013', *European Union Politics* 16(1): 116-138.
- (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) Young, Lori, and Stuart Soroka (2012) 'Affective News: The Automated Coding of Sentiment in Political Texts', *Political Communication* 29(2): 205-231.
- (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) Lemke, Matthias, and Gregor Wiedemann (2016) *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*. Wiesbaden: Springer VS.

Presentations:

- | | |
|---------------------------|--|
| Alexander Busold | Young, Lori, and Stuart Soroka (2012) 'Affective News: The Automated Coding of Sentiment in Political Texts', <i>Political Communication</i> 29(2): 205-231. |
| Jan Jakub Chromiec | Rauh, C. (2015) 'Communicating supranational governance? The salience of EU affairs in the German Bundestag, 1991-2013', <i>European Union Politics</i> 16(1): 116-138. |
| Daniel Brüggman | Laver, Michael, Kenneth Benoit, and John Garry (2003) 'Extracting Policy Positions from Political Texts Using Words as Data', <i>American Political Science Review</i> 97(2): 311-331. |