Department of Sociology
THE UNIVERSITY OF CHICAGO

SOCIOLOGY 40133

Content Analysis

Friday 9:30 – 12:20pm
Spring 2015-2016
Classroom: Social Sciences Research 401
http://chalk.uchicago.edu/

James A. Evans
Office: Bldg. Social Science 420
Tel.: 834-3612; jevans@uchicago.edu
Office Hours: Tuesday 9:00-11:00am

TA: Chris Graziul; graziul@uchicago.edu
Office Hours: Tuesday 1:00-3:00pm
Basement of Stuart Hall

I. The Course

A vast expanse of information about what people do, know, think, and feel lies embedded in text, audio, and images—from transcribed interviews, snapshots, YouTube and the academic cannon to constitutions, poetry, and idling online chat. With the rise of literacy, and more recently of computers, cellphones, scanners and the internet, there are both greater supply and demand for textual and image information, providing us with access to a greater variety of texts, images, video and sound that reach further into the human condition, past and present. This class will begin by considering the nature of text and images and the possibilities for (and limits to) analyzing society and culture through their content. We will spend each session cultivating skills for close, interpretive content analysis (so-called "qualitative content analysis"/QCA) or counting and verifying discovered meaning across text and images (classic "content analysis"), while simultaneously exploring computational methods that could be used to facilitate discovery and assessment of meaning on larger scales. These include introductions to approaches from natural language processing, information retrieval, signal processing, network analysis, and machine vision.

The course is designed to assist those seeking to enhance or expand a small-scale coding project (think interviews, field notes, snapshots, footage), those hoping to efficiently summarize large samples of text and images, or those in search of tools to sift through quantities of documents to find ones meriting closer analysis. The course balances an empirical paper with experimentation using techniques. Students will be evaluated based on their participation, a presentation/report on a chosen method and a final paper.

II. Readings and Computational Tools

Readings will be circulated in class and through the course's chalk site.

Computational tools required to fulfill the exploration assignments (below), may be accessed/developed in a number of ways. Each student will be given access to the Research Computing Center's machines, where software relevant to several of the assignments will be available and data can be stored. These will be accessible via student's CNET ids. Students may also or instead choose to download relevant software onto their individual laptops or other machines for personal use and experimentation. Alternatively, students may choose to develop software consistent with course readings and discussions. Typically students take some combination of these approaches.

III. Course Requirements

A. READING and DISCUSSION (10%)

Content analysis is a seminar. Students are expected to read and reflect on the assigned readings before class. Participation in class discussion is expected of all class members. I ask that students read the assigned reading and at least one additional piece from the optional list (or beyond it) that relates to their exercise that week (see below).

B. EXERCISES (30%)

Students are also expected to complete six short memos intended to invite experimentation with the techniques we consider. At least three of these will involve qualitative coding and interpretation of content, and at least three others experimentation with computational approaches to coding and analysis. Grading will follow the simplified scheme: 3 (~A), 2 (~B+), 1 (~B-), 0 (~F). I will drop the two lowest memo grades and so students will be required to produce between 5 and 8 memos over the course of the quarter for full credit.

For the three qualitative memos, students are to turn in the designated memo and associated materials *at the beginning of class on the day we discuss them*. For these memos, students will engage in the practicum, as described on the syllabus associated with each course, then produce a brief (double-spaced 3-5 page) report of the results and evaluation, along with the detailed materials described. These memos should (1) sketch the process of performing their assignment, (2) present results from a preliminary analysis/interpretation of resulting data using content associated with their anticipated final project topic, (3) qualitative or quantitative assessment of its validity and reliability, (4) aspects that could benefit from formalizing and computationally scaling up along with algorithmic strategies to do so.

For the three computational memos, students are to turn in the designated memo and associated materials *at the beginning of class the week after we discuss them* (or related methods). Students can either use existing programs or write their own and turn in a brief (double-spaced 3-5 page) report of their results and evaluation, along with detailed materials (e.g., computer code; descriptive output, graphical representation, etc.) These memos should (1) sketch the mechanics of using their method, (2) present results from a preliminary ("toy") analysis with it using content associated with their anticipated final project topic, (3) evaluate its drawbacks and (4) identify scope conditions for its beneficial deployment following qualitative demonstration of its validity. Students are encouraged to work on these in groups (although you will turn them in individually) and to key assignments to the substance of the final project.

Students will be prepared to present and discuss detail from these memos in class.

C. WIKIPEDIA ASSIGNMENT (15%)

In groups, students will pick an area of computational content analysis and either: 1) adds and populates a section on a computational method that reviews, explains and evaluates how that method has been used for content analysis; 2) updates the Content Analysis page to transcend the text books from which its drawn to consider newer methods and analysis; 3) creates one or more lists of relevant topics, issues, papers, methods and/or code, software libraries or programs Content Analysts can draw upon.

