

Narrative and computational text analysis in business and economic history

Gregory Ferguson-Cradler^a

^aInland Norway University of Applied Sciences, Institute for Law, Philosophy and International Studies, Lillehammer, Norway

ARTICLE HISTORY

Compiled April 5, 2021

ABSTRACT

Recent calls from within economics for increased attention to narrative open the door to possible cross-fertilization between economics and more humanistically-oriented business and economic history. Indeed, arguments for economists to take narratives seriously and incorporate them into economic theory have some similarities with classic calls for a revival of narrative in history and abandonment of ‘scientific’ history. Both share an approach to explaining social phenomena based on the micro-level. This article examines how new methods in computational text analysis can be employed to further the goals of prioritizing narrative in economics and history but also challenge a focus on the micro level. Through a survey of the most frequently-used tools of computational text analysis and a overview of their uses to date across the social sciences and humanities, this article shows how such methods can provide economic and business historians tools to respond to and engage with the ‘narrative turn’ in economics while also building on and offering a macro-level corrective to the focus on narrative in history.

KEYWORDS

Narrative, natural language processing, topic models, machine learning, digital humanities

1. Introduction

Recent work in economics has suggested the need for what might be described as a ‘narrative turn’ in the discipline. No less prominent an economist than Robert Shiller has argued forcefully that economic theory which considers only functional or material factors is unlikely to be able to account for economic change in the past nor predict it in the future (Shiller, 2017, 2020). Widespread mental narratives in which people frame themselves and their actions are, Shiller has argued, a causal force on economic outcomes and trends. Thus, an economics that omits consideration of the study of narratives and leaves them out of its models will be missing a key causal variable in understanding and predicting the economy. To this end, Shiller has recently called for economists to ‘best advance their *science* by developing and incorporating into it the *art* of narrative economics’ (Shiller, 2020, xv, emphasis in the original). This, he added, will require engagement with other disciplines and other types of arguments and evidence that economists are less accustomed to working with.

The call for renewed attention to narrative in economics is in some respects evocative of a rallying cry by another preeminent representative of their discipline to give more room to narrative: the great early modernist Lawrence Stone in his classic 1979 article ‘The revival of narrative’ (Stone, 1979). To be sure, the goals and disciplinary standards advocated by Shiller and Stone were radically different and, to an extent, even orthogonal to each other. While Shiller’s appeal is to adopt and accommodate study of narratives into the structures and forms of quantitative economic modeling, Stone sought to escape from the straitjacket of quantitative history of his day and embrace narrative as the preferred method and subject matter of historical inquiry. Yet Shiller’s call for ‘a quantitative narrative economics’ and the return to narrative history largely taken up in the wake of Stone share an approach to social inquiry that privileges the individual, for Shiller, and the very local, micro-level in history following Stone. They share the idea that attention of historians and economists ought to be aimed at the micro-level, with particular focus on individual agency and contingency. For Stone and those coming after him, this amounted to giving up some of the larger macro-focus of historians on context to concentrate on the single, unusual, and microhistoric. Shiller his part, has promoted a probabalistic microfoundational study of discourse and narrative in society on the model of virus spread.

Computational methods of text analysis provide tools to further both agendas. At the same time, the assumptions inherent in these methods challenge the methodological individualism of narrative economics and embodied in much of narrative turn literature following Stones intervention. While use of computers to analyse texts stretches back to the immediate postwar period and now boasts more than twenty years of extensive development and application in the social sciences, it is still *terra incognita* in many disciplines, a technology whose promise remains firmly in the future tense (Blevins, 2016). These techniques have clear implications for both Shiller’s and Stone’s ideas of the kinds of economics and history that are desirable and are at the core of the classical historical question of explaining change over time.

This essay, therefore, has a triple purpose. First, it surveys new ideas about the study of narrative in economics compared and contrasted to Stone’s classical call for return to narrative, as well as more recent arguments for a return *from* narrative in historical scholarship. Second, it surveys the most important methods of computational text analysis, considering the epistemological and ontological assumptions contained within such approaches to text and suggesting how such methods allow for compensation for key limitations shared by both Shiller’s proposed narrative economics and history following the return to narrative. Computational text analysis can identify and help visualize patterns, trends, and emergent properties (a type of ‘macrofoundations’) in large-scale study of narrative and discourse. Finally, it overviews concrete applications of these tools across a broad range of social scientific and humanistic disciplines, suggesting ways they might allow they might be particularly enrich economic and business history.

2. Narrative in economics and history

Robert Shiller and coauthors have long been mindful of the role of expectations and ‘non-rational’ background for economic decision-making, famously dubbed ‘animal spirits’ by John Maynard Keynes.¹ Since the 1980s Shiller has written extensively on

¹Some historians of economic thought have debated how central the idea was to Keynes’ though and *The General Theory* in particular (Barens, 2011).

the confidence and narrative thought processes of business leaders, most famously culminating in his book *Irrational Exuberance* that discussed the nature of bubbles and overheated markets (Shiller, 2015). Ideas of bubbles, crashes, and the disputed rationality motivating such events has long been grist for economic historians perhaps most famously in Charles Kindleberger’s (1978) *Panics, manias, and crashes* that set out a Minskyian framework for understanding the cycle of hype, overheating, and crash in financial markets illustrated with numerous examples over several centuries of financial history. Key to much of this work was the question of rationality and how economic models might account for this seemingly irrational behavior or if such behavior had to be considered exogenous, though recent work from sociologists and historians in particular has sought to understand such events without casting judgement on their overall rationality (Dale, 2004; Deringer, 2018; Goldgar, 2008; Temin & Voth, 2004).²

In *Animal spirits*, a co-authored work with George Akerlof, Shiller more explicitly embraced an approach moving away from seeing rationality as driving economic behavior and toward a focus on non-rational, proposing that study of animal spirits be at the center of economic theory rather than treat them as exceptional (Akerlof & Shiller, 2010, 3). To this end the authors pointed to numerous factors of clear and obvious import to any participant in economic activity that had been studiously ignored by economic theory including fairness, corruption, and stories. As they argued, feelings, popular stories, and framings are excluded as ‘non-factual’ by economists, who focus on hard and quantifiable categories which can be used to populate models of constrained optimization. The argument was that stories, ideas, cultural understandings *do* move markets: ‘animal spirits drive the economy; not the other way around’ (Akerlof & Shiller, 2010, 54, 173). The intervention was thus as ontological and epistemological as theoretical – stories and understandings are factors with causal power and to ignore them is to leave vital elements outside the analysis. Similarly, the authors noted that just as their thematic attention and evidence differed from much economic theorizing, so too did their method – the book relied on history and stories and mostly stylized use of statistics with no heavy mathematical apparatus (Akerlof & Shiller, 2010, 168).

Both authors have continued work in this vein with Shiller’s *Narrative economics* 2020 the most recent example. In the vein of his previous work, Shiller argued in this book that ‘narratives’ have causal force in economic trends and events and should be incorporated into models. While he defined a narrative as a ‘contagious story with potential to change how people make economic decisions’ (Shiller, 2020, 3), it often seems to encompass something more expansive, including ‘constellations of meaning’ Shiller (2020, 29,86). Building on a number of historical instances of influential ‘narratives’ from Bitcoin, depression-era angst, job-stealing robots to stock and real estate bubbles (and drawing on a huge amount of material aided by a team of research assistants most historians could only dream of), Shiller repeatedly suggests and deductively motions toward ‘narratives’ either deepening or initiating economic events. While he makes an appeal for economics to venture outside disciplinary boundaries and learn from neighboring social sciences such as anthropology and history (though, surprisingly, not literary studies where narratives are perhaps most studied), the text goes back and forth on the extent to which a turn to narrative represents a new method or simply new evidence base for economics. The book itself is narrative in form, much like *Animal spirits* and the only quantitative evidence in the main text are simple

²Heterodox economists, following Minsky, have tended to give narrative greater pride of place (Dow, 2010; Tuckett, Holmes, Pearson, Chaplin, et al., 2020) as have scholars of central banking (Abolafia, 2010; Holmes, 2013) and economic sociologists, especially in the new literature on ‘fictional expectations’ (Beckert, 2016; Beckert & Bronk, 2018).

word frequency analyses. Thus, wrote Shiller, there can be ‘no exact science to understanding the impact of narratives on the economy’ though at the same time noting ‘we need to keep the true scientific method in mind even when trying to use an essentially humanistic approach’ (Shiller, 2020, 271-2).

Aside from word frequency counts the only math in the book comes in the appendix to the book that discusses briefly an epidemiological model that Shiller suggests might be used to study the rise and fall of contagious narratives in a population of individuals. The model proposes a situation wherein the rate at which the healthy population gets sick is determined by the contagiousness of the disease, modeled as the number of all possible pairs of infected and susceptible individuals multiplied by a constant parameter, capturing how easily the disease spreads. The rate at which the sick population gets better is modeled as a recovery coefficient times the sick population. Finally, the rate of change of the contagious population is simply the difference between the first two equations: the change in the healthy population becoming sick minus the rate at which the sick recover (and acquire immunity). No one dies in the model. Shiller argues that narratives can be understood to behave similarly. They are, thus, short term conditions, which every member of society is equally susceptible to, based on a fixed infectiousness parameter. Contagion is, meanwhile, based on probabilistic transmission through encounters of individuals. ‘Recovery,’ or no longer falling under the sway of a certain narrative, is also a probabilistic process modeled as a fixed parameter across a large population after which the person is no longer susceptible to said narrative.

Conceptualizing narratives as pandemics commits him to some assumptions. The fact that he sees narratives as contagious with fixed infection and recovery coefficients suggest the model might be useful for study of passing economic fads or short-lived trends. Indeed, using keyword search to look for and study these narratives, this is largely what Shiller uncovers. While he seems at times interested in deeper meaning- or world-making, the approach he outlines in both the mathematical model and evidence base only takes the passing and transient into account. Second, as Shiller is well aware, the issue of causation is implied but not conclusively shown (Shiller, 2017, 996). Finally, as the epidemiological model and anecdotal nature of the historical evidence make clear, Shiller’s approach is individualist and probabilistic. Becoming infected with a narrative occurs on the individual level as the result of stochastic processes. Thus, the spread of narrative frames is but the accumulated effect of individual infections. The model and historical method do not allow for emergent properties or varying contexts that might make narrative spread more or less likely or uneven across different subpopulations. Similarly, the historical method used in the book treats historical cases of narrative contagion disembedded from the institutions, laws, and other specifics of their time and place. We thus miss the mechanism by which ‘contagion’ spreads, namely why people in certain places and times come to find certain narratives so appealing or convincing.

In subsequent work Aklerof has also turned to narrative and ‘frames’ of meaning. In joint work with Edward Snower, the two economists approach cognitive categories at a somewhat higher level of sophistication than Shiller, connecting literatures in cognitive science, psychology, and neuroscience to argue that cognitive categories guiding human perception are not based on objective reality but subjective experiences and impressions that are subject to change. Many of these categories, they note, are stable over long time periods. Thus, Aklerof and Snower motivate their attention to narrative as something that is longer lasting than a fad but is constantly revised, renegotiated, and mobilized for economic decision-making. They thus call for these characteristics

to be accounted for with economic models not simply treated as exogenous shocks (Akerlof & Snower, 2016).

