

# **Quantitative Textanalyse in den Politikwissenschaften: Grundlagen, Methoden und aktuelle Anwendungen**

---

Universität Potsdam,  
Hauptseminar, WS 2014/15  
Dienstags, 12-14 Uhr, Raum 3.06.S24

Dr. Christian Rauh  
Wissenschaftszentrum Berlin für Sozialforschung  
[christian.rauh@wzb.eu](mailto:christian.rauh@wzb.eu)

## **A. Kursbeschreibung und Lernziele**

Politik findet durch und in Texten statt. Reden und Presserklärungen, Debatten und Parteiprogramme, Gesetze und Positionspapiere oder aber Zeitungsberichte und Blogs verraten viel über inhaltliche Prioritäten, Positionen und die Macht politischer Akteure. Solche Texte liefern die relevante Information aber häufig nur in unstrukturierter und interpretationsbedürftiger Form. Das Seminar vermittelt den TeilnehmerInnen deshalb die Methoden, mit denen sie systematische, vergleichende Informationen zu eigenen Fragestellungen aus größeren Textmengen gewinnen können.

Der Kurs gliedert sich dazu in drei Blöcke. In einem ersten Teil besprechen wir die Grundlagen der Inhalts- und insbesondere der Textanalyse anhand einzelner Lehrbuchkapitel. Themen sind der Aufbau eines typischen Forschungsprojekts, das Sampling relevanter Texte, mögliche Kodierungsverfahren und die Qualitätskriterien inhaltsanalytischer Forschungsdesigns. Im zweiten Block geht es dann um klassische politikwissenschaftliche Anwendungen von Textanalysen etwa zur Parteienpositionierung, zur Salienz politischer Themen oder zur Gewinnung von Protestdaten aus der Medienberichterstattung. Darauf aufbauend werden im dritten Block moderne Verfahren zur (halb-) automatisierten Textanalyse vorgestellt, mit denen aktuell etwa der Einfluss von Interessengruppen, der Wandel politischer Interpretationsmuster oder aber die Salienz internationaler Politik in nationalen Parlamentsdebatten untersucht wird.

Die TeilnehmerInnen erörtern so die Stärken und Schwächen einzelner Ansätze an konkreten Beispielen und erhalten zudem praktische Tipps zu relevanten Analyseverfahren, Datenquellen und Werkzeugen.

## B. Kursorganisation und Leistungsnachweise

Der Kurs richtet sich an Master StudentInnen, die erste Kenntnisse in quantitativen politikwissenschaftlichen Methoden und Forschungsdesigns vertiefen wollen und ggf. eigene Textanalyseprojekte anstreben. Teilnahmevoraussetzungen sind die regelmäßige vorbereitende Auseinandersetzung mit der zumeist englischsprachigen *Pflichtliteratur*, die aktive *Mitwirkung* im Seminar sowie mindestens eine *Präsentation* eines wissenschaftlichen Aufsatzes. Für einen Leistungsschein ist zudem bis zum 31.03.2015 eine Hausarbeit verlangt, in der entweder ein textanalytisches Projekt fachlich kritisiert oder aber ein eigenes textanalytisches Forschungsdesign zu einer selbstgewählten Fragestellung – etwa zur Vorbereitung von Masterarbeiten – entwickelt wird. Weitere Details zu den Leistungsnachweisen besprechen wir im Kurs.

Pflichttexte für die einzelnen Sitzungen sind in der nachfolgenden Kursübersicht mit (PL) markiert. Die für die Präsentationen vorgesehenen Texte sind mit (PR) gekennzeichnet; gerne könne Sie aber auch selbst gewählte Literatur oder eigene Projekte vorstellen sofern sie zum Inhalt der jeweiligen Sitzung passen. Freiwillige weiterführende Informationsquellen sind mit (WF) vermerkt. Als einführende Literatur empfehle ich die folgenden Bücher, stelle Ihnen aber die wichtigsten Kapitel mit der übrigen Literatur auch per Moodle zur Verfügung:

Krippendorff, Klaus (2004) *Content Analysis: An Introduction to Its Methodology*. 2<sup>nd</sup> Edition. London: Sage Publications.

Neuendorf, Kimberly (2002) *The Content Analysis Guidebook*. SAGE Publications.

Am Ende der Übersicht finden Sie zudem einige weiterführende Online-Ressourcen zu einzelnen Sitzungen.

