

MORS 520, Spring 2009:
METHODS AND APPLICATIONS OF STRUCTURED TEXT ANALYSIS

Prof. Klaus Weber
Management & Organizations
Kellogg School of Management
Jacobs 392
847 491 2201
klausweber@northwestern.edu

Class Meets: Wednesdays, 9am-12pm, Jacobs 387

Overview

Much of what people and organizations do, think, feel and communicate is captured in texts—from transcribed interviews, news and corporate reports to academic articles, recorded speeches and online chat. With the electronic availability of many of these texts and recent methodological developments, ‘text analysis’ has become an increasingly important part of social scientists’ methodological toolkit.

This seminar offers a (critical) introduction to more structured approaches to analyzing social, cultural and psychological phenomena through textual content. The class focuses on techniques that use formal and quantitative approaches to interpreting texts. It draws primarily on formal semiotic and content analytic approaches, but emphasizes their application to research questions in social and organizational research. The course does not cover traditional ‘qualitative’ analysis, such as ethnomethodology, open coding and grounded theory, although the techniques surveyed in the course can support these approaches. It only briefly introduces computational and artificial intelligence techniques developed in information sciences and linguistics, and it barely touches on deconstruction and literary approaches. This is not because these alternatives are not valid or valuable, but because they are addressed in other courses offered on campus.

The course is structured as a methods seminar that combines discussions of readings and short lectures with problem sets and hands-on labs. But unlike many statistical methods seminars, we go through the entire research process: from study design and data collection, to analysis and presentation of findings, to publication and reviewing. Hence, some class sessions take the format of a substantive seminar while others have the feel of a methods class. The course is equally useful for those seeking to do a small-scale coding project (e.g., based on interviews, key documents or free format survey responses) and those hoping to analyze and interpret large bodies of text (e.g., organizational reports, media coverage of an issue, or scientific discourse). You may work on your own text corpus throughout the quarter or I can help you construct one for the exercises. By the end of the course, you should be able to evaluate and perform different types of text analysis and have a sense of how to use the method in your own research.

There are no prerequisites for this course beyond basic computer and statistical literacy. Readings will be distributed electronically via blackboard.

Required book:

Krippendorff (2003). Content analysis: An introduction to its methodology. Beverly Hills, CA, Sage.

Assignments

1) Exercises and Class Discussion – 30%

About half the class sessions are run as seminars. Students are expected to read and reflect on all assigned readings before class. I expect everyone attending class to participate in the discussion. To facilitate this, I expect you email me by 6pm on the day before class a memo that includes: a) at least one discussion question related to the readings and b) at least one application or extension of the method discussed. This applies to classes 2-10. Note: you are allowed to miss two memos over the course of the quarter without reason.

You will also complete several small exercises that explore the techniques we read about. These must be turned in within one week of the day we discuss the method (see class schedule below). I encourage you to key these assignments to the substance of your research paper.

2) Paper Review – 10%

You will write a critical review of a paper that uses one of the methods discussed in this course. This can be a published paper or a working paper of your own choice. The review must be written as if you were an ad hoc reviewer for a journal – critical, constructive and realistic. Keep it brief but insightful.

3) Methods Review - 20%

Groups of 2-3 students select a substantive body of literature in which text analysis has been used. For example, you could choose research on negotiations, social movements, corporate governance, science and technology, or social identity. You will produce a 10 page report that surveys the use of text analysis methods, identifies the main approaches, provides a bibliography of some exemplary pieces, critically evaluates the use of text analysis in the context of alternatives and the sophistication of current studies, and identifies directions for methodological or theoretical advances. Each team will present their key findings and facilitate the class discussion in a 45 minute section during one of the last four classes.

4) Research Paper – 40%

You will write a 15-20 page final paper that reports a text analysis based on tools explored over the course of the quarter. The paper can be the methods and results section of an ongoing research project, or a preliminary analysis for a more nascent project. The paper must include all ingredients of a research paper: motivation, theory, research design, methods, results and implications. The non-methodological parts can be brief, though. There is considerable flexibility around this assignment and I encourage you to talk to me about it early on. The idea is to help you develop your research, not to add burdens.

List of Software Options

<http://academic.csuohio.edu/kneuendorf/content/cpuca/ccap.htm>

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

Other Basic Textbooks

Neuendorf, Kimberly A. 2001. *The Content Analysis Guidebook*. Beverly Hills: Sage.