In an interesting substantive exploration into the power of narrative to provide an alternate guide to decision-making, Aklerof and Snower argue that examination of changing narratives in early Soviet history is required to account for the behavior of historical actors. The account makes considerable room for institutions and particularities of historical time and place. Furthermore, the authors make clear that narratives are entwined in and can be utilized in power relations. Narratives, in this conceptualization, also work at macro-levels by remaking and coordinating social environments and contexts in which people operate. The result is a treatment of Soviet history that makes an attempt to show cause by qualitatively approaching events and isolating those that could not have happened without a narrative guiding action (Akerlof & Snower, 2016).

While Aklerof and Snower argue that this is a topic absent in modern mainstream economics, it was fundamental to classical American institutionalism, which emphasized interactions between people – rather than between people and goods – as the core of economic activity (Commons, 1931; Hodgson, 2006). Narratives thus might be a way to study the spreading and establishment of norms that institutions depend on and explain actions and behaviors that are not rational or utility-maximizing. By moving preference up from the exclusively individual level to a macro-social level, attention to narrative might enable the researcher to follow changes in norms and institutions. In a perceptive piece on classical institutionalism, Vyacheslav Volchik has argued that narrative economics complements institutionalism given that narratives ‘structure interrelations between economic actors’ thus offering important insights into the study of preference, long-term persistent effects (path dependency), entrepreneurship and innovation, as well as help to guide public policy in building sustainable and open institutions (Vol’chik, 2017).

Some forty years earlier came another call for a shift to narrative in the discipline of history. In a now-classic article, Lawrence Stone famously stated that he detected a turn away from structuralist and frequently quantitative analysis in history back to description, focusing on people and contingency rather than structures or social laws or generalizations. A reaction to the materialist economic history of Marxist historians, the quantitative structuralism of the Annales schools, and mathematical deductivism of cliometrics, Stone argued that history now appeared to be returning to telling stories of individuals and events, which had long been downplayed in Braudel’s famous turn-of-phrase as merely ‘crests of foam that the tides of history carry on their strong backs’ (Braudel, 1972, 21). This new kind of history would focus on getting into historical actors’ heads and understanding how people understood their world (Stone, 1979).

Stone’s argument is in some ways reminiscent of Shiller regarding the push to focus on thinking, self-perception, and framing as causal to history or the economy. Stone’s intervention, however, drove in the opposite direction. In his telling, which was professed to be descriptive but was implicitly deeply normative (Hobsbawm, 1980), rather than adopt narrative into a form to be subsumed into the quantitative methodological apparatus of the discipline, Stone sought to make the discipline’s methods more like the evidence whose inclusion he desired. History, he wrote, should make its case by taking narrative and cognitive evidence and telling these stories. Stone seemed to revel in the excitement of new, narrative, non-quantitative and largely non-causal history inspired by theoretical works of the likes of Clifford Geertz, Mary Douglas, and Norbert Elias and embodied in the histories of Carlo Ginsberg, Peter Brown, and Keith

Thomas. ‘If I am right in my diagnosis,’ he wrote – and whatever criticisms one may have of the article there can be little doubt that he was quite right – ‘the movement to narrative by the ‘new historians’ marks the end of an attempt to produce a coherent scientific explanation of change in past’ (Stone, 1979, 19). If Shiller suggests that economists should focus on results and causal outcomes of narrative first and foremost for prediction and downplays the mechanism of how and why certain narratives spread, Stone advocated the precise opposite: mechanism before outcome.

Like many now-classical programmatic statements, Stone’s article is less polemical than generally remembered. While deeply critical of the mathematical and computational methods that had been widespread in the historical profession in the 1970s, he underlined the limits of his critique. Historians should not abandon numbers nor stop counting altogether. The demand for precise numbers in place of hand-waving words such as ‘more’, ‘less’, and “generally” and the requirement that individual cases be shown to be representative based on statistical evidence had ‘undoubtedly improved the logical power and persuasiveness of historical argument’ (Stone, 1979, 10). Stories must not only be interesting; they should also be ‘typical’ (Stone, 1979, 22). It was the opaque methods of cliometrics which not only required mathematical fluency but also necessitated the use of computers not easily available to poorly-funded researchers to perform (or verify) that Stone expressed frustration with.

Stone’s essay has widely been accredited as marking of the beginning of the linguistic or cultural turn in English-speaking history departments.³ There is a sense, however, in which Shiller’s project and what became of Stone’s are strangely parallel. In both accounts, scholars are pushed to focus on the individual or the small group, in epidemiological models or at the local, frequently microhistorical archival research level that became the hallmark of deeply-interpretive, cultural history from the 1980s on. Individual humans and thought processes in individual human minds are emphasized before more macro-level societal processes. To be sure, Stone spoke of culture, understandings, and the then-trendy ‘mentalit.’ But in a discipline that was famously abandoning master-narratives at exactly the same time, focus came to rest on individual agency and contingency. Stone further advocated changing of historians’ focal point ‘from the circumstances surrounding man [sic], to man in circumstances; ... in the prime sources of influence, from sociology, economics and demography to anthropology and psychology; in the subject-matter, from the group to the individual... in the methodology, from group quantification to individual example’ (Stone, 1979, 23-24). Indeed, in the aftermath of Stone’s article, microhistories of the individual, often those on the margins of society, became one of the most celebrated genres of the ‘new cultural history.’ As economic historian Jan de Vries has described it, the return to narrative was a “return” to a different sort of narrative than Stone could have suspected – a petite narrative, immersed in a sauce of thick description and shaped by a postmodern philosophical posture that elevated subjective experience and the identity of the historical subject’ (de Vries, 2018, 321). With ‘many histories but no history,’ the discipline became more and more local and individually-focused, macro- and meso-levels were obscured and, with it, any chance of seeing generative emergent qualities of narrative at a higher level (de Vries, 2013, 2018; Sewell Jr, 2005).

Influential voices have begun pushing back toward more structure and less contingency in history. William Sewell, himself a major participant in the linguistic turn, has urged historians to integrate more generalizable and theoretically-informed notions

³For reasons of space this article cannot discuss the related history of the split, especially in US-based scholarship, between cliometrics and more culturally-focused economic and business history. See the excellent overview in N. R. Lamoreaux (2016).

of agency and structure, specifically encouraging historians to tell ‘eventful histories’ that integrate events and contingency with structures. Sewell has thus argued for work that uses the tools of the cultural turn but ‘abandon[s] the current taboo in historical thought on arguments that might be construed as economic determinism’ (Sewell Jr, 2010, 166). To do so, historians must look at new ways of asking questions and new topics of research, but also pay attention to neighboring disciplines of economic sociology, political economy, and [quantitative] economic history. The latter, he writes, ‘now offer[s] a wide range of insights into, and methods for studying, the history of economic life more broadly’ (Sewell Jr, 2010, 170).

De Vries proposed what he calls ‘nanohistory’: history based on granular archival research that leads to ‘patterns, trends and regularities,’ while pushing historians focusing on the micro-level to place their histories in the context of larger theories or frameworks (de Vries, 2019, 34). This sort of ‘playing with scales,’ as de Vries termed it, is something that becomes possible, indeed is often baked into the method, of computational text analysis based on large corpora that can easily encompass many authors, over long periods of time, and across significant geographical or social distance.⁴ In a related argument, Naomi Lamoreaux has pushed for a move to the center between the disciplines of economics and history – use of formal modeling and hypothesis testing informed by deep historical cultural and institutional knowledge (N. Lamoreaux, 2015). Here too, both approaches are inherently implicated in many computational text methods. Exactly how and in what ways this is so are the subject of the following section.

3. Tools of computational text analysis

3.1. *Computers and narrative*

One of the things that now jumps out from Stone’s text is his characterization of computer technology as expensive, demanding to use, opaque, and featuring data storage that was extremely difficult to share or obtain. The revolution in personal computers in the ensuing years has turned the situation on its head. Not just is the hardware now almost trivially accessible, but algorithms have been written and text corpora digitized at a rapid pace over the recent decades. The very capacity of computers to analyze huge amounts of text pushes the approaches trumpeted by Shiller and Stone precisely toward the nanohistorical approach that can examine more macro-level structures and trends in narrative making. Using what one group of historians has called a ‘macro-scope’ (Graham, Milligan, & Weingart, in press), computational text analysis is not just an exciting new tool offering new evidence bases, though it is certainly this as well. Much more, it contains methodological and ontological assumptions and implications that fruitfully challenge how economic and business historians might think about and analyze narratives and cognitive framings.

The following discussion of digital methods is high-level. It concentrates on how computational techniques represent and *transform* text documents rather than how calculations and optimization algorithms function (Schmidt, 2016). A small degree of mathematical formalism is introduced with the aim of supplementing the discussion and elucidating such transformations. This is not a how-to kit in text analysis and,

⁴Indeed, in a recent survey of ‘big data’ in economic history, Gutmann, Merchant, and Roberts (2018) treat quantitative text analysis as one of several avenues in which economic historians will increasingly be engaging with huge quantities of data enabling new sorts of analysis and requiring new tools.

thus, issues of implementation will not be directly discussed. While outside the scope of this article, brief overview of these issues and citations to useful tutorials and textbooks that can serve as jumping-off points for study of the technical details of computational text analysis is provided in the penultimate section.

Inter-disciplinarity is one of the most interesting and perhaps attractive aspects of computational text analysis, yet it also presents challenges. Fundamentally, the goals and approaches of across the social sciences and humanities, not to mention computer science and information retrieval, often differ significantly in basic and important ways. Many economists approach research questions motivated by the quest to predict or infer causes. This can be at odds with inferential and ideographic approaches. Computer scientists and computational linguists often seek to improve computational search algorithms or typing completion. They thus also have little explicit desire to ‘understand’ a corpus but are rather interested in designing ways to visualize its structure, find and extract information, and predict out of sample observations. These differences represent both potential pitfalls and productive frictions.

3.2. *Dictionary methods*

No doubt the simplest method of quantitative text analysis is to simply count words. As evidenced even by such a methodologically sophisticated economist as Shiller, looking at changes in word frequency over time is a powerful tool for both visualizing and understanding the structure of linguistic change. It often is the case that simple word frequencies are the extent of computational analysis necessary for an analysis, as in the recent example of Levingston (2020) who uses word counts over time to show changing topics of interest on the US Federal Reserve Open Market Committee as a supplement to predominantly qualitative intellectual history.

