Toward a computational hermeneutics

John W Mohr, Robin Wagner-Pacifici and Ronald L Breiger

Abstract

We describe some of the ways that the field of content analysis is being transformed in an Era of Big Data. We argue that content analysis, from its beginning, has been concerned with extracting the main meanings of a text and mapping those meanings onto the space of a textual corpus. In contrast, we suggest that the emergence of new styles of text mining tools is creating an opportunity to develop a different kind of content analysis that we describe as a computational hermeneutics. Here the goal is to go beyond a mapping of the main meaning of a text to mimic the kinds of questions and concerns that have traditionally been the focus of a hermeneutically grounded close reading, a reading that focuses on what Kenneth Burke described as the poetic meanings of a text. We illustrate this approach by referring to our own work concerning the rhetorical character of US National Security Strategy documents.

Keywords

Hermeneutics, content analysis, close reading, Big Data, Kenneth Burke, National Security

Thin reading: The first century of content analysis

Content analysis describes a set of procedures for transforming texts, which are written by and intended to be read by people, into numerical datasets that are read by computers and intended to be interpreted with formal methods.[1] The goals of a content analysis vary from project to project, but mostly social scientists have sought to use these methods to measure the presence of a set of key meanings and to map the distribution of those meanings across the space of a textual corpus.[2]

Scholars have been using these methods to analyze texts for over a hundred years but a significant leap in technical sophistication occurred in the interdisciplinary crucible of the Second World War. Scholars like Harold Lasswell, who had written his dissertation at the University of Chicago on the propaganda campaigns of the First World War, worked on behalf of the US government to create new textual analysis procedures that could be used to gather information from newspapers and other strategic textual corpora. Lasswell, who served as director of the Experimental Division for the study of Wartime Communications at the US Library of Congress, led a staff of brilliant young social scientists in developing a suite of methods for systematically reading large textual corpora in such

a way that critical bits of information could be extracted and a measure of informational reliability could be calibrated (Lasswell et al., 1949). After the war, Lasswell signed on to help direct a project at Stanford's Hoover Institute that used the same suite of methods to study 20,000 newspaper editorials, sampled from the "prestige" papers of five countries—France, Germany, Russia, the US and the UK (between 1890 and 1945).[3] The goal of the project was to map the changing symbolic frames of domestic and international politics and to compare these mappings across the dominant nation states in the years leading up to the war. It is also a useful exemplar of the logic of analysis that came to define the field.

Lasswell and his Stanford colleagues began by laying out their data categories as a set of pre-defined keywords, phrases, and concepts. Next, they wrote out careful instructions (and a set of decision rules) for coding each item.[4] Finally, a team of human coders,

University of California, Santa Barbara, CA, USA
The New School, New York, NY, USA
University of Arizona, Tucson, AZ, USA

Corresponding author:
John W Mohr, University of California, Santa Barbara 3407 Social Sciences & Media Studies Santa Barbara, CA 93106-9430, USA.
Email: mohr@soc.ucsb.edu

following these procedures, read the corpus while searching for 206 place names and 210 key symbols reflecting major political ideologies of the times, concepts like "Nationalism, Nazism, Neutrality, and Nonintervention" (Lasswell et al., 1952: 43). Here, and throughout the history of modern content analysis, the technical procedures were designed to carefully and reliably pare away the complexity of textual information into a small set of core informational units that could be reliably measured and mapped. In short, content analysis, from the start, has been focused on the goal of capturing a set of primary ideas, usually those constituting the manifest meaning of a textual corpus, which is to say, that which is expressed in plain view and about which there is little or no dispute.[5]