D. FINAL PROJECT (45%)

Students will perform a substantial content analysis for social insight based on approaches and tools developed/explored over the course of the quarter. The project can be primarily qualitative or quantitative (or equally both). If qualitative, however, it must incorporate at least one quantitative technique developed during the quarter. If quantitative, it must validate its inferences with qualitative interpretation and assessments. These projects may be performed as a group (of no more than 3 students), but the work should be proportional to the effort demonstrated in the project. The motivation, process and findings from this project will be presented on a Poster at

a class reception on Wednesday, June 8 from 4:30pm. Students will also document this project in a detailed 20 page research appendix that contains methodological justification and description, descriptive data and analysis, interpretation of findings and conclusion.

IV. Calendar of Reading Assignments. Each topic is followed by an approach to scaling it up in brackets.

Week 1.

Apr.1: *General Introduction;*
Semantics, semiotics and the measurement of meaning

Evans, James and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory". *Annual Review of Sociology* 42 (in press).

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Malden, MA: Blackwell, remarks/aphorisms 1, 5, 11, 23, 25, 65, 67, 71, 199, 240.

Danesi, Marcel and Paul Perron. 1999. "The Field of Cultural Semiotics" and "The Signifying Order" in *Analyzing Cultures*. Bloomington: University of Indiana: 39-101

Franzosi, Roberto. 2004. "A Worlde of Wordes" from *From Words to Numbers: Narrative, Data and Social Science*. Cambridge University Press: 193-236.

Geertz, Clifford. 1973. Beginning of "Religion as a Cultural System" in *The Interpretation of Cultures*. Basic Books: 87-95.
http://monoskop.org/images/5/54/Geertz_Clifford_The_Interpretation_of_Cultures_Selected_Essays.pdf

Optional:

Suchman, Mark. 2003. "The Contract as Social Artifact." *Law & Society Review* 37(1):91-142.

Geertz, Clifford. 1973. "Religion as a Cultural System" in *The Interpretation of Cultures*. Basic Books: 87-126.
http://monoskop.org/images/5/54/Geertz_Clifford_The_Interpretation_of_Cultures_Selected_Essays.pdf

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Malden, MA: Blackwell, remarks 1-32, 65-71, 197-205, 240-282, 340-345 (pp. 1-14, 27-29, 68-70, 75-82, 93-94).

Week 2.

Apr.8: *Accounting for Words, Phrases, Concepts and Symbols [Corpus Linguistics]*

Timmermans, Stefan and Iddo Tavory. 2012. "Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis." *Sociological Theory* 30(3) 167–186.

Barley, Stephen R. 1983. "Semiotics and the Study of Occupational and Organizational Cultures" *Administrative Science Quarterly* 28:393-413.

Michel, Jean-Baptiste et al. 2010. "Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* express, December 16.

Gentzkow, Matthew & Jesse M. Shapiro. 2007. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78(1): 35–71.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, B. Pang, and John Kleinberg. 2012. Echoes of Power: Language Effects and Power Differences in Social Interaction. Proc. 21st Int. Conf. World Wide Web, pp. 699–708. New York, NY, USA: ACM.

Manning and Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press: Chapter 3 (linguistic foundations), 5 (collocations), 7 (word sense disambiguation); 8.5 (semantic similarity).

Optional:

Glaser, Barney G. 1965. The Constant Comparative Method of Qualitative Analysis. *Social Problems* 12(4): 436-445.

Spradley, James P. 1979. *The Ethnographic Interview*. Steps 5-6. New York: Holt, Rinehart and Winston: 92-119.

Strauss, Anselm L. and Juliet Corbin. 1990. *Basics of Qualitative Research*. chapters 5-6. Newbury park, CA: Sage: 61-95.

Lofland, John and Lyn H. Lofland. 1995. Chapter 6. *Analyzing Social Settings*. Wadsworth Publishing Company: 99-122.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2013 "Measuring Economic Policy Uncertainty." *Chicago Booth Paper* #13-02.

Danescu-Niculescu-Mizil, Cristian, West R, Daniel Jurafsky, Jure Leskovec, and Charles Potts. 2013. "No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities." Proc. 22nd Int. Conf. World Wide Web (WWW), pp. 307–18. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang, Martin Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449:713-716.

Hearst, Marti A. 1997. Untangling Text Data Mining. Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics.
http://www.ischool.berkeley.edu/~hearst/papers/acl99/acl99-tdm.html

Bollen, Johan, Huina Mao, Xiao-Jun Zeng. 2010. Twitter mood predicts the stock market. arXiv:1010.3003v1.

Leavitt, Stephen and Dubner. *Freakonomics* excerpt from chapter 2: 62-69.

Stone, Philip J. et al. 1966. *The general inquirer; a computer approach to content analysis*. Cambridge, MA: MIT Press: Chapters "General Inquirer" and "Construction of Categories", 67-71, 85-92, 113-168.

Mohr, John. 1998. "The Classificatory Logics of State Welfare Systems: Towards a Formal Analysis." in *Public Rights, Public Rules: Constituting Citizens in the World Polity and National Policy*, edited by Connie McNeely. New York: Garland.