One of the major uses of computational text analysis has been to classify texts by comparing the vocabulary of a given text to a pre-compiled list of words. This method, based on simple word counts and involving no statistical inference at all, requires the researcher to define a list of words (a dictionary) for a category she wishes to map across texts and a function that transforms individual occurrences into a number. One common implementation is for negative and positive sentiment detection. The researcher begins with two word lists (dictionaries): one of words associated with positive sentiment and the other of negatively charged words. These lists can be defined by researchers themselves or borrowed from existing compilations.⁵ A computer then counts how many times words in the dictionaries appear in a text. At its simplest, the researcher can simply add up the total number of positive words and subtract the total number of negative words, perhaps divided by number of words to account for texts of differing lengths. This yields an index number for each text, positive indicating presence of more positive words than negative and vice versa.⁶

A recent simple example is Baker, Bloom, and Davis (2016) who created an index of economic policy uncertainty by looking at the frequency at which articles containing certain words appeared in the US’s ten leading newspapers. In another, Moreno and Caminero (2020) sought to locate and identify which publications banks use to report on climate change-related issues, arguing that dictionary-methods have a distinct advantage over statistical and machine learning models in their ease of interpretation

⁵One danger of working with pre-defined dictionaries are obviously that words can have different meanings in different contexts. In finance, for instance, ‘liability’ has connotations that differ significantly from general speech (Loughran & McDonald, 2011).

⁶For an overview, see Jurafsky and Martin (2014, ch. 19).

and the possibilities they offer for leveraging of the researcher’s domain knowledge. Other, more complex methods have involved implementing dozens of dictionaries for multiple sentiments and using a mathematical technique for dimension reduction to come up with a single index for sentiment (Tetlock, 2007). In a recent overview of text analysis methods in economics, Gentzkow, Kelly, and Taddy (2019) find this to be ‘by far the most common method in the social science literature using text to date.

An example from the discipline of history is that of Cameron Blevins who used a randomized method to select sample articles from the archive of a nineteenth-century Texas newspaper and then manually counted frequency of geographical places outside of Texas in order to estimate the amount of page space dedicated to different geographical categories over time and better understand spatial representation (Blevins, 2014). Alsvik and Munch-Møller (2020) presented both raw frequency counts of words related to representation of women in Norwegian historiography, as well as dictionaries to measure the degree of attention historians have paid to the ‘private’ and ‘public sphere’. The authors found a notable increase in references to women from the 1970s to the 1990s together with a increase in frequency of words associated with the private sphere, noting also that these trends are more pronounced in women historians than men. Klingenstein, Hitchcock, and DeDeo (2014) also used a dictionary method (in combination with more complex analytics taken from information theory) based on a historical dictionary (*Roget’s Thesaurus*) that provides word lists (i.e., dictionaries) for over one thousand categories which themselves were then organized in the original historical dictionary into some one hundred broader categories. The authors then proceeded to divide accounts of trials based on indictment into two categories – violent and non-violent – examining which word categories were mostly highly associated with the two different types of crime and how these changed over time.

Other possibilities of dictionaries include the ability to locate not just certain topics or sentiments but the style of discourse. For instance, counting up use of modal verbs might give broad indications of speaking styles. Fig. 1 graphs the average number of modal verbs per 1,000 words spoken by party in sessions of the Norwegian Stortinget from 1998–2018 with 95% confidence interval error bars. Several interesting patterns come out here – the modal verb ‘ville’ (want, (impersonal) will) is generally most often used, but members of the government use ‘skulle’ (should, (personal) will) more frequently and members of Senterpartiet and Sosialistisk venstreparti use them roughly equally. The Green party uses the ‘kunne’ (can) unusually infrequently and ‘ville’ more often (though subject to a large error interval due to relative infrequency of statements). A starting point rather than ending, a closer look at patterns such as this could uncover different ways people or parties formulate their statements, with changes perhaps occurring over time or subject matter.

Furthermore, the use of minutes or verbatim transcripts from official proceedings of state often offers the possibility of analysis of verbal but non-language data such as cheers, shouts of dissent, interruptions or laughing, all often scrupulously recorded. One could well imagine analysis looking at which topics drew most reactions of what sort, how certain types of non-language reactions changes over time, or if certain groups are inclined to use certain non-language means to interact. For example, FOMC transcripts of the US Federal Reserve Bank record instances of laughter of participants, thus the analyst can easily extract both who brings committee members to laughter and patterns in laughter over time. As Figure 2 shows in a very basic count, laughter seems to have increased significantly over the years in FOMC meetings, peaking right

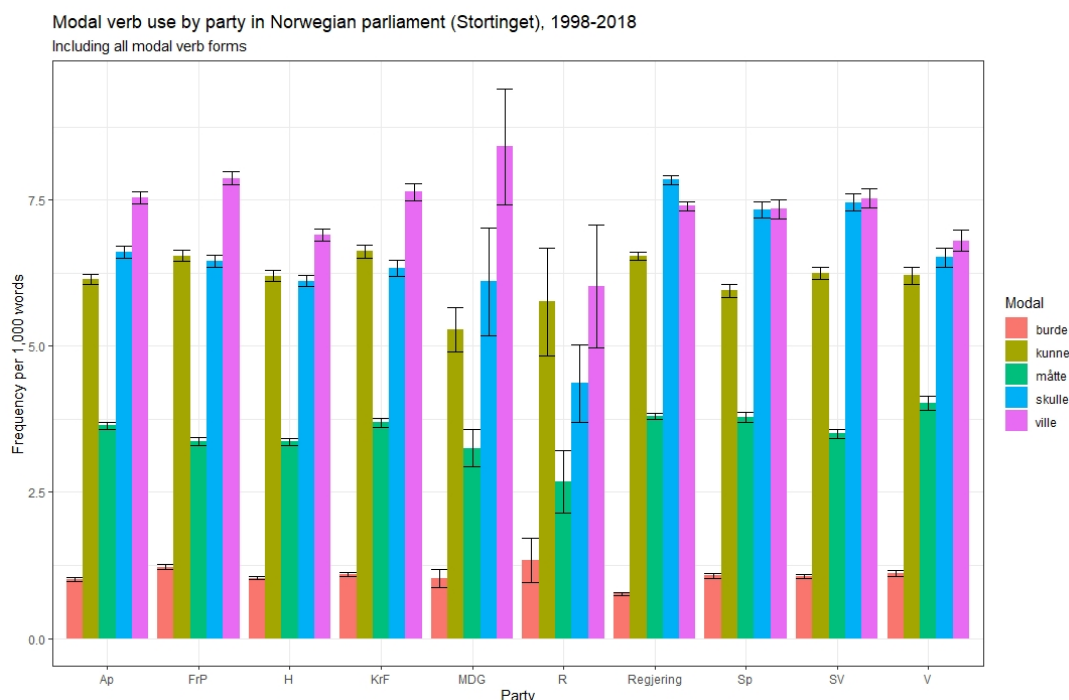


Figure 1. Modal verb use by party in the Norwegian parliament.

before and during the global financial crisis.⁷ Men on the FOMC, who dominate also in terms of words spoken, drew significantly more laughs per word than women, though in 2014 – perhaps not coincidentally the first year under a female Fed chair – the counts near equality.⁸ This sort of non-linguistic analysis can blend humanistic approaches to power, decision-making, and group dynamics with technical discussions of the Fed’s role in the modern economy.

While simple and, frequently, unsatisfactory as a sole source of evidence, dictionary methods along with word and phrase frequency counts provide output that is exceptionally easy to interpret and can be easily run over huge corpora of text. As such they can easily be used in tandem with close reading, giving researchers clues of where and what to read, and giving them evidence of representativeness. Indeed, in his famous article Stone expressed the hope that a move to narrative history would not mean that historians retreated for quantitative summaries of their evidence to amorphous and inexact generalizations like ‘greater’, ‘more’, ‘less’, and so on (Stone, 1979, 10). He would have been disappointed. Dictionary methods are an easy way for qualitative researchers to give precisely this sort of quantitative background to close readings. These data points have the advantage of being ‘measurable and verifiable’ (Alsvik & Munch-Møller, 2020, 214). They also enable iterative methods of visualization of words over large corpora that can direct the researcher’s focus for close reading, which in turn can lead to refinement of dictionary items subsequently spurring close reading in other places and so on in a process some scholars have called ‘blended reading’ or ‘computational grounded theory’ (Nelson, 2020; Stulpe & Lemke, 2016).

Literary scholars have also embraced the idea that ‘representativeness’ must be

⁷Such analysis, of course, assumes consistency in recording laughter in transcripts over the years.

⁸Unsurprisingly, other counts also show that Fed chairs tend to be funnier than non-chairs.

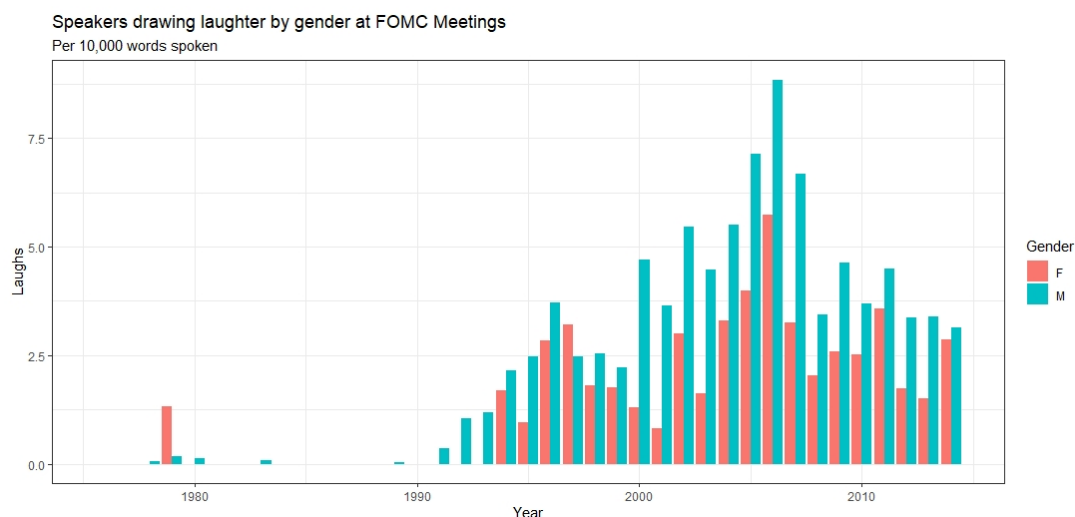


Figure 2. Instances laughter in FOMC meetings from 1976-2014.

proven rather than assumed. Franco Moretti has famously called for reading of the ‘great unread,’ the 99.5% of novels from human history that do not become part of the canon and are almost entirely beyond the reach of human readers and literary scholars (Moretti, 2013, ch. 3). Not just does this broaden horizons to consider a much wider swath of literature as produced and experienced than a narrow and *post facto* determined canon but, as Matthew Jockers has argued, *not* doing this dooms literary historians to cherry-picking of evidence to support pre-defined hypotheses, which ‘seems inevitable in the close-reading scholarly tradition’ (Jockers, 2013, 47). Both Jockers and Moretti give examples of plotting and following frequencies of individual words or groups of words over long periods of time in literary history. Jockers, for example, measures use of words such as ‘the’ or ‘beautiful’ for insight into how British and American use of such words has diverged or converged over time (Jockers, 2013, 106). Other approaches include tracking use of words over the course of novels to look at how style and plot related to certain words have changed. Moretti considered word length and common words appearing in titles of eighteenth-century novels connecting his findings to historical economies of book printing, publication and review (Moretti, 2013, 180-1). Moretti has argued forcefully that computational analysis provides different tools to answer the same questions and conduct the same analysis as literary scholars have always done. ‘This is a quantitative study: but its units are linguistic and rhetorical. And the reason is simple: for me, formal analysis is the great accomplishment of literary study, and is therefore also what any new approach - quantitative, digital, evolutionary, whatever - must prove itself against: prove that it can do formal analysis, better than we already do. Or at least: equally well, in a different key. Otherwise, what is the point?’ (Moretti, 2013, 204).