The pursuit of this primary goal has meant that the subtleties of expression, the complexities of phrasing and the more nuanced meanings of textual corpora are discarded, with only the best, most efficient units of meaning being extracted and preserved as data. Even as the social sciences advanced and computing power exploded over the next half century, this core logic of content analysis methodologies has lived on.[6] During the 1950s, content analysts began to "focus on counting internal contingencies between symbols instead of the simple frequencies of symbols" and they began to worry more about "problems of inference from verbal material to its antecedent conditions" (Pool, 1959: 2). This era introduced the use of co-occurrence matrices that increasingly came to be used as a poor-man's measure of semantic structure.[7] In the 1960s, pre-coded computer dictionaries were compiled and shared among researchers in common domains of inquiry (Stone et al., 1966). By the 1970s and 1980s networks of causal assertions were being constructed by analysts closely reading transcripts of policy deliberations (Axelrod, 1976), and factor analysis and latent structure analysis technologies were being used to help excavate implicit meaning structures from a variety of textual datasets (e.g. Namenwirth and Lasswell, 1970; Weber, 1987). In the 1980s, hand-coding procedures were modified so as to record information about the semantic grammars that linked key terms together in relational sets (Franzosi, 1989, 1990). By the 1990s many new types of relational and, especially, network style methodologies were being applied to textual data demonstrating new ways to unpack implicit meaning and communication structures (Abbott and Hrycak, 1990; Bearman and Stovell, 2000; Breiger, 2000; Carley, 1994; Cerulo, 1988; Ennis, 1992; Martin, 2000; Mohr, 1994, 1998; Tilly, 1997). And yet, throughout all these important advances, the core logic of the field did not change. The goal is to extract the main bits of communicative content from the corpus, to apply formal methods to extract the principal components of the meaning structures (or the communication structures) and to map those onto the textual space of the corpus.

Close reading: The method of hermeneutic analysis

But, as any close reader will tell you, there is always more than one way to read a corpus. For humanists and scholars who specialize in the 'close reading' of texts it is the very complexity of their meanings, their nuanced peculiarities of style, their inherent multivocality, their complex layeredness and organized incoherence, indeed, it is precisely those things that are not so easily found on the surface of texts, that makes them worthy of close study in the first place. It is hardly a surprise that content analysis projects, which have traditionally focused on capturing manifest meanings of textual corpora, have been of little interest to scholars who come from more hermeneutic disciplines. Those qualities of a text that are of greatest interest to a close reader are the very things that traditional content analysis projects seek to cleave away.[8]

In contrast, when confronting a text an experienced close reader will begin with a sense of the basic semantics, syntax, rhetorical forms, and genre of the given text or corpus, along with a knowledge of the text's historical context. But a skilled reader must also have a sense of the positionings, form-takings and flows of the text—its creature-like qualities. A close reader will keep an eye on the text's diachronic pulse, its fluidity, its starts and stops, expansions and contractions (managed through such things as repetition or recursivity or parentheticals...), accelerations and slow-downs. A close reader will move back and forth between the particulars and the whole and between the text and the context(s) and she will attend to any unexpected turn of phrase, to the anomalous character, to the anachronistic appearance of a feature from another genre and so bring a more general "sense of the textual" to the reading. This will include a sense of the language(s) used, a sense of the genre in which the language is embedded, and a sense of the typical flows or relays of these texts with other texts and other symbolic mediations. This type of reading is enormously difficult and time-consuming and, as Paul Ricoeur taught us, always provisional, as the thing about texts is that once they leave the hands of the writers, they launch themselves into unknown and unpredictable shaping contexts and interpreting readers. Nevertheless—sound and illuminating interpretive readings can be made and made to stick through the deployment of an effective hermeneutic practice.

The literary theorist Kenneth Burke described these as two different kinds of interpretations, a semantic and a poetic. According to Burke, a semantic interpretation seeks to clarify and to specify the precise and manifest

communicative intention of a text, in much the same way that a postal system seeks to establish a clear and unambiguous mapping of written addresses and geophysical destinations so that mail can be efficiently mapped to its proper destination. Thus for Burke, the *semantic ideal* has "the aim to evolve a vocabulary that gives the name and address of every event in the universe" (Burke, 1941: 141). But, as Burke also explains, "The address, as a counter, works only in so far as it indicates to the postal authorities what kind of operation should be undertaken. But it assumes an organization. Its meaning, then, involves the established procedures of the mails, and is in the instructions it gives for the performance of desired operations within this going concern" (Burke, 1941: 140).

