DIGITAL SCHOLARSHIP, DIGITAL CLASSROOMS:

New International Perspectives in Research and Teaching

Proceedings of the Gale Digital Humanities Day at the British Library

May 2, 2019

GALE
A Cengage Company

Digital Scholarship, Digital Classrooms

New International Perspectives on Research and Teaching

Digital Scholarship, Digital Classrooms

New International Perspectives on Research and Teaching

Proceedings of the Gale Digital Humanities Day at the British Library
May 2, 2019

GALE
A Cengage Company

GALE
A Cengage Company

Digital Scholarship, Digital Classrooms: New International Perspectives on Research and Teaching

For product information and technology assistance, contact us at
Gale Customer Support, 1-800-877-4253.

Gale
27500 Drake Rd.
Farmington Hills, MI, 48331-3535

ISBN: 978-0-4182-9146-7 (print)

Printed in the United States of America
1 2 3 4 5 6 7 23 22 21 20 19

Table of Contents

Introduction

Seth Cayley, VP Gale Primary Sources

"Founding myths" are an integral part of the technology sector. Every successful company has some version of a story about how it emerged by accident, or eureka moment, or how it was built from the ground up while the founder lived in a shipping container.

Examples of founding myths include Apple being started in Steve Jobs's garage or Bill Gates dropping out of Harvard to start Microsoft. These are "myths," not in the sense that they are false, but in that they simplify complex histories and downplay the important roles of other people and institutions.

Apple's co-founder Steve Wozniak himself has confessed "the garage represents us better than anything else, but we did no designs there.… There were hardly ever more than two people in the garage and mostly they were sitting around kind of doing nothing productive."[1] Similarly, the "drop-out" narrative that surrounds Bill Gates obscures his privileged background (along with his co-founder Paul Allen) attending one of the most prestigious private high schools in America, where he was lucky to have had access to computers.

The consequence of these myths is to amplify the ideal of the solo (usually white and male) genius.

Digital humanities has its own founding myth. Around 1949, the Jesuit Roberto Busa approached the CEO of IBM, Thomas J. Watson Sr., and convinced him to fund and resource his planned *Index Thomisticus*, a tool that Busa intended would allow the user to carry out textual analysis of the complete works of St. Thomas Aquinas. In Busa's own telling of the story, it is presented as a David and Goliath scenario, with the humble Jesuit priest persuading a high priest of capitalism to support a project of pure scholarship.[2] From such unlikely beginnings a whole movement of "humanities computing" was born. As Julianne Nyhan has observed, this underdog story has helped bind the digital humanities community together as a group, differentiated from mainstream humanities.[3]

Without denigrating Busa's unquestionable achievements, there are problems with the Busa founding myth. The scholar Tito Orlandi notes that there were other scholars working in the fields of translation and archaeology who should be considered amongst the pioneers of what is now known as digital humanities.[4] Also, Busa did not work alone. As forthcoming work will highlight, he relied on a team of (female) punched card operators in the 1950s and 1960s to make his vision a reality.[5] By focusing on Busa, we diminish both the importance of teamwork in his success, and we also ignore the wider international scholarly environment that contributed to the development of digital humanities.

Digital humanities in 2019 is much more diverse than the Busa founding myth implies. It is a community that emphasises collaboration, transcends narrow disciplines, and is truly international in outlook. At the time of writing, the project *Oceanic Exchanges*, which traces global information networks in historical newspapers, includes 39 researchers working across 6 national groups.[6] Recent Alliance of Digital Humanities Organizations conferences have made conscious efforts to invite participants from the Global South, and in 2021 Tokyo will be hosting the first such conference to be held in Asia.[7]

While, as Nyhan says, the Busa myth may have been a "useful fiction,"[8] digital humanities is moving beyond its confines and into the mainstream. It is notable that a significant amount of the community conversation is about how to *teach* digital humanities. In the last couple of years, a steady number of papers and books have been published discussing how to integrate digital humanities into the classroom and curriculum.[9]

My personal journey into digital humanities began in 2012, when I supported a project headed by Dallas Liddle to do distant reading of thousands of editorials from the pages of the *Times*. By analysing changing patterns of word counts, Dallas's research challenged long-held beliefs about the slow pace of change in Victorian newspapers … all without specifically requiring him to read any of the actual editorials.[10] The project challenged my own assumptions about what "history" could be. Text could be treated as data. I was hooked.

Since then, Gale has undergone a culture shift to make the underlying data of Gale Primary Sources archives more accessible for such projects around the world. We have provided support to a number of digital humanities projects, looking at ideas as disparate as Tudor networks of power (led by Ruth Ahnert[11]), eighteenth-century printers' ornaments (Fleuron, led by Hazel Wilkinson[12]), and tracing the way news was reprinted across the Anglophone world in the nineteenth century (Scissors and Paste, led by Melodee Beals[13]). In 2018 we launched the Gale Digital Scholar Lab, a cloud-based text and data environment specifically designed to lower the barriers to entry for researchers of all levels to carry out digital humanities workflows.[14]

The time therefore felt right to bring together scholars, librarians, and students to learn about the latest research and trends in digital humanities from a diverse range of international speakers. *Digital Scholarship, Digital Classrooms* offers a selection of the eight papers that were presented at the resulting Gale Digital Humanities Day at the British Library, held on 2 May 2019.

The day was split into four distinct sessions: (1) Literature & Distant Reading; (2) Computers Reading the News; (3) Digital Humanities in the Classroom; and (4) Institutional Support and Infrastructure for digital humanities. Each session was designed to put a spotlight on different themes in digital humanities, with the first two focusing on the latest research, and the latter two exploring issues around teaching and the role of the library.

In this volume you will find four of the papers that were presented on that day.

In the first paper, "On Principles and Values: Mining for Conservative Rhetoric in the London *Times*, 1785–2010," **Joris van Eijnatten**, Professor of Cultural History at Utrecht University, provides a textbook example of how using n-grams and other textual analysis tools can lead to new and interesting discoveries. In his work on moral language and historical newspapers, Joris explores the development of conservative rhetoric in *The Times Digital Archive*. He charts the rise and fall of certain phrases in political reporting, such as the shift from "conservative principles" to "conservative values." Placing these findings in historical context, he notes that "political opinions began to be couched explicitly in moral terms after the 1960s," and that new forms of rhetoric "emerged from a tension with modernity" provoked especially by the permissive society of the 1960s.

Carrying out such research is impossible without the right infrastructure in place. In "Reflections on Infrastructures for Mining Nineteenth-Century Newspaper Data," **Julianne Nyhan**, Associate Professor of Digital Information Studies and Deputy Director of the University College London (UCL) Centre for Digital Humanities, discusses her team's experience of trying to get a large-scale text-mining project off the ground. She highlights how even if researchers have access to data, and are at well-resourced institutions such as UCL, there can be a succession of unexpected challenges to begin being able to analyse that data. Despite having support from the university's Research IT Services group, Julianne notes that "the computing infrastructure was not really set up to support Humanities research," and at a cost of £350 per day, each additional day it took to mount the data drained the project budget. She closes her piece with recommendations for how universities should address the "digital turn," and the potential role of public-private partnerships.

The topic of **Ryan Cordell's** paper is "Teaching Humanistic Data Analysis." Ryan teaches both graduate and undergraduate courses at Northeastern University and has reflected deeply on his own teaching philosophy.[15] In his words: "One of my primary goals as a teacher is to work against notions of the 'digital native' and the ruinous pedagogical consequences following from that designation.… I believe the most pressing scholarly questions in the coming years will require true interdisciplinarity. By this I mean not individual scholars who dabble in many

disciplines, but groups of scholars who can contribute their various expertise to a sustained and substantial enterprise. It is this future that I hope to prepare my students to meet with creativity and verve. Doing so requires not a pedagogy founded on particular technologies or tools, but on capacious scholarly and technological imaginations." Ryan's paper expands on these ideas, arguing that it is not simply programming skills that we need to teach, but a mindset for thinking about data—in his words, "programmatic thinking."

Continuing with the classroom theme, **Sarah Ketchley**, an Egyptologist and digital humanities specialist based at the University of Washington, discusses her own experiences of leading undergraduates through her Introduction to Digital Humanities course. Sarah talks through the learning goals and pedagogical challenges she faced, along with the methodological and technological solutions she used to overcome those challenges. Using her Autumn 2018 class as a case study, she explains how her students used the Gale Digital Scholar Lab to interrogate their data sets and use the resulting analysis to create impressive digital exhibits. Sarah concludes by reflecting on the learning outcomes and how teaching digital humanities helps students develop transferrable skills for life beyond university: "the group was able to clearly identify a range of specific skills they were developing and using in class as being relevant and 'marketable' to employers."

Other speakers on the day included: **Mark Algee-Hewitt**, Director of the Stanford Literary Lab, who gave an overview of the Lab's "Microgenres" project, which looks at the way in which novelists draw on disciplines such as philosophy, history, and natural science with their narratives; **Tomoji Tabata**, Associate Professor of Corpus Linguistics at the University of Osaka, who explained his use of stylometry to analyse eighteenth- and nineteenth-century literature, and revealed to the audience that Charles Dickens's favourite phrase was "his hands in his pockets"; **Melodee Beals**, Lecturer in Digital History at the University of Loughborough, who discussed unexpected uses of heritage and historical data; and **Lisa McIntosh**, Director of Access Services at the University of Sydney, who explained how her university library is developing its infrastructure to support the latest research. Although their papers are not published in this volume, I encourage you to seek out their work.

I would like to extend my thanks to all the speakers for sharing their projects and ideas.

NOTES

1. https://www.washingtonpost.com/news/morning-mix/wp/2014/12/05/steve-wozniak-refutes-apples-creation-story-the-garage-is-a-bit-of-a-myth/.
2. R. Busa, "The Annals of Humanities Computing: The *Index Thomisticus*," *Computers and the Humanities* 14, no. 2 (1980), 84.
3. Julianne Nyhan and Andrew Flinn, *Computation and the Humanities: Towards an Oral History of Digital Humanities* (Springer International Publishing, 2016), 270.
4. Julianne Nyhan and Andrew Flinn, 79.
5. Julianne Nyhan and Melissa Terras, "Uncovering 'Hidden' Contributions to the History of Digital Humanities: The *Index Thomisticus'* Female Keypunch Operators," Digital Humanities 2017 (Montreal, QC, Canada). See also the authors' paper presented at DH2017: https://discovery.ucl.ac.uk/id/eprint/10052279/9/Nyhan_DH2017.redacted.pdf.
6. Oceanic Exchanges Project Team. *Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840–1914*. 2017. DOI 10.17605/OSF.IO/WA94S. Available at osf.io/wa94s.
7. List of ADHO conference locations available at http://adho.org/conference.
8. Julianne Nyhan and Andrew Flinn, 80.
9. For example, Claire Battershill and Shawna Ross, *Using Digital Humanities in the Classroom: A Practical Introduction for Teachers, Lecturers and Students* (Bloomsbury, 2017), and Anna Wing-bo Tso, editor, *Digital Humanities and New Ways of Teaching* (Springer, 2019).
10. Dallas Liddle, "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' the *Times* using *The Times Digital Archive*," *Journal of Victorian Culture* 17, no. 2 (June 2012), 230–237.
11. https://ahrc.ukri.org/research/case-study-archives/tudor-networks-of-power-1509-1603/.
12. https://fleuron.lib.cam.ac.uk/.
13. http://scissorsandpaste.net/.
14. https://www.gale.com/intl/primary-sources/digital-scholar-lab.
15. https://ryancordell.org/statements#teaching.

On Principles and Values: Mining for Conservative Rhetoric in the London *Times*, 1785–2010

Joris van Eijnatten
Professor of Cultural History
Utrecht University, the Netherlands
j.vaneijnatten@uu.nl

Abstract: *This article explores the nature of conservative rhetoric in the London-based* Times *between 1785 and 2010. What kind of language did writers in this newspaper consciously use to express a moral opinion that might be qualified as "conservative"? Such moral languages can be traced over time by examining commonly used word clusters, which in turn reflect ideas—in this case, "conservative" ideas. Word clusters can be highly iterative and often limited in size, which make them ideal for a computer-assisted analysis over a longer period of time. Methodologically, the article employs two proven, easy-to-use text mining techniques: n-grams (especially bigrams) and word embeddings. It traces a number of bigram phrases over the whole period, the most important of which are "conservative principles," "conservative values," "traditional values," and "permissive society."*

The term "conservative principles" figured as a moral and political expression especially in the nineteenth century and the first half of the twentieth century. Major changes in conservative rhetoric occurred mainly during and after the 1960s. New terms began to be used to denote moral positions, especially after the emergence of the so-called permissive society. At the same time, however, it became increasingly difficult to identify a specifically conservative rhetoric. First, the tension between tradition and modernity, which conservatives were bound to address, was part of a much larger public debate. Second, all moral languages became "moralised," as it were: great store began to be put by words such as "values" and "tradition." The result was a confluence of left- and right-wing rhetoric, and the concomitant submergence of ideological differences.

Keywords: conservatism ■ moral language ■ the London *Times* ■ digital history ■ principles ■ values

CONSERVATIVE RHETORIC

On September 25, 1997, Michael Portillo, a Tory politician who had been Secretary of State for Defence in John Major's government since 1995 but had just lost his seat, wrote a book review for the *Times*. He discussed *Is Conservatism Dead?*, a book by two conservative intellectuals, philosopher John Gray and member of parliament David Willetts. At one point Portillo quoted Willetts as having said that "where Mr. Blair offers soft soap about 'values,' Conservatives have 'principles.'"[1] Apparently, Willets had identified a difference in the political rhetoric used by Tories and Labourites (or at least the Blairites among them). The Left talked about values, the Right about principles. Whether Willetts actually said this is a moot point. But interestingly, a historical relationship between conservatism and the word "principles" does, in fact, exist.

The *Oxford English Dictionary* defines "principle" as "a general law or rule adopted or professed as a guide to action; a settled ground or basis of conduct or practice; a fundamental motive or reason for action, esp. one consciously recognized and followed."[2] One of the *OED*'s example sentences is a quotation from the novel *Coningsby* (1844) by the arch-conservative Benjamin Disraeli: "Before I support Conservative principles … I merely wish to be informed what those principles aim to conserve."[3] "Principles" have a distinct nineteenth-century feel to them;[4] acting on a principle meant doing something that followed unambiguously from a deeply rooted moral conviction. Principles could be contrasted among each other, thus making clear the possibility of different courses of action, some of which were right and some of which were wrong, given the ethical status of the starting point. Disraeli used the phrase "conservative principles" no less than thirteen times in *Coningsby*, at one point contrasting it to "concessionary principles."[5] The latter were wrong, because one does not concede principles, and the actions they logically gave rise to could only be unfortunate.

This article deals with the use of conservative rhetoric in the *Times* between 1785 and 2010.[6] What kind of language did writers consciously use to express a moral opinion that might be qualified as conservative in nature?[7] The *Times* itself has a reputation for inclining towards the centre-right, so this source may be less likely to portray conservative rhetoric through stereotypes. The latter do occur, of course, for example in a Letter to the Editor from 2001, in which a reader noted ironically that the prominent Conservative Iain Duncan Smith was not just "a 'never' on the euro" but also "in favour of corporal punishment, hanging and foxhunting."[8] This article also examines conservative rhetoric (which is not the same as the rhetoric of Conservatives or Conservatism) as a "moral language," that is as a form of rhetoric that presupposes fundamental judgements about what is good and what is bad in its attempts to justify social and political action. The importance of such ethically informed languages can scarcely be underestimated, for they influence the decisions and inspire the actions that have brought into being the world as we know it.

Conservative rhetoric is a useful entry into the languages of the past, but only one among many. In what follows I am not primarily interested in conservatism per se, or in arguing that conservatism is a unified political language, the lineage of which can be traced across more than two centuries to the ideas of Edmund Burke;[9] I am first and foremost interested in word clusters that make up the moral languages of the past. These clusters are highly iterative and often limited in size; this enables them to figure as popular memes or sound bites. As such they point to larger semantic fields, repositories of ideas and beliefs that have normative connotations and vary over time in complex ways. Because these memes are iterative, computers can be used to trace them. This article makes ample use of computer-assisted analysis, but I have focused on content rather than methodology. Central to what follows are two proven, easy-to-use digital techniques: n-grams (especially bigrams) and word embeddings. At the same time, the digital analyses are contextualised through a traditional reading of the original material. The methodology has been outlined briefly in the notes.

THE DECLINE OF CONSERVATIVE PRINCIPLES

The term "principles" was apparently significant to Tory history, and we can regard it as a first instance of conservative rhetoric—and an obvious place to start. To illustrate historical usage of the bigram "conservative principles," we can plot simple frequency counts of the bigram as it cropped up in the *Times* (figure 1).[10]

Evidently, the term was in continuous use during the more than two centuries covered by the newspaper. Given that newspapers tend to cover political goings-on, most of the usage will have concerned political journalism. We can test this by comparing the *Times* with debates in the British Parliament between 1800 and 2005 (figure 2).

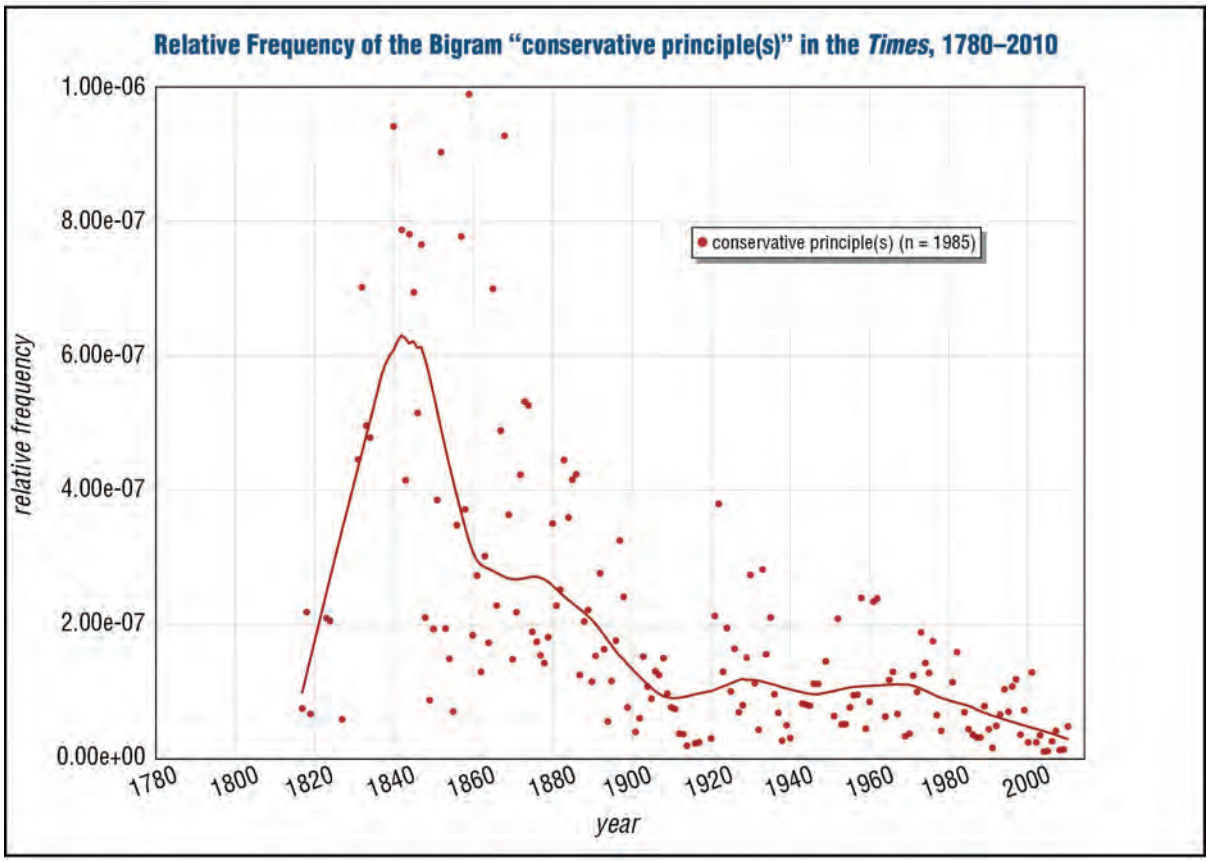


Figure 1. Relative frequency of the bigram "conservative principle(s)" in the Times*, 1780–2010; n = 1,985. Relative frequency is the number of hits in relation to the total number of words per year. The curve is a so-called "smoothing" curve based on a local polynomial regression function ("loess") which allows us to better see patterns.*

The bigram occurred more often in the House of Lords than in the House of Commons but the pattern is quite similar to that of the *Times*. Nineteenth-century MPs were particularly fond of the term; the pattern shows some ups and downs but also a steady decline since 1850.

What, then, were the "conservative principles" to which the *Times* referred in over 1,780 instances between 1785 and 2010? The general pattern is related in part to the passing of fundamental laws. The ups in figures 1 and 2 correspond more or less to the Reform Acts of 1832, 1867, and 1884, the Education Act of 1870, and the so-called Montagu-Chelmsford Reforms of 1919. In other words, when liberal reform was on the agenda, conservative principles were called for. The high relative frequency of the bigram between 1820 and 1850 was due to reports about candidates running for members of parliament. Their reliability was often affirmed during platform addresses and dinner speeches, when orators assured that they were of sound "conservative principles." This kind of journalism disappeared in the second half of the century.