Franzosi, Roberto. 2004. *From Words to Numbers: Narrative, Data, and Social Science*. Cambridge: Cambridge University Press.

Manning, Christopher, and Hinrich Schutze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.

Week 1: Frameworks

Topics

Frameworks of meaning in semiotics and communications; syntactic, semantic, pragmatic meaning
Key concepts: Signs – texts – genres – discourses; text – speech – language; signification ‘systems’
History of text analysis: Example of organizational research

Readings

Chandler, Daniel. n.d. *Semiotics for beginners*. Chapters 1,2 and 12 (Introduction, Signs, Encoding-Decoding).
Online Book: <http://www.aber.ac.uk/media/Documents/S4B/semiotic.html>.
Peirce, Charles Sanders. 1955[1898]. "Logic as semiotic: the theory of signs." Pp. 98-119 in *Philosophical Writings of Peirce*, edited by Justus Buchler. New York: Dover.
Krippendorff. 2004. Chapter 2 (Conceptual Foundations)
Duriau, Reger & Pfarrer. 2007. "A content analysis of the content analysis literature in organization studies."
Organizational Research Methods. 10: 5-34.

Comments

The first class introduces basic terminology and foundations of analyzing textual data. We'll also cover much of the logistics for the quarter. This class is run more like a regular seminar: Some background from the instructor and discussion of assigned readings. Chandler's online book is an excellent introductory resource while Peirce's seminal article is on the heavy side. Krippendorff provides a useful framework of the important issues from a practical perspective and Duriau gives some sense of the state of affairs in organizational behavior.

Assignments

None.

Week 2: Structured "Qualitative" Analysis

Topics:

Semiotic tools: commutation tests, semiotic chains and squares, paradigmatic analysis

Readings:

Mick, David G. 1986. "Consumer research and semiotics: Exploring the morphology of signs, symbols and significance." *Journal of Consumer Research* 13(2): 196-213.
Feldman, Martha S. 1995. Strategies for interpreting qualitative data. Thousand Oaks, CA: Sage. Chap.2
Barley, Stephen R. 1983. "Semiotics and the study of occupational and organizational culture." *Administrative Science Quarterly* 28:393-413.
Weber, Klaus, Kate Heinze, Michaela DeSoucey. 2008. "Forage for thought: Mobilizing codes in the movement for grass-fed meat and dairy products. *Administrative Science Quarterly* 53: 529-567

Comments:

This class covers primarily ideas and techniques from structural semiotics. Mick is the sweeping review, Feldman a methodological cookbook. We'll discuss the papers by Barley and Weber et al as examples of these techniques.

Assignments

Description of text corpus for final research paper (1pg max)

Week 3: Quantitative Content Analysis

Topics

Overview, terminology, varieties of content analysis.

Readings

Roberts, Carl W. 2000. "A conceptual framework for quantitative text analysis." *Quality & Quantity* 34:259-274.
Carley, Kathleen. 1993. "Coding choices for textual analysis: a comparison of content analysis and map analysis." *Sociological Methodology* 23:75-126.
Krippendorff. 2004. Chapter 4, 12 and 14 (logics of content analysis, computer aids, a practical guide)
Dohan, D. and M. Sanchez-Jankowski (1998). "Using computers to analyze ethnographic field data: Theoretical and practical considerations." *Annual Review of Sociology* 24: 477-498.
AutoMap1.2 Software <http://www.casos.cs.cmu.edu/projects/automap/>

Comments

This class begins a block of four sessions that run through quantitative content analysis step-by-step. The first half is a conceptual discussion of core issues and different varieties, the second includes a more hands-on introduction, via a demo of the Automap, a free software suite developed by Kathleen Carley. If you will use Automap for your research paper and exercises, please download and install the software prior to class. You may use any other software package, though, and I will provide an overview of alternative software options. All in-class demos illustrate general concepts and procedures and are not software specific.

Assignments

Description of motivation and research question for research paper, plus constructs to be measured through text analysis (2pg max).

Week 4: Process of Content Analysis I

Topics

Research design, data sources and sampling strategies, text capture, pre-processing text documents.
Units of text and unit sampling; development of standard and custom dictionaries.