Burke argues that human experiences are more complex than this because humans are suspended in elaborate webs of overlapping meanings. Burke explains, "when you have isolated your individual by the proper utilizing of the postal process, you have not at all adequately encompassed his 'meaning.' He means one thing to his family, another to his boss, another to his underlings, another to his creditors, etc." (p. 142). It is this complex multiplicity of layered meanings that brings us toward a poetic reading of a text. Thus, a poetic interpretation is not concerned with the thinning out of meaning, but on the contrary, with the filling out of meaning. It is not arrived at by neutral analysis but instead through the expression and experience of passion and attitude. Burke writes, "(t)he semantic ideal envisions a vocabulary that *avoids* drama. The poetic ideal envisions a vocabulary that *goes through* drama…The first seeks to attain this end by the programmatic elimination of a weighted vocabulary at the start (the neutralization of names containing attitudes, emotional predisposition); and the second would attain the same end by exposure to the maximum profusion of weightings" (Burke, 1941: 149). If the first century of content analysis was focused on semantic interpretation, we expect that the next century will focus on the poetic.

Thick reading: The new age of computational hermeneutics

The arrival of Big Data is changing the way that social scientists and humanists analyze texts. Most obviously we have seen a transformation in the scale and the breadth of digitized textual corpora that are becoming available for analysis and this changes the kinds of questions that then come into focus (e.g. Goldstone and Underwood, 2012; Jockers, 2013; Jockers and Mimno, 2013; Lazer et al., 2009; Liu, 2013; Mayer-Schonberger and Cukier, 2013; Michel et al., 2011; Moretti, 2013; Tangherlini and Leonard, 2013, see

also cases described by Bearman in this issue). But researchers are also making fundamental changes in how they use text analytic methodologies to measure the meanings and character of textual corpora. Here we discuss the emergence of one such strand of text mining sensibilities that we call computational hermeneutics.[9]

The central idea of a computational hermeneutics is that all available text analysis tools can and should be drawn upon as needed in order to pursue a particular theory of reading. Here the most important impact of Big Data is the expansion of new types of algorithmic and computational tools for reading texts. Instead of restricting ourselves to collecting the best small pieces of information that can stand in for the textual whole in the manner that Lasswell's project illustrates, contemporary technologies give us the ability to instead consider a textual corpus in its full hermeneutic complexity and nuance. This is what creates the opportunity for a new style of computational hermeneutics. But with this change, so do the research questions shift in a fundamental way. Now we must ask, given the complexity of the textual whole, how can we extract those various poetically meaningful components (or structurally intertwined sets of poetically meaningful components) that would be of greatest use to whatever interpretive intention we bring to the corpus? Put differently, how can we begin to focus on whatever combination of measurable textual features that we would most want to attend to as a focused close reader of this text?

This is not just a methodological question, it is very much a theoretical question. Before we can ask what component of textual expression we would want to extract, we must have a theory of the text within which the concept of a component makes sense, a component of what? What is this meaningful whole?[10] This is the kind of interpretive endeavor that we think has become possible in the age of Big Data, and it has created the opportunity for building a different kind of computational hermeneutics.

An example: The changing rhetorical logic of US National Security Strategy

It was George W Bush's preventive war doctrine that first led us to examine the corpus of National Security Strategy (NSS) reports of the United States executive branch. An initial close reading of several of the NSS reports suggested that there were some elements of these texts that a close reading, with its exclusively hermeneutic tools, could not illuminate but only dimly discern. The intuition was about latent networks—a deep structure of the international order (from the US point of view, of course) that involved interactions (exchanges, recognitions, performative speech acts like exhortations, expositives, exercitives, hailings, callings-out)

and also political-cognitive mappings and structures. One possible approach anticipated a typologization of these networks along familiar social structural lines: family, clan, bureaucracy, corporation, schoolyard, comprised of stock characters—patriarchs, elders, bullies, friends, partners, upstarts, middle managers. And the hope was that a more formal computational reading of the multiple texts would be able to discern if and how these networks took shape and behaved across the compiled set of texts.