The term "conservative principles" was used mostly in the context of politics. Only in a few instances was it employed in the *OED*'s sense of "a primary element, force, or law which produces or determines particular results" in culture and society.[11] In 1907, for example, human activity was presented as "the result of two complementary principles," imitation, the conservative principle, and initiation, the progressive one.[12] In 1930 Keith Feiling, an Oxford historian and well-known conservative thinker, tried to ground conservatism ideologically in a metaphysics of sorts. According to Feiling, the basic conservative principle is one of continuity, which manifests itself in both "prolonged material strength" and an "inherent spiritual quality." Because national strength

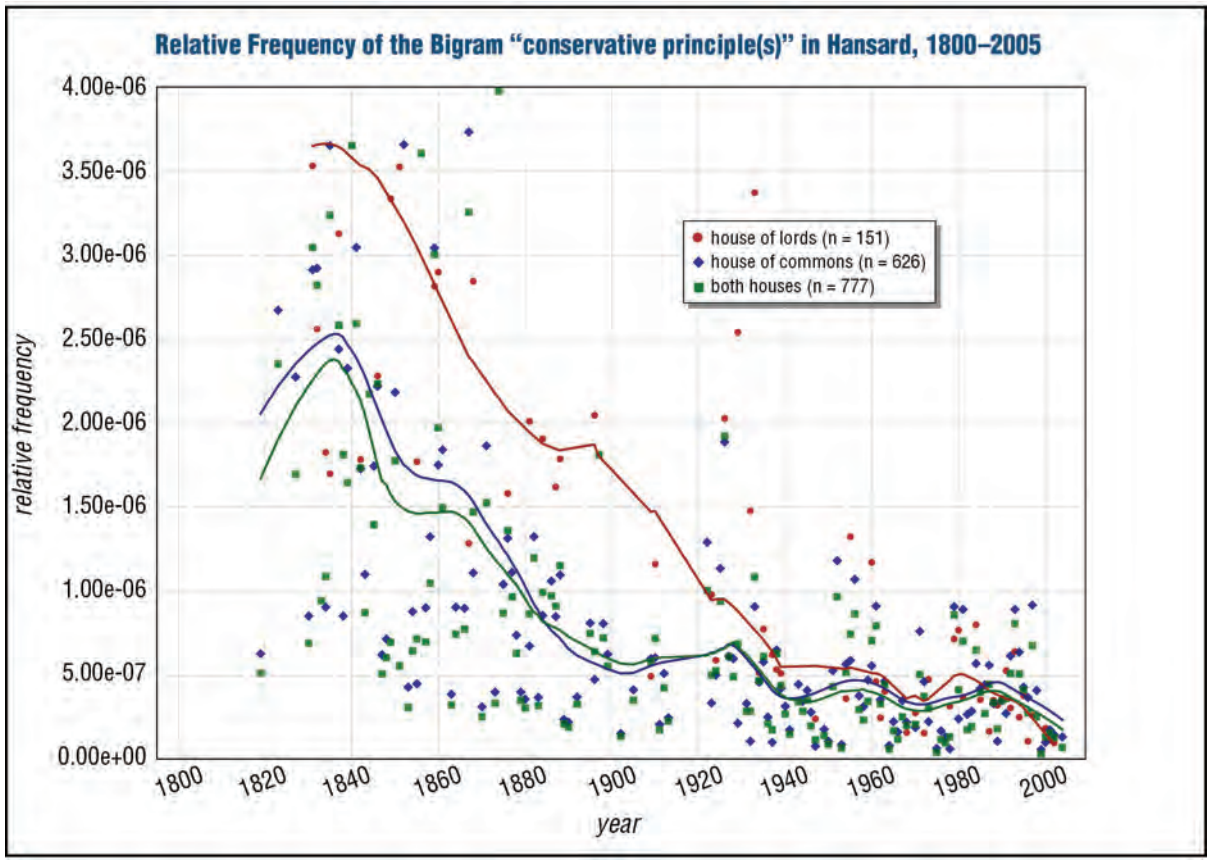


Figure 2. Relative frequency of the bigram "conservative principle(s)" in Hansard, 1800–2005; n = 777. Relative frequency is number of hits in relation to total number of words per year.

is diverse in its forms, "diverse quality and not gross volume is a conservative criterion of national well-being." One way of securing such qualitative differences is through the preservation of class, which must therefore be defended "against the alternative of obliterating it in a universal democracy." Despite his predilection for difference and inequality, Feiling stressed a paternalistic obligation to ensure the well-being of all. What distinguished conservatism from socialism was not concern for the collective, but its aim, which was not equality but "the lasting values of life."[13]

Between 1785 and 2010 the more strictly political usage of the bigram in the *Times* falls into roughly three phases. Before 1830 the term "conservative principles" occurs infrequently and when it does it refers to foreign policy in the context of revolutionary thought and action. Newspaper reports from 1823 mention the "conservative principle of legitimacy" (in France) and "the conservative principle of social order" (in Spain).[14] Such statements reveal the central, counter-revolutionary thrust of early conservative thought. Indeed, the idea of "law and order" would remain a core conservative tenet. It basically entailed the rejection of all ideologies that proposed rapid, radical change, ranging from the French Revolution and nineteenth-century Liberal reform to communism and Labour Party socialism.

Phase 2 (1830–1950) is also in part negatively related, not so much to revolution as to reform and the Reform Movement. One critic helpfully simplified conservative rhetoric by defining its principles as the mere "denial of all comprehensive reform."[15] More justice is done to the conservative rhetoric of the time by summarizing its principles under four related headings: throne, constitution, church, and empire. Law and order remains a central tenet, often under the aegis of the constitutional settlement achieved during the Glorious Revolution (1688–1689)—which was exactly what most Conservatives wished to conserve. In 1841 the *Times* put "the main

principle of Conservatism" in a nutshell by describing it as "the preservation of the ancient constitutional balance of power." The objective of the principle was "to vindicate the integrity and freedom of action of all the classes and constituents which compose the legislative authority, each within its separate limits, without either suffering any one class to encroach upon the rest, or permitting the whole together to isolate themselves from the body of the people."[16]

This was the age of what has been called Benjamin Disraeli's "one-nation conservatism"—a phrase, incidentally, that only began to be used regularly in the *Times* in the 1970s.[17] The idea was a corporate state in which the King, the Lords, and the Commons maintained a peaceful balance, each estate exercising its own rights and privileges to the benefit of the whole. The argument was sometimes flavoured with references to historical or national identity. This "ancient principle and practice of the constitution" was grounded in "the collected wisdom of many individuals."[18] It preserved "that constitution of law, of institutions, of usages, of habits, and manners, which have contributed to mould and form the character of Englishmen."[19] To this constitutionalist and legalist mix were added as further conservative principles the union of church and state and the maintenance of the integrity of the British Empire.

Religion played a significant role in all of this. Society, suggested a clergyman at a conservative dinner in Birmingham in 1838, was best preserved by educating and enlightening the poor, imparting "lessons of moral duty, a wholesome spirit of contentment, a spirit of Christian and not slavish obedience." This enabled the less fortunate to "fill their allotted station as worthy members of society, and grateful sons of the church of Christ."[20] Such paternalism was the necessary corollary of representation by proxy in a situation where suffrage was limited. Hence one critic's definition of conservative principles as "investing a particular individual sitting upon a throne with absolute power" and "a few hundred privileged families (…) with power no less absolute."[21] What nineteenth-century conservatives really meant, of course, was "a fair representation of different classes and different interests," rather than giving "control of the country to a numerical majority."[22]

After World War I, conservative principles in the *Times* were framed in particular in opposition to socialism and the Labour Party. In conservative rhetoric "the Conservative principle of comradeship and the good of all classes alike" was contrasted "with the Socialist desire to limit the expression of comradeship to one class alone."[23] A group of conservative "Die-Hards" included the empire in their rather standard list of principles; its global presence allowed Britain to secure "to peoples less advanced than ourselves the priceless gift of just and civilized government."[24] Stanley Baldwin, one of the more significant conservative leaders of the period, likewise reiterated the four "eternal principles" of religion, throne and constitution, empire, and the people's welfare, laid down repeatedly by Disraeli (and now promoted by the Primrose League).[25]

The third phase (1950–2010) saw a decline in use of the bigram "conservative principles," suggesting that political leaders were looking for ways to give new content to the expression. Winston Churchill still quite traditionally discussed the "free and flexible working of the laws of supply and demand," and thought that this principle ought to be balanced by "compassion and aid for those who, whether through age, illness or misfortune, cannot keep pace with the march of society."[26] But in Edward Heath's 1965 programme "Putting Britain Right Ahead," change and reform had been fully internalised and paternalism seemingly cast aside. The Empire was on the wane and social rank had become less obviously important; in this situation it would be wise to recognise the "reality of Britain's position in the modern world." What now qualified as conservative principle was the idea that "the individual human being" was "the mainspring of change, modernization and reform."[27]

The 1970s and 1980s retained the individualism but saw a reorientation towards cultural conservatism. With the rise of the New Right the bigram "conservative principles" began to mean less than before as a political rallying cry. Tellingly, the "Reaganite" (and Thatcherite) neoliberal

"creed of strong international leadership, free trade, smaller government, lower taxes, deregulation and cultural conservatism" was now referred to as a set of "neo-conservative principles."[28] This quotation is from the 1990s, however, when both Reagan and Thatcher had already relinquished power. In general, these years had witnessed the rise of the American moral majority and a turn to right-wing politics under George Bush Sr., whose programme "Empower America" was presented as an alternative to "the isolationist, protectionist 'paleo-conservatism' championed by Pat Buchanan."[29]

Times had changed. Principles were out. In the 1990s John Major tried to revamp traditional "one nation conservatism," while Michael Portillo advised the Tory leadership to "reassociate" the Tory party with "compassion and magnanimity."[30] That was a clear indictment of Thatcher's neoliberalism. Letters to the editor confirmed the doubts entertained by some conservative-minded readers concerning Thatcher's politics. "For too long the Conservative Party has been hijacked by free-market dogmas, spawned in American universities by securely tenured academics," wrote one reader of the *Times*. He spurned an economic doctrine that clearly owed "more to 19th-century liberalism than any recognisably Conservative principles." What the Tories needed was "a compassionate, flexible and inclusive politics in keeping with the best Conservative traditions."[31] If the increasingly confusing word usage in the *Times* is anything to go by, the conservatives of the 1990s needed a new rhetoric.

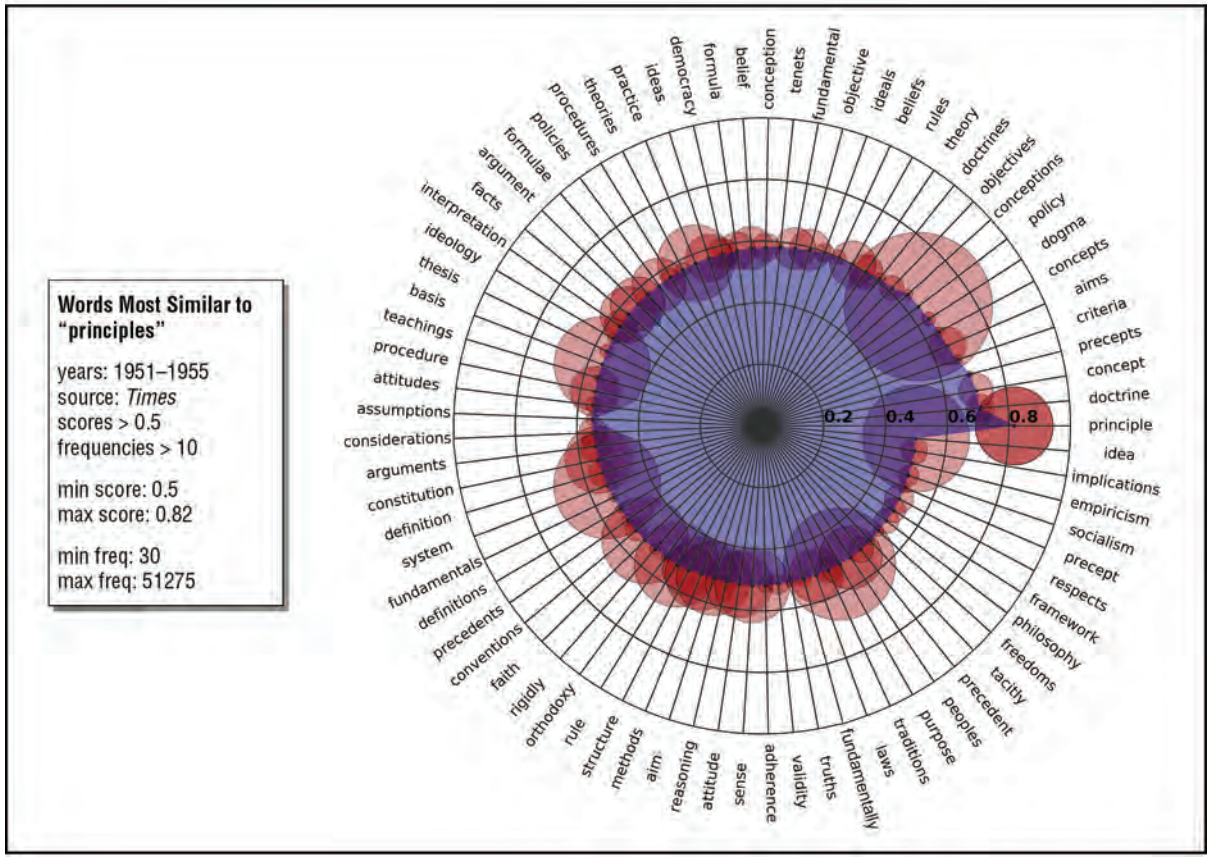


Figure 3. Words most similar to "principles," based on unigram embeddings of the Times *(1951–1955). The radial graph shows the similarity score of each word in relation to the word "principles," for all values equal to or higher than 0.5; the higher the score, the greater the similarity. The red circles reflect the occurrence of each word in the dataset for the period in question; words with frequencies lower than 30 have been discarded.*

Top Twenty Words Most Similar to "principles"

1851-1855	1901-1905	1951-1955	2001-2005
doctrine	doctrine	doctrine	tenet
basis	policy	concept	doctrine
theory	idea	precept	belief
truth	proposition	criterion	concept
dogma	notion	aim	value
system	theory	dogma	norm
policy	tenet	policy	framework
opinion	axiom	conception	philosophy
tendency	method	objective	ideology
tenet	conception	theory	ideal
maxim	formula	rule	ethos
notion	ideal	belief	constitution
dictate	belief	ideal	policy
interest	shibboleth	fundamental	precept
Idea	dictate	tenet	concept
consideration	tradition	formula	approach
wisdom	dogma	idea	objective
reasoning	usage	practice	purpose
rules	consideration	procedure	rule
assumption	teaching	argument	notion

Table 1. Top twenty words most similar to "principles," based on word embeddings for four periods; duplications resulting from plural forms have been removed.

FROM PRINCIPLES TO VALUES

Until World War II the bigram "conservative principle" had been a common enough expression, a turn of phrase preferred by those who spoke for conservatism, and recognized as such by their critics. But were there other ways in which conservative writers could articulate basic beliefs that expressed a similar sense of moral purpose? Were phrases analogous to the bigram "conservative principles" in use, in which a noun referring to something fundamental (such as "principles") was qualified by an adjective referring to something ideological (such as "conservative")? One might expect "conservative beliefs" or "conservative tenets" to have been used, but was this, in fact, the case?

Ideally, we would be able to detect phrases similar to "conservative principles" automatically by using bigram embeddings.[32] The top synonym for "conservative principles" in the years

Number of Occurrences of Most Similar Words Preceded by "conservative" for All Years (1785–2005)

bigram	occurrence	range
conservative policy	5407	1830-2010
conservative interest	2340	1830-2010
conservative basis	1257	1840-2010
conservative element	778	1845-2010
conservative opinion	684	1835-2010
conservative approach	368	1930-2010
conservative attitude	366	1880-2010
conservative rule	352	1840-2010
conservative tradition	234	1850-2010
conservative tendency	192	1830-2010
conservative value	216	1910-2010
conservative method	155	1885-2010
conservative philosophy	151	1930-2010
conservative aim	118	1880-2010
conservative idea	117	1855-2010
conservative legislation	91	1840-2010
conservative system	87	1830-2010
conservative assumption	80	1930-2010
conservative doctrine	80	1830-2010
conservative ideal	78	1890-2010

bigrams	occurrence	range
conservative law	70	1830-2010
conservative convention	60	1865-2010
conservative faith	56	1880-2010
conservative argument	53	1885-2010
conservative belief	49	1925-2010
conservative standard	39	1830-2010
conservative practice	35	1930-2005
conservative ideology	30	1970-2010
conservative standpoint	25	1885-2010
conservative notion	13	1840-2005
conservative dogma	11	1870-1995
conservative objective	11	1970-2005
conservative orthodoxy	10	1980-2010
conservative concept	6	1980-2000
conservative definition	6	1970-2010
conservative interpretation	6	1970-2005
conservative theory	6	1970-1990
conservative constitution	5	1975-2005
conservative structure	5	1970-1990
conservative tenet	5	1840-1870

Table 2. Number of occurrences of most similar words preceded by "conservative" for all years (1785–2005), including a rough indication of the year range of their occurrence.

2001–2005 transpires, however, to be "winning elections," while the bigrams "hard-line minority" and "tasted blood" follow close behind. Although evidently related to politics, these rather unspecific results indicate that the bigram "conservative principles" simply does not occur often enough for phrase embeddings to make much sense. The bigrams most similar to "conservative

principles" for the years 1876–1880 (when the bigram occurred more often) are more meaningful, with such phrases as "liberal principles" and "political convictions." The latter sounds promising: however, a bigram such as "conservative convictions" occurs only a dozen times in the whole corpus.

The unigram "principles" occurs with much greater frequency (more than 300,000 hits between 1785 and 2010) so that it should make more sense to use unigram embeddings to find out which words were given a similar meaning to "principles." For brevity's sake we will look at the top 100 most similar bigrams in four brief periods only, skipping half a century at a time: 1851–1855, 1901–1905, 1951–1955 and 2001–2005. The results for the third period (1951–1955) are shown by way of illustration (figure 3).

Table 1 shows the top twenty synonyms per period.

Were any of these words used in the context of conservative rhetoric? Examining bigrams in which "conservative" is the qualifying adjective is the most straightforward way to test this. By combining each of the top one hundred most similar words for each period (i.e., 400 words in total) with the adjective "conservative" and subsequently removing all duplicates (including the plural form of each word) we obtain the result in table 2.

The first ten bigrams do not convey the strong sense of ethical worth carried by the term "principle." The bigrams that do convey this connotation, and also occur with meaningful frequency over a longer period, are "conservative tradition(s)," "conservative value(s)," "conservative philosophy," "conservative idea(s)," and "conservative belief(s)." While conserva-

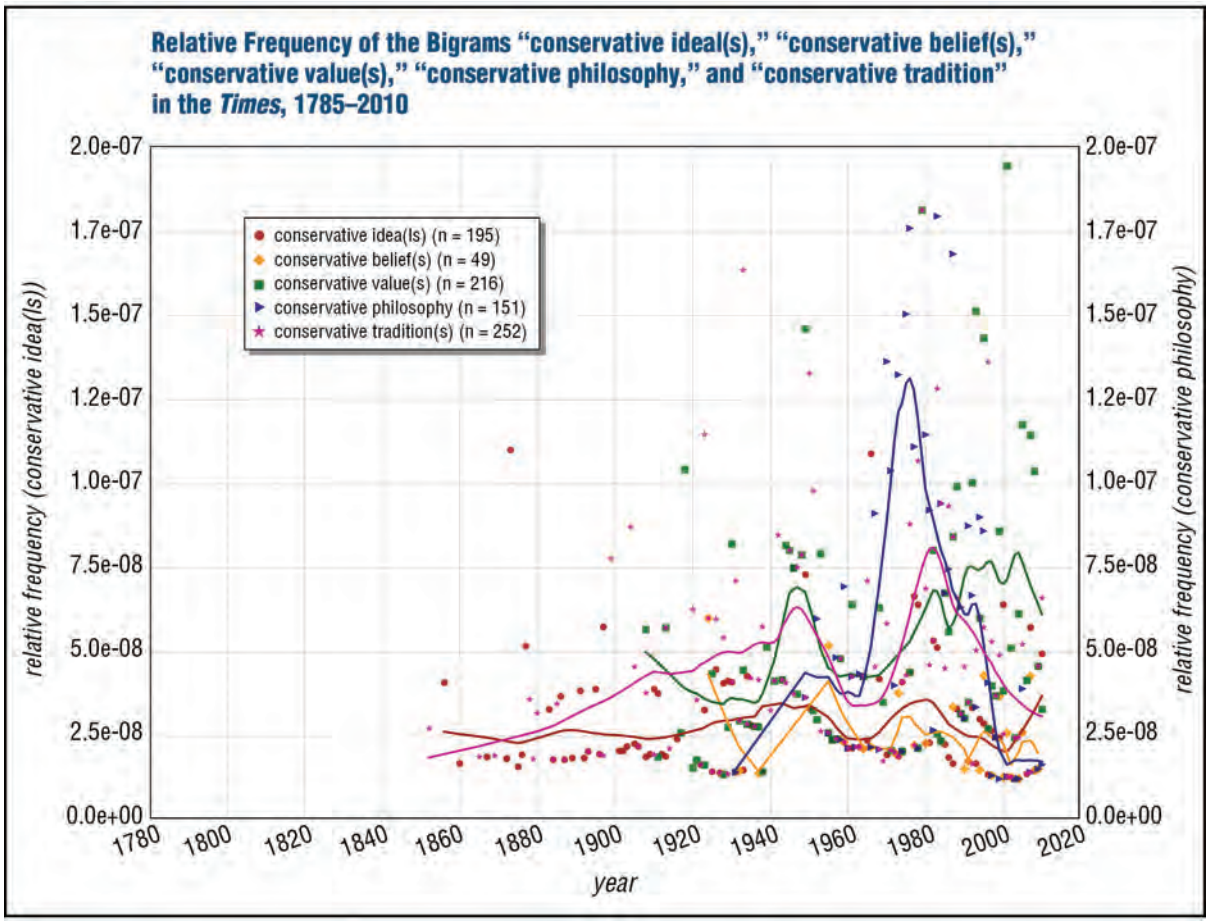


Figure 4. Relative frequency of the bigrams "conservative ideal(s)" (n=195), "conservative belief(s)" (49), "conservative value(s)" (216), "conservative philosophy" (151), and "conservative tradition" (252) in the Times, *1785–2010.*

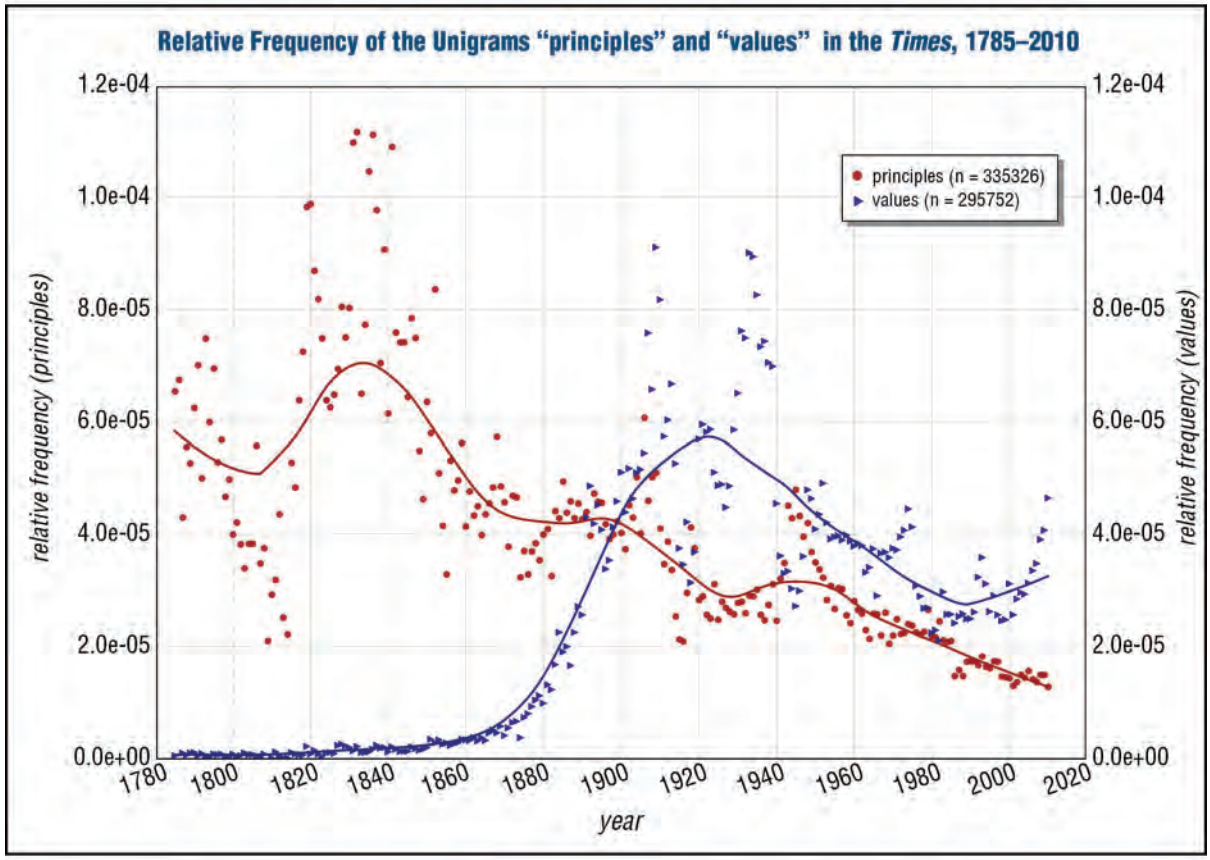


Figure 5. Relative frequency of the unigrams "principles" (n=335,326) and "values" (n=295,752) in the Times*, 1785–2010.*

tive traditions, ideas, and ideals were already present in the nineteenth century (figure 4), the more ethically imbued bigrams are largely a twentieth-century affair.