## **C. Detaillierte Kursübersicht**

### ***Block I: Grundlagen der quantitativen Inhalts- und Textanalyse***

#### **1. Sitzung (14.10.2014):**

##### Kursorganisation und kurze Geschichte der Inhaltsanalyse

- (PL) Krippendorff, Klaus (2004) Content Analysis: An Introduction to Its Methodology. London: 2<sup>nd</sup> Edition. London: Sage Publications: Chapter 1 "History" (pp. 3-17).
- (PL) Neuendorf, Kimberly (2002) The Content Analysis Guidebook. SAGE Publications: Chapter 2 "Milestones" (pp. 27-46)

#### **2. Sitzung (21.10.2014):**

##### Definitionen und typischer Aufbau von Inhaltsanalysen

- (PL) Krippendorff, Klaus (2004) Content Analysis: An Introduction to Its Methodology. London: 2<sup>nd</sup> Edition. London: Sage Publications: Chapter 2 "Conceptual foundation" (pp. 19-43).
- (PL) Neuendorf, Kimberly (2002) The Content Analysis Guidebook. SAGE Publications: Chapter 1 "Defining content analysis" (pp. 1-26)
- (PL) Neuendorf, Kimberly (2002) The Content Analysis Guidebook. SAGE Publications: Chapter 3 "Beyond description" (pp. 47-52)

#### **3. Sitzung (28.10.2014):**

##### Festlegung und Auswahl relevanter Texte und Analyseeinheiten

- (PL) Neuendorf, Kimberly (2002) The Content Analysis Guidebook. SAGE Publications: Chapter 4 "Message Units" (pp. 71-93)
- (PL) Krippendorff, Klaus (2004) Content Analysis: An Introduction to Its Methodology. London: 2<sup>nd</sup> Edition. London: Sage Publications: Chapter 5 "Unitizing" (pp. 97-110) and Chapter 6 "Sampling" (pp. 111-124).

#### **4. Sitzung (04.11.2014):**

##### Gütekriterien der Textkodierung – Reliabilität und Validität

- (PL) Neuendorf, Kimberly (2002) The Content Analysis Guidebook. SAGE Publications: Chapter 6 "Measurement techniques" (pp. 111-25 & 132-37)
- (PL) Krippendorff, Klaus (2004) Content Analysis: An Introduction to Its Methodology. London: 2<sup>nd</sup> Edition. London: Sage Publications: Chapter 11 "Reliability" (pp. 211-23 & 241-50) and Chapter 13 "Validity" (pp. 313-38).

## ***Block II: Klassische, manuelle Textanalysen in den Politikwissenschaften***

### **5. Sitzung (11.11.2014)**

#### Parteipositionen im Wahlkampf – das Comparative Manifestos Projekt

- (PL) Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge & Michael McDonald (2006) *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990–2003*. Oxford: Oxford University Press: Introduction (pp. xvii-xxiii) and Appendices I-II (pp. 150–94)
- (PL) 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.
- (PR) Pappi, Franz Urban; Susumu Shikano (2004) 'Ideologische Signale in den Wahlprogrammen der deutschen Bundestagsparteien 1980 bis 2002', *Arbeitspapiere – Mannheimer Zentrum für Europäische Sozialforschung*: 76.
- (PR) Green-Pedersen, C. (2007) 'The growing importance of issue competition: The changing nature of party competition in Western Europe', *Political Studies* 55(3): 607–628.
- (PR) Mikhaylov, et al. (2012). 'Coder Reliability and Misclassification in the Human Coding of Party Manifestos'. *Political Analysis* 20(1):78–91.

### **6. Sitzung (18.11.2014)**

#### Die Verteilung politischer Aufmerksamkeit – das Comparative Policy Agendas Projekt

- (PL) 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.
- (PL) John, P. (2006) 'The policy agendas project: a review', *Journal of European Public Policy* 13(7): 975–86.
- (PL) Alexandrova, P., M. Carammia, S. Princen, and A. Timmermans (2014) 'Measuring the European Council Agenda: Introducing a new approach and dataset', *European Union Politics* 15(1): 152–67.
- (PR) Mortensen, P. B., C. Green-Pedersen, G. Breeman, L. Chaqués-Bonafont, W. Jennings, P. John, A. M. Palau, and A. Timmermans (2011) 'Comparing government agendas. Executive Speeches in the Netherlands, United Kingdom, and Denmark', *Comparative Political Studies* 44(8): 973–1000.
- (PR) Alexandrova, P. and A. Timmermans (2013) 'National interest versus the common good: The presidency in european council agenda setting', *European Journal of Political Research* 52(3): 316–38.