Our first attempt to analyze this corpus of documents focused on their rhetorical form (Mohr et al., 2013). We were interested in rhetoric because it is a style of textual analysis that has deep roots in the history of hermeneutic studies and because it is constructed according to a series of fairly well understood, relatively formal properties and principles that make it a field of investigation that is amenable to the sorts of structured investigations that were of interest to us. The study of rhetoric is also a convenient way to link textual analysis to consequential matters because when rhetoric is deployed effectively in significant fields of social action, such as in the world of National Strategic Security institutions, it can become some of the most materially powerful types of institutionalized speech activity. This is because rhetorical logics can become enacted things in the world, by being re-deployed through bureaucratic forms. Not necessarily as formal operating orders, but as rhetorical frames that undergird a broader discursive framing of the international scene. And so these are texts that matter as the public face of a configuration of rhetorical framings about the nature of international order.

Our corpus included 11 NSS statements (published from 1990 to 2010) and we drew upon Kenneth Burke's insights about how to perform a rhetorical analysis (Burke, 1941). Burke proposes to identify the dramatic logic of a text by attending to how events are characterized within what he called the dramatistic pentad—this includes five terms, "what was done (act), when or where it was done (scene), who did it (agent), how he did it (agency), and why (purpose)" (Burke, 1945: xv). We asked how might we apply new computer based tools to read the corpus just as Kenneth Burke would have us read it? In our first effort, we employed topic models as a way to sort the corpus into different thematic arenas (or "scenes" in Burke's terminology).[11] We then used Named Entity Recognition (NER) tools to identify different agents and semantic grammar analysis to identify the actions taken by these agents. We present this as graphs of actor–action–actor semantic networks broken out by topic. This gives us a way to begin to visualize the shifting rhetorical logics that moved across time and across US presidential administrations.

Figure 1 shows how this approach captures the Clinton administration's framing of the problem of terrorism in 1996. This was a frame that focused on the end of the Cold War. Key agents in the frame included Russia, China, Ukraine, Belarus, and Kazakhstan. Control was the primary action that was invoked and the objects of concern were weapons of mass destruction, nuclear tipped missiles, and nuclear proliferation. Local hot-spots were also included in this discussion—Bosnia, South Africa, and Guatemala, places where terrorism might flare up. Figure 2 shows how the same topic was discussed by the Bush administration a decade later. The frame has shifted dramatically toward the Middle East. Key agents in the discussion of terrorism now include Iraq, Iran, Syria, Egypt, Saudi Arabia, Israel, the Palestinians, and Saddam. From the Clinton administration's concerns with 'controlling' the agents of terror, we have moved into the Bush administration's focus on 'removing', 'commanding,' 'building,' and 'expanding'. The overall goal is also much broader, from controlling the proliferation of materials, the focus has shifted toward creating opportunities for people to live in freedom so that terrorism will find no footing. In short, by focusing in on what Burke describes as the grammar of motives we are able to use automated methods to mimic a particular style of close reading and, in doing so, to reveal interpretations that are not immediately visible on the surface of these texts.