"Beliefs" are not very prominent, but, interestingly, "philosophy" becomes rather popular after about 1970 and "values" after 1980. The latter is probably the bigram that comes closest to the moral sense of "conservative principles." We will therefore turn to "conservative value(s)" as a phrasal soundbite that became more common to conservative rhetoric once conservative principles had become all but extinct.

THE RISE OF CONSERVATIVE VALUES

The word "values" began to be used regularly towards the end of the nineteenth century, when the unigram "principles" was already on the wane (see figure 5).

As an element of bigrams, however, values were then predominantly if not exclusively connected to economy, finances, and trade, as in "closing values" or "market values." The *OED* suggests that the ethical rather than economic or monetary connotation of values, as "the principles or moral standards held by a person or social group; the generally accepted or personally held judgement of what is valuable and important in life," originated in the USA.[33] But it became a common enough word in British English. And, *pace* Willetts, it was not a term restricted to left-wing rhetoric.

The first relevant use of the bigram "conservative values" stems from an account of the proceedings of a "Conservative Conference" in 1961. "I hope this conference will put Conservative values to the test," Lord Home, foreign secretary under Harold Macmillan, was reported as saying, "and see that those values give a clear lead to the people."[34] True to the spirit of pre-war Toryism, the values he mentioned implied that actions should "all stem from the principles of Christian religion," be "steeped in loyalty to the Crown, the constitution and the law" and "inspired by the desire to serve all."

Like conservative principles, conservative values more often than not referred to political values: those values supported by adherents to the Conservative Party or analysed by conservative political philosophers. Chris Patten, the later governor of Hong Kong, was the person who (if we follow the *Times*) in 1982 initiated in-depth talk about conservative values, in an article on "Why Mrs Thatcher should join the real Tories" and thus "return to traditional Conservative values." The terms he used in this connection included such conservative evergreens as proportion and balance, values that implied pragmatism but in Patten's view also community. Man, he argued, is "a social animal who can reach his full stature only in groupings greater than himself, such as his country, his church or his family."[35]

Margaret Thatcher would echo this position six years later, although she magnified the family as the locus of conservative values. That distinctive community, moreover, was a territory presided over specifically by women. "For the family," she posited in a celebrated speech to the Conservative Women's Conference in 1988, "is the building block of society. It is a nursery, a school, a hospital, a leisure centre, a place of refuge and a place of rest. It encompasses the whole of society. It fashions our beliefs. It is the preparation for the rest of our life. And women run it."[36] Rather than women, however, it was the "traditional" family that became central to conservative rhetoric. In newspaper accounts between 1980 and 2010, the bigram conservative values was employed in three contexts, which all point to a tendency to harness the "traditional" to the cause of conservatism.

The first context concerned the younger generation of the "post-permissive society," who apparently no longer found "radical protest" attractive. Instead, they favoured conservative values like "old-fashioned patriotism, religious belief and support for 'law and order' policies," or, alternatively, took a more conservative stance regarding marriage and homosexuality.[37] Such assessments would prove to be incorrect, but they were relatively common in the 1980s. The second context was the USA, where Reagan and the "New Right" worked to "restore" conservative values and remove "the evils of abortion, bussing, pornography, reverse discrimination and institutionalized atheism."[38]

The third context was British politics. In her 1988 speech on women as time-tested managers, Thatcher had mentioned self-reliance, personal responsibility, good neighbourliness, and generosity as "the Conservative values." Barring the neoliberalism she espoused[39] her ideas remained popular. Every now and then someone would hark back to "the old Conservative values of freedom, nation, family and responsibility," on which Tory politician Keith Joseph had built Thatcherite policies in the 1970s.[40] John Gray however, made short shrift of Thatcherism. In a more or less prophetic article called "Why Tories should vote Labour" he lamented the way Tories had come to spurn the time-honoured objectives of traditional conservatism, "the nurturing of communities and the renewal of civic institutions in Britain." Thatcher's neoliberalism had only increased "risk and uncertainty in the everyday lives of ordinary people." Unfortunately, conservatives now flouted the "norms of fairness and decency which are deeply ingrained in British culture." Thatcherism had given rise to "a divided and demoralised society in which crime is endemic, and family life neglected and fractured."[41]

In the decades that followed no Tory politician, whether pro or anti-Thatcher, could afford not to mention what he or she regarded as the traditional (and invariably nuclear) household. "We believe in the family," reiterated David Cameron time and again. But as a journalist observed in 2005, Labour now too believed in families. Together with all the other conservative values Cameron put on his list—personal responsibility, lower taxes, high standards of health and

education, limited government, national sovereignty—the traditional family appealed to the Left as much as it did to the Right. Values, apparently, had not been the best choice for replacing principles.

MORALISING POLITICS

All this conservative talk about "conservative values" was of course highly normative, if not downright moralistic. That raises the question as to the broader linguistic context in which these words were being used: did values arise in another context than principles? Again, we can use word embeddings to find which meaningful associations with "conservative" existed between 1785 and 2010. Because usage of the word "conservative" itself in the *Times* is almost exclusively focused on politics ("Conservatives" and "Conservative Party" are extremely common) we need to identify a broader set of words worth investigating. These words need to have relatively stable meanings over the whole period, to facilitate comparisons.

I selected nine words: "conservative" (the central term), "constitutional" (as a political term associated with conservative), "traditional" and "customary" (as related in meaning to the adjective conservative), "Christian," "religious," "spiritual" and "moral" (indicating the sources of religious and secular morality conservative writers will often have referred to), and "civilised" (signalling the quality of being moral, again a word one might expect conservatives to be fond of).[42] For each of these words, in each of the three periods 1901–1905, 1951–1955, and 2001–2005, I generated the top fifty most similar words; for each of the these fifty words I again calculated another top fifty most similar words (figure 6).

This resulted in 22,500 words in total (9 x 50 x 50), with the number of unique words ranging from 5,000 to 7,500. The data was then visualised as a network, in which the unique words figure as nodes and the "similarity scores" (showing the degree of similarity between words) as the edges.

By forcing the network to cluster automatically into three groups, a rough pattern emerges.[43] In the first period (1901–1905, figure 7), two clusters are visible, which, judged by their semantic import, are clearly "political" (blue) and "civilisational" or "moral" (orange). Half a century later (1951–1955, figure 8), three clusters are visible. Again, we see a political and a civilisational cluster, and a dark green cluster relating to words most similar to "traditional" or "customary." The surprise is in the third period (2001–2005, figure 9), where the same algorithm with the same settings identifies only one cluster, which I have identified as "civilisational," based on the semantic import of the majority of words.

This approach is imprecise and needs to be explored further, but we can draw one obvious conclusion. The pattern appears to indicate what we might call the historical "moralisation" of politics: between 1955 and 2001 the semantic relations between the "political" and the "civilisational" became stronger. The transition we have described so far from conservative principles to conservative values fits into this pattern. This suggests that, if we want to fully explore the nature of conservative rhetoric, we must do more than look only at "conservative values"; we should examine value-laden phrases in which a political qualifier (such as "conservative") does not necessarily appear. A logical choice to test the occurrence of such phrases is simply to look at the bigram "traditional values," since the conservation of tradition is what conservatism is supposed to be about, and values now lay at the heart of political discourse. The bigram indeed proves to be meaningful. Where "conservative value(s)" had an incidence of 165, "traditional value(s)" rocketed after 1950, with a total score of 2,095 (figure 10).

Somewhat surprisingly perhaps, the bigrams related to national identity ("national," "British," and "English" values) all lag far behind. Let us take an extensive look at the bigram "traditional values."

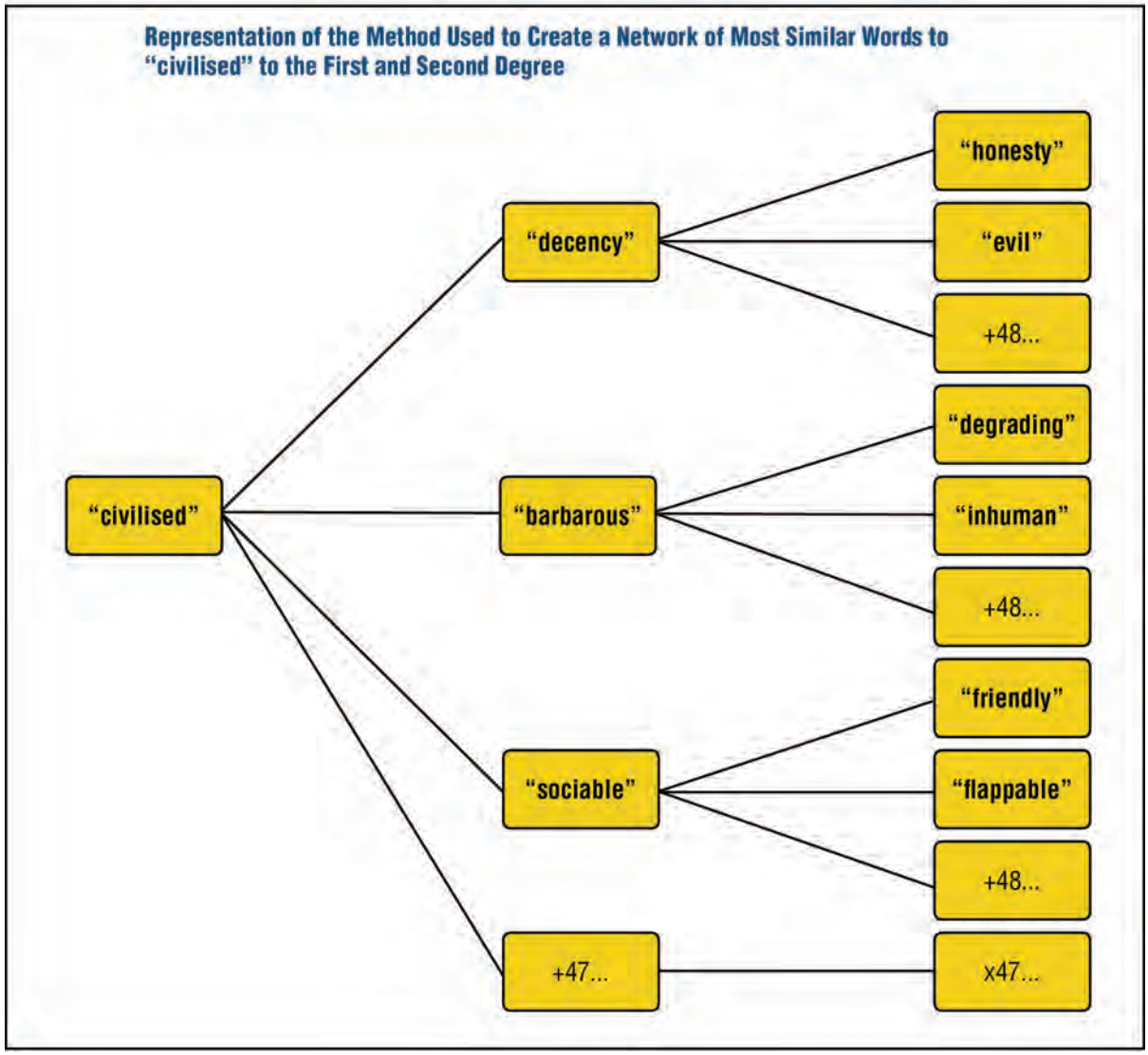


Figure 6. Representation of the method used to create a network of most similar words to the first and second degree; the example is based on the word "civilised."

TRADITION VS. MODERNITY

In which context did the bigram "traditional values" occur in *the Times*? It materialised in the first half of the twentieth century, for example in 1934 when Thorold Coade (1896–1963), headmaster of Bryanston School in Dorset, described education as "handing on the culture of the past to the rising generation." This approach, he argued, implied character building, "the perpetuation of traditional values," disciplined scholarship, organised games, and the prefect system.[44] Whether Coade voted for the Conservative Party is not directly relevant here; the point is that the values he appealed to can be called conservative. And underlying this kind of appeal to things traditional was an opposition to, or rather a tension with, modernity.

Sometimes this friction was expressed in overtly ideological terms. Bolshevism, Adolf Hitler was reported as saying in 1935, sacrifices to its theory of a classless society "millions of human beings and incalculable cultural and traditional values," achieving very little in the process.[45] But more often the bigram "traditional values" was employed to indicate a problem rather than a

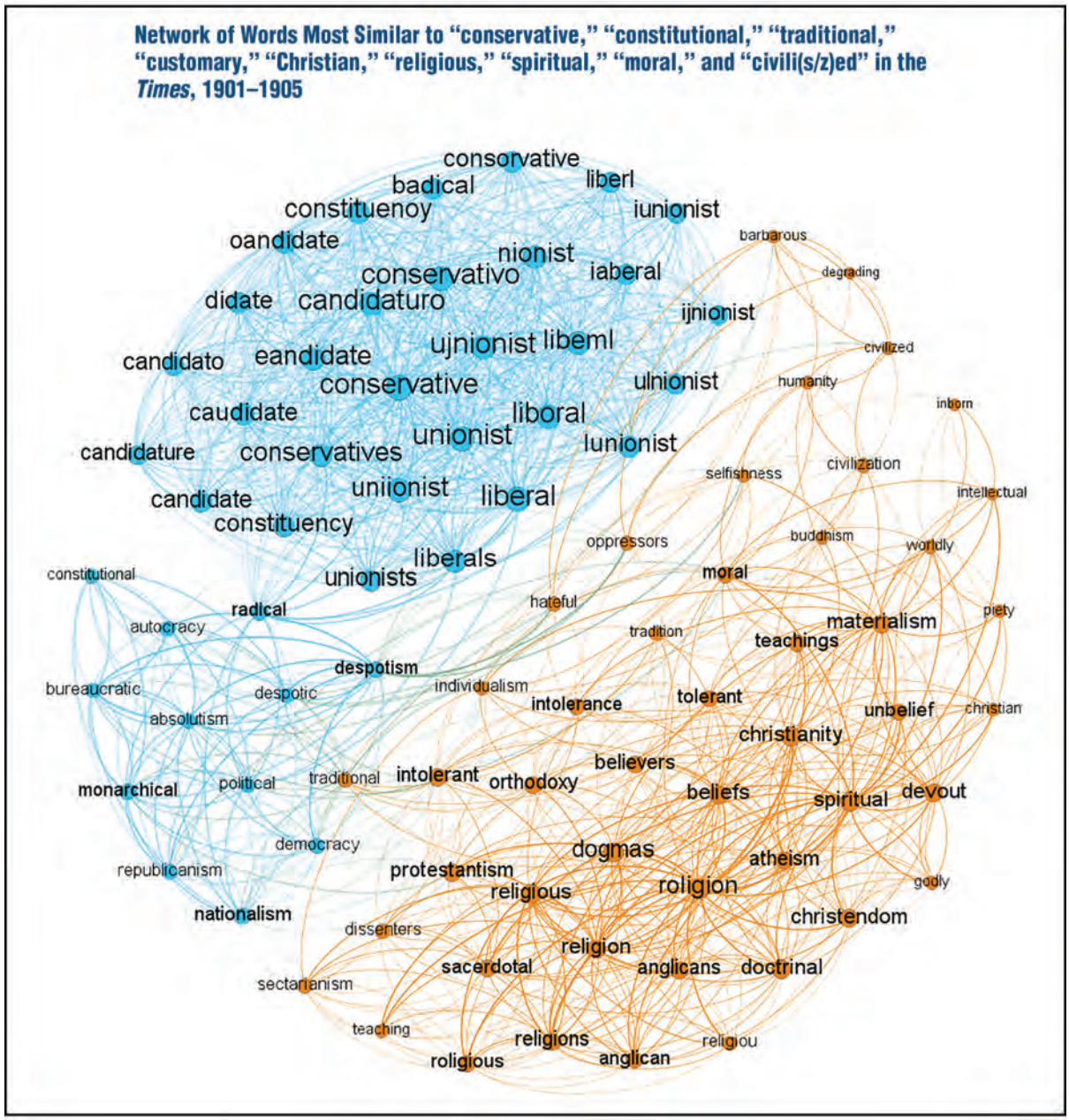


Figure 7. Network of words most similar to "conservative," "constitutional," "traditional," "customary," "Christian," "religious," "spiritual," "moral," and "civili(s/z)ed," to the first and second degree. Based on unigram embeddings of all words in the Times*, 1901–1905 (no OCR correction).*

doctrine. That the twentieth century was an age of unprecedented, rapid change is perhaps a cliché; but it did witness the passing of a traditional way of life for many people. Most were uncomfortable with that change, whether they were explicitly conservative or not (see figure 11).

In 1970 in Japan the novelist Yukio Mishima, who, admittedly, was "dedicated to the restoration of full imperial rule" and in that sense an outright reactionary, even committed *seppuku* over the loss of traditional values.[46] In the *Times* this discourse about tradition and modernity arose especially in the 1950s. Pakistan's new constitution, for instance, needed to harmonise secular law with the Quran. The *Times* noted: "The task of reconciling traditional values with modern needs has been complex and delicate."[47]

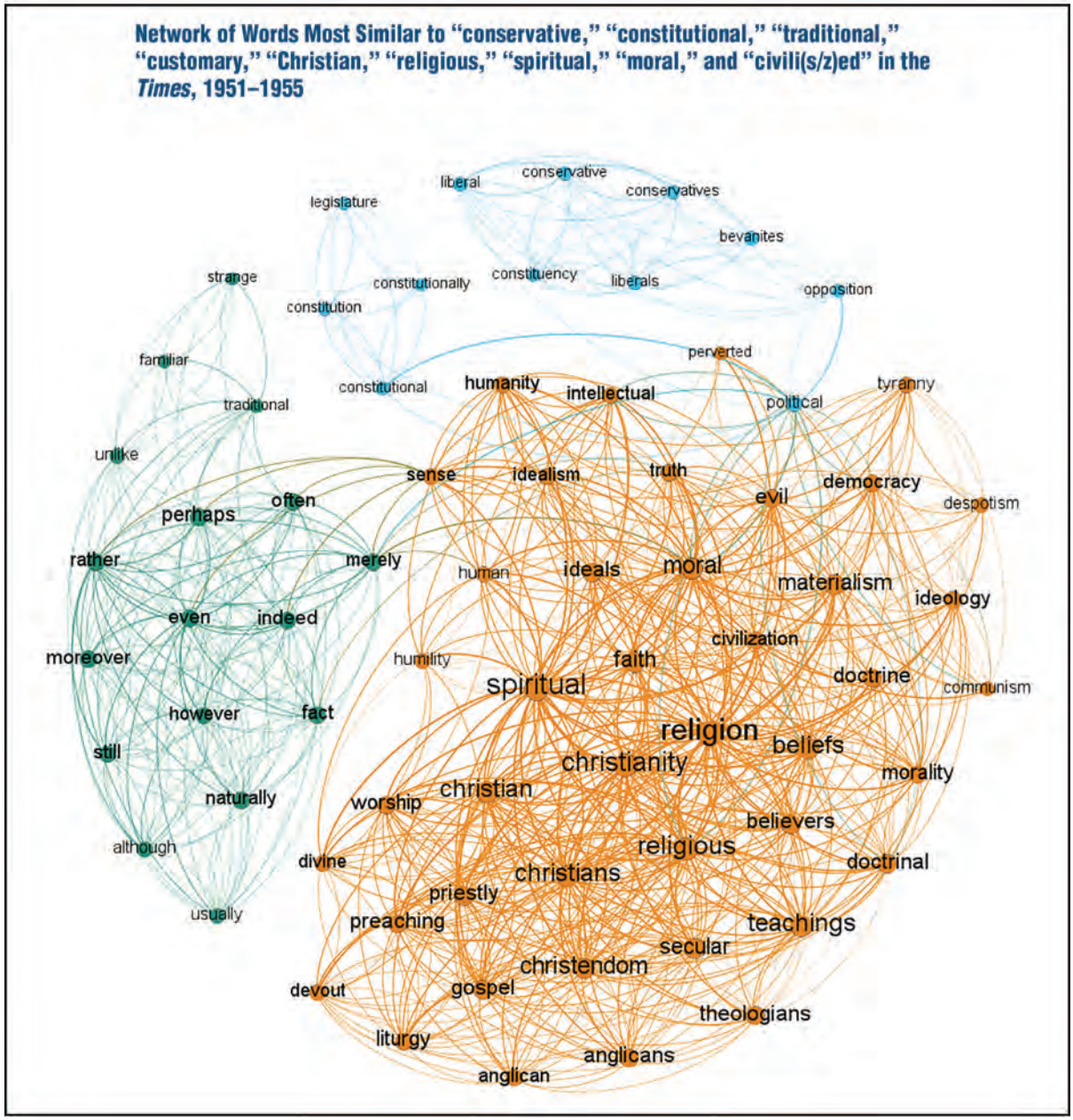


Figure 8. Network of words most similar to "conservative," "constitutional," "traditional," "customary," "Christian," "religious," "spiritual," "moral," and "civili(s/z)ed," to the first and second degree. Based on unigram embeddings of all words in the Times*, 1951–1955 (no OCR correction).*

