

Methods in Social Research

Michaelmas Term 2014

Week 8 – December 1

Methods for text analysis

Evelyn Ersanilli



When to do text analysis?

Topics

- *What* people(/companies/organisations) are talking or writing about
- *How* they are talking or writing about it
- The *relationship* between *real-world events* and *coverage* of these events in media, parliament, social media
- How any of the above *changed* over a certain *time period*

Role of theory

- Inductive (from data to theory)
- Deductive (from theory to data: hypothesis testing)

When to do text analysis?

Research questions

- How do newspapers in Ukraine report on emigrants?
- How has the framing of human trafficking as a policy problem shifted over time?
- Do left-leaning newspapers report more positively about migration than right-leaning newspapers?
- How strong is the relation between the number of news items on immigration and the volume and composition of immigrant inflow?

When to do text analysis?

Strengths

- Unobtrusive
- Non-reactive ?
- Possible to revisit texts to collect additional information
- Relatively objective way of measuring changes over time (retrospectively)

Weaknesses

- Sampling can be challenging (access to sources, choosing appropriate search terms, too many 'hits' to analyse)
- Social media: difficult to determine author
- When using media reports as source of events: media do not cover everything

Types of text

- Parliamentary debates
- Policy documents
- Speeches
- Textbooks
- Newspapers
- Magazines
- Cartoons
- Advertisements
 - *products*
 - *vacancies*
 - *immigration services*
 - *personal ads*
- Websites
 - *discussion forums,*
 - *company or government sites*
 - *personal profile sites (facebook, myspace)*
- Twitter feeds
- Diaries
- Letters
- Lyrics
- (Interview transcripts)
-

Types of text analysis

- Discourse Analysis (lecture Bastian Vollmer)
- Content Analysis
- Frame Analysis (Snow & Benford, 1988)
- Political Claims Analysis (Koopmans & Statham, 1999)
- ...

Content Analysis

‘an approach to the analysis of documents and texts that seeks to quantify content in terms of predetermined categories and in a systematic and replicable manner’ (Bryman, 2012:290)

Content analysis assumes the meaning of words is more or less constant. This allows aggregating occurrences.

A codebook is drafted before data collection, but can be adapted during data collection.

The codes assigned to a document should be independent of the coder; i.e. they should be *reliable*

Q&A: Validity and reliability

Measurement quality determined by:

- *Reliability*: consistency of measurement
- *Validity*: accurately capture what it is supposed to measure

Example: automated Content Analysis

Boomgaarden & Vliegenthart (2009) How News Content Influences Anti-Immigration Attitudes: Germany, 1993–2005', European Journal of Political Research, 48(4): 516-542

Research question: Does variation in news coverage exert an influence on variation in anti-immigration attitudes?

Role of theory: deductive (test 8 hypotheses)

Example: automated Content Analysis

Analysing news coverage

Text type: newspapers

Target population: German newspapers in 1993-2005

Sample: all articles from three German daily newspapers

Operationalisation of news coverage: number of articles on immigration and immigrant integration

Measurement: computer assisted content analysis with a search string containing 31 (combinations of) words relating to immigration and immigrants

automated Content Analysis

Search string:

(diskrim! OR (hass OR feind! OR abneigung w/5 ausländer OR immigr!
OR zuwander! OR einwander!) OR einbürger! OR ((kurs! OR
unterricht) w/5 (sprach! OR einbürgerung! OR ausland! OR
immigrant! OR asyl OR zuwand!)) OR sprachkur! OR sprachunter! OR
einbürgerungskur! OR sprachunterri! OR Familienzusammenführung
OR Scheinehe OR Zwangsheirat OR Zwangshochzeit OR
Zwangsverlobung OR Scheinhochzeit OR Immigra! OR Zuwanderu! OR
Ausland! OR Einwand! OR Multikult! OR (ghett! w/5 Ausländer OR
immigr! OR zuwander! OR einwander!) OR Muslim! OR Islam! OR
asyl! OR Aufenthaltsgenehm! OR Staatsbürgerschaft! OR (asyl w/5
abweis! OR abgewies!) OR (ehefrau w/5 ausland) OR Flüchtling! OR
Ehrenmord! OR (Hassprediger! OR Haßpred!) OR (terror w/5
fundament!) OR (menschenhandel OR Menschenhändl!))

Example: automated Content Analysis

Reliability: computer programme

Validity: *“for present purposes it presents an appropriate strategy since we use computer-assisted content analysis solely to measure salience and not tone or direction of the news”* (2009: 524-525).

Validity of measurement here depends on accuracy of the search string: does it capture all articles on immigration and immigrant integration

Example: Content Analysis 'by hand'

Michalowski (2011) 'Required to assimilate? The content of citizenship tests in five countries', Citizenship Studies, 15:6-7

Research questions: To what extent are citizenship tests liberal and to what extent do they reflect national differences in perceptions of membership?