In a way, most if not all historians already engage in dictionary methods when they do keyword searches. As Laura Putnam has persuasively argued, the digital turn to being able to search online collections of historical periodicals, printed sources, archival catalogues, and, sometimes, even the archives themselves, fundamentally change the ways historians investigate their sources and the kinds of claims they can make. Doing deep reading directed by digital keyword search decontextualizes research, argued Putnam. (Putnam, 2016) ‘Discovery via algorithm offers instant reward. In doing so,

it deprives you of experiential awareness of just how rare mentions of your term were, of how other issues crowded your topic out in debates of the day. It erases the kind of *sitzfleisch*-based test of statistical significance on which our discipline has implicitly relied’ (Putnam, 2016, 392, emphasis in the original). Thus, as Putnam and Tim Hitchcock have argued, historians are already using digital methods, they are simply doing so haphazardly and passively. It is ‘not only about being explicit about our use of keyword searching – it is about moving beyond a traditional form of scholarship to data modelling and to what Franco Moretti calls “distant reading”’ (Hitchcock, 2013, 19).

3.3. Document similarity

Another area of comparatively simple computational analysis is that of document comparison. There are numerous ways to measure the similarity of documents. One recent example comes from nineteenth-century American legal history, where Funk and Mullen (2018) study adoption of civil procedure codes by new states in the American West. New York’s code was known to have been widely copied in states across the expanding union but the process had never been explicitly examined by historians. The authors divided a corpus of state codices into overlapping five-word chunks and located all identical chunks found across multiple state codes and calculated ratios of identical to non-identical text between states. In this way the authors could track borrowing of text between states and map which states borrowed what text from which other states in a process they dubbed ‘algorithmic close reading’ (Funk & Mullen, 2018, 154). One can imagine using similar methods in other legal settings such as corporate charters, economic regulations, and other areas of interest to the economic and business historian.

Another more general and commonly-used method is to measure the total similarity of two texts by comparing their vocabularies. A text document is reduced to a list of numbers, called a vector. To do this, a computer first reads all the documents in the corpus and creates a list of all unique words in the corpus. Each document is then represented as a list of numbers corresponding to how many occurrences of each unique word it features. The vector will have as many entries as there are unique words in the corpus.

Such raw frequency counts are often transformed into term frequency-inverse document frequency (TF-IDF) scores. TF-IDF indices are sometimes interesting in their own right as they are designed to capture both the frequency of a term in a certain document and, at the same time, its overall unusualness across the corpus. Term frequency (TF) is the frequency of the term in the given document. Inverse document frequency (IDF) represents how frequent a word is across the entire corpus and is calculated as the total number of documents in the corpus divided by number of documents containing the given word. TF-IDF multiplies TF and IDF together giving an index that reflects both commonness in a given document and uncommonness across a corpus as a whole.

Cosine similarity utilizes TF-IDF by measuring how similar document word vectors are to each other, where a document is represented as a list of TF-IDF word frequencies. Basically stated, cosine distance measures the angle separating two text vectors. If two texts have no common words, cosine distance will be zero. If the documents have exactly the same number of the same words (or some multiple thereof) cosine distance

will be one.⁹

While this technique makes a number of assumptions, including disregarding word order, and considers similarity only in vocabulary, it has the potential to give useful information. For example, Figure 3 shows two cosine similarity matrices comparing total vocabulary used in statements in the German Bundestag over two periods between 1970 and 2017. The cosine distances correspond with our *a priori* assumptions of what should be expected, with some interesting features. Statements in the first period seemed to have more in common with one another with the exception of the Greens, who appear in cosine similarity like the outsiders they were. In the subsequent three decades similarity between statements of the SPD, CDU and FDP significantly decreased. Clearly this is not the end of analysis but could lead a researcher to explore further questions quantitatively (such as when the greatest differences occurred or which words and speakers contribute most to divergences) as well as qualitatively through close-reading. While not frequently used in text analysis, this is a method that is fairly simple and requires no statistical inference or black box algorithms. One can easily imagine applications for the business and economic historian comparing statements of politicians across party or across time, state planning documents, corporate reports, or enabling comparisons between, for instance government regulations and publications by corporate lobbyists.¹⁰

3.4. Topic modeling

Topic modeling, while a decidedly old method in the machine learning community, has in recent years become an increasingly widespread and popular means of analyzing text in the social sciences and humanities. Topic models belong to class of model called ‘generative models’ that are based on an assumption of how output is generated and then work backwards to find the parameters of the model that is most likely to have generated the observed data. Topic models assume a data generating process – i.e. the writing of a text – as follows. A writer is taken to have a probability distribution of topics she will write about in a specific document. She proceeds to write the document by, at every word occurrence, selecting a topic randomly from the document topic distribution and then selecting a word from within that topic according to a separate probability distribution giving probabilities of individual words occurring in the given topic. Every document can, therefore, be said to be made up of a ‘mixture’ of topics and topic models are, thus, also referred to as mixed-membership models. A topic model algorithm, therefore, estimates a number of parameters, two of which are of particular interest to the researcher: the topics present in the corpus (in estimated percentage terms for each individual document) and the words (and their probabilities) that make up these topics. We can call these ‘topic prevalence’ and ‘topical content,’ respectively (Roberts, Stewart, & Airolidi, 2016). The topic model algorithm observes only the text, thus topic prevalence and topical content are unobserved (or ‘latent’) structures that the model takes to exist and then seeks to estimate by, at its simplest, taking the text, looking at the structural assumptions just described and asking what parameters of the assumed data generating process are most likely to have produced the given corpus.

⁹For a simple introduction to and helpful visualizations of representing texts as vectors and similarity measurements, see Jurafsky and Martin (2014, ch. 6).

¹⁰See Jockers (2013) for discussion of use of cosine similarity in a literary studies context.

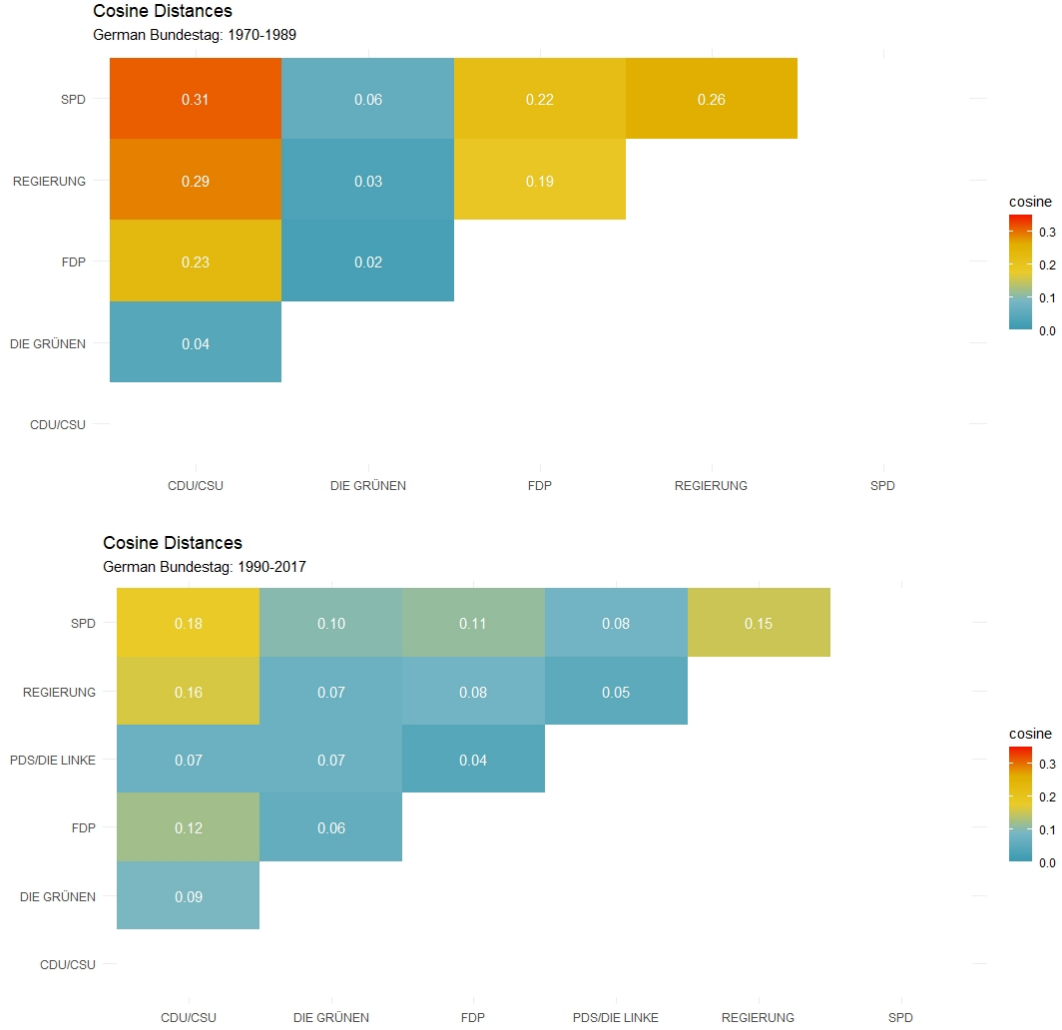


Figure 3. Comparison of cosine similarity by party.

Formally, the generative process takes place in four steps.¹¹ First,

$$\vec{\beta}_k \sim \text{Dir}_V(\eta) \quad (1)$$

denotes the probability distribution of the topical content for each topic. It is the probability, given for each topic, that the writer picks a certain word to express that topic. The vector beta ($\vec{\beta}$) is a list of probabilities, one for every unique word in the corpus (denoted V) that a given word will be selected in the given topic. Every topic (denoted k) has its own beta vector of word probabilities. In the original and many subsequent topic model algorithms, these vectors are assumed to be drawn from a Dirichlet distribution, thus leading topic models to be frequently called ‘latent Dirichlet allocation’ (LDA), though today some leading topic models (such as the structural topic model discussed below) do not use Dirichlet distributions.

¹¹This section hews most closely to the explanation in Blei and Lafferty (2009). For other helpful high-level overviews see Blei (2012a, 2012b). A technical description of the original algorithm can be found in Blei, Ng, and Jordan (2003).