Modernity related to tradition in various ways. This association often took place on the level of a society or state, as in the case of Pakistan's constitution. Sometimes modernity was equated with Western society in particular, as in a report on Southeast Asia, where traditional values were "being swept aside by the mass products of the West."[48] Whether the tradition/modernity drama was framed positively or negatively or neutrally, it played out everywhere, from Zambia, Ceylon, Canada, and Sardinia to Zaire, Iran, and Morocco.[49] Tanzania's *ujamaa* socialism was represented not as the negation but rather as the transformation of traditional values: "solidarity, collective ownership and work-sharing" simply became "democracy, egalitarian sharing of

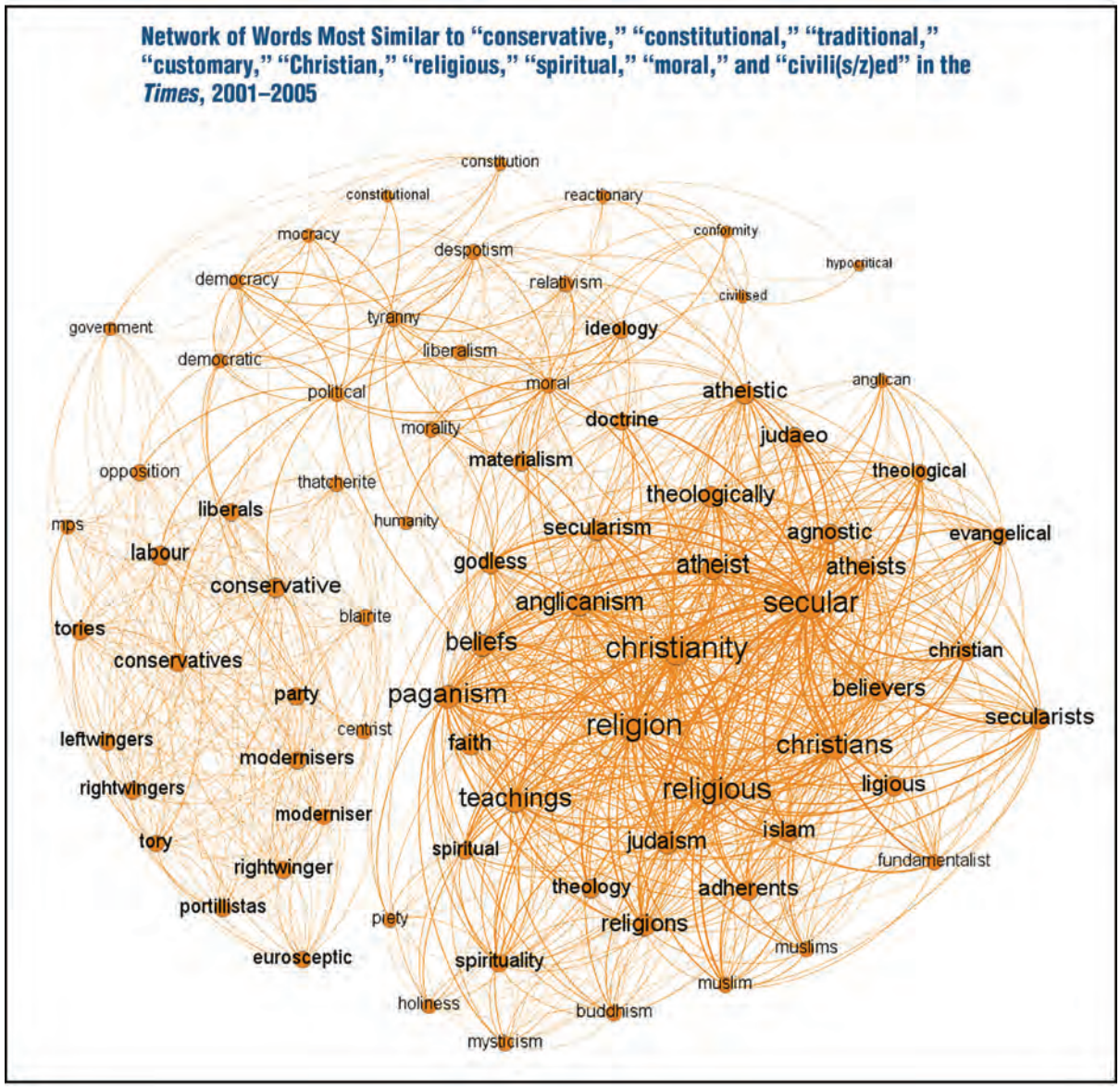


Figure 9. Network of words most similar to "conservative," "constitutional," "traditional," "customary," "Christian," "religious," "spiritual," "moral," and "civili(s/z)ed," to the first and second degree. Based on unigram embeddings of all words in the Times*, 2001–2005 (no OCR correction).*

resources and cooperation."[50] The exception, apparently, was Chile, where in 1970 traditional values still remained intact. "These values uphold constitutional and orderly government, a respect for legality, politeness and hospitality between people, education, and a fierce regard for all things Chilean."[51]

Business values were an important domestic motif. In 1965 the *Times* gave managers a wake-up call. Technological change was leading "to the evolution of a new science-based culture in industry in direct conflict with traditional values." Man's natural reaction, however (and that of a British man particularly), was "to cling more resolutely to his old and once successful customs and habits."[52] Traditional values had made Britain great, economically speaking. But old-fashioned qualities—a "high degree of technical competence and complete integrity and impartiality"—were now no longer an asset if British business failed to modernise.[53]

Another obvious theme was the metropolitan city as the seat of modernity contrasted to the countryside as the bulwark of traditional values. Western France tended to vote Roman Catholic and conservative, demonstrating that "provincial life and opinion is less excitable, more down to earth, and more suggestive of an underlying stability."[54] Gender roles were yet another thread in the *Times*. The French feminist Evelyne Sullerot, arguing for women's participation in civic life, portrayed tradition as a masculine concern. "In an increasingly technological society, it is reassuring for men to keep women as their link with traditional values, bending over stewpots as their mothers and grandmothers did."[55]

When gender was combined with reverence for the rural, nostalgia tended to triumph over emancipation. "It is encouraging," wrote one (male) art critic, "to think that the traditional values are being preserved somewhere." The Mazowsze Song and Dance Company from Poland championed the cause. It represented "all the old-fashioned virtues," with "coy and pretty" women in voluminous petticoats modestly taking care not to reveal their undergarments while dancing the mazurka with "active and attentive" men carrying agricultural tools.[56] To some extent such exuberant reviews are timeless: tradition is always good and nostalgia always sells well, even when it comes expensive. A hotel in Keswick, Cumbria, offered a pricey return to traditional values, including freshly "ironed morning papers, evening dress codes, old sixpences for tipping and ancient copies of Punch."[57] At another in Sidmouth, Devon, guests might enjoy "the gentle sounds of cricket on the village green, a stroll along the promenade, or a Traditional English afternoon tea on the lawn."[58] Meanwhile, the Hamper People's gift baskets celebrated "traditional values in a modern world."[59] The popularity of traditional values in advertising was a sign that things were changing.

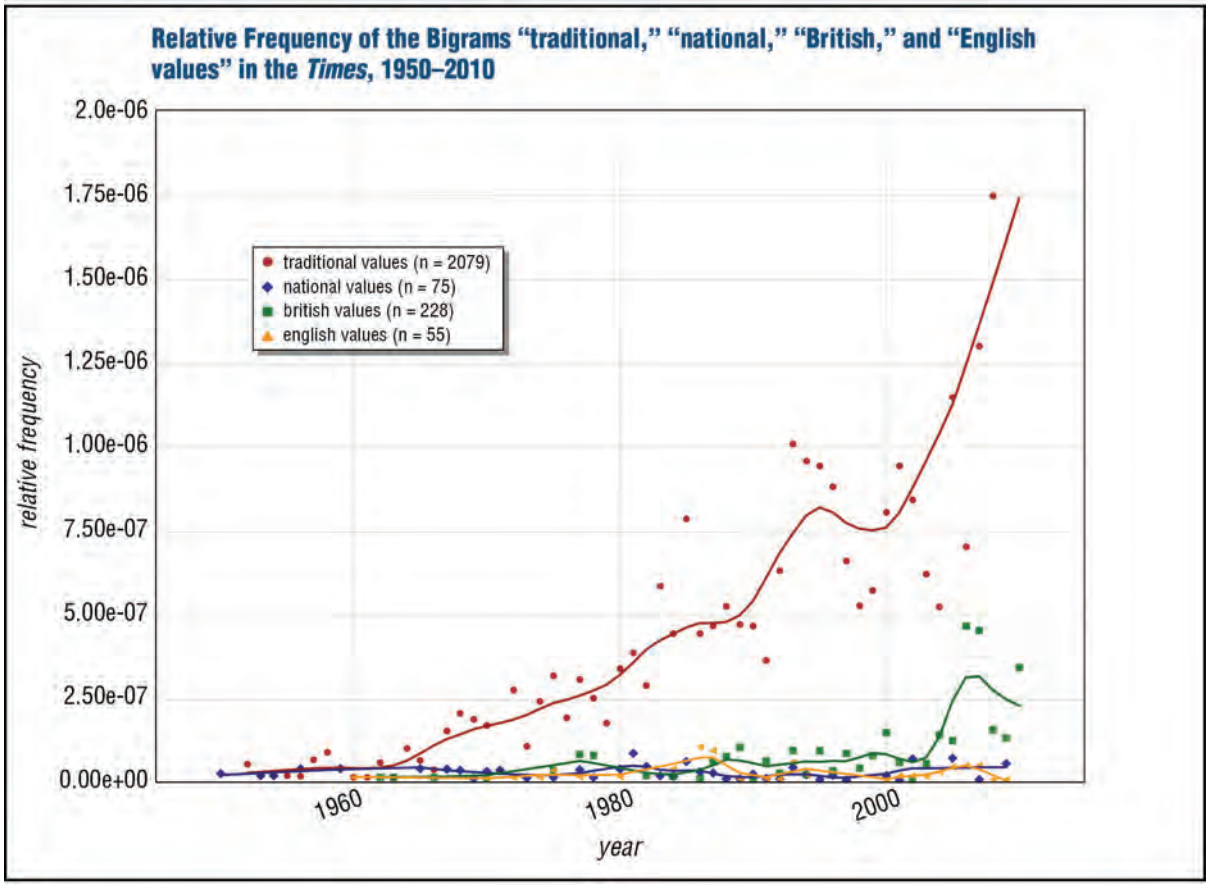


Figure 10. Relative frequency of the bigrams "traditional" (n=2,095), "national" (412), "British" (252), and "English values" (61) in the Times*, 1950–2010.*

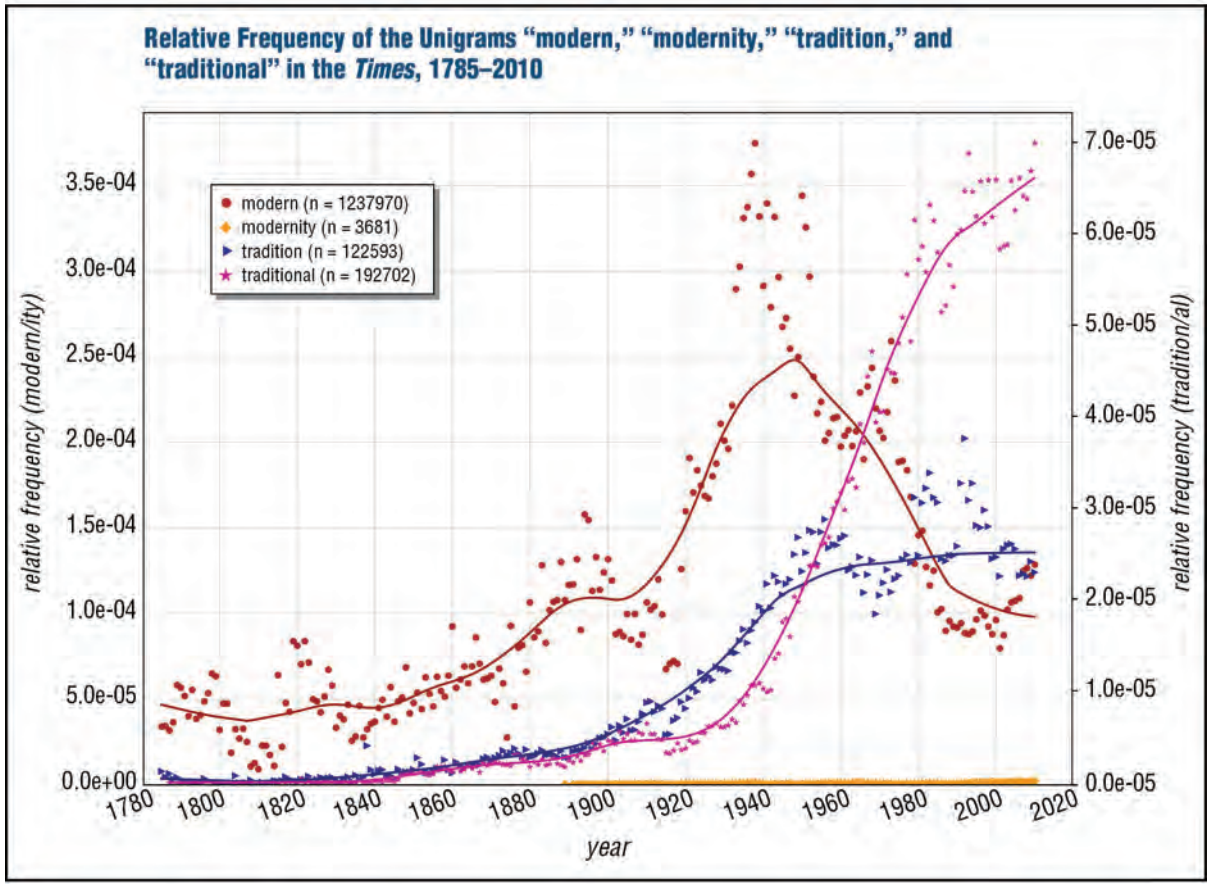


Figure 11. Relative frequency of the unigrams "modern," "modernity," "tradition," and "traditional" in the Times*, 1785–2010. Note that* y axes have different scales.

By the early 1980s traditional values had turned from a liability into an asset. This is evident from classified ads, which now began to invoke them indiscriminately. Boarding schools paraded their attachment to traditional values.[60] Car manufacturers summoned the moral past. A reviewer applauded the Series 200, a "breed of Rover that combines traditional values with more than a little panache."[61] The construction company Mowlem successfully blended "traditional values and skills with a visionary approach to today's challenges."[62] The Isle of Man seduced companies with "a way-of-life that emphasises traditional values of friendliness and independence" (and much lower taxes).[63] Next to the Post Office and the mature 1997 Chateau Cissac, companies in real estate (Hillier Parker), law (Walker, Smith & Way; Reynolds Porter Chamberlain), housing (Beazer), medicine (Bayer), yacht building (Westerly), and consultancy (ADTI Consulting; Connaught) all presented their ability to reconcile tradition with modernity as a unique selling point until 2010, and probably beyond.

In the tradition-vs.-modernity debate anybody could take in any position; it was largely a debate about socio-cultural change. So when did the friction between tradition and modernity mutate into a conflict between conservatism and progressivism? The *Times* included some attempts at dispassionate analysis, which pointed to a fundamental ideological transformation. Long before the re-emergence of Western populism and the protests of a disenfranchised middle class, radical Liberal journalist Nesta Wyn Ellis warned that "fascism" was lurking around the corner. A sense of "social dislocation resulting from the breakdown of traditional values" encouraged "patriotism and racial identity" especially among "a seemingly beleaguered bourgeoisie of small shop-keepers, business people and professionals whose status and security are thus at

risk."[64] But in the minds of those who attached importance to traditional values, the main culprit was not immigration or globalisation, but the protest generation. Modernity, "sixties" style, was responsible for undermining the traditional values.

The Governor-General of Canada, Georges Vanier, and his wife Pauline had already founded the Canadian Conference of the Family in 1964 (later the Vanier Institute of the Family), to recapture the traditional values and spiritual heritage so essential to society.[65] But once all hell had broken loose, once "long hair, short skirts and obscene entertainment" had brought about the wholesale "ruin of traditional values,"[66] right-wing commentators and politicians from Italy to Britain began to climb the barricades themselves. Sometimes comments were extremely nuanced. The conservative editor of the *Times*, William Rees-Mogg, famously impressed his readership by taking British justices to task for harshly prosecuting "Mr. Jagger" on a drug charge. "Who breaks a butterfly on a wheel?," ran the editorial's headline: "If we are going to make any case a symbol of the conflict between the sound traditional values of Britain and the new hedonism, then we must be sure that the sound traditional values include those of tolerance and equity."[67] But society was changing, and it was changing fast.

AGAINST THE PERMISSIVE SOCIETY

Already in the 1960s the self-styled "guardians of tradition and morality" had been railing against the "permissive society," this unexpected, head-on collision with authority, this wilful destruction of the "spiritual realities on which all healthy society is based." To prevent traditional values from being permanently lost, "we shall all have to learn that the fear of the Lord is the beginning of wisdom."[68] Unsurprisingly, the conservative rejection of what was seen as surrender to unrestrained self-gratification was often expressed in religious terms. Church of England bishops eagerly entered into the debate, although some, like Russell Barry, wisely took the middle ground. "In the present mood of public opinion it is thought that an absolute, normative morality is a violation of individual freedom." If progressives undermined traditional values, moralists "equated absolute moral norms with inflexible and unchanging moral rules."[69]

The moralists, religious or otherwise, were there to stay. Religious organisations arose to preserve what they thought had been lost. Pro Fide, an organisation opposed to "progressive" influences in the Roman Catholic Church in Britain, reasserted traditional values.[70] The Order of Christian Unity "pledged to defend traditional values of British life";[71] Family Solidarity upheld "traditional values and culture," which came down to a condemnation of family planning and divorce.[72] Once the AIDS epidemic became really serious, persons of a conservative but unrealistic persuasion (m/f) had a field day. As one reader of the *Times* advised: the "traditional values of chastity and fidelity within a permanent relationship and the high ideal of one partner for life should be actively promoted."[73] Chaste friendship, in the spirit of Plato, Aristotle, Cicero, Thomas Aquinas, and "the English Abbot Ailred," was the solution to AIDS presented by another letter-writer worried about the loss of traditional values.[74]

A Manchester University professor had remarked in 1973 that "ideological warfare" had broken out between conservative radicals and Marxist progressives. The former believed in high culture and the traditional values of the past; the latter thought all cultures to be of equal worth but especially appreciated the "primitive" ones.[75] The warfare lasted a while. When Britain's Chief Scout, the decorated ex-army officer Michael Walsh, suggested in 1982 that the Scout Association might benefit form a "return to traditional values" (better turnout, sportsmanship, politeness, and a preference for backwoodsmanship rather than rubbing brass) a controversy ensued.[76] The phrase "permissive society" was highly charged with moralistic connotations, including those related to sexuality, and in the 1980s and 1990s was especially linked with ideologies that would have elicited negative comments from a conservative point of view (see table 3).