Role of theory: Mostly deductive (test hypotheses)

Example: Content Analysis ‘by hand’

Analysing citizenship tests

Text type: citizenship exams

Sample: all questions from citizenship tests in the USA, Austria, UK (sample questions), Germany, and the Netherlands (guidelines)

Example: Content Analysis ‘by hand’

Operationalisation of liberalism: focus on subjects that refer to the ‘overlapping consensus’ (that is to say, basic civic rights and freedoms) and to the political system [...] about facts and the knowledge of ‘what is right’ (2011: 752)

Measurement:

- inductively determined thematic categories
 - 1) politics, history & geography,
 - 2) economy and the provision of public goods and services
 - 3) traditions and public moralwith 14 thematic sub-categories
- ‘What is right’ vs ‘what is good’ (Rawlsian definition of liberalism)

Example: Content Analysis ‘by hand’

Reliability: “The coding scheme and classification rules were refined after a first round of feedback and then given out to others a second time” (p 766, note 4)

Validity: quality of operationalisation of liberalism, process of establishing thematic categories

Example: Content Analysis ‘by hand’

Table 5. Coding examples of ‘what is right’ or ‘what is good’.

	Right	Good
The candidate gets informed about unwritten rules (through observation, asking for advice, asking for support or correction, etc.)		x
Who can ask for a divorce in Germany? (a) only the husband, (b) the wife’s parents, (c) only the wife, >(d) the wife and the husband	x	
The candidate interacts with colleagues and superiors in an adequate way		x
The candidate deals with agreement and disagreement in a socially accepted manner		x
What is a German law? (a) you must not smoke on the street, (b) women must wear skirts, (c) you must not hit children, (d) women must not drink alcohol	x	
The candidate knows that women and girls are expected to live a (financially) independent life.		x

Originates with Erving Goffman

Popularised and elaborated in Social Movement Organisation (SMO) research (Snow et al 1986; Snow & Benford, 1988, Benford & Snow 2000), now also common in analyses of policies and political discourse

Framing:

“assign[ing] meaning to and interpret[ing] relevant events and conditions in ways that are intended to mobilize potential adherents and constituents, to garner bystander support , and to demobilize antagonists” (Snow & Benford, 1988:198)

“contentious in the sense that it involves the generation of interpretive frames that not only differ from existing ones but that may also challenges them” (Benford & Snow, 2000:614)

Frames consist of three elements

1. *Diagnosis*: what is the problem?
2. *Prognosis*: how should the problem be resolved?
3. *Call for action*: what courses of action are suggested?

Frame analysis often looks at

- *Frame diffusion*: how do frames move across movements, cultures or time
- *Resonance*: how effective are frames in mobilising support for action?
- *Changes* to the frames used by certain actors
- *Stance*: who uses which frame

Frames can be determined inductively or deductively

Frame Analysis vs discourse analysis

Discourse Analysis

Mainly developed by linguists (Van Dijk, Fairclough and Wodak)

Looks at underlying (implied) meaning of what is said.

Looks at words, semantics, metaphors, categorisation, stereotyping, rhetorical devices such as the attribution of negative or positive traits to certain groups.

Frame analysis

Mainly developed by social scientists (Goffman, Snow & Benford)

Looks at what is said.

Looks at the line of argument (diagnosis, prognosis, call for action)

Frame Analysis - example

Roggeband, C.M. and Vliegenthart, R. (2007). 'Divergent Framing: The Public Debate on Migration in the Dutch Parliament and Media, 1995-2004'. West European Politics, 30(3): 524-548.

Research question: What is the relation between framing of migration in parliament and the media?

Role of theory: Deductive (test 6 hypotheses)

Frame Analysis - example

Mixed method study

- 1) manual frame analysis of parliamentary documents
- 2) followed by automated content analysis

The coding scheme for frames has been adapted (p 529, see also sensitising questions p545), added:

- Causal, final and normative chains to diagnosis
- Standing of speaker/author

NB “standing” is not part of the frame, but information on who uses which frame

Frame Analysis - example

Multicultural Frame

*This frame, which is often presented as the prototypical Dutch cultural frame, sees (cultural) diversity as an asset that enhances the quality of society. Policy interventions to reduce inequalities in participation in education, labour and politics should respect and foster diversity. Dialogue and civil participation are viewed as a central means to create more understanding and respect towards cultural diversity. In this frame, the **diagnosis** is little elaborated, it is mainly argued that there is insufficient space for diversity and that discrimination and 'glass ceilings' may prevent migrants from entering the higher echelons. Most emphasis is on the **prognosis**, promoting respect for cultural diversity and creating more space for difference.*

Frame Analysis – group exercise

For the Roggeband and Vliegenthart article, discuss in groups of 3:

1. What is their sample?
 - I. Political realm
 - II. Media realm
2. How are frames ‘measured’? Are they determined inductively or deductively?
3. How do the authors establish reliability & validity of their frame measurements?