## **7. Sitzung (25.11.2014)**

### **Proteste und andere politische Signale – die Erhebung von Ereignisdaten aus Medienberichten**

- (PL) Koopmans, Ruud and Dieter Rucht (2002) 'Protest Event Analysis', In: Bert Klandermans and Suzanne Staggenborg (eds.) *Methods of Social Movement Research*. Minneapolis: University of Minnesota Press: 231-59.
- (PL) Earl, J., A. Martin, J. D. McCarthy, and S. A. Soule (2004) 'The use of newspaper data in the study of collective action', *Annual Review of Sociology* 30(1): 65-80.
- (PR) Uba, Katrin, and Fredrik Uggla (2011) 'Protest Actions against the European Union, 1992-2007', *West European Politics* 34(2): 384-393.
- (PR) Rauh, Christian, and Gerald Schneider (2013) 'There is No such Thing as a Free Open Sky: Financial Markets and the Struggle over European Competences in International Air Transport', *Journal of Common Market Studies* 51(6): 1124-40. (plus Web Appendix A6!)
- (WF) Hutter, Swen (2014) *Protesting Culture and Economics in Western Europe: New Cleavages in Left and Right Politics* Minneapolis: University of Minnesota Press: esp. Appendix A (pp. 147-65).
- (WF) Woolley, John T. (2000) 'Using Media-Based Data in Studies of Politics', *American Journal of Political Science* 44(1): 156-73.

## **8. Sitzung (02.12.2014)**

### **Politische Akteure im öffentlichen Raum – Claims Analyse und verwandte Verfahren**

- (PL) Koopmans, R. & Statham, P. (1999) 'Political Claims Analysis: Integrating Protest Event and Political Discourse Approaches', *Mobilization* 4: 203-221.
- (PL) Marc Helbling and Anke Tresch (2011): 'Measuring Party Positions and Issue Salience from Media Coverage: Discussing and Crossvalidating New Indicators', *Electoral Studies* 30(1): 174-183.
- (PL) Schneider, Steffen (2010) 'Empirische Legitimationsforschung'. In: F. Nullmeier, D. Bigeon, J. Gronau, M. Nonnhoff, H. Schmidtke and S. Schneider (eds) *Prekäre Legitimitäten. Rechtfertigung von Herrschaft in der postnationalen Konstellation*. Frankfurt/New York: Campus Verlag: 45-67.
- (PR) Adam, Silke, and Barbara Pfetsch (2009) 'Europa als Konflikt in nationalen Medien – Zur Politisierung der Positionen in der Integrationsdebatte'. In: *Politik in der Mediendemokratie, Politische Vierteljahresschrift, Sonderheft 42*.
- (PR) Kriesi, Hanspeter, Edgar Grande, Romain Lachat, Martin Dolezal, Simon Bornschier, and Timotheos Frey (2006) 'Globalization and the transformation of the national political space: Six European countries compared', *European Journal of Political Research* 45(6): 921-956.
- (PR) Nonnhoff, Martin, Jennifer Gronau, Frank Nullmeier, and Steffen Schneider (2009) 'Zur Politisierung internationaler Institutionen: Der Fall G8', *Zeitschrift für Internationale Beziehungen* 16(2): 233-63.

- (WF) De Wilde, Pieter (2013) 'Representative claims analysis: theory meets method', *Journal of European Public Policy*: 278-294.
- (WF) Pieter de Wilde, Ruud Koopmans and Michael Zürn (2014) 'The Political Sociology of Cosmopolitanism and Communitarianism: Representative Claims Analysis', Discussion Paper SP IV 2014-102 Wissenschaftszentrum Berlin für Sozialforschung.

### **Block III: (Halb-) Automatisierte Verfahren politikwissenschaftlicher Textanalyse**

#### **9. Sitzung (09.12.2014)**

##### Grundlagen automatisierter Textanalyseverfahren

- (PL) Krippendorff, Klaus (2004) Content Analysis: An Introduction to Its Methodology. London: 2<sup>nd</sup> Edition. London: Sage Publications: Chapter 12 "Computer Aids" (pp. 257-312).
- (PL) 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.
- (WF) Lowe, Will (2003) 'The Statistics of Text: New Methods for Content Analysis', Paper presented at the *Midwest Political Science Association Conference* April 2003. Available at SSRN: <http://ssrn.com/abstract=2210808> .