Politicians soon tried to capitalise on the discourse about traditional values. In 1970s USA, George McGovern deflected accusations of radicalism by elevating "the restoration of America's traditional values" to the mission of the Democratic Party.[77] The traditional values of "middle

Top Twenty Phrases Most Similar to the Bigram "permissive society" in the *Times*, 1961–2010

	1961-1965	1966-1970	1971-1975	1976-1980	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010
1	foreignness	racial prejudice	permissiveness	uncaring	materialist	feminism	puritans	social revolution	insidious	freedom choice
2	exotically	misguided	morality	mentality	holocaust	permissiveness	feminism	religious differences	institutionalised	public discourse
3	scenes characters	intolerance	nazism	bigotry	telling phrase	intolerance	liberalism	sexual matters	road serfdom	totalitarian regimes
4	opera goer	bigotry	homosexuality	neuroses	sexual freedom	materialism	stalinism	secularism	irrational self	moral superiority
5	inaccessibility	immoral	death wish	atavistic	christian beliefs	liberalism	oppressor	calvinism	dismisses	whole notion
6	characteristic	morals	materialistic	taboos	modern architecture	anarchism	reactionary	sexual liberation	trade unionism	completely justified
7	altogether absent	hypocritical	fanaticism	restlessness	lust power	homosexuality	reactionaries	turmoils	lies root	proper regulation
8	difficult detect	permissive	perversion	inarticulate	liberal democracy	elitism	moralists	solidarity union	miscegenation	cultural relativism
9	regional characteristics	shameful	perverted	world weary	detestation	marxism	traditionalists	permissiveness	simply repeat	incapable
10	appreciate true	dogmatic	elitist	fatalism	agnosticism	family values	fascism	western culture	islamophobia	sense fairness
11	form expression	permissiveness	revolting	callousness	grossest	hypocrisy	puritanism	xenophobia	jobbery	persuasions
12	habits mind	ignorance	atheist	repressed	moralism	egalitarianism	(camille) paglia	atmosphere fear	totalitarian state	become burden
13	antecedent	misunderstood	womanhood	work ethic	sentimentality	authoritarianism	heretics	matriarchy	religious toleration	presumption
14	affluent society	revulsion	totalitarianism	egotism	tibetan people	passionately	sense shame	rastafarianism	woolly minded	restriction
15	avoid becoming	hypocrisy	neurosis	miseries	israel diaspora	reactionary	cruelties	apprehensions	superstition	permissiveness
16	great religions	facts life	revulsion	inbred	spiritualism	evils	triumphalist	provisionals	occultism	confusing
17	primordial	self-righteous	blasphemy	xenophobia	iconoclasm	bigotry	peasantry	anti-american	preoccupation	weak vulnerable
18	abrupt changes	homosexuality	sexuality	human spirit	semitism	selfishness	homosexuality	feminism	immigration crime	context explanation
19	horror films	ignorant	bigoted	craving	hardly blamed	tyranny	bad loser	radicalisation	strangled birth	opaquely
20	becomes aware	vindictive	self-righteous	hatreds	irreligious	feminists	marxists	sectarianism	tim luckhurst	imaginative ways

Table 3. Top twenty phrases (stop words removed) most similar to the bigram "permissive society," based on bigram embeddings per five-year period in the Times*, 1961–2010. Words/phrases related to morality are in red; words related to ideologies (mostly "ism" and "ist") in blue; words related to sexuality in green; all other words in black. Camille Paglia (listed in the column for 1991–1995) is a critic of mainstream feminist ideas about sexuality; Tim Luckhurst (column 2001–2005) wrote about sex education as a* Times *columnist. Recognisable too (in the column for 2001–2005) is Friedrich Hayek's* Road to Serfdom *(1944).*

America" were identified elsewhere as "patriotism, religiosity, independence, optimism and self-improvement,"[78] those of Nixon's supporters as "law and order, personal morality, sobriety, self-help and patriotism."[79] The right-wing phenomenon that eventually caught everyone's attention was Ronald Reagan, who conquered America on a programme of "God, country, family and traditional values."[80]

This was also the heyday of the American Coalition for Traditional Values, to which many conservative evangelicals in the USA subscribed. To push back homosexuality, pornography, and other forms of permissiveness, proclaimed its chairman, "we must flood the federal Bureaucracy with Christians."[81] That was in 1983; almost a decade later, the Coalition campaigned against Clinton's "anti-God agenda," supporting televangelist Pat Robertson, who defined feminism as "a socialist, anti-family political movement that encourages women to leave their husbands, kill their children, practise witchcraft, destroy capitalism and become lesbians."[82] This kind of rhetoric was probably difficult to stomach even for hard-line conservatives on the London side of the Atlantic, especially after the Coalition turned its attention to Harry Potter (whose "witches also believe in abortion as a sacred act").[83]

The American turn to the moral Right resonated in Western Europe, which likewise witnessed the ascent of moralised conservative rhetoric. Germany's Christian Democrat leader Helmut Kohl began to stress the need to properly reward women for staying at home, since the family was the bastion of "traditional values and moral qualities."[84] In France, right-wing politician Raymond Barre extolled the traditional values of *travail*, *famille*, *patrie*.[85] In Italy, the right-wing alliance favoured a Christian society, "the primacy of the nation," and traditional values in language and dress.[86] In Britain, a rejuvenated conservative rhetoric on "decent, healthy traditional values" made itself felt in the 1980s, when the House of Lords passed an amendment of the Education Act. It prescribed that sex education should instill into the young "moral considerations and the value of family life."[87] Not long after that, Thatcher started her "clean-up Britain crusade" to reinstate the "traditional values of fairness, integrity, honesty and courtesy" and eradicate the "false values of socialism."[88]

The problem for British conservatives (in contrast to American Republicans) was, of course, that their appeal to religion had a tenuous basis in society. There was a time when the Church of England had been called "the Tory party at prayer"; by 2004 the only stalwart defenders of traditional religious values left were Roman Catholics.[89] A journalist had already observed in 1991 that:

> It is likely that a majority of the Tory conference fodder, as they bay for traditional values, as well as most of the Cabinet are atheists. The nostalgia they feel for the English hymnal, the Book of Common Prayer and the Authorised Version is terribly nice and respectable and entirely vacuous in religious terms. The easy way out of this for the atheist Tory is to insist that Christianity is the most important historical and cultural force in our society and must be taught as such. (…) They are aspiring to resurrect religion as a socially unifying force amid the widespread conviction that it is untrue.[90]

That was hitting the nail on the head. It was also the reason for people anxious about the world they lived in to turn to moralised politics: if we no longer believe in Christianity, at least we can value it as a cultural force. Only the already converted could take comfort in the Prince of Wales' expectation that "the survival of civilised values, as we have inherited them from our ancestors, depends on the corresponding survival in our hearts of that profound sense of the sacred."[91]

Nor did the Conservatives monopolise traditional values; the Labour Party, too, identified with them. To Tony Benn, they had included adherence "to the democratic rights of the British people and to the advocacy of socialism by consent."[92] To the trade unionist Bill Sirs they amounted to "democracy, tolerance, fairmindedness and understanding."[93] But these early appropriations of what to all intents and purposes was an expression of conservative rhetoric did not even come close to the unabashed left-wing hijack of conservative moral language in the 1990s, when "New Labour" began to put traditional values "in a modern setting."[94]

TRIVIALISING THE TRADITIONAL

In 1993 the *Times* reported on John Major's famous "Back to Basics" speech in Blackpool. The Prime Minister had made an inspirational, revivalist appeal to "traditional values, family

responsibility, grammar and spelling and crime-free streets." The journalist who wrote this account observed that it struck a chord, even though it was "more notable for its tone and style than its substance." It was not a mere protest against the 1960s and the permissive society. "It was more an invocation of an earlier age, the world of Picture Post and Ealing comedies, a time when Surrey always won the county cricket, an era of black-and-white films rather than Technicolor blockbusters."[95] The Tory right rejoiced. "Basic values," fifties' style, were now on the national agenda, even though a columnist wisely predicted that this new "emphasis on duties and responsibilities, rather than rights" would be mostly rhetorical.[96]

Traditional values, including "self-discipline and fidelity,"[97] were certainly on the agenda, but not in the way Major had anticipated. There were two problems. The first one was that Major's cabinet soon had to deal with a procession of serial adulterers and an assortment of sexual delinquents from within the Tory Party itself. The second problem was that tradition had become more socially acceptable than ever. *Good Housekeeping*, for instance, not just any magazine, had already claimed that male chauvinism and female victimhood were things of the past; the future lay with women as the "new traditionalists."[98] Tradition was certainly no longer the sole prerogative of Conservatives.

This is where Tony Blair came in. In 1994 he issued a statement, "strong on vision and short on pledges," in which he revealed his ambition "to retain Labour's traditional values but put them in a modern setting."[99] A year later the *Times* reproduced part of another speech by Blair:

> The family is important because it is in the family that self-respect and respect for others are learnt. It is in the family that the limits of freedom are first experienced and the roots of responsibility are put down. The family is the antithesis of narrow selfishness. From the family, we build out into broader society. I believe in a moral obligation to help those worse off or weak or unemployed. Here is where the traditional values of the Left, applied sensibly and practically to the modern world, are its strength. It can fashion a new moral purpose for the nation.[100]

If the single, capitalised word "Left" in this quotation were replaced with "Right," a trueblood Tory would have found the message entirely palatable. Labour had defeated the Conservatives on their own ground.

The triumph did not last long. The "hard, struggling no-nonsense voice of Middle Britain" wanted a better life; it might become all gooey about traditional values but it had no interest in hearing politicians talk about them.[101] Admittedly, Blair's rhetoric was unsurpassed. The *Times* tended to present Tony Blair as an extremely talented orator whose speeches managed to make any audience feel that it really belonged. To what, exactly, it was supposed to belong was not always so clear. Experts who analysed Blair's language found a distinct preference for "bringing together apparent opposites, linking them with 'and' or 'but,' and conveying them as though available jointly rather than as alternatives." By running with the hare and hunting with the hounds, rhetorically speaking, Blair was able to compress an all-inclusive message into a single 839-word speech consisting of numerous paired opposites:

- ambition and compassion
- head and heart
- invest and also change
- social justice and opportunity
- unite work with business but pursue social justice
- different but more reasonable
- continue but make sure
- punished but offered rehabilitation
- quickly but properly

Meanwhile, the most famous example of Blair's (and his speech writers') ability to merge left- and right-wing language by conjoining disjunctions remained Labour's sublimated antithesis "traditional values in a modern setting."[102]

One commentator (the erstwhile Marxist Mick Hume) noted that ostensibly right-wing rhetoric was now spurned by the intellectual elite, so that "the champions of traditional values have retreated before the politics of cultural relativism, difference and identity." "Western civilisation has become a dirty expression."[103] But what may have applied to the intelligentsia did not apply to society at large. Despite "the liberal supremacy in media circles, ordinary people" still clung to traditional values. Society revelled openly in them, so much so that some twenty-first-century Conservatives based hope for recovery on a message of "faith, the flag and family."[104]

In the *Times* the bigram "traditional values" (and its corollary "family values," which became so pervasive between 1980 and 2010 that it deserves separate analysis) was now appropriated by anyone who felt the need to express a "conservative" outlook on life. Personal dating advertisements offer a fine illustration of this penchant for trivialities. Ads began to invoke traditional values around 1990. "It could change your life," advertised a "gentleman farmer with sporting Interests," on the look-out for a "young woman of breeding and education with traditional values." A "cantankerous old codger" sought a "strong-minded, smartly turned out merry widow," middle-aged, intelligent, well-educated, good background, and stocked with traditional values.[105] Not long after, references to traditional values, as something single people with a conservative mind-set valued in prospective soulmates, hit the roof (figure 9). The term had become a part of everyday speech.

It is difficult to discern a specific pattern in the associations these ads made with traditional values, other than what one would have expected, given the presumably rather elitist *Times* readership. Male and female advertisers shared upscale hobbies like music, theatre, art, travelling, cooking, and countryside activities. They put great store by their appearance. They expected potential partners to be cultured, refined, successful, professional, well-balanced, sociable, humorous, and intelligent. Very occasionally people identified themselves as being Christian, Greek Orthodox or otherwise religious, but usually companion seekers did not advertise their ideologies. One notable exception was an "attractive writer," a female Oxford graduate who sought "a gentleman of intellect, culture and traditional values" whom she expected to be "kind, interested in politics and soundly right wing."[106] The lady in question apparently had some difficulties in finding a spouse; she began at age 39 in 2007, curiously turned 44 only two years later, and was still at it in 2010.

BEYOND THE CONSERVATIVE

Tracking several strands of conservative rhetoric, from "conservative principles" through "conservative values" to "traditional values," I have tried to map out part of a moral language we might call "conservative." This has led to some interesting discoveries. Specific phrases clearly followed specific historical trajectories. Principles were replaced by values. Political options began to be couched explicitly in moral terms after the 1960s: political rhetoric was moralised, as it were. This played out in the ideological differences between the Left and the Right, a tension sublimated in the 1980s by the assimilation of conservative rhetoric into popular discourse. British society as a whole seemed to have made a spectacular turn to tradition, or at least to a general conservative rhetoric promoting traditional values.

By 2010, this seemingly conservative rhetoric had practically become mainstream. Reading the *Times*, it seemed that small government and national identity were the only options left to develop further a politics that could somehow be called "conservative," and even these options were no longer always identifiably "right-wing." The fact that since the 1960s people with a conservative mind-set have found it difficult to discover a rhetoric that could replace the once self-evident discourse on "conservative principles" is less interesting than the fact that these new forms of rhetoric emerged from a tension with modernity provoked especially by the "sixties." Yet, as time moved on permissiveness itself became a relative notion, so that it was unclear what it was, exactly, that needed to be conserved. But this wrestling with modernity applied not just to conservatism; it presumably applied to all moral languages that grew out of that problematic category, the "modern."

We tend to categorise moral languages as ideologies, as, e.g., "Enlightened," "liberal," "Christian," "nationalist," or "conservative," but these simple labels often do more justice to ordering the present than to understanding the past, and vice versa. What I hope to have made clear is that digital history techniques help us to identify the changing clusters of words that come together in moral languages. This article's focus on principles and values, both conservative and traditional, has allowed us to touch on the complexity of these shifting fields. The challenge is to break them down into the normative, dynamic rhetorical elements that constitute them. That would allow us to appreciate a different, more complex, ambiguous, and varied past, and help us conceive of different futures in a present that seems to have lost its bearings.

NOTES

1. "Right foot forward, march," *Times* (Sept. 25, 1997), 38. All references to the *Times* are through *The Times Digital Archive* (https://www.gale.com/intl/c/the-times-digital-archive) and accessed in April 2019.
2. "principle, n.," *OED* Online (Oxford, Oxford University Press), accessed April 29, 2019.
3. The reference is to Disraeli, *Coningsby, or The New Generation*, Vol. III, Book viii, Chapter iii.
4. This does not just apply to the English language. Cf. for example the Dutch word for principles, *beginselen*, which is still in use but has a nineteenth-century flavour; see "beginsel" in http://gtb.ivdnt.org/, accessed April 8, 2019.
5. Disraeli, *Coningsby*, Vol. I, Book iii, Chapter v.
6. For this research the full dataset of the *Times* was used, made available by Gale as XML files, which were then converted to CSV for further processing. No further OCR correction was done.
7. For a history of conservative ideas, see E. H. H. Green, *Ideologies of Conservatism. Conservative Political Ideas in the Twentieth Century* (Oxford: Oxford University Press, 2002).
8. "Conservative Values Past, Present and for the Future," *Times* (July 23, 2001), 15.
9. The recent literature includes: Emily Jones, *Edmund Burke and the Invention of Modern Conservatism, 1830–1914. An Intellectual History* (Oxford: Oxford University Press, 2017); Richard Bourke, "What Is Conservatism? History, Ideology and Party," *European Journal of Political Theory* 17 (2018), 449–75.
10. N-grams were calculated per year on the basis of all newspapers. Only n-grams with a frequency higher than 4 over the whole corpus (1784–2010) were taken into consideration.
11. See note 2.
12. "British Association," *Times* (Aug 2, 1907), 9 (on physiological inquiry in London).
13. "Conservatism," *Times* (May 9, 1930), 10, a review of Feiling's *What Is Conservatism?* (1930). On Feiling: Reba N. Soffer, *History, Historians, and Conservatism in Britain and America: From the Great War to Thatcher and Reagan* (Oxford: Oxford University Press, 2009), 86–109.
14. "Private Correspondence," *Times* (Feb. 13, 1823), 3; "Despatch from the Count de Nesselrode," *Times* (Jan. 20, 1823), 3.
15. "House of Commons," *Times* (Nov. 21, 1837), 2.
16. "London, Tuesday, August 3, 1841," *Times* (Aug. 3, 1841), 4.
17. The trigrams "one nation," "tory," "tories," "toryism," "conservatism," "conservative(s)," and "party" occur 250 times from 1971 onwards. In "Why Tories Are Unpopular," *Times* (Jan. 12, 1996), 1, Thatcher was reported castigating John Major's pro-European politics as "no-nation Conservatism." On the one-nation tradition, see Robert Walsha, "The One Nation Group and One Nation Conservatism, 1950–2002," *Contemporary British History* 17, no. 2, 69–120; David Seawright, *The British Conservative Party and One Nation Politics* (New York: Continuum, 2010); Stephen Evans, "The Not So Odd Couple: Margaret Thatcher and One Nation Conservatism," *Contemporary British History*, 23, no. 1 (March 2009), 101–21.
18. "The British Constitution," *Times* (Mar. 20, 1832), 3.
19. "The Peel Banquet," *Times* (May 14, 1838), 5.
20. "Conservative Dinner at Birmingham," *Times* (Dec. 20, 1838), 3.
21. "Mr. Bright at Birmingham," *Times* (Apr. 25, 1859), 11.
22. "Election Intelligence," *Times* (Mar. 19, 1867), 12.
23. "Equal Sacrifice in Industry," *Times* (July 2, 1925), 13. On the rejection of socialism, see G. C. Webber, *The Ideology of the British Right, 1918–1939* (New York: St. Martin's Press, 1986).
24. "Manifesto by 'Die-Hards,'" *Times* (Mar. 8, 1922), 14. See N. C. Fleming, "Diehard Conservatism, Mass Democracy, and Indian Constitutional Reform, c. 1918–35," *Parliamentary History* 32, no. 2 (2013), 337–60.
25. "Mr. Baldwin on Defence," *Times* (May 2, 1936), 8. On the Primrose League, Mitzi Auchterlonie, *Conservative Suffragists: The Women's Vote and the Tory Party* (New York: Tauris, 2007).

26. "Prime Minister on Defence of Europe," *Times* (Oct. 12, 1953), 11.
27. "The Conservative Path for 'Putting Britain Right Ahead,'" (Oct. 7, 1965), 18. See Peter Dorey and Mark Garnett, "'The Weaker-Willed, the Craven-Hearted': The Decline of One Nation Conservatism," *Global Discourse: An Interdisciplinary Journal of Current Affairs and Applied Contemporary Thought* 5, no. 1 (2015), 69–91.
28. "Republicans Already Staking Claims for 1996 Nomination," *Times* (Jan. 13, 1993), 11.
29. "Right's Knight Issues Challenge to Bush," *Times* (Dec. 11, 1991), 8.
30. "Right Foot Forward, March," *Times* (Sept. 25, 1997), 38.
31. "Christian Conservatism," *Times* (Dec. 7, 1998), 23.
32. Unigram and bigram embeddings were created mostly per five-year period in all cases with a minimum of 100 million tokens per period after stop word removal, on the basis of the following article genres in the *Times*: "Law," "News," "Sport," "Letters to the Editor," "News in Brief," "Reviews," "Editorials/Leaders," "Business and Finance," "Politics and Parliament," "Arts and Entertainment," "Feature Articles (aka Opinion)," and "Obituaries." Gensim's word2vec (https://radimrehurek.com/gensim/models/word2vec.html) was used with the following settings: size=160, window=10, iter=12, min_count=3, workers=3.
33. "value, n.," *OED* Online (Oxford: Oxford University Press), accessed April 29, 2019.
34. "Soviet Convinced of Risks, Says Lord Home," *Times* (Oct. 12, 1961), 6.
35. "Why Mrs Thatcher Should Join the Real Tories," *Times* (Oct. 5. 1982), 10.
36. "We Are Only in Our Third Term, and a Woman's Work Is Never Done," *Times* (May 26, 1988), 12.
37. "The New Young Embrace the 'Older' Values," *Times* (Nov. 26, 1982), 14; "European Teenagers Value Money above Personal Happiness," *Times* (May 11, 1989), 3.
38. "The Right Hopes It Can at Last Court Favour," *Times* (June 21, 1986), 5.
39. Cf. "MPs in New Group Are 'Steeped in Traditional Tory Values,' says Pym," *Times* (May 15, 1985), 4, on the "Conservative Centre Forward" group.
40. "I Will Not Make the Same Mistake," *Times* (July 4, 2001), 12.
41. "Why Tories Should Vote Labour," *Times* (June 24, 1994), 16; based on Gray's *The Undoing of Conservatism* (1994).
42. Joris van Eijnatten and Ruben Ros, "The Eurocentric Fallacy: A Digital Approach to the Rise of Modernity, Civilization and Europe," *International Journal for History, Culture and Modernity* (2019).
43. The networks were generated in Gephi: filtered on degree > 75; the modularity was set to 3 clusters; no OCR corrections were done. The network layout is based on the Fruchterman-Reingold algorithm.
44. "Co-Education at the Top," *Times* (Oct. 26, 1934), 11.
45. "Herr Hitler's Thirteen Points," *Times* (May 22, 1935), 16.
46. "Japan Keeps the Mask on Mishima," *Times* (June 19, 1985), 12.
47. "Pakistan's New Shape," *Times* (Dec. 29, 1952), 7.
48. "Problems of the New Burma," *Times* (Aug. 22, 1955), 7.
49. "An Alienated Elite," *Times* (Nov. 18, 1968), VII; "Elite Point Tasks Ahead," *Times* (Jan. 8, 1969), II; "Indians on the March—A Crucial Choice," *Times* (Feb. 26, 1969), VII; "Collapse of Farm Economy," *Times* (Apr. 27, 1970), 26; "Mobutu—A Personal Portrait," *Times* (Nov. 24, 1972), III; "Conditions in Iran," *Times* (Sept. 22, 1978), 15; "Moroccans Fast in Daylight but Feast by Night," *Times* (June 14, 1983), 5.
50. "Letter from Dar es Salaam," *Times* (Mar. 4, 1975), XIV.
51. "Order in the Midst of Strife," *Times* (Aug. 29, 1970), I.
52. "Managers Must Be Ready for Revolutionary Changes," *Times* (Sept. 8, 1965), 7.
53. "Consultants: A Major Force in Engineering," *Times* (Aug. 9, 1971), 14.
54. "Traditional West Undergoing Process of Change," *Times* (Nov. 7, 1958), 11.
55. "France's Feminine Feminist," *Times* (May 14, 1969), 11.
56. "Mazowsze," *Times* (July 23, 1976), 11.
57. "Necks Please," *Times* (Jan. 18, 1994), 18.
58. "The Victoria Hotel," *Times* (May 26, 1994), 2.
59. "Hamperpeople," *Times* (May 21, 2001), 19.
60. "Courses and Seminars Review 1981," *Times* (Aug. 17, 1981), 20 (on Buckswood Grange).
61. "Austin Rover," *Times* (Jan. 8, 1985), 7.
62. "John Mowlem & Company Plc," *Times* (Apr. 19, 1985), 17.
63. "Isle of Man," *Times* (Feb. 21, 1986), 23.
64. "Disturbing Signs that Fascism Could Be Just round the Corner in Britain," *Times* (June 1, 1977), 7.
65. "Protecting Canada Family Life," *Times* (June 10, 1964), 14.
66. "Italians Fear Growth of Gangster Violence," *Times* (Sept. 27, 1967), 5.
67. "Who Breaks a Butterfly on a Wheel?," *Times* (July 1, 1967), 11.