Frame Analysis – group exercise

1. What is their sample?

I. Political realm:

P 528

Time frame: analysed all documents presented to and discussed in parliament between January 1995 and the end of 2004. In this period, four different coalitions governed

How sampled: through the search engine Parlando, which gives access to all documents presented to, and discussed in, parliament and the Senate”

II. Media realm

528- 529. five most-read Dutch national newspapers throughout the research period [...] All these newspapers are available electronically via LexisNexis for the period we have studied, except for De Telegraaf, which is only available digitally from 1999 onwards. Using framing in later years compared with the framing by the other newspapers in this period and the data available for these other newspapers before 1998, we controlled for the absence of De Telegraaf between 1995 and 1998.

Frame Analysis – group exercise

2. How are frames ‘measured’? Are they determined inductively or deductively?

Step 1. Sensitising questions on parliamentary documents (appendix A)

P 530. *“This instrument [sensitising questions] was used, in a first inductive phase, for a qualitative analysis on the major policy documents in the field of migration and integration (see Appendix 2).”*

Step 2. Turn each frame into a search string and apply to all parliamentary documents & newspaper articles

P 532. *The search-strings were composed on the basis of the coding of the documents included in our qualitative analysis, complemented with a random sample of both parliamentary documents and newspaper articles. For each of these documents and articles, the coder established which frames were present and which word combinations were indicative of these frames. Although a complete frame contains diagnostic, prognostic and motivational components, we did not take that as an absolute requirement for the presence of a frame.*

Frame Analysis – group exercise

3. How do the authors establish reliability & validity of their frame measurements?

They provided a detailed description all the frames they uncovered. This allows readers to judge the face validity (does it make sense).

“To check the reliability of the search-strings, a coder coded a sample of 30 articles manually. Reliability between computer and coder was established by calculating Hostli’s Intercoder Reliability Formula (1969): Overall reliability was .91, which is more than acceptable” (p.532)

It would have been helpful if they had presented the search strings used to count the frames.

Sampling of newspaper articles

Decisions to make:

- Time period
 - Everything in this time-period
 - Subsample: eg Monday, Wednesday, Saturday papers
- Which newspapers
 - National and/or regional
 - Broadsheet and/or tabloid
 - One or more
- Types of articles
 - Only news
 - Editorials
 - Letters to the editors
 - Interviews
- Search terms (search string)

Sampling of parliamentary documents

Decisions to make:

- Time period
- Types of documents
 - Debates
 - Questions
 - Debates in committees
 - Speeches made outside of parliament
- If bicameral:
 - Both houses
 - Only upper house
 - Only lower house
- Search terms (search string)

Take-home messages

- Text analysis methods can be a particularly productive approach to examine change over time
- There are different approaches to text analysis; which one is most appropriate depends on your research question
- Choices in assembling a corpus (set of texts) can influence outcome of the analysis: think carefully about timeframe, sources and search terms
- Validity of automated coding can be checked by manual recoding
- Reliability of manual coding can be checked by having multiple coders

References

- Benford, R. D. & D. A. Snow (2000) [Framing Processes and Social Movements: An Overview and Assessment](#), *Annual Review of Sociology*, 26:611-639
- Boomgaarden, H. G. & R. Vliegenthart (2009) 'How News Content Influences Anti-Immigration Attitudes: Germany, 1993–2005', *European Journal of Political Research*, 48(4): 516-542
- Bryman, A. (2012) *Social Research Methods*, 4th edition. Oxford: Oxford University Press
- Charmaz, K. (2006) *Constructing grounded theory: a practical guide through qualitative analysis*. London: Sage
- Domínguez, M., Pineda, F., & Mateu, A. (2014) ['Life in a nutshell: evolution of a migratory metaphor in Spanish cartoons'](#), *Media, Culture & Society*, (early view)
- Michalowski, I. (2011) ['Required to assimilate? The content of citizenship tests in five countries'](#), *Citizenship Studies*, 15:6-7, 749-768,
- Parsons, Ch. & Vézina, P-L (2014) [Migrant Networks and Trade: The Vietnamese Boat People as a Natural Experiment](#) IMI Working Paper WP-91-2014
- Roggeband, C.M. and Vliegenthart, R. (2007) ['Divergent Framing: The Public Debate on Migration in the Dutch Parliament and Media, 1995-2004'](#), *West European Politics*, 30(3): 524-548.
- Seawright J. & J. Gerring (2008) 'Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options', *Political Research Quarterly*, 61(2):294-308

References

- Snow, D. A. & R. D Benford (1988) Ideology, frame resonance, and participant mobilization, *International Social Movement Research*, 1:197-218
- Snow, D. A., Rochford, E. B., Jr., Worden, S. K., and Benford, R. D. (1986) Frame alignment processes, micromobilization, and movement participation, *American Sociological Review*, 51 (4): 464–481