Stone, P.J. 1997. "Thematic text analysis: new agendas for analyzing text content." In *Text Analysis for the Social Sciences*, Carl Roberts, ed., Lawrence Erlbaum Associates.

Bird, Stephen, Ewan Klein & Edward Loper. 2008. "Regular Expressions." Appendix from *Natural Language Processing in Python.*

Evans, James. 2010. "Industry Induces Academic Theory to Know Less About More." *American Journal of Sociology* 116(2): 389-452.

Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. "Dictionaries and Tolerant Retrieval." Chapter 3 from *Introduction to Information Retrieval.*

Hirschberg J, Manning CD. 2015. Advances in natural language processing. *Science.* 349(6245):261–66.

Qualitative practicum: Engage in an initial "open coding" of a document, documents or document selection (~15 pages) that links specifically to your own anticipated research project. This will involve your highlighting and labeling *things* in the document - lexical, semantic, syntactic, pragmatic, etc. - that illuminate the referenced content, the author(s), their context, mental models, the Zeitgeist, an (imagined) audience, etc. You can code with pencil/pen on paper, in the margins of (or embedding xml tags in) an electronic document, or in a text analysis program (e.g. Atlas/ti; N*Vivo; Nud*IST; Weft—freeware). Also, draw a simple schematic of your codes and their relationship to one another (e.g., taxonomic tree, network, Venn Diagram). Turn in text, linked codes and figure along with your memo.

Computational possibilities: Word frequencies; Conditional frequencies; Statistically significant collocations; Distinguishing or Important words and phrases (Wordls!); POS-tagged words and phrases; Lemmatized words and phrases; Dictionary-based annotations.

Week 3.

Apr.15: *Discovering higher-level Patterns [Dimension Reduction – Clusters and Topics]*

Davis, Fred. 1959. The cabdriver and his fare: Facets of a fleeting relationship. *American Journal Sociology* 65:158-165.

Evans, James, Gideon Kunda, Stephen Barley. 2004 "Beach Time, Bridge Time and Billable Hours: The Temporal Structure of Temporal Contracting" *Administrative Science Quarterly* 49:1-38.

Manning and Schütze. 1999. Chapter 15 ("Topics in Information Retrieval"); Chapter 13 ("Clustering").

Grimmer, Justin and Gary King. 2011. "General purpose computer-assisted clustering and conceptualization." PNAS (Feb. 3).

Blei, David. 2012. "Probabilistic Topic Models". Communications of the ACM 55(4):77-84.

Nelson, Laura K. 2015. "Political Logics as Cultural Memory: Cognitive Structures, Local Continuities, and Women's Organizations in Chicago and New York City." Working Paper.

Taddy, Matt. 2013. "Multinomial Inverse Regression for Text Analysis". *J. Am. Stat. Assoc.* 108(503):755–70.

Optional:

Mohr, John and Petko Bogdanov. 2013. "Introduction—Topic models: What they are and why they matter." Topic Models and the Cultural Sciences. 2013. ***Poetics*** 41(6): 545-569

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding." *Poetics* 41(6):570-606.

Spradley. *The Ethnographic Interview*. Steps 8, 10: 132-154, 173-184.

Strauss and Corbin. *Basics of Qualitative Research*. Chapter 7: 96-115.

Lofland and Lofland. *Analyzing Social Settings*. Portion of Chapter 9: 193-203.

Miles, Matthew B. and A. Michael Huberman. *Qualitative Data Analysis: An Expanded Sourcebook*, second ed. Thousand Oaks: Sage: 245-262; Diagramming chapters throughout (esp. 262-287).

Lofland and Lofland. *Analyzing Social Settings*. Chapter 7: 123-148.

Strauss and Corbin. *Basics of Qualitative Research*. Chapter 12: 197-223.Mohr, John W. and Petko Bogdanov. 2013. "Introduction—Topic models: What they are and why they matter." *Poetics* 41(6):545-569.

McFarland, Daniel A., Daniel Ramage, Jason Chuang, Jeffrey Heer, Christopher D. Manning, and Daniel Jurafsky. 2013. "Differentiating language usage through topic models." *Poetics* 41(6):607-625.

Topic Models and the Cultural Sciences. 2013. ***Poetics*** 41(6): 545-770.

Blei, David M. and John D. Lafferty. 2007. "A Correlated Topic Model of *Science*." *The Annals of Applied Statistics* 1(1):17-35.

Suchman, Mark. 2006. "The Contracting Universe: Law Firms and the Evolution of Venture Capital Financing in Silicon Valley." Working paper.

Blei, David M., Andrew Y. Ng and Michael I. Jordan. 2003. "Latent Dirichlet allocation." *Journal of Machine Learning Research* 3:993-1022.

Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. "Scoring, term weighting and the vector space model." Chapter 6 from *Introduction to Information Retrieval.*

Landauer, Thomas K. 2007. "LSA as a Theory of Meaning" chapter 1 in T. K. Landauer, W. Kintsch, and D. S. McNamara, eds., *Handbook of Latent Semantic Analysis.* Mahwah, NJ: Lawrence Erlbaum Associates: 3-35.

Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. "Flat Clustering" and "Hierarchical Clustering." Chapters 16 and 17 from *Introduction to Information Retrieval.*

Qualitative practicum: Engage in a kind of "taxonomic" or structured coding on a sample of documents (~20 pages) related to your final project. You may open-code (as before), but now with an effort to creating a connected, coherent set of meaning clusters or more structured "scheme" where the codes have a relationship to one another, and where they capture features that cover the range of phenomena most relevant to the question animating your project (this may predominantly "emerge" from your reading of the texts or result from an interaction of what is "available" in the texts with your own pre-existing questions or intuitions.) This will involve both adding and dropping

suggestive but irrelevant codes from your scheme. Second, create a grouping, taxonomy/schematic (or update the one you may have turned in from last time), with an eye to "filling it in" or—adding all of the features, dimensions, sub-components, or connections between larger components to make it a thorough, structural description of the phenomenon at hand (and deleting those that distract/run orthogonal to such description). In your memo, include a description of the relevance of your typology to the question you anticipate you will ask for your final project (this may, of course, change). Then turn in the 10 pages of code and the structure in addition to the memo.

Computational possibilities: Document, code or word Clustering; Community Detection; LSA Topic Modeling (supervised or unsupervised).

Week 4.

Apr.22: *Discovering higher-level Patterns [Dimension Reduction – Semantic Spaces]*

Osgood, Charles E. 1952. "The nature and measurement of meaning." Psychological Bulletin, Vol 49(3), May 1952, 197-237.

Jurafsky, Daniel and James H. Martin. 2015. Speech and Language Processing. Chapter 19 ("Vector Semantics")

Widdows, Dominic. 2004. Geometry and Meaning, chapters 4 and 5 ("Measures of Similarity and Distance" and "Word-Vectors and Search Engines"). Center for the Study of Language and Information.

Kulkarni, Vivek, et al. "Statistically significant detection of linguistic change." Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015.

Kemp, Charles and Joshua B. Tenenbaum. 2008. "The discovery of structural form." *PNAS* (Aug 5;105(31):10687-92.

Optional:

Mikolov T, Chen K, Corrado G, Dean J. 2013a. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781

Goldberg Y, Levy O. 2014. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv:1402.3722

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global Vectors for Word Representation." EMNLP. Vol. 14. 2014.

Taddy, Matt. Document Classification by Inversion of Distributed Language Representations. arXiv:1504.07295.

Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. "Scoring, term weighting and the vector space model." Chapter 6 from *Introduction to Information Retrieval.*

Landauer, Thomas K. 2007. "LSA as a Theory of Meaning" chapter 1 in T. K. Landauer, W. Kintsch, and D. S. McNamara, eds., *Handbook of Latent Semantic Analysis.* Mahwah, NJ: Lawrence Erlbaum Associates: 3-35.

Widdows, Dominic. 2003. *Geometry and Meaning*, chapter 6 "Exploring Vector Spaces in One and More Languages" OR chapter 7 "Logic with Vectors: Ambiguous Words and Quantum States."

Pocklington, Richard and Michael L. Best. 1997. "Cultural Evolution and Units of Selection in Replicating Text." *Journal of Theoretical Biology* 188(1): 79-87.

Qualitative practicum: Same as before (but don't do the Qualitative Practicum both last week and this).

Computational possibilities: Latent Semantic Indexing; Word Vector Similarity; Word embedding models; Cluster Graph Induction.

Week 5.

Apr.29: *Sampling, Crowd-Sourcing and Reliability [Modeling Accuracy]*

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to its Methodology*. Thousand Oaks, CA: Sage: "Sampling" and "Reliability" 111-124; 211-256.

Morning, Ann. 2008. "Reconstructing Race in Science and Society: Biology Textbooks, 1952-2002." *American Journal of Sociology* 114(s1): S106-S137.

Roscigno, Vincent J. and Randy Hodson. 2004. "The Organizational and Social Foundations of Worker Resistance." *American Sociological Review* 69(1): 14-39.

Dawid, A. P., and Skene, A. M. 1979. "Maximum Likelihood Estimation of Observer Error-rates using the EM Algorithm." *Applied Statistics* 28(1): 20-28.

Rzhetsky, Andrey, Hagit Shatkay and W. John Wilbur. 2009. "How to Get the Most out of Your Curation Effort." *PLoS Computational Biology* 5(5): e1000391: 1-13.

Optional:

Kalton, Graham. 1983. *Introduction to survey sampling*. Beverly Hills : Sage Publications: 1-96.

Henry, Gary T. 1990. *Practical Sampling*. Newbury Park: Sage Publications.

Uebersax, JS and W. M. Grove. 1993. "A latent trait finite mixture model for the analysis of rating agreement." *Biometrics* 49(3): 823-835.