Readings

Murmann et al. 2007. "Automatic coding of printed materials". *International Journal of Humanities and Arts Computing* 1: 151-187
Krippendorff. 2004. chapters 5 & 6 (unitizing, sampling)
Stone, Philip J. et al. 1966. The general inquirer; a computer approach to content analysis. Cambridge, MA: MIT Press: 67-71, 85-92, 113-168.
General Inquirer: <http://www.wjh.harvard.edu/~inquirer/>; Kansas Event Data System (KEDS): <http://web.ku.edu/keds/>, Linguistic Inquiry and Word Count: <http://www.liwc.net/>, DICTION: <http://www.dictionsoftware.com/> (focus on built-in categories)

Comments

Like last week, this class is more hands-on. It covers the construction of the two main ingredients of content analysis: creating a clean corpus of "raw text" to be analyzed and defining what textual elements are of theoretical interest, i.e., coding dictionaries. The first part of the class addresses mundane steps and decisions in structuring a text corpus that can have serious implications. The second weighs the pros and cons of using custom or standard dictionaries, and the procedures involved. Take a look at the General Inquirer or LIWC or Diction (widely used standard dictionaries) or KEDS (a so-called parser) and assess their usefulness.

Assignments

Description and example of capture and conversion process (source format to digital text files (3pg max))

Week 5: Process of Content Analysis II

Topics

The coding process, developing scales and variables from text coding, assessing reliability and validity

Readings

Krippendorff. 2004. Chapters 7 & 11 (coding, reliability)

Krippendorff. 2007. Testing the Reliability of Content Analysis Data: What is Involved and Why (in Krippendorff, K. & Bock, M, A. The Content Analysis Reader)

Andrew F. Hayes, and Klaus Krippendorff (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures* 1: 77-89.

Comments

The class addresses issues involved in moving from a text corpus to variables that measure constructs or categories of interest. The concepts of reliability and validity of measurement are relevant just as they would be in survey research, though some issues are different. We discuss approaches from simple key word counts to complex scales and spend some time on inter-rater reliability (where the raters can be computers or human coders). Read the book chapters and skim both articles - there is some redundancy between all three but the short pieces are less technical than the book.

Assignments

Description of pre-processing steps, sampling and unitizing on a small sample of text, plus draft coding dictionary. (5pg max)

Week 6: Process of Content Analysis III

Topics

Data reduction techniques (e.g., scaling, clustering, latent semantics)

Visual and numerical presentation, interpretation.

Practical issues in publishing

Readings

Krippendorff. 2004. Chapter 10 (analytical/representational techniques)

Mohr, John W. 1998. "Measuring meaning structures." *Annual Review of Sociology* 24:345-370.

Wade, James B, Joseph F Porac, and Timothy G Pollock. 1997. "Worth, words, and the justification of executive pay." *Journal of Organizational Behavior* 18:641-664.

Pennebaker, James W, and Laura A King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*. 77(6): 1296-1312

Comments

This class moves back to a more seminar-style format. The first two readings provide an overview of different approaches to simplify and present information extracted from texts in articles and book. As these techniques are not specific to text analysis, there will be no demos or exercises. I will select two empirical papers for discussion based on student interests.

Assignments

Description and interpretation of coding and scale reliability analysis for data sample. (3pg max)

Week 7: Types of Content Analysis I - Single Elements

Topics

Words, themes, constructs, categories and variables; Examples of linguistic elements: Speech acts, rhetorical topes

Readings

Chandler, Daniel. n.d. Semiotics for beginners. Online Book, University of Aberdeen, UK:

<http://www.aber.ac.uk/media/Documents/S4B/semiotic.html>. Chap.8 (Rhetorical Tropes)

Gibson, Cristina B, and Mary Zellmer-Bruhn. 2001. "An intercultural analysis of the concept of teamwork." *Administrative Science Quarterly*. 46(2): 274-303.

Austin, J L. 1962. *How to do things with words*. Boston, MA: Harvard University Press. (excerpts)

Twitchell, Douglas P, and Jay F Nunamaker. 2004. "Speech act profiling: A probabilistic method for analyzing persistent conversations and their participants" *Proceedings of the 37th Intl. Conference on System Sciences*.

Fiss, Peer C, and Paul Hirsch. 2005. "The discourse of globalization: framing and sensemaking of an emerging concept." *American Sociological Review* 70:25-52.