Topic prevalence over documents in a corpus is denoted as

$$\vec{\theta}_d \sim \text{Dir}_k(\vec{\alpha}) \quad (2)$$

whereby topics are assumed to be randomly selected by the writer. Theta ($\vec{\theta}$) is a vector, or list, of probabilities that every topic (denoted by letter k) has of appearing in a given document, referred to above as topic prevalence. Each document (denoted d) has a theta vector associated with it. These are the two distributions, β and θ , we are interested in estimating.

To do so, however, the model needs to connect them to the text itself, the observed structure. This is the role of the following two structural assumptions. First,

$$Z_{d,n} \sim \text{Mult}(\vec{\theta}), Z_{d,n} \in \{1, \dots, K\} \quad (3)$$

says that for each individual word (n) in each document (d), the writer (or text generator) decides what topic (k) the word will have based on $\vec{\theta}$, which again is the probability of each individual topic appearing within a given document.

Next, given the topic a given word instance is associated with, the exact word is chosen, based on $\vec{\beta}$, the list of probabilities of each word appearing in the selected topic. Thus,

$$W_{d,n} \sim \text{Mult}(\vec{\beta}_{Z_{d,n}}), W_{d,n} \in \{1, \dots, V\}. \quad (4)$$

defines how the exact word will be chosen from one of the total unique words in the corpus (V) based on the distribution of topic content β in the topic selected in previous step.

This is, to put it mildly, not how human beings speak or write texts. The fact that topic models, for better or for worse, do not parse text through anything resembling human cognition is something that must always be kept in mind. However, as the great statistician George Box famously quipped, ‘all models are wrong, but some are useful’ (cited in Graham et al., in press, ‘Topic modelling by hand’). There have been a wide range of situations in which topic models have been found to be quite useful.¹² Furthermore, as this mapping of the topic model architecture has shown assumptions are based on both individual document-level topic prevalence but also corpus-level topic content, thus connecting micro and macro levels.

An example is shown in Figure 4, showing one way to visualize a topic model – in this case one fit on all statements made in the Russian Duma from 1995 to 2018 containing the word electricity. This means of visualization – which graphs the topic prevalence ($\vec{\theta}$) for selected topics (which have been named by hand) by year – shows declining prevalence over time for some issues (electricity disconnections, a serious problem in the 1990s), several waves of debate regarding the fate of the post-Soviet electricity monopoly RAO EES, peaks of topic prevalence surrounding regulation – and liberalization – of the electricity market in the mid 2000s and more recent increase in electricity-heat coupling and markets and payment for electricity. This both confirms basic expectations for what sorts of issues Russian legislators would have been discussing at different time periods and poses a number of questions that might warrant closer examination. Again, this would not be the end of an analysis but

¹²In certain situations, however, topic models do seem to perform poorly, such as the case of poetry or non-standard language (Binder, 2016; Rhody, 2012).

a possible beginning or, alternately, a means to visualize and quantify topics identified and analyzed through close reading.

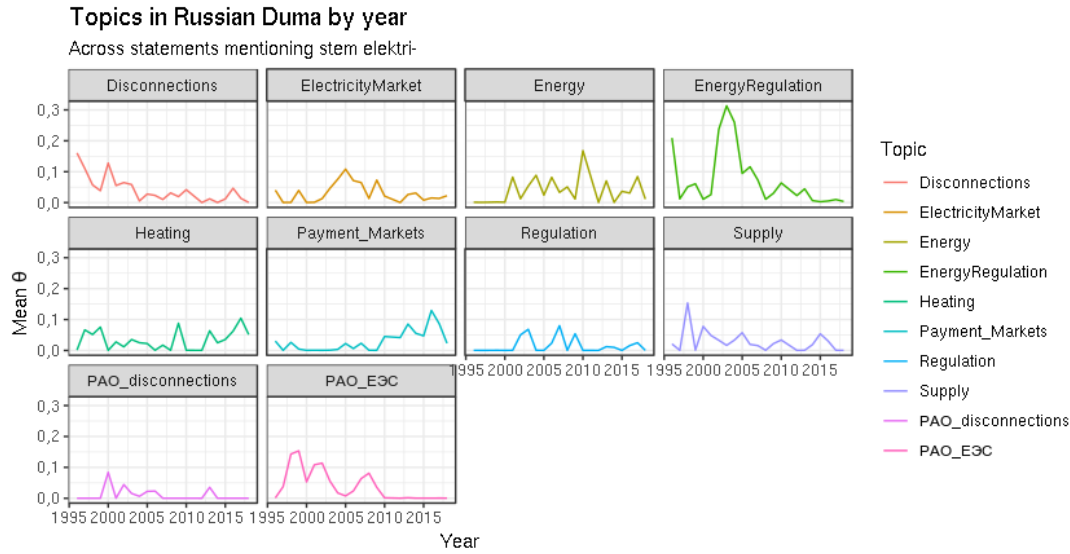


Figure 4. Topic models of statements in the Russian Duma mentioning the word stem electricity.

A number of variations on this original model and estimation algorithm have been offered since LDA was originally proposed. The original authors designed a ‘correlated’ topic model which takes into account the obvious fact that some topics are correlated with others. This team also created a ‘dynamic’ topic model that allows for content of topics to change across time, or possibly any other variable (Blei & Lafferty, 2006). Related to the correlated topic model is the structural topic model, which allows for incorporation of metadata or document covariables which are allowed to play a role in estimating topics, based on the intuitive notion that co-variables – such as author or date – are likely to play a role in deciding the distribution of topics. Each document is thus modeled to have its own prior distribution over topics and topic distribution over words that is related to its covariates (Roberts et al., 2016; Roberts, Stewart, & Tingley, 2019).

Hidden in the technical details are some consequential questions both for social scientific epistemology and method. One frequently discussed issue is the fact that the topic model has no way to figure out how many topics it should look for in a corpus. This must be supplied by the researcher and is, generally, seen as something that should involve domain expertise. For social scientists of a more positivist bent this presents an issue – there might be said to be one ‘true’ number of topics in a document, which was responsible for the data generating process. Indeed, optimization algorithms exist to automate setting of the number of topics (Mimno & Lee, 2014). For some in the more quantitative social sciences, one of the strengths of topic modeling is that it eliminates human intervention from the algorithm, providing a method that is completely automated and, thus, completely ‘objective’ and easily replicated (Grajzl & Murrell, 2019; Soto, 2019; Wehrheim, 2019).¹³

¹³This ideal of objectivity as an approach to understanding and representing nature that most fully removes human intervention bears resemblance to Daston and Gallison’s description of ‘mechanical objectivity,’ whereas the humanist and social scientist approach might be likened to ‘trained judgement’, with the expansion and caveat that interpretive study of text often admits of multiple possible objective realities (Daston & Galison,

For humanists or more interpretively-inclined social scientists, however, the data generating process – writing and speaking – is more multivalent and polyphonic. As such, it is not just reasonable but also perhaps desirable to view topic models as not underlying any ‘true’ hidden structure but as helping the researcher find and visualize relationships and patterns that might not be visible with the human naked eye (Gentzkow, Kelly, & Taddy, 2019, 549). Similarly, Roberts et al. (2016) and Newman and Block (2006) argued that machine learning discipline’s desire for optimization and prediction is not shared by many humanists and social scientists and, indeed, frequently at odds with it. Formally ‘sub-optimal’ solutions often yield better results from topic models than those whose objective function has been strictly and globally optimized. From this standpoint, the fact that topic models allow for some hand-tuning is a feature rather than a bug.

The use of topic models has skyrocketed in the recent years. All told, one might summarize that topic models are put to two main uses (though these are by no means exhaustive nor mutually exclusive): classification to serve as independent or dependent variables in regression or other causal inference analysis; and topic models that serve to summarize and visualize large corpora.

In one interesting example of the former, with parallels to Shiller’s project, Cerchiello and Nicola (2018) take text from financial news media over several years to map the frequency of certain topics over time and across different countries. With topic models fit, they proceed to calculate Granger (predicted) causality – the relationship between lagged values of independent variables on dependent variables – to argue that prevalence of a topic in one country caused rises in prevalence in others. At the same time, the authors note potential drawbacks, including the fact that the topic model found topics for some events (the Madoff ponzi scheme or LIBOR manipulation scandals) before the scandals had actually occurred.¹⁴ Other examples include studies examining the relationship between topics of trade bills and likelihood of interest group lobbying; the relation between topics discussed by representatives of parliament and the personal background of the speaker; and the effect of new transparency rules on the topic distribution at meetings of the Federal Open Market Committee of the US Federal Reserve Bank (Geese, 2020; Hansen, McMahon, & Prat, 2018; Kim, 2017). In all these cases, topic models yielded a quantified topic variable that could then be used in econometric analysis.¹⁵

Outside of economics scholars are more likely to use topic models as a way exploring and visualizing corpora, as well as documenting change. One early treatment of topic models in the more interpretive social sciences came in a 2013 special edition of the journal *Poetics* that discussed topic modeling as a tool in historical sociology. Ian Miller demonstrated topic models on a corpus of primary documents from Qing China, focusing on four topics associated with different sorts of banditry and social unrest in the eighteenth and nineteenth centuries. He interpreted this as a type of historical ‘crime rate’ from a time and place where there are few other proxies for such information. He also argued that topic models can be used to explore how crime and banditry were understood in the Qing Empire and mapped the waxing and waning of different

2007).

¹⁴For a thoughtful inquiry into ways the assumptions of topic models can misrepresent topics in a humanistic context, see Schmidt (2012).

¹⁵Using computers to automatically classify text are an abiding concern of many social scientists, including many economic historians, and by no means are topic models the most common way of doing so (Gutmann et al., 2018, 283–285). Other widely used methods include text regression (Dittmar & Seabold, 2019; Gentzkow & Shapiro, 2010; Gentzkow, Shapiro, & Taddy, 2019), wordscores (Benoit & Laver, 2007), and wordfish (Slapin & Proksch, 2008), as well as semantic embedding to be discussed below.

conceptions of crime over two centuries (Miller, 2013). Another article analyzed shifts in language from press coverage of arts funding in the US. The authors argued that topic modelling is particularly well adapted to sociological ‘frames,’ which they defined as the ‘semantic contexts that prime particular associations and interpretations of a phenomenon in a reader’ (DiMaggio, Nag, & Blei, 2013, 578).¹⁶ Another study used topic models to identify cognitive frames on the FOMC around the period of the financial crisis. Combined with qualitative close readings over a smaller sample of FOMC meeting minutes, the authors argued that the members of the FOMC predominantly understood the events of the early financial crisis in a ‘macroeconomic’ frame (rather than a ‘finance and banking’ frame) that led to certain interpretation (and ultimately misinterpretation) of events leading up to the financial crisis. This caused the Fed to downplay or explain away certain types of evidence and indicators that might have provided advanced warning signs of impending danger (Fligstein, Stuart Brundage, & Schultz, 2017).