Qualitative practicum: Engage in a selective coding exercise similar to Roscigno and Hodson (2004), either based on a code taxonomy from theory or emergent from a qualitative or quantitative exercise. First, develop a "coding instrument," a short series of questions and instructions, based on your coding taxonomy, designed to guide a novice coder. Second, use this instrument yourselves and recruit two other people to join you in coding a previously uncoded 5 (or more) pages of document(s)—a document or documents which you have not yet read but which cover an area similar to that explored in your previous project. Follow your coding questions and instructions as closely as possible. Third, evaluate the intercoder reliability of the three coders with two (or more) agreement measures, and then reflect on the most reasonable approach for determining the "true" code. Finally, in your memo, describe the extent to which your instrument captured what was of greatest interest in the new document(s), your assessment of what it missed, the nature and distribution of reliability estimates and your preferred approach for code "combination." Then turn in 1) the instrument; 2) the 3 sets of 5 pages of code; and 3) the memo. I also encourage you to glance at Hodson's protocol and codesheet.http://www.sociology.ohio-state.edu/rdh/weprotocol.htmhttp://www.sociology.ohio-state.edu/rdh/WE-CODESHEET.doc

Computational possibilities: Field a coding experiment with "the crowd" via Mechanical Turk (http://www.mturk.com). Then use a probability model to predict the most likely outcome. (You

can use the free software library PyAnno, described at: http://pypi.python.org/pypi/pyanno and available in the "Links to software/tutorials" on this chalk site).

Week 6.

May.6: *Classifying Concepts, Accounts, and Documents [Machine Classification]*

Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. "Text Classification and Naïve Bayes", "Vector Space Classification," and "Support Vector Machines." Chapters 13-16 from *Introduction to Information Retrieval.*

Witten, Ian H. and Eibe Frank. 2011. "Ensemble Learning" Chapter 8 from *Data Mining: Practical Machine Learning Tools and Techniques*: 351-371.

Hopkins, Daniel J. and Gary King. 2010. A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science* 54(1): 229-247.

Optional:

Manning and Schütze. 1999. Chapter 16 *Text Categorization.*

Witten, Ian H. and Eibe Frank. 2005. "Algorithms: The Basic Methods" and "Implementations: Real Machine Learning Schemes." Chapter 4 and 6 from *Data Mining: Practical Machine Learning Tools and Techniques*: 83-142; 187-283.

Fan, Jung-Wei and Carol Friedman. 2007. "Semantic Classification of Biomedical Concepts Using Distributional Similarity." *Journal of the American Medical Informatics Association* 14(4): 467-177.

Qualitative practicum: No qualitative practicum opportunity this week.

Computational possibilities: Based on qualitative code assessments from week 2, 3 or 4 (or an independently coded/annotated sample built just for this assignment), (1) build one or more statistical classification models (e.g., support vector machine, inverse regression, neural network, tree-based classifier, naïve Bayes, etc.) that predict qualitative codes on part of your sample of coded texts, (2) test it/them on another portion of that sample (or engage in a cross-validation exercise), (3) predict codes on an uncoded portion of that sample using your model(s), and (4) qualitatively evaluate the predicted codes, and (5) quantitatively enumerate them in terms of true positives, true negatives, false positives, false negatives and at least two metrics based on these quantities.

Week 7.

May.13: *Extracting Claims, Arguments, Events [Natural Language Processing]*

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.

Mohr, John W., Robin Wagner-Pacifici, Ronald L. Breiger, and Petko Bogdanov. 2013. "Graphing the grammar of motives in National Security Strategies: Cultural interpretation, automated text analysis and the drama of global politics" *Poetics* 41(6):670-700.

Jurafsky, Daniel & James H. Martin. 2000. *Speech and Language Processing.* Singapore: Pearson Education, Inc.: Chapter 8 ("part of speech tagging") Chapter 15.5 ("Semantic Analysis: Information Extraction"): 577-586.

Bird, Stephen, Ewan Klein & Edward Loper. 2008. "Partial Parsing and Interpretation." Chapter 7 from *Natural Language Processing in Python*.

Friedman, Carol, Pauline Kra and Andrey Rzhetsky. 2002. "Two biomedical sublanguages: A description based on the theories of Zellig Harris." *Journal of Biomedical Informatics* 35: 222-235.

Rzhetsky et al. 2004. "GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data." *Journal of Biomedical Informatics* 37(1): 43-53.

Optional:

Ignatow, Gabriel. 2004. "Speaking Together, Thinking Together? Exploring Metaphor and Cognition in a Shipyard Union Dispute." *Sociological Forum* 19(3): 405-433.

Qualitative practicum: Qualitatively extract statements of a given logical or syntactic type (e.g., Cause-Effect pairs; Subject-Verb-Object triplets; Adjective-Noun Phrase pairs; etc.) from a sample of text(s) and assemble them into a spreadsheet or database. Next, evaluate patterns of these claims across your text(s) relative to your broader questions and interests.

Computational possibilities: Template-based information extraction using classification modeling (e.g., Event extraction; Key-value pair extraction); Semantic parsing, Syntactic parsing (e.g., Dependency parsing). With any of these, as assemble extracted statements into a database and analyze patterns relative to your research questions and interests.