Conclusion: Toward a computational hermeneutics

In this short essay we have sought to suggest that the age of Big Data is important not only because it presents us with opportunities to use larger and more comprehensive datasets, but also because it gives us the opportunity to change the way that we formally engage with and interpret textual corpora. Rather than seeking to extract small amounts of critical information that we hope is representative of the manifest meaning of a text, as has been true throughout the history of content analysis, the new age of computational hermeneutics provides us with a chance to pursue deeper, subtler and more poetic readings of textual corpora. Instead of focusing on the main communicative intentions of a text, we are now able to push toward the kind of close reading that has traditionally been conducted by hermeneutically oriented scholars who find not one simple uncontested communication, but multiple, contradictory and overlapping meanings. Instead of just content, we are now able to focus on style, and on the ways in which texts are embedded in broader literary conversations.[12] In this sense, style is substance.
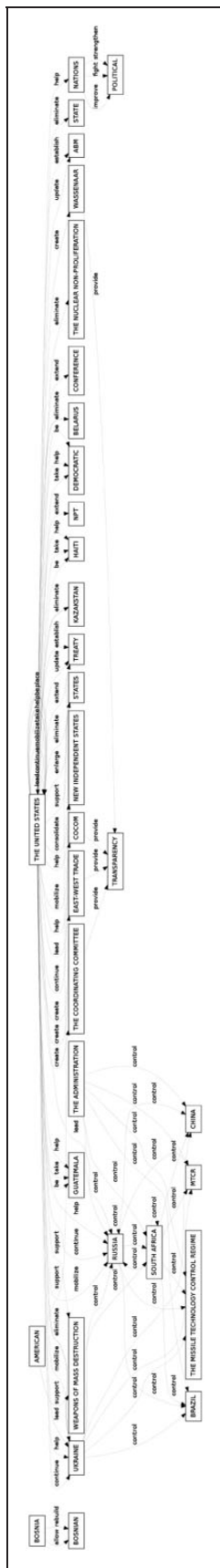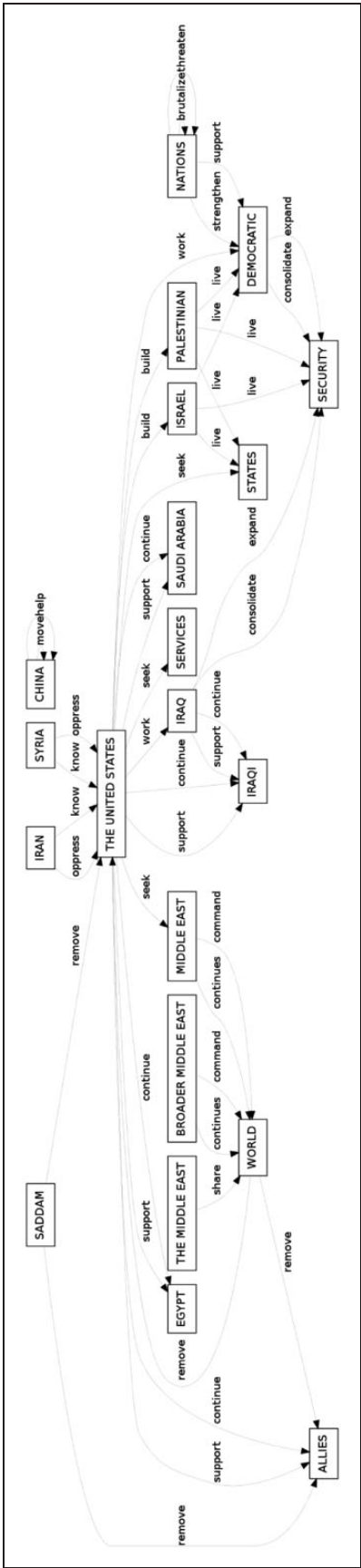
Figure 1. Actors and acts in terrorism discourse, NSS 1996 (Clinton).

Figure 2. Actors and acts in terrorism discourse, NSS 2006 (GW Bush).

Our own work in this domain is just beginning. For one thing, we have not yet completed Burke's mandate for rhetorical analysis. In the paper discussed here, we focused on analyzing just three elements from Burke's theory of the dramatistic pentad (actors, acts and scenes). In new work we hope to fill in the other two elements from Burke's theory (agency and purpose). We are exploring ways of using "named-entity recognition" (NER) as a way of coding what Burke describes as the problem of "agency" (specifying what means or instruments were used in carrying out an act) and looking to see whether sentiment analysis can provide us with a way of coding Burke's last rhetorical element, the "purpose" of the act. From here, there are many directions to proceed. For example, Burke highlights the elements of textual "friction" in his pentadic ratios (Act-Scene; Act-Agent; Agent-Purpose). This friction expresses the points of ambiguity or contradiction or uncertainty in texts where there is not a seamless alignment of all the pentadic elements—agents are not "at home" in their scene; acts don't have a clear purpose and so forth. We wish to press on this fundamental Burkean insight—a kind of textual uncertainty principle.