68. "Slippery Slope of a Permissive Society," *Times* (Dec. 27, 1969), 7.
69. "The Paradox of Christian Freedom," *Times* (July 8, 1972), 16.
70. "Ten Priests Guilty of Heresy, Group Says," *Times* (Mar. 19, 1971), 2.
71. "The Need for Better Research into Family Life," *Times* (Mar. 28, 1977), 14.
72. "Irish Divided by Divorce," *Times* (Apr. 30, 1986), 11.
73. "Fight against AIDS," *Times* (Dec. 8, 1986), 13.
74. "AIDS and Morals," *Times* (Jan. 28, 1987), 17.
75. "Parents May Win Right to Appoint School Staff," *Times* (Jan. 8, 1973), 4.
76. "Threat to Grant Regrettable," *Times* (Dec. 3, 1982), 4.
77. "Democrats Given Blunt Warning," *Times* (Aug. 10, 1972), 5.
78. "Voices that Have Become an American Institution," *Times* (Aug. 29, 1973), 14.
79. "End of a Presidency," *Times* (Aug. 9, 1974), 17.
80. "Election 84," *Times* (Nov. 2, 1984), 7.
81. "Moral Majority Awaits Its Greatest Triumph," *Times* (Nov. 5, 1984), 7.
82. "Evangelicals Rally to Beat the 'Sinful,'" *Times* (Oct. 19, 1992), 11.
83. "Spin Cycle," *Times* (Aug. 23, 2001), 14.
84. "CDU Offers the Traditional Remedies," *Times* (Feb. 8, 1983), 5.
85. "Jovial Defender of Tradition," *Times* (Mar. 12, 1986), 40.
86. "Italy Moves to the Right in Local Elections," *Times* (Apr. 18, 2000), 21.
87. "Putting Love into Sex Education," *Times* (June 4, 1986), 13.
88. "Thatcher Champions the Family," *Times* (May 26, 1988), 1 and 24.
89. "Is the Pope a Tory?," *Times* (Apr. 3, 2004), 6.
90. "Finding God in the Classroom," *Times* (Aug. 5, 1992), 10.
91. "Prince Blames 'God of Technology' for the Loss of Values," *Times* (Apr. 30, 1997), 6.
92. "Socialism by Consent Pledge in Statement," *Times* (Apr. 3, 1981), 2.
93. "Steel Union Faces Call for Review of Affiliation," *Times* (Apr. 13, 1981), 2.
94. "Blair Says Tradition Is Bedrock of Change," *Times* (June 8, 2000), 4. Blair's list of values: "Respect for the old for what it has still to teach; respect for others, honour, self-discipline, duty, obligation, the essential decency of the British character."
95. "A Party Eager for that Old-Time Religion," *Times* (Oct. 9, 1993), 8.
96. "How It Looks," *Times* (Jan. 1, 1994), 16.
97. "Major Fights to Salvage Back-to-Basics," *Times* (Jan. 7, 1994), 1.
98. "Housewife, Cinderella or Superstar?," *Times* (July 10, 1991), 12.
99. "Blair Stakes His Claim to Downing Street," *Times* (June 24, 1994), 8.
100. "Is Labour the True Heir to Thatcher?," *Times* (July 17, 1995), 17.
101. "No 10 Told to Cut the Spin as Poll Lead Dwindles," *Times* (June 12, 2000), 10.
102. "He Won't, They Can't, and We Know It," *Times* (June 16, 2001), 22. The analysis was performed by *Encarta World English Dictionary*.
103. "What Do They Know of England Who Only England Loathe?" *Times* (Oct. 1, 2001), 16.
104. "Tory Right Rallies around Faith, Flag and the Family," *Times* (July 25, 2005), 24.
105. "Announcements & Personal," *Times* (June 2, 1990), 13; "Announcements," *Times* (Sept. 1, 1990), 13.
106. "Multiple Classified Advertising Items," *Times* (Mar. 21, 2008), 74.

Reflections on Infrastructures for Mining Nineteenth-Century Newspaper Data

Julianne Nyhan
Associate Professor of Digital Information Studies
University College London (UK)
j.nyhan@ucl.ac.uk

Tessa Hauswedell
Research Associate
University College London (UK)
t.hauswedell@ucl.ac.uk

Ulrich Tiedau
Associate Professor in Dutch
University College London (UK)
u.tiedau@ucl.ac.uk

Abstract: *In this study we compare and contrast our experiences (as historians and as digital humanities and information studies researchers) of seeking to mine large-scale historical datasets via university-based, high-performance computing infrastructures versus our experiences of using external, cloud-hosted platforms and tools to mine the same data. In particular, we reflect on our recent experiences in two large transnational digital humanities projects:* Asymmetrical Encounters: E-Humanity Approaches to Reference Cultures in Europe, 1815–1992*, which was funded by a Humanities in the European Research Area grant (2013–2016) and* Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories 1840–1914*, which was funded through the Transatlantic Partnership for Social Sciences and Humanities 2016 Digging into Data Challenge (2017–2019). As part of the research for both these projects we sought to mine the OCR text of nineteenth-century historical newspapers that had been mounted on UCL's High-Performance Computing Infrastructures from Gale's TDM drives. We compare and contrast our experiences of this with our subsequent experiences of performing comparable tasks via Gale Digital Scholar Lab. We contextualise our experiences and observations within wider discourses and recommendations about infrastructural support for humanities-led analyses of large datasets and discuss the advantages and drawbacks of both approaches. We situate our discussions in the aforementioned infrastructural scenarios with reflections on the human experiences of undertaking this research, which represents a step change for many of those who work in the (digital) humanities. Finally, we conclude by discussing the public and private sector research investments that are needed to support further developments and to facilitate access to and critical interrogation of large-scale digital archives.*

Keywords: digital infrastructures ■ text mining ■ historical newspaper collections ■ high-performance computing ■ critical cultural heritage ■ digital humanities ■ *The Times Digital Archive* ■ the *Times* of London

INTRODUCTION

In this talk we will compare and contrast our experiences as historians, digital humanists, and information studies researchers of seeking to mine large-scale, digitised historical newspaper collections via university-based high-performance computing infrastructures versus our experiences of using external, cloud-hosted tools to mine that same data.[1]

We situate these reflections in the context of two large transnational projects in which we have participated. *Asymmetrical Encounters: E-Humanity Approaches to Reference Cultures in Europe, 1815–1992*, was funded by a Humanities in the European Research Area grant (2013–2016) and included partners from the United Kingdom, Netherlands, and Germany. It was coordinated by the University of Utrecht. With digital technologies it carried out a longitudinal analysis of large digital newspaper and magazine archives and asked how the cultural aspects of European identity changed between 1815 and 1992 (AsymEnc, n.d.). We also draw on the experiences we have gained so far in the project *Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840–1914* (OcEx), funded by the Transatlantic Platform of the Digging into Data programme (2017–2019). The project brings together researchers in Finland, Germany, Mexico, Netherlands, the United Kingdom, and the United States, and is coordinated by Northeastern University. It "examine[s] patterns of information flow across national and linguistic boundaries and to link research across large-scale digital newspaper collections" (Oceanic Exchanges, n.d.).

In this talk, and in line with literature such as that of Susan Leigh Star and Karen Ruhleder (1996) and American Council of Learned Societies (2006), we define infrastructure as comprising, and being shaped by, a complex set of interacting dynamics that enfold not only physical structures but also social, cultural, and institutional processes and contexts. Indeed, Wolfgang Kaltenbrunner has argued that:

> Infrastructure occurs when the tensions between globally valid standards and local contexts, as well as between automated technological processes and tasks performed by human actors can be successfully resolved. An important consequence of this definition of infrastructure is that it develops incrementally—it is not created, it evolves. (Kaltenbrunner 2015, 211)

Other definitions of digital research infrastructure use the analogy of a "digital ecosystem," which provides "services that are built around communities" (Blanke, Kristel, Romary 2015). In our case, the community needs we are concerned with are those of the humanities and specifically those of the field of history and heritage. Accordingly, in this paper we do not limit our observations to the computational infrastructures that we used to undertake our research but we also reflect on the institutional, technical, legal, sociocultural, and labour organisation factors that sometimes aided, and sometimes impeded our work but which are relevant to consider in the context of humanities research. After setting these out, we close with some reflections on evolving infrastructures for undertaking large-scale data mining of cultural heritage sources and artefacts and the future directions they may take.

We begin by describing our experiences of seeking to computationally analyse *The Times Digital Archive*, an online, full-text facsimile of more than 200 years of the [London] *Times*, "one of the most highly regarded resources for eighteenth-, nineteenth-, and twentieth-century news coverage, with every page of every issue from 1785 to 2010" (Gale, n.d.).

ACQUIRING MINING ACCESS TO *THE TIMES DIGITAL ARCHIVE*

Following on from the newspaper digitisation programmes by national libraries and commercial companies that began in the late 1990s (see, e.g., Gooding 2014 for a history), we have seen an exponential increase in the creation and global availability of digital newspaper archives. Large quantities of historical newspapers have been digitised through the public hand (e.g., the Library

of Congress), commercial companies (e.g., Gale/Cengage), and public-private partnerships (e.g., the British Newspaper Archive involving the British Library). Despite the many gaps in coverage that exist (Hobbs 2013; Milligan 2013), and issues like variable OCR quality of transcriptions (Smith and Cordell 2018; Tanner et al 2009), the existing collections of national libraries, commercial companies, and, in some cases, public-private partnerships now offer an abundance of material for researchers and other communities to interrogate (see Milligan 2019). Through the interfaces to the materials that providers offer, researchers can run simple and advanced keyword searches of the material. Those who are interested in carrying out more complex queries or investigating the metadata that makes structural and other elements of those digital surrogates (and information about them) machine readable, also have various possibilities open to them. The APIs[2] and data dumps[3] that providers like Europeana make available allow researchers to take copies of open-licenced data and query or transform it according to their needs. For those who wish to work with licenced data and metadata, both Gale and ProQuest provide text and data mining hard drives to paying users (on the history of this in the context of Gale see Fyfe 2016).

Mining *The Times Digital Archive* has been an important aspect of the projects that we have engaged in, and for this it was necessary to get access to the underlying full-text data and metadata of the collection. While our university library had a subscription to *The Times Digital Archive* for years, the collection was accessible to us via the standard search form only. Through a chance encounter with a colleague from Gale/Cengage in 2014, we became aware of the possibility of acquiring a copy of the *Times* data on an external hard drive, a so-called Text and Data Mining Hard Drive or TDM Hard Drive. On it was approximately 2 TB of *The Times Digital Archive*, then covering the years 1785 through 2010 (meanwhile extended to 2013), represented as one XML file per issue plus one TIFF image per page. We paid a moderate three-digit sum for the hard drive and this was mostly for the work that Gale/Cengage did to reformat the data so we could perform text and data mining on it, and also for the delivery of the files to the University College London (UCL) library who made it available to us. As we understand it, the licence essentially positioned the hard drives as extensions to the subscription that UCL libraries already had with Gale/Cengage, and for the duration of those subscriptions only, which explains the reasonably moderate fee. Gale granted UCL a "royalty-free non-exclusive, non-transferrable [sic], non-sublicensable, worldwide right, subject to terms and conditions of the Agreement … to enable Authorised Users to access materials solely for purposes of performing Text and Data Mining activities for non-commercial research purposes" (Gale licence).

Before moving on to recall the steps we next had to take to get the data mounted, and to seek to mine it, we will pause to reflect on the advantages and disadvantages of obtaining the data in this way. Firstly, the advantages: concerns about the risks of monetising cultural heritage through digitisation projects that created gated content have been raised and convincingly made in recent years (e.g., Prescott 2014; Darnton 2010). The *New Renaissance Report of the Comité des Sages* has, for example, cautioned of the potential of creating a digital dark age should the balance between public, private, and public-private digitisation give way (Comité des Sages 2010). Yet, as also made clear in that report and as wide experience has shown, there is much to be gained from mutually beneficial public-private partnerships. Away from the institutional and collection-level benefits of this approach, as individual researchers who needed to secure nineteenth-century newspaper text and metadata to work with, we found the experience with Gale/Cengage to be efficient and straightforward. As university-based researchers our instinct had been to attempt to secure access to the data we needed through our university-based infrastructures and other public institutions. That private providers could facilitate the first level of access to the sources alerted us to the distributed nature of the infrastructures we require to do this research. That we were privileged enough to have the funding required to leverage this route must be noted too. That said, there were particular constraints to working in this way. Secondly, then, to the disadvantages: we sought the TDM drives in the context of a multinational project yet

the licence that we signed was conceptualised along national lines and linked to the revenue stream of one university—extending the licences along international lines proved more problematic. As such, our working solution could not scale to the transnational space that the project occurred in.

MOUNTING AND QUERYING

In any case, we had managed to acquire the full text and associated metadata of *The Times Digital Archive* and we in UCL were ready to do something with it. Making the link to UCL Research IT Services (RITS) was also serendipitous and came about during the course of a Digital Excursion that was organised by the UCL Centre for Digital Humanities.[4] UCL RITS[5] was at that time quite keen to show value to the institution outside of the usual data rich disciplines, who routinely undertake computationally intensive research, with which it often works. Their stated aim was to "enable researchers beyond those in science and engineering" (UCL 2017) and at the time, they had just worked with a humanities project for the first time ever: "Enabling complex analysis of large-scale digital collections collaboration between the British Library and UCL," which had been funded as part of the JISC Research Data Spring (Terras 2015). The UCL Research Computing Platforms Service, one of the services of UCL Research IT Services, offers "[s]upport … [for] computationally intensive research at UCL through provision of specialist platforms for high performance and high throughput computing.[6] They make three computing clusters, (Legion, Myriad, and Grace) available to the wider UCL community, who do not pay for the time it takes to run scripts or for storage. Beyond those tasks the service charges about £350 per day to research projects to, for example, write and run bespoke text mining scripts. It was agreed that, with the help of our colleagues in Research IT Services, we could mount *The Times Digital Archive* data and run queries on it with Legion, "a mixed-use cluster hosted in UCL's Bloomsbury data centres."[7]

The first step was to upload the data to a protected account in the integrated Rule-Oriented Data System (iRODS) of UCL Research Data Services (see Wilson 2017), where we were given a 5 TB account. Almost from the outset, though, it became apparent that the system was not originally set up to support humanities data and research. Instead of the massive and highly structured data that the computationally driven disciplines often query (e.g., Stevens 2013), as already mentioned, our data comprised numerous small XML files and TIFF files, grouped into multiple directories. As a result, it took days for it to be uploaded. Once uploaded it was necessary for Research IT Services to rechunk the data so we could merge the XML files for individual newspaper issues into monthly, quarterly, or yearly bundles. We understand that this was necessary not only to support the longitudinal analysis that we hoped to perform on the data but also from a computational perspective. It was also necessary for a parser class for the XML schema to be added but a boon was that the code that had resulted from the JISC Research Data Spring "Enabling complex analysis" project (mentioned above) could be reused.

Once the data was finally mounted it was possible to reuse this code and other specifically written R and Python scripts to mine the data. We continued, however, to note that the computing infrastructure was not really set up to support humanities research. For example, the idea behind the reference data sets on Legion is that they shouldn't be changed once stored. As David Smith and Ryan Cordell note, "the proportion of erroneous words in nineteenth-century newspapers can exceed 40% error rates can be even higher for other languages and earlier periods" (2018, 5). In a situation where our licence would have allowed us to undertake, for example, some form of post-correction of the OCR transcripts, or even some combination of automated and manual correction, the time barrier involved in remounting the data would have been significant enough to deter us.

Moving on to the experience of seeking to query the material, although we were immensely grateful to UCL Research IT Services, we found our efforts to use university-based high-performance computing (HPC) to perform text mining to be a fraught experience from a number of perspectives. The example that we will now present relates again to the peculiarities of the cultural heritage datasets that we sought to work with and the questions that we sought to bring

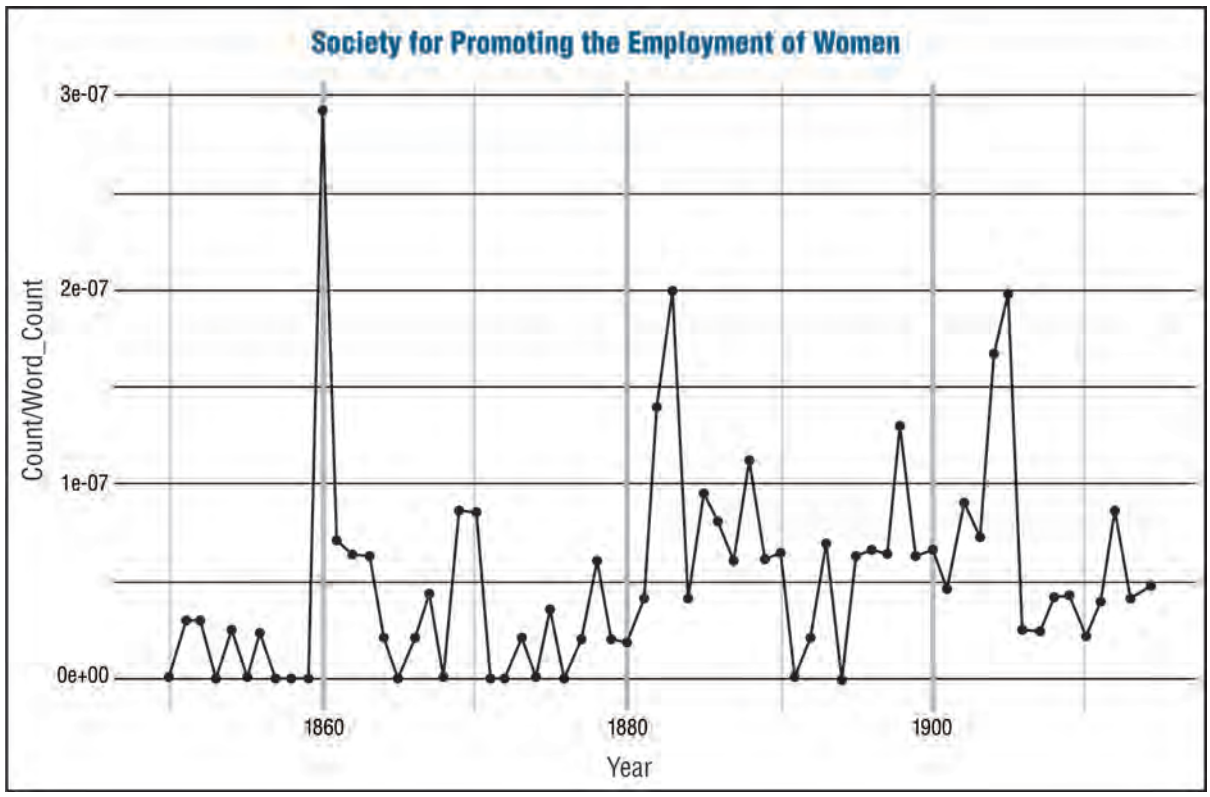


Figure 1. Plot showing the normalised occurrences of the phrase "Society for Promoting the Employment of Women" from 1850 through 1914 in the TDA dataset.

to bear on them. As stated above, digitised newspaper text is relatively unstructured and tends to be riddled with OCR errors. This can be especially challenging when dealing with multilingual sources according to language-specific rules, which are challenging for both the humanities researchers and the research software engineers in terms of the linguistic and computationally linguistic skill sets required. Even simple queries on monolingual texts can involve an element of unpredictability, since even this might retrieve an unmanageable amount of results or not return any meaningful results, confounding expectations and necessitating either further refining or widening the search parameters.

For example, we wished to use the TDA hard drives to conduct research on how female emigration was reported in nineteenth-century newspapers. We were especially interested in how these papers discussed and reported the emigration of women to British colonies of the time, like South Africa, New Zealand, and Australia. Via our initial queries, we searched for articles containing the names of the female emigration societies that were operating during the second half of the nineteenth century. During this time, emigration was increasingly being facilitated, assisted, and managed by charitable societies who saw it as their mission to send women overseas to populate the countries they had colonised (Constantine 1991, 95; Bush 1994). Our initial queries concentrated on plotting the occurrence of mentions of these emigration societies in newspaper articles during the period 1850 through 1914. To support this, we researched and drew up a list of names of emigration societies and sent this to the software engineers with whom we were working. Given the length of the names of these societies (for example, "British Women's Emigration Association) and the vagaries of OCR quality, each of the names was manually assigned a defined "error tolerance" based on word density per title. Our aim here was to avoid, on the one hand, underfitting of the results whilst, on the other, ensuring we did not introduce too many false

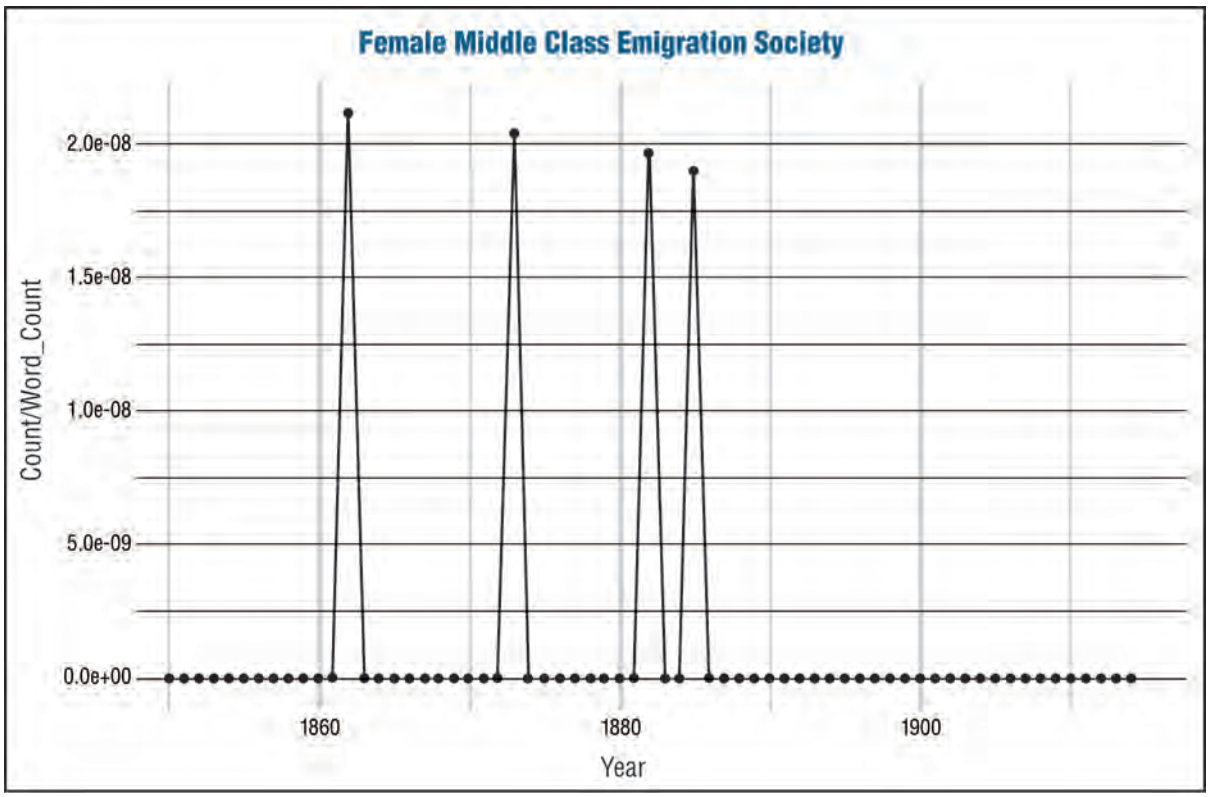


Figure 2. Plot showing the normalised occurrences of the phrase "Female Middle Class Emigration Society" from 1850 through 1910.

positives in the process. When we received back the first plots, however, there was huge variance in the results for each of the emigration societies.

The results for the "Society for Promoting the Employment of Women" indicated that the society was mentioned and discussed in newspaper articles across the decades. The plot for the "Female Middle Class Emigration Society" included hardly any mentions from the mid-1880s. Yet, the existing historical research about this society indicates that it was especially active from the 1880s onwards (Chilton 2007). Why was this not borne out by the plots? It became clear to us that it would be necessary to refine and recalibrate the assigned error tolerances in order to produce more reliably representative results. However, without clear knowledge about the OCR quality for each year and issue of the TDA, it was clear that this process would involve several iterations through our list of about 25 emigration societies.