****Pitch Texts/Context for Final Papers****

Week 8.

May.20: *Content Relations [Network Analysis]*

Carley, Kathleen. 1993. "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis." *Sociological Methodology* 23:75-126.

Carley, Kathleen. 1994. "Extracting Culture Through Textual Analysis." *Poetics* 22:291-312.Aral, Sinan and Marshall Van Alstyne. 2011. "The Diversity-Bandwidth Trade-off."*American Journal of Sociology* 117(1): 90-171. http://www.jstor.org/stable/10.1086/661238

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.

Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. "Link Analysis." Chapter 21 in *Introduction to Information Retrieval.*

Easley, David and Jon Kleinberg. 2010. Chapter 4.4 ("Tracking Link Formation in Online Data") and 13 ("The Structure of the Web") from *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Complete preprint on-line at http://www.cs.cornell.edu/home/kleinber/networks-book/

Optional:

Carley, Kathleen. 1997. "Network Text Analysis: The Network Position of Concepts." in *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, edited by Carl Roberts. Mahwah, NJ: Lawrence Erlbaum Associates.

Batagelj, Vladimir and Andrej Mrvar. Density-based approaches to Network analysis: Analysis of Reuters Terror News Network. Working Paper.

Pearl, Judea. 2009. "Introduction to Probabilities, Graphs and Causal Models." Chapter 1 in *Causality: Models, Reasoning and Inference*: 1-40.

Qualitative practicum: No qualitative practicum opportunity this week.

Computational possibilities: Network analysis and visualization of the relationship between qualitative codes, named-entities, statements, references, or visual motifs (e.g., centrality and centralization, partitioning, proximity, topological motifs like triangles).

Week 9.

May.27: *Beyond Text*

1) *Discourse, Conversation, and Music [Sound/Signal Processing]*
2) *Scenes, Visual events and Art [Machine Vision]*

Discourse:

Goffman, Erving. 1981. "Replies and Responses." *Forms of Talk* Philadelphia: University of Pennsylvania Press: 5-77.

van Dijk, Teun A. 1992. "Discourse and the Denial of Racism." *Discourse & Society* 3:87-118.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. "Echoes of power: Language effects and power differences in social interaction." Proceedings of WWW 2012.

McFarland, Daniel A., Dan Jurafsky, and Craig Rawlings. 2013. "Making the Connection: Social Bonding in Courtship Situations." *American Journal of Sociology*, 118(6):1596-1649.

Pentland, Alex. 2012. "The New Science of Building Great Teams." *Harvard Business Review*, April 2012. http://hbr.org/2012/04/the-new-science-of-building-great-teams/ar/1

Eagle, Nathan and Alex Pentland. 2006. "Reality mining: sensing complex social systems." *Journal Personal and Ubiquitous Computing* 10(4):255-268.

Pentland, Alex and Tanzeem Choudhury. 2003. "Sensing and Modeling Human Networks using the Sociometer." Seventh IEEE International Symposium on Wearable Computers (ISWC'03)

Scenes:

Collins, Randall. 2009. "The Micro-sociology of Violent Confrontations" and "Confrontational Tension and Incompetent Violence" (parts of Chapter 1 and Chapter 2) from *Violence: A Microsociological Theory*: 29-82.

Ogale, Neeti A. "A survey of techniques for human detection from video" https://www.cs.umd.edu/grad/scholarlypapers/papers/neetiPaper.pdf

Girshick, Ross B., Pedro F. Felzenszwalb and David McAllester. 2011. "Object Detection with Grammar Models" to appear in NIPS.

Malisiewicz, Tomasz, Abhinav Gupta, and Alexei A. Efros. 2011. "Ensemble of Exemplar-SVM's for Object Detection and Beyond." ICCV: 1-8.

Optional:

Browse:
Pentland, Alex. 2014. "Social Physics: How Good Ideas Spread—The Lessons from a New Science." Penguin.

http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html

Qualitative practicum:

Perform qualitative conversation analysis on a recorded interaction between two or more social actors that involves (1) identification and coding of each conversational "move", and (2) discovery of larger or higher-order movements in the conversational event (and/or interpretation the event as a whole). Students can alternatively analyze a recorded audio event, like a piece of music or a recorded concert, and attend to the same dynamics described above (i.e., code each "move" and identify higher-order patterns that might inscribe institutions).

OR

Engage in an initial "open coding" of image(s) or video(s) (the equivalent of 15 pages). This will involve highlighting and labeling *things* in the image(s) and video(s) – indices, icons symbols, etc. - that illuminate the visual content. You can code with pencil/pen on images or draw schematics of a video in motion. Also, draw a simple schematic of your codes and their relationship to one another (e.g., taxonomic tree, network, Venn Diagram). Turn in images, linked codes and figure along with your memo.

Computational possibilities:

Quantitative assessment of pitch, timbre, frequency, and duration of conversational moves in relation to one another; Network analysis of conversational moves; Music classification or clustering.

OR

Object (e.g., human bodies, faces) detection and classification; Scene understanding; Image clustering; Object classification and tracking in video; Image/Video clustering.