Our practical goal in this research has been to understand how these kinds of rhetorical logics are used to create a grounding of normalcy within which acts of strategy across the international order will be perceived as legitimate, rational and powerful. Once we are able to more effectively map these rhetorical elements in the current corpus, we are intrigued to see whether we will also be able to track how these rhetorical frames flow across broader institutional domains—how they are mimicked, changed, and contested by others. But our overall goal is to move to embrace the full complexity of the textual and to do so in a way that takes advantage of new opportunities for computational analysis. Ultimately we hope to employ these kinds of computational methods to assemble a sort of rhetorical interpretation machine. But more broadly we hope to have suggested that as these new kinds of computational methodologies continue to proliferate, so too does the need for skillful close readers, scholars who bring a sophisticated understanding of textuality, hermeneutics, and theories of close reading to bear. Without this sort of theorizing, the new computational methodologies will be severely hobbled.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

1. Of course, this is not always true. For example, when Wagner-Pacifici learned content analysis in graduate school in the early 80s from George Gerbner, there was no expectation that the analysis would necessarily be carried out by computers—it was human readers, category makers, counters, collaters, and analysts all the way down at that point. Thus, while content analysis can indeed avail itself of computational readings, it's not required.
2. Some of the classic statements defining the state of the art from the early years of content analysis include: Berelson (1952); Holsti (1969); Holsti et al. (1968); Gerbner et al. (1969); Krippendorff (1980); Lasswell et al. (1949, 1952); Namenwirth and Weber (1987); North et al. (1963); Pool (1959); Stone et al. (1966); Weber (1985).
3. This was one of several large content analysis projects that Lasswell organized as a part of the RADIR program at Stanford (see Lasswell et al., 1952; Rogers, 1994).
4. See Pool (1952: 20) for a detailed description of the coding procedures.
5. Bernard Berelson's classic definition says as much, "Content analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication" (Berelson, 1952). Indeed, one of the key technologies developed in this period was a measure for calculating inter-coder reliability which begins from the assumption that two readers will, assuming no error, have the same interpretation of the same text.
6. It operates as a form of what Bourdieu (2004) calls scientific doxa (see also Thagard, 1992).
7. Other research by psychologists such as Charles Osgood, (Osgood et al., 1957) who was interested in the measurement of meaning using semantic differential techniques, also played a role here. Note that Osgood was among a small group of scholars who were invited to the Stanford Center for Advanced Study in the Social Sciences in 1962 to help develop new methods of content analysis using state-of-the-art scaling and factor methodologies in an effort to create new tools for policy analysis (North et al., 1963).
8. Ted Underwood's essay for this special issue describes a similar experience. Underwood writes, "I think literary historians were mostly right, for instance, to ignore statistical models in the twentieth century. The modeling methods that prevailed for most of that century were best suited to structured datasets with relatively few variables, and that isn't the form our subject usually takes . . . in literary history, problems of historical incommensurability are in effect the discipline's central subject, so quantitative methods have found little foothold."
9. The idea that there is a natural complementarity between hermeneutic theories of interpretation and the practical theories of computational text analysis goes back a long

way in the literature. A classic essay by Mallery et al. (1986) from the MIT AI Lab explicitly linked traditional theories of hermeneutics to the efforts of computer scientists to create more effective AI programs. Mallery et al. highlighted the important similarities between the philosophical concept of the "hermeneutic circle or the notion that understanding or definition of something employs attributes which already presuppose an understanding or a definition of that thing" and the computer science concept of "bootstrapping: a process which uses a lower order component (a bootstrap component) to build a higher order component" (1986: 2). They also draw attention to earlier usage of the concept of computational hermeneutics, the earliest that we know of, highlighting a collaboration between Alker et al. (1985) and Lehnert et al. (1983) who focused on "investigations of the affective structure of texts and in their concern with systematic rules for identifying the genres of narrative." In 1987, Robert Philip Weber scanned the content analysis horizon and reported that, "At this writing, developments in 'Computational Hermeneutics' hold great promise for natural language understanding and social science applications" (1987: 193). He then proceeded to list nearly a dozen other articles in this stream.

10. Consider, for example, what the "social" means in the context of a formal analysis of texts. It is quite unclear. Do the texts under analysis representationally reproduce the social world outside? If so, how? Do they discursively reproduce social structures, cognitive structures, and social interactions among specified textual agents? These all implicate complex theories of the social and that means we need to have some understanding of what we expect from a text before we are able to know how to read the text. How else will we know what to attend to?

11. For us, topic models are useful because like other projects that focus on capturing continuities of thematic content across time (e.g. Heuser and Le-Khac, 2012 or Rule et al., forthcoming), the goal is to not find exactly the same discussion every year but rather to identify a group of discussions that nonetheless fall under the same general topic. On topic models, see Blei et al. (2003); Blei (2011, 2012); Mohr and Bogdanov (2013).