#### **10. Sitzung (16.12.2014)**

##### Inhaltsanalyse als Klassifikationsproblem – Wörterbuch-basierte Verfahren

- (PL) Neuendorf, Kimberly (2002) The Content Analysis Guidebook. SAGE Publications: Chapter 6 "Measurement techniques" (pp. 125-32)
- (PL) Rauh, Christian (2015) 'Communicating supranational governance? The salience of EU affairs in the German Bundestag, 1991-2013', *European Union Politics* 16(1). (plus Web Appendix pp. 1-2)
- (PR) Young, Lori, and Stuart Soroka (2012) 'Affective News: The Automated Coding of Sentiment in Political Texts', *Political Communication* 29(2): 205-31.
- (PR) M. Laver & J. Garry (2000). 'Estimating Policy Positions from Political Texts', *American Journal of Political Science* 44(3): 19-34.
- (PR) Reinhard, J., J. Biesenbender, and K. Holzinger (2014, May). Do arguments matter? argumentation and negotiation success at the 1997 Amsterdam intergovernmental conference. *European Political Science Review* 6, 283-307.
- (PR) Dodds, Peter Sheridan, and Christopher M Danforth 'Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents', *Journal of Happiness Studies* 11(4)

## **11. Sitzung (06.01.2015)**

### Komplexere Klassifikationsverfahren: Topic Modelle und Machine Learning

- (PL) Blei, David (2012) 'Probabilistic topic models', *Commun. ACM* 55(4): 77-84.
- (PL) Quinn, Kevin, Burt Monroe, Michael Colaresi, Michael Crespin, and Dragomir Radev (2010) 'How to Analyze Political Attention with Minimal Assumptions and Costs', *American Journal of Political Science* 54(1): 209-228.
- (PL) Hopkins, Daniel, and Gary King (2010) 'A Method of Automated Nonparametric Content Analysis for Social Science', *American Journal of Political Science* 54(1): 229-247.
- (PR) Ceron, Andrea, Luigi Curini, Stefano M Iacus, and Giuseppe Porro (2013) 'Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France', *New Media & Society* 16(2): 340-358.
- (PR) Quinn Albaugh, Stuart Soroka, Jeroen Joly, Peter Loewen, Julie Sevenans, and Stefaan Walgrave (2014) 'Comparing and Combining Machine Learning and Dictionary-Based Approaches to Topic Coding', *Unpublished Manuscript*.
- (PR) Grimmer, J. (2010, December). A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis* 18(1): 1-35.
- (WF) Cardie, Claire, and John Wilkerson (2008) 'Text Annotation for Political Science Research', *Journal of Information Technology & Politics* 5(1): 1-6.
- (WF) Scharkow, M. (2013) 'Thematic content analysis using supervised machine learning: An empirical evaluation using German online news', *Quality and Quantity* 47(2): 761-773.

## **12. Sitzung (13.01.2015)**

### Die Positionen von politischen Sprechern oder Autoren: Skalierungsverfahren

- (PL) Laver, M., K. Benoit, and J. Garry (2003). Extracting policy positions from political texts using words as data. *American Political Science Review* 97(2): 311-31.
- (PL) Slapin, J. B. and S.-O. Proksch (2008) 'A scaling model for estimating Time-Series party positions from texts', *American Journal of Political Science* 52(3): 705-22.
- (PR) Klüver, H. (2009) 'Measuring interest group influence using quantitative text analysis', *European Union Politics* 10(4): 535-49.
- (PR) Proksch, Sven-Oliver, and Jonathan Slapin (2009) 'How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany', *German Politics* 18(3): 323-344.
- (PR) Bernauer, Julian, and Thomas Bräuninger (2009) 'Intra-Party Preference Heterogeneity and Faction Membership in the 15th German Bundestag: A Computational Text Analysis of Parliamentary Speeches', *German Politics* 18(3).
- (WF) Lowe, Will (2008) 'Understanding Wordscores', *Political Analysis* 16(4): 356-71.