}Many digital humanities projects that work with noisy data stress that the process should involve iterative stages of cleaning or otherwise refining data, generating and adapting queries, and further refining the working hypothesis of the research (see Schöch 2013). Yet, remember that this work is costed at £350 a day. Even in impressive-looking major research grants (in the high six-digit area), when institutional overheads are taken off and the project budget is split between multiple project partners, resources are tight. Indeed, it quickly becomes apparent that the luxury of undertaking the desired iterative approach may exceed project funds. In this way, our experience of working on projects like this is that they proceed by trial and error. When the TDM drives arrive one can feel like a world of opportunity is opening up, given the TBs of content they offer. Yet, as we have summarised, grappling with the institutional, logistical, financial, bureaucratic, and other hurdles that follow take a significant amount of time that cannot adequately be reflected in the published outputs of the project.

GALE DIGITAL SCHOLAR LAB

From a number of perspectives, then, cloud-hosted tools like Gale Digital Scholar Lab offer an immense step forward. The struggles outlined above in terms of securing and negotiating access to the underlying data; mounting the data; thinking through the peculiarities or limitations of the data and how this might be accounted for computationally; running queries on the data and seeking to work iteratively so that emerging insights can be integrated are, to varying extents (and providing that your institution has the appropriate subscriptions), well provided for.

Gale Digital Scholar Lab (DSL) provides an interface that is highly usable: a number of core text-mining routines can be applied to the data: clustering, n-grams, sentiment analysis, and topic modeling. Named entity recognition and part of speech tagging are also supported and the annotated data can be downloaded for further transformation, like network analysis. The lab is straightforward to use and thus invites explorative research and playful engagement with text. Put in a form that popularises these approaches, it offers a low-barrier entry into digital scholarship–especially for students unfamiliar with text mining. The explanations about the tools that Gale DSL provides are concise and helpful; they do not overburden the more casual and inexperienced user. For these reasons, the platform lends itself for teaching purposes. Much potential can also be noticed in the way that Gale DSL can help to further embed digital forms of enquiry in the scholarship of people who might not otherwise have undertaken elements of this research. In this way Gale DSL contributes to the task of raising the visibility of digital humanities research across the disciplines.

In addition, Gale Digital Scholar Lab also goes some way to meeting the needs of the more advanced user and researcher. Amongst its offerings, we note, is the ability to create custom content and to personalise folders with subsets of relevant articles in a way that is easy and intuitive to manage and organise. Searches can be retained to work with these personalised corpora. Upon getting trial access to the lab, we were concerned that it offered access to licenced material only, and about the new silos of open versus licenced content analysis that the lab might cement. Yet, as we understand it, the possibility of adding external, open licence collections will follow (of course, some would rather not see open licenced materials used within a proprietary platform such as this, but that is another issue).

Moreover, because the original primary source document and OCR text are presented next to each other, it is immediately possible to compare the two and assess the OCR against the original source. Additionally, the user receives detailed metadata information together with an OCR confidence rating, which provides them with some guidance as to how to interpret and treat the results they have generated.

Information about OCR confidence is usually missing from standard digital newspaper search interfaces, but incorporating it in the Gale Digital Scholar Lab will be welcome news for researchers as it gives a better understanding of the collection in question. This openness about the quality of transcription and the provenance of the data is, we believe, crucial to fostering a more critical engagement with digital sources and we hope that more providers will consider implementing such functionalities going forward.

Further, the ability to export tabular data allows more advanced users to take control of their results and to use these for additional visualisations or simply to document their workflow. This is an important aspect for those researchers who are seeking to publish their results and require their findings to be reproducible and repeatable. Again, as it currently stands, when researchers rely on standard search interfaces, it is often difficult to document the workflow in a satisfactory way that meets these requirements. The Gale Digital Scholar Lab, therefore, goes some way towards helping scholars achieve these standards because it can facilitate transparency about how the results of their text mining analysis has been arrived at.

Nevertheless, and we do not mean to imply that this is part of Gale/Cengage's proposition, the availability of cloud-based tools like Gale Digital Scholar Lab do not obviate the need to use high-performance computing or other infrastructures to undertake deep dive text mining–led research.

To discuss our experiences with the Gale DSL in a bit more detail, we will return to our previous example about female emigration societies. We ran a search of the TDA across entire documents (one can choose from, e.g., document title or place of publication) restricted to the years 1850 through 1914. We further restricted our search to content types labelled as "newspapers and periodicals." Exact matches of the phrase "Society for Promoting the Employment of Women" numbered 54 (or 58 if not an exact match). This was starkly different from the number of results that we had obtained from the TDA drives (162 matches). The difference is apparently caused by the error rates that the lab sets, which are different from the error rates that we had assigned when working with the TDA drives. It was not possible for us to change the Gale Digital Scholar Lab's error rates, though the use of regular expressions is permitted in searches. This may offer one way of forcing an adjustment of the lab's error tolerances, but it is not an approach that would scale well to the running of a large number of queries.

In terms of the results that were returned, we found the lab's interface and presentation of metadata helped us to swiftly gain an overview of the kind of material that comprised those 54 occurrences of the search term. We could immediately see the document types in which the results occurred: article (34), advertisement (10), and letter to the editor (10). A breakdown of subjects is also given: 11 in total with the top 3 being working women (9), associations (5), chambers of commerce (1). Articles can also be clustered under those subject headings. One illustrated work, specifically a table, was also returned. The timeframe of the articles that the keywords occurred in was from 1860 through 1913. The OCR confidence levels varied widely, from in excess of 90% down to the low 40%. After adding the documents to a content set, a list of 10 of the authors of those documents also became available. When it came to interrogating those results, we found that we could use the lab to generate some views of our data that may offer some interesting ways of reading it. For example, we could generate a chart of sentiment across the years 1860 to 1914 that might further avenues of investigation (see figure 3).

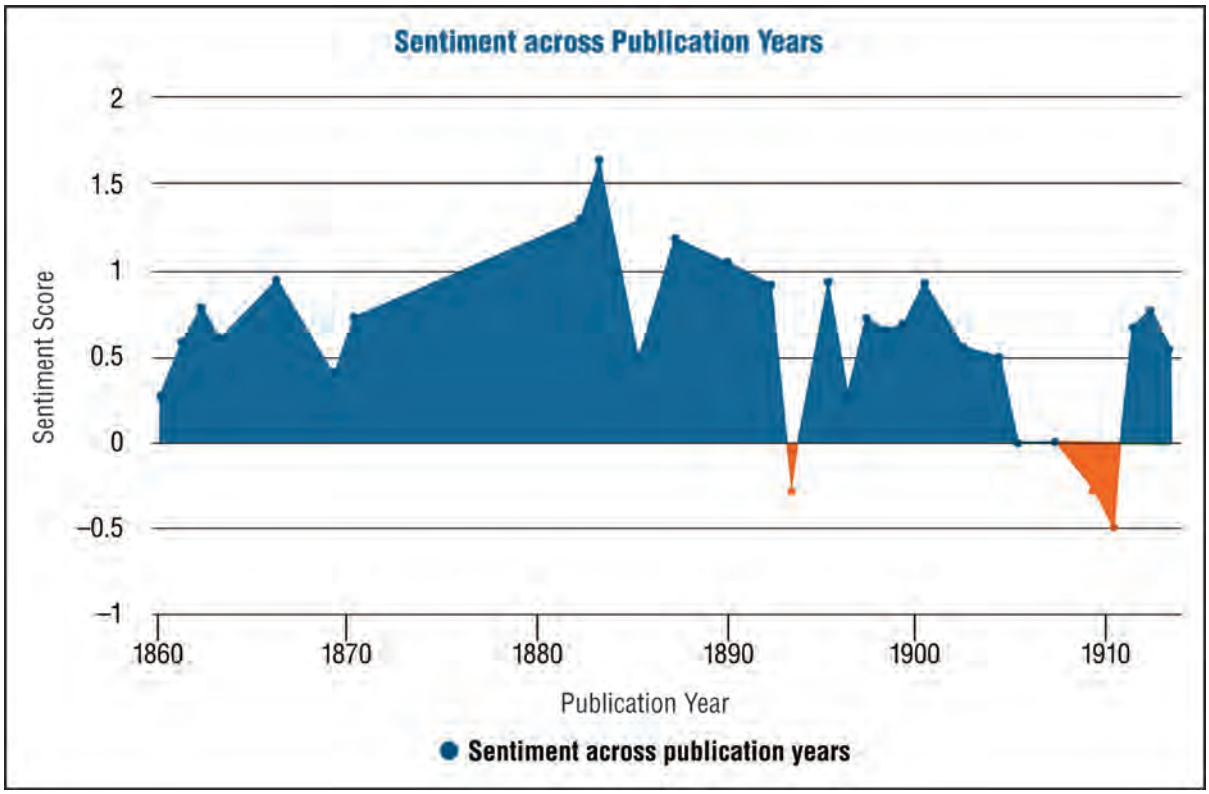


Figure 3. Plot showing the sentiment across the publication years 1860 through 1914.

However, the lab does not currently offer the routines that would allow us to plot the data in the way that we require for our research. Indeed, within the Digital Scholar Lab it is not possible to write one's queries and so the usefulness of the lab for the research we had hoped to undertake is limited. That said, we certainly can imagine a scenario where researchers could use the lab as a site to explore and critically evaluate their datasets, and that this process could then inform and enhance their subsequent use of costly and more bespoke HPC-based interrogation of their data.

With regard to the lab, we would also point out that currently it does not seem possible to use the output of one routine as the input of another so that cascading digital workflows can be built up. Custom scripting on HPCs like those offered by our own Research IT Services (or commercial alternatives like Amazon Web Services) are needed for that. Also, at present, Gale Digital Scholar Lab's support for collaborative work is limited. So, while it will be most useful in some contexts, like teaching on introductory modules and for individual users, it is currently not scaled towards sustained text mining work undertaken by large-scale and geographically dispersed project teams. These projects would need dedicated functionalities for larger group accounts in which results can be stored and accessed by different team members, who might require some bespoke functionalities.

Yet of course, we are aware that no one tool can meet the needs of every type of researcher, let alone every type of research group. It might well be the case, therefore, that cloud-based platforms will function as an intermediary that researchers can trust to perform the "standard" text mining functionalities to exacting and reliable standards. For the great majority of users these will be sufficient, while complex operations that require custom scripting will continue to be performed in other settings.

CONCLUSION

We will close with some general reflections on our experiences of seeking to use university-based high-performance computing infrastructures to mine large-scale, commercially digitised nineteenth-century historical newspaper collections. We are certainly not the first researchers to communicate the difficulties that we have encountered when attempting to undertake this work (and it is in this context writ large that wider pan-European initiatives like CLARIN operate (Wynne 2013) as well as the open-source, web-based Voyant Tools (Rockwell and Sinclair 2016)). Yet, even though, "DH needs have been growing more complex, as humanists—like scholars in many other fields—tackle research questions with larger-scale data than was previously possible" (Dombrowski and Lippincott 2018), the infrastructures required to support this kind of computationally intensive work in the humanities are evolving unsteadily. What to do? The recommendations of the "Enabling large scale data-analysis" project mentioned above included the following:

1. Invest in research software engineer capacity to deploy and maintain openly licensed large-scale digital collections from across the GLAM sector to facilitate research in the arts, humanities, and social and historical sciences
2. Invest in training library staff to run these initial queries in collaboration with humanities faculty, to support work with subsets of data that are produced, and to document and manage resulting code and derived data (Terras et al. 2018, 463).

These recommendations are useful and we hope that they will be taken up. Nevertheless, it seems important to also draw attention to a broader point of research strategy and foresight regarding computationally intensive humanities work that we believe our experiences and these recommendations speak to.

King's College London's digital lab is a rarity in that it employs research software engineers on permanent contracts, in a dedicated centre, with clear professional progression paths, many of whom have a wealth of experience of working as and with humanities and digital humanities researchers to model and implement research software for cultural heritage and arts and humanities sources.[8] Elsewhere, this investment is much more uneven. So, yes, the case needs to be made to engage research software engineers to work with open data sets, but a much wider case needs to be made too about the crucial and more wide-ranging role that research software engineers (and the infrastructures they participate in and enable) will play as research in the humanities becomes ever more digital. At the present time, for the humanities researcher it can be difficult even to find out that research software engineering posts exist in their university and, given the often-unclear career paths for these posts in universities, and the much better conditions that research software engineers can attain outside of the university, these posts can have a high turnover rate. This is unfortunate for the individuals involved and difficult in terms of the management of limited-term grants.

As such, we believe that our experiences in the difficulties of seeking to mine large-scale, digitised cultural heritage collections, whether open or gated, is symptomatic of the struggle of many universities to adequately respond to the wider investment proposition that is required and opened by the digital turn. To return to Kaltenbrunner's discussion of infrastructure and how it involves the interlacing of the global and the local, it seems that the difficulties faced in local contexts can accordingly have implications of global reach. The difficulties faced when seeking to mine large data sets can have global implications in terms of those wider conversations and contexts that humanities scholars may be impeded from contributing to. The difficulties of doing this work limits not only the questions that humanities researchers can ask of sources at scale; it also reduces the visibility of the humanities, and potentially impedes initiatives to develop partnerships between universities and other external actors like commercial companies and the creative economy. These difficulties also limit our opportunities to pursue and shape the economic proposition of digital cultural heritage in the wider digital economy.

Returning, then, to Gale Digital Scholar Lab, perhaps one of the most interesting aspects of it is how it invites us to think about innovation and enterprise in digital humanities, and the potential—and of course pitfalls—of working with industry in order to find perhaps less familiar and extra-university-based ways to facilitate crucial aspects of the infrastructures that are needed to undertake, for example, computationally intensive interrogation of large-scale digital cultural heritage data. Mutually beneficial collaborations undertaken in this space may result in new ways to convince universities of the value proposition of roles like research software engineers, of the need to further invest in the infrastructures that can support digital research in the humanities, and of the multifaceted impacts of digital humanities. Through the private-public partnerships that have been undertaken to digitise cultural heritage materials much has been learned, by both public and private bodies, of the benefits and pitfalls of working together. We must continue to draw on these experiences as we think through the kinds of infrastructures and access we require to facilitate the digitally led interrogation of large-scale, open, and licensed data cultural heritage collections.

NOTES

1. The views that we present in this paper are our personal views as individual researchers; they should not be taken to express the views or opinions of University College London.
2. https://pro.europeana.eu/resources/apis [accessed May 1, 2019].
3. https://pro.europeana.eu/post/experimental-text-dumps-from-europeana-newspapers [accessed May 1, 2019].
4. https://www.ucl.ac.uk/digital-humanities/.
5. https://www.ucl.ac.uk/research-it-services/ [accessed May 1, 2019].
6. https://www.ucl.ac.uk/isd/services/research-it-services.
7. https://wiki.rc.ucl.ac.uk/wiki/Cluster_Computing.
8. See https://www.kdl.kcl.ac.uk/.

REFERENCES

American Council of Learned Societies. 2006. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. Accessed June 13, 2009. https://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf.

Anderson, Sheila, and Tobias Blanke. 2012. "Taking the Long View: From e-Science Humanities to Humanities Digital Ecosystems." *Historical Social Research* 37, no.3: 147–64.

AsymEnc (website). Accessed May 1, 2019. http://asyenc.eu.

Blanke, Tobias, Conny Kristel, and Laurent Romary. 2015. "Crowds for Clouds: Recent Trends in Humanities Research Infrastructures." Accessed July 8, 2019. http://arxiv.org/abs/1601.00533.

Bush, Julia. 1994. "'The Right Sort of Woman': Female Emigrators and Emigration to the British Empire, 1890–1910." *Women's History Review* 3, no. 3: 385–409.

Chilton, Lisa. 2007. *Agents of Empire: British Female Migration to Canada and Australia, 1860s–1930*. Toronto: University of Toronto Press.

Comité des Sages. 2010. *The New Renaissance. Report of the 'Comité Des Sages.'* Reflection Group on Bringing Europe's Cultural Heritage Online. https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/final_report_cds_0.pdf.

Constantine, S. 1991. "Empire Migration and Social Reform 1880–1950." In *Migrants, Emigrants, and Immigrants: A Social History of Migration*, ed. by Colin G. Pooley and Ian Whyte. New York: Routledge, pp. 62–87.

Darnton, Robert. 2010. *The Case for Books: Past, Present, and Future*. New York: PublicAffairs.

Dombrowski, Quinn, and Joan Lippincott. 2018. "Moving Ahead with Support for Digital Humanities." *EDUCAUSE Review*, March 12, 2018. Accessed July 8, 2019. https://er.educause.edu/articles/2018/3/moving-ahead-with-support-for-digital-humanities.

Fyfe, Paul. 2016. "An Archaeology of Victorian Newspapers." *Victorian Periodicals Review* 49, no. 4: 546–77. https://doi.org/10.1353/vpr.2016.0039.

Gale. "The Times Digital Archives 1785–2013." https://www.gale.com/intl/c/the-times-digital-archive.

Gooding, Paul. 2014. "Search All about It: A Mixed Methods Study into the Impact of Large-Scale Newspaper Digitisation." PhD diss., University College London.

Hobbs, Andrew. 2013. "The Deleterious Dominance of *The Times* in Nineteenth-Century Scholarship." *Journal of Victorian Culture* 18, no. 4: 472–97. https://doi.org/10.1080/13555502.2013.854519.

Kaltenbrunner, Wolfgang. 2015. "Scholarly Labour and Digital Collaboration in Literary Studies." *Social Epistemology* 29, no. 2: 207–33. https://doi.org/10.1080/02691728.2014.907834.

Milligan, Ian. 2019. *History in the Age of Abundance?: How the Web Is Transforming Historical Research*. Montreal: McGill-Queen's University Press.

Milligan, Ian. 2013. "Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010." *The Canadian Historical Review* 94, no. 4: 540–69.

Oceanic Exchanges (website). Accessed May 1, 2019. https://oceanicexchanges.org.

Prescott, Andrew. 2014. "I'd Rather Be a Librarian: A Response to Tim Hitchcock, 'Confronting the Digital.'" *Cultural and Social History* 11, no. 3: 335–41. https://doi.org/10.2752/147800414X13983595303192.

Rockwell, Geoffrey, and Stéfan Sinclair. 2016. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, MA: The MIT Press.

Schöch, Christof. 2013. "Big? Smart? Clean? Messy? Data in the Humanities." *Journal of Digital Humanities* 2, no. 3 (Summer 2103). Accessed June 10, 2019. http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

Smith, David A., and Ryan Cordell. 2018. *A Research Agenda for Historical and Multilingual Optical Character Recognition*. Boston: Northeastern University; NULab; The Andrew W. Mellon Foundation. Accessed June 8, 2019. https://ocr.northeastern.edu/report/.

Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7, no. 1 (March 1996): 111–37. https://pdfs.semanticscholar.org/9cfc/d2dfe7927451f2c39617e6ac0aa499fd2edb.pdf.

Stevens, Hallam. 2013. *Life Out of Sequence: A Data-Driven History of Bioinformatics*. Chicago: University of Chicago Press.

Tanner, Simon, Trevor Muñoz, and Pich Hemy Ros. 2009. "Measuring Mass Text Digitization Quality and Usefulness." *D-Lib Magazine* 15, no. 7/8 (July/August 2009). https://doi.org/10.1045/july2009-munoz.

Terras, Melissa M. 2011. "The Rise of Digitization: An Overview." In *Digitisation Perspectives*, edited by Ruth Rikowski, 3–20. Rotterdam: Sense Publishers.

Terras, Melissa M. 2015. "Bluclobber, or: Enabling Complex Analysis of Large Scale Digital Collections." UCLDH Blog. Accessed May 7, 2015. https://blogs.ucl.ac.uk/dh/2015/05/07/bluclobber-or-enabling-complex-analysis-of-large-scale-digital-collections/.

Terras, Melissa, James Baker, James Hetherington, David Beavan, Martin Zaltz Austwick, et al. 2018. "Enabling Complex Analysis of Large-Scale Digital Collections: Humanities Research, High-Performance Computing, and Transforming Access to British Library Digital Collections." *Digital Scholarship in the Humanities* 33, no. 2: 456–66. https://doi.org/10.1093/llc/fqx020.

UCL. 2017. "Partners in Time: HPC Opens New Horizons for Humanities Research." Research IT Services, October 24, 2017. Accessed 8 July 2019. https://www.ucl.ac.uk/research-it-services/news/2017/oct/partners-time-hpc-opens-new-horizons-humanities-research.

Wilson, James A. J. 2017. "The Research Data Storage Service at UCL – A LEARN Case Study." In *LEARN Toolkit of Best Practice for Research Data Management*, Leaders Activating Research Networks (LEARN), pp. 78-81. https://doi.org/10.14324/000.learn.00.

Wynne, Martin. 2013. "The Role of CLARIN in Digital Transformations in the Humanities." *International Journal of Humanities and Arts Computing* 7, no. 1–2: 89–104. https://doi.org/10.3366/ijhac.2013.0083.

Teaching Humanistic Data Analysis

Ryan Cordell
Associate Professor of English
Northeastern University, USA
r.cordell@northeastern.edu

Abstract: *This paper advocates for a "data humanism" that is distinct from both computer science and data science. Seeking to simply teach data science skills to humanities students misses the unique contributions our fields and our students might make to broader conversations about data and culture. The most important reason to imagine a distinct data humanism is that doing so in turn imagines students who will go into the world outside the academy and change it, rather than simply slotting into existing frameworks. This data humanism should be exploratory, iterative, and dialogic, and takes as its goal directing scholars' attention to unexpected places in and outside digitized collections, while raising new questions about those collections and their absences. The paper concludes by offering four concrete recommendations for fostering humanistic engagement with data in the undergraduate and graduate classroom. These comprise starting with creativity, teaching using humanistic rather than pre-packaged data, foregrounding the investigation of corpora over computational methods, and likewise foregrounding the inculcation of a mindset for approaching data over any particular method.*

Keywords: data ■ computational analysis ■ programming ■ pedagogy ■ data science ■ digital humanities

INTRODUCTION: HUMANITIES DATA

Lorraine Daston and Peter Galison's 2017 book *Objectivity* attempts to trace the emergence of objectivity as a concept, ideal, and moral framework for scientists during the nineteenth century. The work focuses primarily on shifting ideas about scientific images during the period. In the eighteenth and early nineteenth centuries, Daston and Galison argue, the scientific ideal was "truth-to-nature," in which particular specimens are primarily useful for the ways in which they reflect and help construct an ideal type: not this leaf, specifically, but this type of leaf. Under this regime scientific illustrations did not attempt to reconstruct individual, imperfect specimens, but instead to generalize from specimens and portray an ideal type.