12. The living proof of this is vividly on display in the Digital Humanities today where a robust and solidly hermeneutic project has been brilliantly coupled with a broad range of new computational tools, some of which have been borrowed from the computational sciences and some of which have been invented by the Digital Humanists themselves. In addition to the work of Rachel Buurma and Ted Underwood in the current special issue, see, for example, Moretti's work (2013) and the online working papers put out by his Stanford Literary Lab. There are many other good examples as well. See, for example, Ramsay, 2011 and the papers collected in Schreibman and Unsworth (2004) and in Gold (2012). Alan Liu (2013) provides a useful mapping of the field.

References

Alker HR, Lehnert WG and Schneider DK (1985) Two reinterpretations of Toynbee's Jesus: Explorations in computational hermeneutics. In: Tonfoni G (ed) *Artificial Intelligence and Text Understanding. Quaderni di Ricerca Linguistica* 6: 49–94.

Abbott A and Hrycak A (1990) Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology* 96: 144–185.

Axelrod R (ed) (1976) *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton, NJ: Princeton University Press.

Bearman P and Stovel K (2000) Becoming a Nazi: A model for narrative networks. *Poetics* 27: 69–90.

Berelson B (1952) *Content Analysis in Communication Research*. New York, NY: Free Press.

Blei DM (2011) *Introduction to Probabilistic Topic Models*. Princeton, NJ: Princeton University.

Blei DM (2012) Topic modeling and digital humanities. *Journal of Digital Humanities* 2(1): 8–11.

Blei DM, Ng AY and Jordan MI (2003) Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.

Bourdieu P (2004) *Science of Science and Reflexivity*. Chicago: University of Chicago Press.

Breiger RL (2000) A tool kit for practice theory. *Poetics* 27(2–3): 91–115.

Burke K (1941) *The Philosophy of Literary Form*. Berkeley: University of California Press.

Burke K (1945) *A Grammar of Motives*. Berkeley: University of California Press.

Carley KM (1994) Extracting culture through textual analysis. *Poetics* 22: 291–312.

Cerulo K (1988) Analyzing cultural products: A new method of measurement. *Social Science Research* 17: 317.52.

Ennis JG (1992) The social organization of sociological knowledge: Modeling the intersection of specialties. *American Sociological Review* 57: 259–265.

Franzosi R (1989) From words to numbers: A generalized and linguistics-based coding procedure for collecting textual data. *Sociological Methodology* 19: 263–298.

Franzosi R (1990) Computer-assisted coding of textual data: An application to semantic grammars. *Sociological Methods and Research* 19(2): 225–257.

Gerbner G, Holsti OR, Krippendorff K, et al. (1969) *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques*. New York: John Wiley and Sons.

Gold MK (2012) *Debates in the Digital Humanities*. Minneapolis: University of Minnesotta Press.

Goldstone A and Underwood T (2012) What can topic models of PMLA teach us about the history of literary scholarship? *Journal of Digital Humanities* 2(1): 40–49.

Heuser R and Le-Khac L (2012) A quantitative literary history of 2,958 nineteenth-century British novels: The semantic cohort method. *Literary Lab Pamphlet 4*, May 2012, Stanford Literary Lab.

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

Holsti OR, Loombs JK and North RC (1968) Content analysis. In: Lindzey G and Aronson E (eds) *The Handbook of Social Psychology*. Cambridge, MA: Addison-Wesley.

Jockers ML (2013) *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.

Jockers ML and Mimno D (2013) Significant themes in 19th-century literature. *Poetics* 41(6): 750–769.

Krippendorff K (1980) *Content Analysis; An Introduction to its Methodology*. Beverly Hills, CA: Sage Publication.

Lasswell H and Leites N, et al. (1949) *Language of Politics: Studies in Quantitative Semantics*. New York, NY: George W. Stewart Publisher.

Lasswell HD, Lerner D and Pool IS (1952) *The Comparative Study of Symbols: An Introduction*. Palo Alto, CA: Stanford University Press.