In a recent article discussing ways in which communication and media studies can build on the idea of narrative economics as expressed by Shiller, Aklerof and Snower, Müller, von Nordheim, Boczek, Koppers, and Rahnenführer (2018) argued that LDA offers means for media scholars to interact with and contribute to theories of macroeconomic expectations building. They, too, find close correspondence between the idea of a topic identified by a topic model with ‘latent organizing principles active in the background,’ which give ‘coherence and meaning to symbols,’ in other words: a sociological frame (Müller et al., 2018, 567).

A somewhat similar use comes from one of the few studies to date in economic history to undertake topic modeling. Wehrheim (2019) estimated a model fit on text of articles from the *Journal of Economic History* from 1941 to 2016. He found the model accounts well for known and expected structure of the corpus and changes over time, and is especially suited to quantification and visualization of the ‘cliometric revolution’ in economic history in the 1960s.

Historians might be interested in diachronic change not just from one topic to another, but within individual topics. In a recent implementation of dynamic topic models, Guldi (2019) fit a model to a corpus of parliamentary debates from the British Empire over the course of the eighteenth century to conduct what she called a “‘distant reading’ of the history of the infrastructure of the British Empire.” Guldi found unexpected prominence of topics related to infrastructure and historical figures generally overlooked in historiography of the British Empire. The topic model also enabled discovery and mapping of competing discourses about technology in the Empire (Guldi, 2019, 92).

Another example that might be of particular interest to economic historians is that of Farrell (2016) who used topic models on a corpus of reports, press releases, web pages, meeting minutes, conference transcripts and other texts from organizations that voice skeptical views on climate change. Farrell first examined the total content of the large corpus and then, employing network theoretic methods, mapped a network connecting climate-skeptical organizations and the topics they addressed. This led to the additional finding of a clear and statistically testable relationship between funding and topics the organizations took up. Still other uses of topic models include, among many others, mapping discourses in medieval Christian and Islamic political thought (Blaydes, Grimmer, & McQueen, 2018), study of aviation accidents to find clusters of commonly repeating issues (Kuhn, 2018), mixed-methods demography (Chakrabarti

¹⁶For an discussion of computational text analysis in sociological theory building, see (Evans & Aceves, 2016).

& Frye, 2017), perceptions of climate change (Tvinnereim, Fløttum, Gjerstad, Johannesson, & Nordø, 2017), diverging discourses in historical feminist movements (Nelson, 2020), and sentiment analysis on topic models to better understand the tone of discourse regarding renewable energy politics in Germany (Dehler-Holland, Schumacher, & Fichtner, 2020).

It is precisely the idea of ‘frames’ for cognitive thinking, highlighted especially by Aklerof and Snower, that topic models allow a researcher to estimate. Rather than having to rely, as Shiller does, on ‘narratives’ and that are keyword searchable and, thus, more likely to capture fads or manias, topic models map distributions over collections of words. They are thus able to offer something more like what Aklerof and Snower or Stone aim for, cognitive frames or understandings whose content can be analyzed synchronically and diachronically. At the same time, topic models suggest how the method employed and discussed by Stone, as well as Aklerof and Snower, can be expanded from the micro-level, single-archive, single-milieu, and often single-individual close reading to a meso level – the kind of ‘nanohistory’ proposed by de Vries, where close readings can be aggregated and placed with larger evidentiary contexts, allowing for larger and more generalizable meso- and macro-scale claims. As some of the most successful instantiations to date have shown, topic modeling at its best is not a single-run process but one best incorporated into a process of iteration, whereby topic models guide close readings, which in turn offer refinement for further model runs, and so forth (Nelson, 2020). Finally, as many have likewise pointed out, topic models and the assumptions they incorporate are not guaranteed to output useful or reliable results. In the words of a team responsible for ‘structural’ topic model, researchers must ‘verify, verify, verify’ (Grimmer & Stewart, 2013). We might add, ‘iterate, iterate, iterate’ and ‘supplement, supplement, supplement.’ The promise of topic models is not to reproduce human close reading at scale but to do something qualitatively different and best realized in more interpretative social science through the process of interacting with and probing model results.

3.5. *Word embedding*

The final method to be surveyed here is one developed only within the last decade and now host to some of the most exciting computational text and corpus linguistics studies in social science. This method is based on the intuition that words which share similar contexts often share similar meanings. Thus, rather than look for words that co-occur over multiple documents as in topic modeling, word embedding seeks to give words that appear in similar contexts similar values while also giving them values dissimilar to words that do not appear in the same contexts. The values are not single numbers but lists of numbers (again, vectors) which can be thought of as values along numerous different dimensions. They, thus, denote displacement in direction and magnitude in multi-dimensional space. As such, word embedding is also referred to as ‘vector semantics.’ The intuition is that word vectors denoting words of similar meanings should be close to each other in direction, something that can be measured using a number of different tools, most frequently cosine distance as already discussed.¹⁷ The two most widely used algorithms to find word embeddings are word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) and GloVe (Pennington, Socher, & Manning, 2014).

Word embedding gives analysts the ability to look at words that share synonymous

¹⁷For a deeper, yet readable, introduction to vector semantics, see Jurafsky and Martin (2014, ch. 6).

or related meanings and to track how distances between words and nearest neighbors change between different corpora. Tracking changes over time has been one of the most notable successes of the method. Kulkarni, Al-Rfou, Perozzi, and Skiena (2015), for example, showed how the word ‘gay’ can be mapped as most closely related to words ‘daft,’ ‘flaunting,’ and ‘sweet’ in texts from the first decade of the twentieth century, ‘witty,’ ‘bright,’ and ‘frolicsome’ in the mid-century and finally to ‘lesbian’ and ‘homosexual’ by the 1990s. Others have used word embeddings to find relationships between frequency of a word and its shift in meaning over time, as well as investigating different reasons for why words shift context over time, suggesting that methods building on word embedding can isolate change in meaning occurring for cultural reasons and those constituting what they call ‘linguistic drift’ (Hamilton, Leskovec, & Jurafsky, 2016a, 2016b).

Representing meaning through position in space, though perhaps not intuitive, has surprising parallels with how humans generally understand meaning. The creators of the first major word embedding algorithm word2vec found that treating words as vectors allowed for mathematical operations such as, most famously, adding the vectors for ‘king’ and ‘woman’ and subtracting ‘man,’ which leads to a place in geometric spaces closest to ‘queen’.¹⁸ The research team found similar results for a range of word analogies, such as $\overrightarrow{Paris} - \overrightarrow{France} + \overrightarrow{Italy} \approx \overrightarrow{Rome}$ (Mikolov et al., 2013). This works because word embeddings are purely relational, thus differences between vectors are themselves vectors that can have semantic meanings. Being vectors, the offsets between word vectors can correspond to dimensions of human meaning, here gender, meaning that similar offsets are observed separating king-queen, man-woman, heir-heiress, aunt-uncle, and so on.

Recent work in machine learning has looked at biases in word embeddings. One well-known study has shown that a word2vec model trained on news texts finds that $\overrightarrow{man} - \overrightarrow{computer\ programmer} + \overrightarrow{woman}$ leads most closely to the vector $\overrightarrow{homemaker}$ (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016). The team found a number of other gender biases encoded into the embeddings that stirred discussion in the machine learning community about ways to ‘de-bias’ word embeddings (Caliskan, Bryson, & Narayanan, 2017; Gonen & Goldberg, 2019).

For social scientists, however, the fact that word embeddings encode so much cultural and often implicit meaning – including biases and stereotypes – make them a tool of great interest. In one of the most innovative studies using word embeddings to date, Kozlowski, Taddy, and Evans (2019) sought to find and analyze an ‘affluence’ dimension in twentieth-century word embeddings. The authors found a number of word pairs exhibiting a similar offset (e.g. rich/poor and affluence/poverty) and then took the arithmetic mean of these offsets to find the most general vector dimension of affluence that allowed them subsequently to measure how similar other words were to the meaning of affluence in vector space. Accordingly, when looking at names of sports, they found that terms pointing most closely in the direction of the affluence vector (‘golf’ and ‘tennis’) are ones commonly associated with wealth, whereas words pointing most exactly in the opposite direction of the affluence vector were ‘boxing’ and ‘camping.’ The authors then tested how well word-embedding dimensions corresponded with human intuition by comparing algorithmic with human judgement about valences in words. They found that the algorithm generally matched with human intuition, though not in all areas. Race, in particular, was a difficult category

¹⁸In vector notation, $\overrightarrow{king} + \overrightarrow{woman} - \overrightarrow{man} \approx \overrightarrow{queen}$, just as a human would generally suggest ‘queen’ in answer to the question: man:woman as king:_____?

for word embedding to code. This serves to underline the importance of verification. They then analyzed how meanings related to social class have changed since the early twentieth century in regard to a number of dimensions along which sociological theory has treated class: affluence, employment, status, education, cultivation, morality, and gender. In addition to allowing for the mapping and visualization of interrelated meanings interwoven into ideas about class, productive in and of itself, the authors' findings also challenged certain findings of sociological theory. Whereas theorists of recent decades have come increasingly to believe that education influences perceptions of class by raising individuals' cultural cachet rather than through the skills it imparts, in general public perception the opposite occurred as education became increasingly closely connected with affluence unmediated by cultivation or cultural capital.

Other recent implementations of vector semantics include Ash, Chen, and Naidu (2019), who sought to quantify the extent to which judicial opinions in different cases reflected thinking coming from the subdiscipline of 'law and economics'. They did this by training word2vec on their corpus and then computing an average word vector for each case as a whole by computing averages of all word vectors present in the case, weighed by inverse frequency that seeks to privilege rare words in a similar manner as in TF-IDF. They then took the cosine distance between the average word vector for each case and the average word vector of a dictionary of terms associated with law and economics. The measure showed a roughly similar trend to a basic dictionary frequency-counting method. These measures were then used as dependent variables for regressions investigating the relationship between judges' attendance of a well-known law and economics seminar and, *inter alia* law and economics language in subsequent judicial opinions.

Another interesting application of the method uses word embeddings to compile a dictionary, with all words clustered around 'uncertainty' (anxiety, war, fear) to compose an uncertainty dictionary. The authors subsequently fit a topic model to find segments of bank earnings calls that had high proportion of two particular topics (interest rates and housing) and then measured the level of uncertainty evinced in discussion of these two topics using the dictionary identified by word embedding (Soto, 2019). One difficulty of using relational embeddings is that distances can only be measured from one word to another. There is not one stable, meaningful point that all words can be compared to and which gives information about relations with each other. Gurciullo and Mikhaylov (2017) have attempted to get around this by constructing indices of cosine distances for words used in a base year that can then be compared to each other.