Fiona Lee, Christopher Peterson and Larissa Z. Tiedens. 2004. Mea Culpa: Predicting Stock Prices From Organizational Attributions. *Personality & Social Psychology Bulletin*. 30(12):1636-1649

Comments

Most current research uses text analysis to selectively measure one or few constructs, themes, or categories that are derived from social science theories. In this session, we discuss and evaluate this approach with the example of a few specific studies. Ask yourself: What is the quality of the text analysis performed? How is it reported in the paper? What additions would you propose and why? What alternative approaches could the authors have considered? In addition to content elements derived from social science theories, we also examine content coding based on more linguistic aspects of the text: rhetorical tropes, speech acts, and other linguistic elements.

Assignments

Paper review (2pg max)

Week 8: Types of Content Analysis II – Sets of Elements

Topics

Vocabularies, repertoires, discourses; Field overview presentations

Readings

Weber, K. 2005. "A toolkit for analyzing corporate cultural toolkits." *Poetics*. 33:227-252

Abrahamson, Eric and Donald C Hambrick. 1997. "Attentional homogeneity in industries: the effect of discretion." *Journal of Organizational Behavior* 18:513-532.

Jones, Candace, and Ruth Livne-Tarandach. 2008. "Designing a frame: rhetorical strategies of architects". *Journal of Organizational Behavior*. 29: 1075 – 1099.

Farkas, George, and Kurt Beron. 2004. "The detailed age trajectory of oral vocabulary knowledge: differences by class and race". *Social Science Research*. 33(3): 464-497.

Comments

An alternative to picking out particular aspects of a text corpus is to look at aggregate dimensions that describe its totality, such as how many words or concepts it contains, or how those sets are structured and change. Evaluate the application of ideas around sets of text elements in then example papers. What challenges and opportunities do you see? What theories could they help illuminate? The latter part of the class showcases the first presentations of "field overviews" from student projects.

Assignments

Methods review presentations.

Week 9: Types of Content Analysis III – Relations and Structures

Topics

Concept and word associations, maps, networks, narratives; Field overview presentations.

Readings

- Carley, K Network text analysis: the network position of concepts (in Roberts)
- Hill & Carley. 1999. "An approach to identifying consensus in a subfield: The case of organizational culture." *Poetics* 27:1-30
- Corman, Stephen R., Timothy Kuhn, Robert D. McPhee, Kevin J. Dooley. 2002. Studying Complex Discursive Systems: Centering Resonance Analysis of Communication. *Human Communication Research* 28(2): 157-206.
- Lee, Peggy M, and Erika Hayes James (2007). "She'-e-os: Gender effects and investor reactions to the announcements of top executive appointments." *Strategic Management Journal*. 28: 227-241.
- Nadkarni, S. and V. K. Narayanan (2007). "Evolution of collective strategy frames in high and low velocity industries." *Organization Science* 18(4): 688-710.
- Franzosi, Roberto. 1994. "From words to numbers: a set theory framework for the collection, organization, and analysis of narrative data." *Sociological Methodology* 24:105-136.

Comments

Instead of treating each element in a text as simply a unit, relational approaches emphasize that associations and syntactic connections between elements are informative and allow access to additional insight embedded in texts. The approaches of Carley and Corman et al are varieties of applying network analysis to texts. Each comes with an empirical illustration. Read either Nadkarni & Narayanan (cognitive mapping) or Franzosi (narrative).

Assignments

Methods review presentations.

Week 10: Types of Content Analysis IV – Relation of Text to Context

Topics

Social distribution of texts and language, correspondences, lattices;

Readings

- Breiger, Ronald L. 2000. "A tool kit for practice theory." *Poetics* 27:91-115.
- Breiger & Mohr. 2004. "Institutional Logics from the Aggregation of Organizational Networks: Operational Procedures for the Analysis of Counted Data." *Computational & Mathematical Organization Theory*, 10:17-43.
- Kennedy, Michael T. 2008. "Getting counted: Markets, media, and reality." *American Sociological Review*. 73: 270-295.
- Chappell, H. W. J., et al. (1997). "Monetary policy preferences of individual FOMC members: A content analysis of the memoranda of discussion." *The Review of Economics and Statistics* 79(3): 454-460.
- Weber, Klaus. 2009. Cultural repertoires and organizational fields. *Working Paper*

Comments

The last variety of text analysis directly examines how texts as cultural/communicative products relate to social context, such as relationships, identities and structures.

Assignments

Methods review presentations.

!! The research paper is due one week after the last class !!!