Lazer D, Pentland AS, Adamic L, et al. (2009) Life in the network: The coming of age of computational social science. *Science* 323(5915): 721–723.

Lehnert WC, Alker HR and Schneider DK (1983) The heroic Jesus: The affective plot structure of Toynbee's Christus Patiens. In: *Proceedings of the sixth international conference on computers and the humanities* (eds SK Burton and Dd Short). Rockville, MD: Computer Science Press.

Liu A (2013) The Meaning of the Digital Humanities.*PMLA* 128 (2013): 409–423.

Mallery JC, Hurwitz R and Duffy G (1986) Hermeneutics: From textual explication to computer understanding? In: *MIT Artificial Intelligence Laboratory Memo No. 871*, Cambridge, MA.

Martin JL (2000) What do animals do all day? *Poetics* 27: 195–231.

Mayer-Schonberger V and Cukier K (2013) *Big Data: A Revolution that Will Change How we Live, Work and Think*. London: John Murray.

Michel JB, Shen YK, Aiden AP, et al. (2011) The Google Books Team, Pickett JP, Hoiberg D, Clancy D, et al. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014): 176–182.

Mohr JW (1994) Soldiers, mothers, tramps and others: Discourse roles in the 1907 New York City Charity Directory. *Poetics: Journal of Empirical Research on Literature, the Media, and the Arts* 22: : 327–357.

Mohr JW (1998) Measuring meaning structures. *Annual Review of Sociology* 24: 345–370.

Mohr JW and Bogdanov P (2013) Introduction—Topic models: What they are and why they matter. *Poetics* 41(6): 545–569.

Mohr JW, Wagner-Pacifici R, Breiger R, et al. (2013) Graphing the grammar of motives in U.S. National Security strategies: Cultural interpretation, automated text analysis and the drama of global politics. *Poetics* 43(6): 670–700.

Moretti F (2013) *Distant Reading*. London: Verso.

Namenwirth JZ and Lasswell HD (1970) *The Changing Language of American Values: A Computer Study of Selected Party Platforms*. Beverly Hills, CA: Sage.

Namenwirth JZ and Weber RP (1987) *Dynamics of Culture*. Boston, MA: Allen and Unwin.

North RC, George Zaninovich OHM and Zinnes DA (1963) *Content Analysis: A Handbook with Applications for the Study of International Crisis*. Evanston, IL: Northwestern University Press.

Osgood CE, Suci G and Tannenbaum P (1957) *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.

Pool IS (1952) The Prestige Papers: A Survey of their Editorials. *Hoover Institute Studies, Series C: Symbols*. No. 2. Palo Alto, CA: Stanford University Press.

Pool IS (ed) (1959) *Trends in Content Analysis*. Urbana: University of Illinois Press.

Ramsay S (2011) *Reading Machines: Toward an Algorithmic Criticism*. Chicago: University of Illinois Press.

Rogers E (1994) *A History of Communication Study: A Biographical Approach*. New York, NY: The Free Press.

Rule AR, Cointet JP and Bearman PS (2015) Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014. *PNAS*. 112(35): 10837–10844.

Schreibman RS and Unsworth J (eds) (2004) *A Companion to Digital Humanities*. Malden, MA: Blackwell Publishing.

Stone PJ, Dunphy DC, Smith MS, et al. (1966) *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press.

Tangherlini TR and Leonard P (2013) Trawling in the sea of the great unread: Sub-corpus topic modeling and humanities research. *Poetics* 41(6): 725–749.

Thagard P (1992) *Conceptual Revolutions*. Princeton, NJ: Princeton University Press.

Tilly C (1997) Parliamentarization of popular contention in Great Britain, 1758-1834. *Theory and Society* 26(2–3): 245–273.

Weber RP (1985) *Basic Content Analysis*. Beverly Hills, CA: Sage.

Weber RP (1987) Measurement models for content analysis. Chapter 8. In: Zvi Namenwirth J and Philip Weber R (eds) *Dynamics of Culture*. Boston, MA: Allen and Unwin.

This article is part of a special theme on *Colloquium: Assumptions of Sociality*. To see a full list of all articles in this special theme, please click here: http://bds.sagepub.com/content/colloquium-assumptions-sociality.