Week 10.
June.3:

****Turn in Draft Introduction and Methods Sections of Papers****

Week 11.
June.8: Content Fair and presentation of projects

****Turn in Final Introduction and Appendices due Friday, June.10 at 5pm****

Selected Content Analysis Bibliography:

Texts / Surveys:

Berelson, Bernard. 1952. *Content Analysis in Communication Research*. Glencoe, IL: The Free Press.

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

Holsti, Ole. 1969. *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley.

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to its Methodology*. Thousand Oaks, CA: Sage.

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

Osgood, Charles, George Suci, and Percy Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.

Popping, Roel. 2000. *Computer-assisted Text Analysis*. London: Sage.

Smith, Charles ed. 1992. *Motivation and Personality: Handbook of thematic content analysis*. Cambridge: Cambridge University Press.

Theory:

Caws, Peter. 1988. *Structuralism: The Art of the Intelligible*. Atlantic Highlands, NJ: Humanities Press International.

Clark, Herbert, and Eve Clark. 1977. *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace Jovanovich.

Collins, Allan, and Elizabeth Loftus. 1975. "A Spreading-Activation Theory of Semantic Processing." *Psychological Review* 82:407-428.

D'Andrade, Roy. 1995. *The Development of Cognitive Anthropology*. Cambridge: Cambridge University Press.

DiMaggio. 1997. "Culture and Cognition." *Annual Review of Sociology* 23:263-287.

Klimoski, Richard, and Susan Mohammed. 1994. "Team Mental Model: Construct of Metaphor?" *Journal of Management* 20:403-437.

Mitkov, Ruslan ed. 2003. *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.

Rapoport, Anatol. 1969. "A System-Theoretic View of Content Analysis." in *The Analysis of Communication Content*, edited by George Gerbner, Ole Holsti, Klaus Krippendorff, William Paisley, and Philip Stone. New York: John Wiley & Sons.

Sowa, John. 1984. *Conceptual Structures*. Readomg. MA: Addison-Wesley.

Thomas, Sari. 1994. "Artifactual Study in the Analysis of Culture: A Defense of Content Analysis in a Postmodern Age." *Communication Research* 21:683-697.

Weber, Robert. 1990. *Basic Content Analysis*. Newbury Park: Sage.

Wittgenstein, Ludwig. 2001. *Philosophical Investigations*. Malden, MA: Blackwell.

Wittgenstein, Ludwig. 2004. *Tractatus Logico-Philosophicus*. London: Routledge.

Qualitative Content Analysis Method:

Glaser, Barney G. 1978. *Theoretical Sensitivity*. Mill Valley, CA: The Sociology Press.

Glaser, Barney G. and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Lofland, John and Lyn H. Lofland. 1995. Chapter 6. *Analyzing Social Settings*. Wadsworth Publishing Company: 99-122.

Miles, Matthew B. and A. Michael Huberman. *Qualitative Data Analysis: An Expanded Sourcebook*, second ed. Thousand Oaks: Sage: 245-262.

Spradley, James P. 1979. *Participant Observation*. New York: Holt, Rinehart and Winston.

Spradley, James P. 1979. *The Ethnographic Interview*. Steps 5-6. New York: Holt, Rinehart and Winston.

Strauss, Anselm. 1987. *Qualitative Analysis for Social Scientists*. New York: Cambridge University Press.

Strauss, Anselm L. and Juliet Corbin. 1990. *Basics of Qualitative Research*. chapters 5-6. Newbury park, CA: Sage.

Weitzman, Eben and Matthew Miles. 1995. *Computer Programs for Qualitative Data Analysis*. Thousand Oaks, CA: Sage.

Standard Content Analysis Method:

Breiger, Ronald. 2000. "A Tool Kit for Practice Theory." *Poetics* 27:91-115.

Brent, Edward, Alan Thompson, and Edward Mirielli. 1995. "Disambiguating Verbal Comments in Social Interaction: A Computer Model of Meaning." *Journal of Mathematical Sociology* 20:109-125.

Carley, Kathleen. 1993. "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis." *Sociological Methodology* 23:75-126.

Campbell, Donald, and Donals Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56:81-105.

Carley, Kathleen. 1997. "Network Text Analysis: The Network Position of Concepts." in *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, edited by Carl Roberts. Mahwah, NJ: Lawrence Erlbaum Associates.

Carley, Kathleen, and Michael Palmquist. 1992. "Extracting, Representing, and Analyzing Mental Models." *Social Forces* 70 (March):601-636.

Cronbach, Lee, and Paul Meehl. 1955. "Construct Validity in Psychological Tests." *Psychological Bulletin* 52:283-302.

Franzosi, Roberto. 1989. "From Words to Numbers: A Generalized and Linguistics-Based Coding Procedure for Collecting Textual Data." *Sociological Methodology* 19:263-298.

Franzosi, Roberto. 1990. "Computer-Assisted Coding of Textual Data: An Application to Semantic Grammars." *Sociological Methods & Research* 19:225-257.