### **13. Sitzung (20.01.2015)**

#### Automatische Gewinnung von Ereignisdaten – Parsing-basierte Verfahren

- (PL) Schrodtt, Philip A. and Deborah J. Gerner (2012) 'Chapter 2: Fundamentals of Machine Coding', In: Schrodtt, Philip A. and Deborah J. Gerner (eds.) *Analyzing International Event Data: A Handbook of Computer-Based Techniques*. Unpublished book manuscript, available at: <http://eventdata.parusanalytics.com/books.html>
- (PL) Leetaru, Kalev and Philip A. Schrodtt (2013) 'GDELT: Global Data on Events, Location and Tone, 1979-2012', Paper presented at the *International Studies Association meetings*, San Francisco, April 2013.
- (PR) Hammond, Jesse, and Nils B Weidmann (2014) 'Using machine-coded event data for the micro-level study of political violence', *Research & Politics* 1(2):
- (WF) King, G. and W. Lowe (2003, June). An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization* 57 (03), 617-642.

### **14. Sitzung (27.01.2015)**

#### Praktische Aspekte – Datenquellen, Textaufbereitung und Analysesoftware

- (PL) 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', Paper presented at the 1<sup>st</sup> *European Workshops on International Studies (EWIS)* Tartu, Estonia, 5-8 June 2013.
- (PL) Alexa, Melina and Cornelia Zuell (2000) "Text Analysis Software: Commonalities, Differences and Limitations: The Results of a Review." *Quality and Quantity* 34(3):299-321.
- (PL) Alonso, Sonia, Andrea Volkens, and Braulio Gómez (2012) *Content-analyzing political texts*. Madrid: CIS. Kapitel 1 (esp. pp. 24-30)
- (WF) Jackman, Simon (2006) 'Data from the Web Into R', *The Political Methodologist* 14(2): 11-15

### **15. Sitzung (03.02.2015)**

#### Zusammenfassung und Feedback



## **D. Weiterführende Online Ressourcen zu einzelnen Sitzungen (freiwillig)**

### 5. Sitzung: Comparative Manifestos

Datensatzbeschreibungen und Daten zum Download:

<https://manifesto-project.wzb.eu/>

Kurzes Anwendungsbeispiel:

<http://democracy.blog.wzb.eu/2013/09/17/and-they-do-differ-an-analysis-of-the-german-parties-election-programs-2013/>

### 6. Sitzung: Comparative policy agendas

Überblick, Codebooks und weitere Ressourcen zu den Policy Agendas Projekten:

<http://www.comparativeagendas.info/>

<http://www.policyagendas.org/>

### 7. Sitzung: Ereignisdaten

Beschreibung, Publikationen und Datensätze des PRODAT Projekts:

<http://www.wzb.eu/de/forschung/beendete-forschungsprogramme/zivilgesellschaft-und-politische-mobilisierung/projekte/prodat-dokument>

### 8. Sitzung: Claims Analyse

Beschreibung und Kodiervorgang des EUROPUB Projekts: <http://europub.wzb.eu/>

### 10. Sitzung: Wörterbuch-basierte Verfahren

Will Lowes JFREQ und Yoshikoder:

<http://conjugateprior.org/software/jfreq/>

<http://conjugateprior.org/software/yoshikoder/>

Der von Young/Soroka verwendete Lexicoder einschließlich Wörterbücher:

<http://lexicoder.com/>

## 12. Sitzung: Skalierungsverfahren

Weiterführende Informationen und Werkzeuge für WORDSCORES:

[http://www.tcd.ie/Political\\_Science/wordscores/](http://www.tcd.ie/Political_Science/wordscores/)

Weiterführende Informationen und Werkzeuge für WORDFISH

[www.wordfish.org](http://www.wordfish.org)

Ein R package, das beide Verfahren implementiert:

<http://conjugateprior.org/software/austin/>

## 13. Sitzung: Automatisierte Ereignisdaten

Alle Informationen rund um KEDS, CAMEO, TABARI:

<http://eventdata.parusanalytics.com/>

Alle Informationen rund um GDELT:

<http://www.gdeltproject.org/>

Anwendungen und kontroverse Diskussionen rund um GDELT:

Twitter: #gdelt

## 14. Sitzung: Praktische Aspekte

Gute Überblicksseite mit vielfältigen Ressourcen:

<http://www.textanalysis.info/>

Nützliche Software zum Edit von Text-Dateien (inkl. Regular Expressions, Batch Edits):

<http://www.digitalvolcano.co.uk/textcrawler.html>

<http://notepad-plus-plus.org/>

Grundlagen der Character Manipulation in R:

[http://en.wikibooks.org/wiki/R\\_Programming/Text\\_Processing](http://en.wikibooks.org/wiki/R_Programming/Text_Processing)

R packages zur automatischen Sprachverarbeitung:

<http://cran.r-project.org/web/views/NaturalLanguageProcessing.html>