Like topic models, semantic embedding rests on fitting and optimizing algorithms that are opaque, yet they represent a transformation that is oddly congruent with human linguistic cognition. Word meaning is encapsulated along multiple axes enabling words to be compared with each other as well as dimensions of commonalities to be found. As used by Kozlowski et al. (2019) to investigate topics treated by sociological theory, one can imagine similar projects to examine narrative understandings and cognitive framings, again at a significantly more sophisticated level than Shiller and at a broader and higher level than possible through close readings while providing evidence for representativeness. Semantic embeddings offer the possibility of comparing changing meaning of words over time, encompassing one of the central ideas of intellectual history of the last decades that ideas and words must be understood within their time and place (Skinner, 1969). To expand and deepen Shiller's proposal, we can imagine investigations into the how gradients of meaning of words such as 'spending' or 'consumption' might complement understandings of recessions and depressions, allowing

for better understandings of economic behavior of individuals, not as individuals either infected or non-infected with a contagious narrative fad but how words have allowed different sorts of cognitive understandings of people’s worlds at the level of society.

3.5.1. Contextualized word embeddings and transformers

One obvious shortcoming of word embeddings is the quite obvious observation that a single word can have different meanings depending on its context. In response to this, NLP researchers have suggested several new methods that are, as of this writing, at the very cutting edge of natural language processing and artificial intelligence. The first such method, called Embeddings from Language Models or ELMo uses a deep learning (neural network) algorithm to look ahead and behind individual words in their sentence context, which is then fed through a pre-trained model for that language to find individual embeddings for every individual appearance of every word in a corpus (Peters et al., 2018). Similarly, a new class of algorithms called transformers implements a method of weighing how much context words affect the meaning of each word within a larger deep learning algorithm hierarchy. Perhaps the most widely used transformer currently is the Bidirectional Encoder Representations from Transformers (BERT) algorithm (Devlin, Chang, Lee, & Toutanova, 2018). Transformers have revolutionized natural language processing, particularly in machine translation as well classification (Smith, 2019). Critically, these algorithms require training on general, large corpora and then provide a model that can be ‘transformed’ onto new texts, perhaps – indeed, ideally – with ‘fine-tuning’ of the model based on new text.

These methods are, as of yet, so new there has been little implementation in the social sciences. To date, papers that have appeared – in political science and finance – have merely analyzed the performance of these algorithms in classification and sentiment analysis relative to other methods, finding that they perform as well and often better than other widely used machine learning algorithms (Masson & Paroubek, 2020; Mishev, Gjorgjevikj, Vodenska, Chitkushev, & Trajanov, 2020; Terechshenko et al., 2020).

There remain significant drawbacks, however. The neural networks used in these architectures require millions and even billions of parameters and demand huge text corpora for training, which in turn necessitates computational facilities difficult for most social scientists to command. While the pre-trained models are freely available and can be fine-tuned more quickly and at much less cost, this makes the assumption that the languages they have been trained on are similar to those that will be fed into them. The off-the-shelf models are trained on contemporary corpora.¹⁹ This is a problematic assumption for historians, to say the least. Algorithms based on neural networks are also famously opaque and indecipherable to humans. Thus, the refrain, as urged by Grimmer and Stewart (2013), “verify, verify, verify” would apply in spades here. At the same time, because contextual word embeddings give more specific outputs (vectors on each individual word *instance*), these could be easier to verify and be more amenable to iterative workflows of algorithmic and close readings. Indeed, this is a case where deep knowledge and ‘*sitzfleisch*’-based feel of language and document sources could end up mattering more, rather than less.

The idea, however, remains an appealing one suggests a number of future avenues of use. Clearly the same word in a text does have different meanings and the ability to map this and compare ranges of meanings of certain words over time is one that con-

¹⁹The English language model was trained on English Wikipedia articles and an online corpus of “free books written by as yet unpublished authors” (Zhu et al., 2015, 3) [devlin2018bert]

textual word embeddings would likely be even better at than the word embeddings so powerfully leveraged by Kozlowski et al. (2019). Reliable contextual word embeddings could allow future researchers to replicate their work and do much more. Secondly, transformer algorithms are being used to do much more than classification. Not only are they at the heart of GPT-3, an AI system that has proved exceptionally good at producing believable human speech (Brown et al., 2020; Manjoo, 2020), so too has it been tested in algorithms to summarize texts, judge similarity between documents (in a much deeper way than cosine similarity), and to judge if the meaning of one text ‘entails’ the meaning of another (Radford, Narasimhan, Salimans, & Sutskever, 2018). The perspective of mapping which parts of documents disagreed or agreed with a certain premise, ability to quickly summarize massive amounts of text, or visualize similarity in meaning (not just vocabulary) could clearly be of interest to economic and business historians.

4. Implementation

The above brushes over issues of tremendous practical importance such as document digitization, web scraping, optical character recognition (ORC), text cleaning, stemming, not to mention choice of software for conducting analysis. If diversity of methods and algorithms in computational text analysis can be overwhelming, the array of practical concerns and modes of implementation can be downright dizzying. As the programming adage goes: there are always multiple means to accomplish the same thing.

One perennial issue that a beginning text analyst will be faced with is the decision about software. Web-based visualization software such as Voyant allow for basic and standardized exploration of machine-readable texts. For guide to text analysis not requiring coding, see Walsh and Horowitz (n.d.). However, to perform most of the methods discussed here require knowledge of a programming language, generally Python or R. The choice depends to some extent on what one is planning to do with the language. Python is extensively used amongst computer scientists, developers, and machine learning researchers. Many cutting edge machine learning algorithms (BERT, for instance) are written in Python or quickly available with Python wrappers. Python has also traditionally boasted leading web scraping tools (Mitchell, 2018). While not generally boasting as cutting-edge algorithms, the R language is more widely used in the social sciences and humanities and offers an exceptionally rich array of software packages for text analysis, as well as for visualization and other data science applications, making it the programming language of choice for many in the social sciences and humanities.

Graham et al. (in press) give a broad overview to numerous approaches. A good (though at the time of writing still in-progress) overall introduction to R for historians is Mullen (2018–). Overviews of text analysis in R, including cleaning and preparing documents are available in Welbers, Van Atteveldt, and Benoit (2017), Silge and Robinson (2017), Jockers and Thalken (2020), Roberts et al. (2019), and Benoit et al. (2018). A broader introduction to R including text and other methods of possible interest to historians are available in Arnold and Tilton (2015) and Wickham and Golemund (2016). For Python, Montfort (2016), Karsdorp (n.d.), and Sinclair, Rockwell, and Mony (2018) introduce social scientists and humanists to basic uses, making particular use of the NLP package `nltk` (Bird, Klein, & Loper, 2009).

5. Conclusion

Digital methods open up new avenues for historians of the economy and business including the means to go back to asking truly big overarching macro-level questions that are both descriptive and causal. Methods surveyed above as well as those currently being developed in NLP offer significant advantage beyond the wide but shallow word searches used by Shiller to locate economic fads, and the deep but close, narrow reading suggested by Stone and taken up in the anthropologically-inspired cultural history since the 1980s. The methods outlined here contain implicit assumptions that push beyond both, giving the researcher the means to gather evidence, show patterns, find regularities from an immeasurably larger range of texts. The suggestions in the text above for possible applications of these methods for business and economic historians are but the tip of the proverbial iceberg.

The narrative turn among a few, but prominent, mainstream economists and simultaneous explosion of new methods in computational text analysis furthermore offers an exceptional possibility for economic and business historians to engage in cross-disciplinary conversations. With feet in both history and economics, deep contextual and historical knowledge and well-honed close reading skills coupled with basic facility with and general competence in compiling and working with quantitative data, economic and business historians are well situated to intervene and participate in fruitful debates, bringing expertise and concern with narrative and text into discussions with economists and presenting quantitative and representative ‘nanohistorical’ arguments in interactions with historians.

References

- Abolafia, M. Y. (2010). Narrative construction as sensemaking: How a central bank thinks. *Organization studies*, 31(3), 349–367.
- Akerlof, G. A., & Shiller, R. J. (2010). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton university press.
- Akerlof, G. A., & Snower, D. J. (2016). Bread and bullets. *Journal of Economic Behavior & Organization*, 126, 58–71.
- Alsvik, O., & Munch-Møller, M. G. (2020). Historiografi møter algoritmer. *Heimen*, 57(3), 201–215.
- Arnold, T., & Tilton, L. (2015). *Humanities data in R: Exploring networks, geospatial data, images, and text*. Springer.
- Ash, E., Chen, D. L., & Naidu, S. (2019). Ideas have consequences: the impact of law and economics on american justice. *Center for Law & Economics Working Paper Series*, 4.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4), 1593–1636.
- Barens, I. (2011). *animal spirits in john maynard keyness general theory of employment, interest and money*. Routledge London.
- Beckert, J. (2016). *Imagined futures: Fictional expectations and capitalist dynamics*. Harvard University Press.
- Beckert, J., & Bronk, R. (2018). *Uncertain futures: Imaginaries, narratives, and calculation in the economy*. Oxford University Press.
- Benoit, K., & Laver, M. (2007). Estimating party policy positions: Comparing expert surveys and hand-coded content analysis. *Electoral Studies*, 26(1), 90–107.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An r package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774.

- Binder, J. M. (2016). Alien reading: Text mining, language standardization, and the humanities. In *Debates in the digital humanities 2016* (pp. 201–217). University of Minnesota Press.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: analyzing text with the natural language toolkit*. ” O’Reilly Media, Inc.”.
- Blaydes, L., Grimmer, J., & McQueen, A. (2018). Mirrors for princes and sultans: Advice on the art of governance in the medieval christian and islamic worlds. *The Journal of Politics*, 80(4), 1150–1167.
- Blei, D. M. (2012a). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Blei, D. M. (2012b). Topic modeling and digital humanities. *Journal of Digital Humanities*, 2(1), 8–11.
- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on machine learning* (pp. 113–120).
- Blei, D. M., & Lafferty, J. D. (2009). Topic models. *Text mining: classification, clustering, and applications*, 10(71), 34.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3, 993–1022.
- Blevins, C. (2014). Space, nation, and the triumph of region: A view of the world from houston. *The Journal of American History*, 101(1), 122–147.
- Blevins, C. (2016). Digital history’s perpetual future tense. *Debates in the Digital Humanities*, 308–324.
- Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in neural information processing systems* (pp. 4349–4357).
- Braudel, F. (1972). *The Mediterranean and the Mediterranean world in the age of Philip II* (Vol. 1). London: Collins.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., . . . others (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186.
- Cerchiello, P., & Nicola, G. (2018). Assessing news contagion in finance. *Econometrics*, 6(1), 5.
- Chakrabarti, P., & Frye, M. (2017). A mixed-methods framework for analyzing text data: Integrating computational techniques with qualitative methods in demography. *Demographic Research*, 37, 1351–1382.
- Commons, J. R. (1931). Institutional economics. *The American economic review*, 648–657.
- Dale, R. (2004). *The first crash: lessons from the south sea bubble*. Princeton University Press.
- Daston, L., & Galison, P. (2007). *Objectivity*. New York: Zone Books.
- Dehler-Holland, J., Schumacher, K., & Fichtner, W. (2020). Topic modeling uncovers shifts in media framing of the german renewable energy act. *Patterns*, 100169.
- Deringer, W. (2018). *Calculated values: Finance, politics, and the quantitative age*. Harvard University Press.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- de Vries, J. (2013). The return from the return to narrative. *Max Weber Lecture*(01).
- de Vries, J. (2018). Changing the narrative: The new history that was and is to come. *Journal of Interdisciplinary History*, 48(3), 313–334.
- de Vries, J. (2019, 11). Playing with scales: The global and the micro, the macro and the nano. *Past & Present*, 242(Supplement 14), 23–36. Retrieved from <https://doi.org/10.1093/pastj/gtz043>
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, 41(6), 570–606.
- Dittmar, J., & Seabold, S. (2019). *New media and competition: printing and europe’s trans-*