Gerbner, George, Ole R. Holsti, Klaus Krippendorff, William J. Paisley, and Philip J. Stone. 1969. *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques*. New York: John Wiley & Sons, Inc.

Gower, J. 1985. "Measures of Similarity, Dissimilarity, and Distance." in *Encyclopedia of Statistical Sciences*, edited by S. Kotz and N. Johnson. New York: Wiley & Sons.

Harris, Zellig. 1964. "Discourse Analysis." in *The Structure of Language: Readings in the Philosophy of Language*, edited by Jerry Fodor and Jerrold Katz. Englewood Cliffs, NJ: Prentice-Hall.

Jepperson, Ronald, and Ann Swidler. 1994. "What Properties of Culture Should We Measure?" *Poetics* 22:259-371.

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

Mohr, John. 1998. "Measuring Meaning Structures." *Annual Review of Sociology* 24:345-370.

Olsen, Mark, and Louis-Georges Harvey. 1988. "Computers in Intellectual History: Lexical Statistics and the Analysis of Political Discourse." *Journal of Interdisciplinary History* 18:449-464.

Roberts, Carl. 1989. "Other Than Counting Words: A Linguistic Approach to Content Analysis." *Social Forces* 68:147-177.

Rosenberg, Stanley, Paula Schnurr, and Thomas Oxman. 1990. "Content Analysis: A Comparison of Manual and Computerized Systems." *Journal of Personality Assessment* 54:298-310.

Widdows, Dominic. 2004. *Geometry and Meaning*. Stanford, CA: CSLI.

Practice:

Anheier, Helmut, and Jurgen Gerhards. 1991. "Literary Myths and Social Structure." *Social Forces* 69:811-830.

Berelson, Bernard, ed., and Morris Janowitz, ed. 1950. *Reader in Public Opinion and Communication*. Glencoe, IL: The Free Press.

Best, Michael. 1997. "An Ecology of Text: Using Text Retrieval to Study A life on the Net." *Artificial Life* 3:261-287.

Burton, Michael, and Lorraine Kirk. 1979. "Sex Differences in Maasai Cognition of Personality and Social Identity." *American Anthropologist* 81 (New Series):841-873.

Carley, Kathleen. 1991. "A Theory of Group Stability." *American Sociological Review* 56:331-354.

Carley, Kathleen. 1986. "Knowledge Acquisition as a Social Phenomenon." *Instructional Science* 14:381-438.

Carley, Kathleen. 1988. "Formalizing the Social Expert's Knowledge." *Sociological Methods & Research* 17:165-232.

Carley, Kathleen. 1994. "Extracting Culture Through Textual Analysis." *Poetics* 22:291-312.

Hodson, Randy. 2005. "Management Behaviour as Social Capital: A Systematic Analysis of Organizational Ethnographies." *British Journal of Industrial Relations* 43, 1 (March): 39-63.

Hodson, Randy. 2004. "Organizational Trustworthiness: Findings from the Population of Organizational Ethnographies." *Organization Science* 15,4 (July-August): 432-445.

Hodson, Randy. 2004. "A Meta-Ethnography of Employee Attitudes and Behaviors." *Journal of Contemporary Ethnography* 33,1 (February): 4-38.

Hodson, Randy. 1998. "Organizational Ethnographies: An Underutilized Resource in the Sociology of Work." *Social Forces* 76,4 (June):1173-1208.

Mische, Ann, and Philippa Pattison. 2000. "Composing a Civic Arena: Publics, Projects, and Social Settings." *Poetics* 27:163-194.

Mohr, John. 1998. "The Classificatory Logics of State Welfare Systems: Towards a Formal Analysis." in *Public Rights, Public Rules: Constituting Citizens in the World Polity and National Policy*, edited by Connie McNeely. New York: Garland.

Mohr, John, and Vincent Duquenne. 1997. "The Duality of Culture and Practice: Poverty Relief in New York City, 1888-1917." *Theory and Society* 26:305-356.

Roberts, Carl. 1997. "A Generic Semantic Grammar for Quantitative Text Analysis: Applications to East and West Berlin Radio News Content from 1979." *Sociological Methodology* 27:89-129.

Rosenberg, Seymour. 1989. "A Study of Personality in Literary Autobiography: An Analysis of Thomas Wolfe's *Look Homeward, Angel*." *Journal of Personality and Social Psychology* 56:416-430.

Shapiro, Gilbert, and John Markoff. 1998. *Revolutionary Demands: A Content Analysis of the Cahiers de Doleances of 1789*. Stanford, CA: Stanford University Press.

Swede, Southworth, and Philip Tetlock. 1986. "Henry Kissinger's Implicit Theory of Personality: A Quantitative Case Study." *Journal of Personality* 54:617-646.

Tilly, Charles. 1986. "European Violence and Collective Action since 1700." *Social Research* 53:160-184.

Tilly, Charles. 1997. "Parliamentarization of Popular Contention in Great Britain, 1758-1834." *Theory and Society* 26:245-273.