- formation after gutenber. Centre for Economic Performance, LSE.
- Dow, S. (2010). The psychology of financial markets: Keynes, minsky and emotional finance. In D. Papadimitriou & L. R. Wray (Eds.), *The elgar companion to hyman minsky* (pp. 246–62). Edward Elgar Publishing.
- Evans, J. A., & Aceves, P. (2016). Machine translation: Mining text for social theory. *Annual Review of Sociology*, 42, 21–50.
- Farrell, J. (2016). Corporate funding and ideological polarization about climate change. *Proceedings of the National Academy of Sciences*, 113(1), 92–97. Retrieved from <https://www.pnas.org/content/113/1/92>
- Fligstein, N., Stuart Brundage, J., & Schultz, M. (2017). Seeing like the Fed: Culture, cognition, and framing in the failure to anticipate the financial crisis of 2008. *American Sociological Review*, 82(5), 879–909.
- Funk, K., & Mullen, L. A. (2018). The Spine of American Law: Digital Text Analysis and U.S. Legal Practice. *The American Historical Review*, 123(1), 132–164.
- Geese, L. (2020). Immigration-related speechmaking in a party-constrained parliament: Evidence from the refugee crisis of the 18th german bundestag (2013–2017). *German Politics*, 29(2), 201–222.
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–74.
- Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1), 35–71.
- Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4), 1307–1340.
- Goldgar, A. (2008). *Tulipmania: money, honor, and knowledge in the dutch golden age*. University of Chicago Press.
- Gonen, H., & Goldberg, Y. (2019). Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. *arXiv preprint arXiv:1903.03862*.
- Graham, S., Milligan, I., & Weingart, S. (in press). *Exploring big historical data: The historian’s macroscope* (2nd ed.). World Scientific Publishing Company. Retrieved from <http://www.themacroscopic.org/2.0/>
- Grajl, P., & Murrell, P. (2019). Toward understanding 17th century english culture: A structural topic model of francis bacon’s ideas. *Journal of Comparative Economics*, 47(1), 111–135.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267–297.
- Guldi, J. (2019). Parliament’s debates about infrastructure: An exercise in using dynamic topic models to synthesize historical change. *Technology and culture*, 60(1), 1–33.
- Gurciullo, S., & Mikhaylov, S. J. (2017, October). Detecting policy preferences and dynamics in the UN general debate with neural word embeddings. In *2017 International Conference on the Frontiers and Advances in Data Science (FADS)* (pp. 74–79). Xi’an: IEEE. Retrieved 2020-05-04, from <http://ieeexplore.ieee.org/document/8253197/>
- Gutmann, M. P., Merchant, E. K., & Roberts, E. (2018). “big data” in economic history. *The journal of economic history*, 78(1), 268–299.
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016a). Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing* (Vol. 2016, p. 2116).
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016b). Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.
- Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2), 801–870.

- Hitchcock, T. (2013). Confronting the digital. *Cultural and Social History*, 10(1), 9-23.
- Hobsbawm, E. J. (1980). The revival of narrative: some comments. *Past & Present*(86), 3-8.
- Hodgson, G. M. (2006). What are institutions? *Journal of economic issues*, 40(1), 1-25.
- Holmes, D. R. (2013). *Economy of words: communicative imperatives in central banks*. University of Chicago Press.
- Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.
- Jockers, M. L., & Thalken, R. (2020). *Text analysis with r*. Springer.
- Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing. vol. 3*. Pearson London London.
- Karsdorp, F. (n.d.). *Python programming for the humanities*. Retrieved from <http://www.karsdorp.io/python-course/>
- Kim, I. S. (2017). Political cleavages within industry: firm-level lobbying for trade liberalization. *American Political Science Review*, 111(1), 1-20.
- Kindleberger, C. P. (1978). *Manias, panics, and crashes: a history of financial crises*. Macmillan.
- Klingenstein, S., Hitchcock, T., & DeDeo, S. (2014). The civilizing process in london's old bailey. *Proceedings of the National Academy of Sciences*, 111(26), 9419-9424.
- Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5), 905-949.
- Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, 87, 105-122.
- Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically significant detection of linguistic change. In *Proceedings of the 24th international conference on world wide web* (pp. 625-635).
- Lamoreaux, N. (2015). The future of economic history must be interdisciplinary. *The Journal of Economic History*, 75(4), 1251-1257.
- Lamoreaux, N. R. (2016). Beyond the old and the new: economic history in the united states. *Routledge handbook of global economic history*, 35-54.
- Levingston, O. (2020). Minskys moment? the rise of depoliticised keynesianism and ideational change at the federal reserve after the financial crisis of 2007/08. *Review of International Political Economy*, 1-28.
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65.
- Manjoo, F. (2020). How do you know a human wrote this. *New York Times, Editorial*, 29.
- Masson, C., & Paroubek, P. (2020). NLP analytics in finance with DoRe: A French 250 million-token corpus of corporate annual reports. In *Proceedings of the 12th language resources and evaluation conference* (pp. 2261-2267).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
- Miller, I. M. (2013). Rebellion, crime and violence in qing china, 1722-1911: a topic modeling approach. *Poetics*, 41(6), 626-649.
- Mimno, D., & Lee, M. (2014). Low-dimensional embeddings for interpretable anchor-based topic inference. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1319-1328).
- Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE Access*, 8, 131662-131682.
- Mitchell, R. (2018). *Web scraping with python: Collecting more data from the modern web.* "O'Reilly Media, Inc."
- Montfort, N. (2016). *Exploratory programming for the arts and humanities*. MIT Press.

- Moreno, Á. I., & Caminero, T. (2020). Application of text mining to the analysis of climate-related disclosures. *Documentos de Trabajo/Banco de España*, 2035.
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Mullen, L. (2018–). *Computational historical thinking: With applications in R*. Retrieved from <https://dh-r.lincolnmullen.com>
- Müller, H., von Nordheim, G., Boczek, K., Koppers, L., & Rahnenführer, J. (2018). Der Wert der Worte—Wie digitale Methoden helfen, Kommunikations- und Wirtschaftswissenschaft zu verknüpfen. *Publizistik*, 63(4), 557–582.
- Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, 49(1), 3–42.
- Newman, D. J., & Block, S. (2006). Probabilistic topic decomposition of an eighteenth-century american newspaper. *Journal of the American Society for Information Science and Technology*, 57(6), 753–767.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- Putnam, L. (2016). The transnational and the text-searchable: Digitized sources and the shadows they cast. *The American Historical Review*, 121(2), 377–402.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
- Rhody, L. M. (2012). Topic modeling and figurative language. *Journal of Digital Humanities*, 2(1).
- Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515), 988–1003.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, 91(1), 1–40.
- Schmidt, B. M. (2012). Words alone: Dismantling topic models in the humanities. *Journal of Digital Humanities*, 2(1).
- Schmidt, B. M. (2016). Do digital humanists need to understand algorithms? *Debates in the Digital Humanities*, 2016, 53.
- Sewell Jr, W. H. (2005). *Logics of history: Social theory and social transformation*. University of Chicago Press.
- Sewell Jr, W. H. (2010). A strange career: The historical study of economic life. *History and Theory*, 49(4), 146–166.
- Shiller, R. J. (2015). *Irrational exuberance: Revised and expanded third edition*. Princeton university press.
- Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4), 967–1004.
- Shiller, R. J. (2020). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
- Silge, J., & Robinson, D. (2017). *Text mining with r: A tidy approach*. O'Reilly Media, Inc.
- Sinclair, S., Rockwell, G., & Mony, M. (2018). *The art of literary text analysis*. Retrieved from <https://github.com/sgsinclair/alta/blob/master/ipynb/ArtOfLiteraryTextAnalysis.ipynb>
- Skinner, Q. (1969). Meaning and understanding in the history of ideas. *History and theory*, 8(1), 3–53.
- Slapin, J. B., & Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3), 705–722.
- Smith, N. A. (2019). Contextual word representations: A contextual introduction. *arXiv preprint arXiv:1902.06006*.
- Soto, P. E. (2019). Breaking the Word Bank: Effects of Verbal Uncertainty on Bank Behavior. *SSRN Electronic Journal*. Retrieved 2020-05-04, from <https://www.ssrn.com/abstract=3347329>

- Stone, L. (1979). The revival of narrative: reflections on a new old history. *Past & Present*(85), 3–24.
- Stulpe, A., & Lemke, M. (2016). Blended reading: Theoretische und praktische dimensionen der analyse von text und sozialer wirklichkeit in zeitalter der digitalisierung. In *Text mining in den sozialwissenschaften* (pp. 17–61). Springer.
- Temin, P., & Voth, H.-J. (2004). Riding the south sea bubble. *American Economic Review*, 94(5), 1654–1668.
- Terechshenko, Z., Linder, F., Padmakumar, V., Liu, M., Nagler, J., Tucker, J. A., & Bonneau, R. (2020). A comparison of methods in political science text classification: Transfer learning language models for politics. *Available at SSRN*.
- Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3), 1139–1168.
- Tuckett, D., Holmes, D., Pearson, A., Chaplin, G., et al. (2020). Monetary policy and the management of uncertainty: A narrative approach. *Bank of England Staff Working Papers*, 870, 1–28. Retrieved from <https://www.ssrn.com/abstract=3627721>
- Tvinnereim, E., Fløttum, K., Gjerstad, Ø., Johannesson, M. P., & Nordø, Å. D. (2017). Citizens preferences for tackling climate change. quantitative and qualitative analyses of their freely formulated solutions. *Global Environmental Change*, 46, 34–41.
- Vol’chik, V. V. (2017). Narrativnaia i institutsional’naia ekonomika. *Zhurnal institutsional’nykh issledovaniï*, 9(4), 132–143.
- Walsh, B. W., & Horowitz, S. (n.d.). *Introduction to text analysis: A coursebook*. Retrieved from <http://walshbr.com/textanalysiscoursebook/>
- Wehrheim, L. (2019). Economic history goes digital: topic modeling the journal of economic history. *Cliometrica*, 13(1), 83–125.
- Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text analysis in R. *Communication Methods and Measures*, 11(4), 245–265.
- Wickham, H., & Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. ” O’Reilly Media, Inc.”.
- Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the ieee international conference on computer vision* (pp. 19–27).