The "truth-to-nature" framework changed, in part, because of the introduction of photography, which nudged scientists toward a new ideal of mechanical objectivity. In early debates about the virtues of illustration versus photography, illustration was touted as superior to the relative primitivism of photography, as technologies such as drawing or engraving simply allowed finer detail than blurry nineteenth-century photography could. Nevertheless, photography increasingly dominated scientific images over the course of the century because it was seen as less susceptible to manipulation, and less dependent on the imagination of the artist (or, indeed, of the scientist). As Daston and Galison explain,

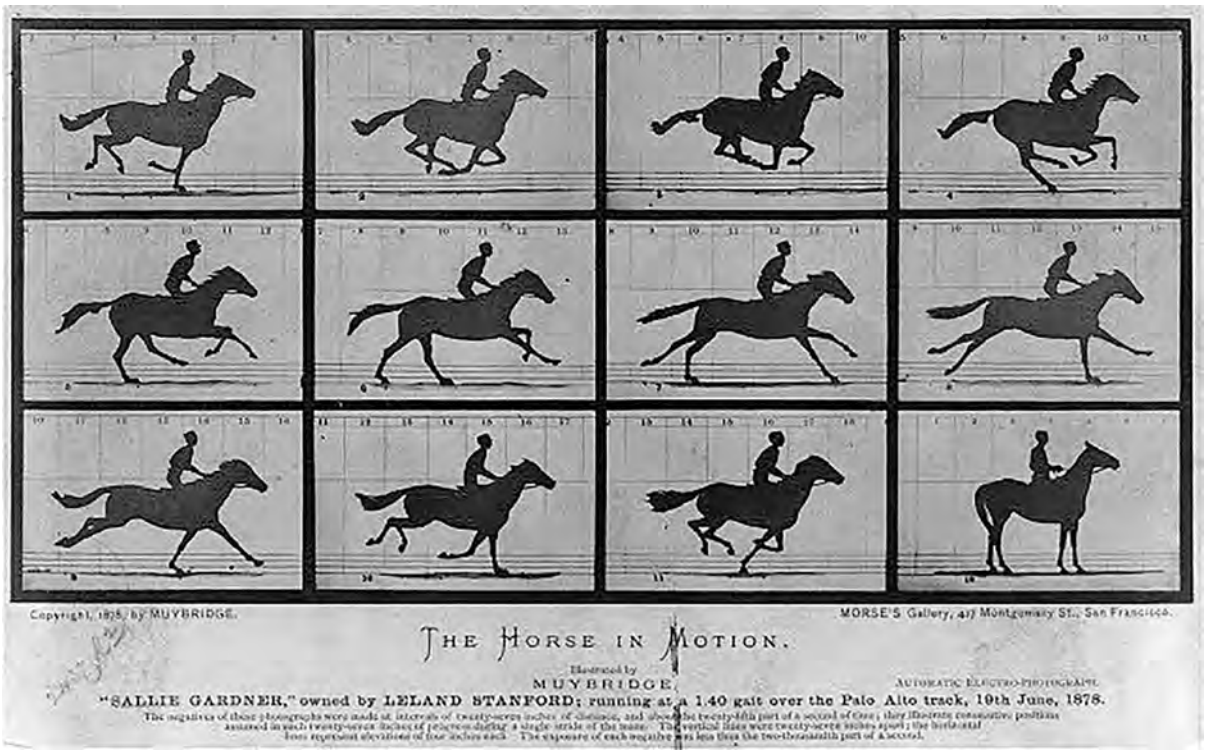
> It is this internal struggle to control the will that imparted to mechanical objectivity its high moral tone.… One type of mechanical image, the photograph, became the emblem for all aspects of noninterventionist objectivity.… This was not because the photograph was more obviously faithful to nature than handmade images—many paintings bore a closer resemblance to their subject matter than early photographs, if only because they used color—but because the camera apparently eliminated human agency. (187)

For scientists increasingly worried that their own stubborn wills would sully the truth of their findings, mechanical means of image production offered a solution.

Eventually, the camera also enabled views of nature impossible for the human eye alone. As the exposure time required for photography was reduced, for instance, it enabled new accounts of motion. Consider Eadweard Muybridge's famous photographs of racing horses, made between 1878 and 1884, which established that horses do indeed bring all four hooves off the ground at certain moments during a gallop. In this and subsequent photo series, the technology of the photograph dramatically recontextualized scientists' conception of nature by allowing them to freeze motion and analyze its constituent "parts" separately. In other words, these photographs convert a fluid motion, or a process, into data: discrete units of observation, measurement, or analysis.

This moment in scientific history recalls a touchstone essay for those engaged in computational text analysis in the humanities. In "Text: A Massively Addressable Object," Michael Witmore argues that "What distinguishes th[e digital] text object from others" is that

> it is *massively addressable at different levels of scale*. Addressable here means that one can query a position within the text at a certain level of abstraction.… The book or physical instance, then, is *one of many levels of address*. Backing out into a larger population, we might take a genre of works to be the relevant level of address. Or we could talk about individual lines of print, all the nouns in every line, every third character in every third line. All this variation implies massive flexibility in levels of address. And more provocatively,



Eadweard Muybridge's "The Horse in Motion," late 1878. COURTESY OF THE LIBRARY OF CONGRESS

> when we create a digitized population of texts, our modes of address become more and more abstract: all concrete nouns in all the items in the collection, for example, or every item identified as a "History" by Heminges and Condell in the First Folio. Every level is a provisional unity: stable for the purposes of address but also stable because it is the object of address. Books are such provisional unities. So are all the proper names in the phone book.

As Witmore acknowledges, "Physical texts were *already* massively addressable before they were ever digitized" through scholarly apparatus such as indices, concordances, or marginal notes. It is the flexibility of address that is new. Suddenly the book, or even the library, needn't be converted into a new form to become data: the book or library can itself be data.

Certainly other forms of humanistic data also existed before computation: consider the library card catalog or subject bibliography as forms of database. Like Muybridge's photographs, however, the computer reshapes our relationship to text by making it radically separable and manipulatable: a unity that can be reshaped into its constituent parts and at many levels, as Whitmore outlines. The primary objects of much humanistic inquiry—and text is only one of these, though it is what I know best and thus what I focus on here—can be considered as discrete units of observation, measurement, or analysis.

TOWARD A DATA HUMANISM

This paper poses and begins to answer two interrelated questions: first, what precisely do we mean by "data analysis" as an area of instruction in humanities courses? and second, how might we distinguish humanistic data analysis from analogous methods in the natural and social sciences? In much the same way that Datson and Galison argue scientists turned toward machines in order to avoid the mistakes of human judgment, in some articulations of digital humanities, computation is invoked as a solution to problems of will that are quite familiar from decades of humanistic scholarship. We all know that the canon of English literature was shaped as much by the identities of its gatekeepers as by the inherent virtues of the books it includes. Feminist, postcolonial, and other theoretical schools have taught us that our biases limit our critical horizons and result in atrophied canons that do not adequately represent human identities or literatures.

Methods such as distant reading, macroanalysis, or cultural analytics are sometimes proposed as methods that can bypass the fallible human wills that constructed existing canons through a species of mechanical objectivity. While human beings choose what to focus on for all kinds of reasons, including conscious or unconscious biases we would seek to mitigate, the computer will look for textual patterns unencumbered by those preconceptions. Under this vision of computational research, the machine is less susceptible to the social, political, or identity manipulations of canon formation.

If we look closer, of course, we see that this position only sets the problem at one remove, in the humans writing the code rather than in the humans assembling the canon. As scholars such as Safiya Umoja Noble,[1] Cathy O'Neil,[2] and Virginia Eubanks[3] have amply shown, computer programs always come laden with their creators' biases and oversights. Whether purposefully or inadvertently, software designed to create business or government efficiencies often reinscribes structural inequalities such as white supremacy, and as scholars such as Ben Fagan (2016) have demonstrated, the same holds true for many mass digitization projects that through their selection processes reinscribe the dominance of mainstream cultures. Given this reality, I would suggest that we cannot hope to automate our way into a more just canon formation. Instead, we should advocate for a humanities data analysis (HDA) that seeks primarily to point our attention to the gaps in our digitization and critical practices. To my mind, the most compelling articulation of this potential remains Lauren Klein's description of topic modeling (2015) as "a technique that stirs the archive" and which, in an iterative relationship, can be in turn stirred by the archive (and as a side note, Klein's book-in-progress with Katherine D'Ignazio, *Data Feminism,*[4] will be an essential touchstone in this area when complete).

In Klein's 2015 account, methods of humanistic data analysis allow researchers and students to articulate connections or paths through a collection that are not apparent through linear reading, while at the same time—and this is essential—the contours of our collections shape the kinds of connections or paths available, such that "domain experts … must be able to probe the semantic associations that the model proposes." In order to make sense of a topic model, the researcher must understand the collection upon which it was trained in order to articulate the set of possibilities available to the model and to notice what possibilities are missing.

The primary outcomes of the computational analyses Klein describes are not statistical models—though they might employ such models along the way—but instead a speculative dialogue between researcher and data. We can trace a similar conception of humanistic data analysis in Stephen Ramsay's book *Reading Machines* (2011a). Ramsay is best known for a provocative three-minute position paper, "Who's In and Who's Out" (2011b) delivered at the 2011 MLA conference—in a roundtable, it must be remembered, in which all participants were asked to draft provocative position papers to spur discussion—in which he argued that digital humanities scholars should know how to code. In *Reading Machines*, however, Ramsay offers a deeply nuanced—and, frankly, more characteristic—vision of what might constitute an "algorithmic criticism" in literary studies. In particular, Ramsay rejects scientistic frameworks for computational work in favor of a dialogic framework:

> If text analysis is to participate in literary critical endeavor in some manner beyond fact-checking, it must endeavor to assist the critic in the unfolding of interpretive possibilities.… (10) However far ranging a scientific debate might be, however varied the interpretations being offered, the assumption remains that there is a singular answer (or a singular set of answers) to the question at hand. Literary criticism has no such assumption. In the humanities the fecundity of any particular discussion is often judged precisely by the degree to which it offers ramified solutions to the problem at hand. We are not trying to solve Woolf. We are trying to ensure that discussion of *The Waves* continues. (15)

In these passages Ramsay emphasizes the distinct hermeneutic underpinnings of humanistic work. In this paper I will follow Klein and Ramsay's frameworks to argue for humanistic data analysis that is primarily exploratory, iterative, and dialogic, with a goal of directing our attention to unexpected places in and outside our digitized collections, while raising new questions about those collections and their absences.

This "data humanism" is not entirely dissimilar to the methods often gathered under the phrase "data science." At least in the United States, new masters programs in data science seem to pop up daily, alongside promises about the rewarding and well-compensated work that will follow such a credential. From the rhetoric alone, data science seems to be something like, but not quite identical to, computer science: programming is required, likely in a statistics-friendly language such as R, but the focus is not on building systems or writing software. Instead, the data scientist is adept at "just in time" programming that explores economic, political, or civic datasets, identifying trends of immediate use to government, corporate, nonprofit, journalistic, or other actors. The key trait of the data scientist seems to be flexibility, as they work across domains and datasets seeking meaningful patterns. Though these qualities are also useful for humanities data analysis, I will propose that data *humanism* must differ, both in substance and methodology, from data analytics as it is imagined and practiced.

Digital humanities scholars have been routinely accused of trying to turn literature, history, and related fields into branches of computer science. This charge is largely untrue. In particular, the interpretive aims of digital humanities (DH) work are quite distinct from computer science (CS), and many claims that conflate the fields stem from mistaken assumptions about what computer scientists do. However, the ways we teach computing proficiency, the specific computational techniques we employ, and the ways we describe the findings of computational work have drawn too exclusively from corpus linguistics, data science, and related social sciences. We have largely failed to develop specifically humanistic approaches to computational data analysis, and the keen lack of such methods has become a particular hindrance to our field's development. Seeking to simply add data science skills to humanities students misses the

unique contributions our fields and our students might make to broader conversations about data and culture. The most important reason to imagine a distinct data humanism is that doing so in turn imagines students who will go into the world outside the academy and change it, rather than simply slotting into existing frameworks. I do not propose data humanism as opposed to data science, but as a complement and (occasional) corrective.

IN PRACTICE

So what might data humanism look like in practice, and how do we inculcate its skills in our students? The remainder of this paper will draw on my experiences teaching data analysis to both undergraduate and graduate students in Northeastern University's English Department, largely through courses such as Reading and Writing in the Digital Age,[5] Technologies of Text,[6] Humanities Data Analysis,[7] and Reading Machines: Technology and the Book.[8] These issues are also very much on my mind as I prepare to teach a version of Humanities Data Analysis,[9] at the 2019 Digital Humanities Summer Institute at the University of Victoria (Canada). The principles I draw from those classroom experiences evidence my own evolving sense of how a data-rich humanities pedagogy might be structured and what goals it might seek to meet, which is to say I will reflect as much on approaches that have failed as on those that have succeeded. The *ideals* I outline are not ones I believe my teaching has yet met, though I seek each semester to work closer to them. In outlining these ideals, I also do not mean to imply that no one in computer science, data science, digital humanities, or cultural analytics follows these practices. Certainly many do; my recommendations are nearly all grounded in examples (which I will try to cite fairly).

1. START WITH CREATIVITY

For many years, I would begin the programming units in my classes with text analysis: word counts, keywords in context, n-grams, topic modeling. After spending several weeks (in a single unit within a larger class) or most of the semester (in an HDA focused class) on these activities, we might devote a final week to a more creative engagement with programming, such as building a poetic Twitter bot. In recent years, however, I have become convinced that this is precisely the wrong way to introduce humanities students to programming or data. This past year, I reversed the order of operations, and I think all of my classes benefited immensely.

In my Reading and Writing in the Digital Age course, we began our unit on data with Giorgia Lupi and Stefanie Posavec's project *Dear Data,*[10] which began as a website and became a beautiful book. Lupi and Posavec spent a year recording a different aspect of their daily lives each week, creating a new way to visualize that data on a postcard, and mailing the postcards to each other; the book compiles these 104 experiments and presents them beautifully. We began here because I wanted my students first to understand that data is a bigger category than digital data, and that the suites of visualization tools built into existing computer programs include only a handful from the infinite possibilities for data visualization. Students were then charged[11] to spend a week recording and then visualizing some aspect of their own data, reflecting both on what they learned from the data, the choices they made in representing it, and the ways those representational choices shaped what they could (and couldn't) learn about the data. An exercise such as this prepares students to encounter computational data work from a productively critical perspective.

This year I also began the programming units in two courses, my undergraduate Technologies of Text and my graduate Reading Machines, with the Twitter bot assignment that used to serve as the cap to this unit. On the whole the exercise is pretty simple: students choose a poem or other short text and then build a program that will substitute words of the appropriate part of speech into the chosen poem: think of the children's game Mad Libs. The results are sometimes nonsensical, sometimes hilarious, and sometimes oddly profound. Practically, students learn a number of important skills through this lesson, including how to manipulate text strings in R, how to query web services using an application programming interface (API), and how to output the results of an R program back to a web service such as Twitter. Most importantly, however, they

learn these skills through a process that is generative and creative, and which leverages their existing skills, such as the ability to understand and analyze a poem: at least well enough to know what substitutions are likely to "work" and be funny. Though simple, this exercise fits into a long tradition of cut-up and experimental poetry with which students are familiar, and it helps them understand programming from the outset as something more creative and expressive than they might have imagined.

The way we introduce data analysis matters. If you survey publicly available "learn to code" courses on sites such as Coursera or Codeacademy, you will notice that, regardless of language, they all begin in nearly the same way: teaching students how to do math. I do not want to imply that math does not enter into humanities data analysis—though I will discuss this further below—but I will argue that, given the huge range of tasks to which programming can be put in 2019, beginning in this way is not necessary and can throw up immediate barriers to humanities students who perhaps see themselves as not skilled at or interested in mathematics.

Most current programs for teaching coding or data analysis presume that the endeavor must be rooted in computer science, but I follow Annette Vee in seeking to decouple programming as a practice from computer science as a discipline, to the benefit of both. In her 2017 book *Coding Literacy*, Vee argues that in much of the world today "computer code is infrastructural," a literacy in the ways it is valued and deployed: "layered over and under the technology of writing, computer code now structures much of our contemporary communications" (3).

> When we consider programming a mode of written communication, it is no longer bounded by the field of computer science. Its roots are no longer solely in math, engineering, and science; they include written communication as well. Decoupling programming from CS not only helps us understand programming as communication but also frees CS from being overly identified with just one of their practices … neither programming nor CS is well served by the idea that they map perfectly onto each other. (41)

Vee shows in her book that "programming never was a domain exclusively for specialists" while advocating that we consider programming a literacy that can be applied in many ways and in many domains. This is also my goal, and is a primary reason that I seek to show students that a humanistic approach to data can have distinct aims from the outset. To expand out a bit: we need new pedagogical resources to teach data analysis for humanities students that begins with questions, data, and programming tasks that will resonate with those students.

2. TEACH USING DOMAIN-SPECIFIC DATA

My next recommendation is straightforward: when teaching students humanities data analysis, we should use humanities data. This point might seem obvious, but it has not been so in practice, in part because it can be harder to practice than it should be. Insisting on teaching with humanistic data leads inevitably to a complementary idea, that we should not shy away from data that is complex and messy. Computer science and data science are often taught using tidy datasets that are designed, quite literally, to help students gain proficiency with particular methods. I think of the "mtcars" data that is built into the R programming environment. It is a relatively small, well-structured table of information about various car models, including details such as their miles per gallon efficiency. The dataset was extracted from the 1974 issue of *MotorTrend* magazine, and is designed to be a resource for learning how to use R's mathematical or plotting functions, and thus is invoked in tutorial after tutorial teaching just these functions. The other most popular built-in datasets for R include "iris," which includes various measurements for 50 flowers in the iris family; "ToothGrowth," which contains the results of an experiment studying the effect of vitamin C on tooth growth; or "USArrests," which includes statistics about violent crime in the US.

There are a number of problems with defaulting to such datasets when teaching humanities students. First, our students often find it difficult to extrapolate from data about cars or tooth growth to the domains they care about. This is especially true for students coming to data analysis with some hesitancy or trepidation, as many stock training datasets available resonate, teleologically, as corporate or governmental. To state an even stronger version of this point, these

datasets look and feel like what our students think data should look and feel like: which is to say, they look and feel like information from domains far removed from humanistic inquiry, and they alienate students rather than welcoming them in.

Second, however, such datasets are ontologically distinct from humanities data, insofar as data drawn from literary, historical, and related domains is defined (at least in part) by its complexity and messiness. This is in fact why many computer science professors are drawn to our data for their own research: it is intriguingly difficult. A few years ago, I co-taught a course titled Bostonography for computer science undergraduates at Northeastern. Our students in that course repeatedly expressed how frustrated but exhilarated they were to be working with what they called "real data"—which is to say, historical data related to the city of Boston—rather than the canned datasets they encountered in their CS classes. The data in our class, we heard, was uniquely challenging but more rewarding than the dull stuff they encountered again and again. Humanities students need to work with this challenging humanities data from the beginning.

Unfortunately it is not self-evident how to teach using humanities data—so much so that perhaps the most common complaint among those teaching such classes is the lack of relatively small, teachable data sets in literary studies, history, or related fields. Many of the datasets one encounters in the field's literature are simply too large for a group of students to work with simultaneously in a classroom setting, and in fact can be frustrating even outside the classroom when it takes many minutes—or much longer—each time a student has to learn that a line of code does not work as it should. The iteration required for learning programming—run the code, see if it works, revise, run it again—cannot happen with very big data, which is why fields such as CS and data science have developed the teaching datasets I described earlier. As more people seek to cultivate data humanism in students, we will need to cultivate also manageable, but still interestingly messy, datasets for use in classrooms.

My own research largely centers on historical newspapers, and I have developed some unexpectedly generative teaching exercises around the Library of Congress's *U.S. Newspaper Directory.*[12] This is not their collection of digitized newspapers, which is far too large for classroom work, but instead their index of metadata about all known newspapers founded in the United States between 1690 and the present. Like R's "cars" dataset, there are many natural experiments to be run here to learn how to compare categories or plot: how many weeklies versus dailies were founded in the 1870s? Which states had the most newspapers at the turn of the century? Can we analyze or visualize the most popular words used in the titles of new newspapers over decades, and what might such analyses tell us about the changing ways editors and readers conceptualized the newspaper across time? Not all of my students are newspaper researchers, but because this data is about something closer to their interests, they know the kinds of questions we might ask about it, and these investigations are more easily transferred into their own domains.

3. FOREGROUND CORPUS OVER METHOD

Asking questions about the data we use in our classrooms leads directly to my next recommendation, which is that in humanities data analysis, teaching our corpora or datasets should precede teaching methods. In her 2018 book *A World of Fiction*, Katerine Bode draws on her book history expertise to critique the ways many computational literary scholars approach the data at the heart of their analyses. In particular, Bode argues that

> in their literary-historical work both [Franco] Moretti and [Matthew L.] Jockers present literary data and digital collections as pre-critical, stable, and self-evident. In conceiving data and computation as providing direct and comprehensive access to the literary-historical record, they deny the critical and interpretive activities that construct that data and digital record and make them available for analysis. (20)

Bode does not claim distant reading scholars fail to disclose the corpora from which their conclusions are drawn, but instead that they often do not adequately describe those corpora: what they include, what they do not. In this way, Bode argues, distant reading looks a lot like new

critical close reading, in that the specific material circumstances of the text at hand—whether that text is a single Keats poem or the HathiTrust library—are discounted or overlooked in favor of an idealized notion of "the text" or "the archive."

Bode's claim extends critiques by scholars such as Johanna Drucker, who argues that the very word "data" implies that the information it labels is "a 'given' able to be recorded and observed" or simply "a natural representation of pre-existing fact" (Drucker 2011). While I am not *quite* advocating with Drucker that we replace the word "data" with "capta"—because "*capta* is 'taken' actively"—I would concur that a data humanism must begin with discussions of data construction rather than data analysis. To make claims from the Library of Congress's *Chronicling America* newspaper archive, as we do in the Viral Texts project,[13] requires understanding how its newspapers were compiled through the National Digital Newspaper Program, which lays out specific selection guidelines for states that participate. These guidelines include ideas of "representativeness" around circulation and influence that ultimately shape the kinds of newspapers included, and the kinds excluded, from the overall collection.

Returning to classroom conversations about the Library of Congress's *Newspaper Directory*, I would note that conversations about what we might learn analyzing it lead naturally to rich conversations about the construction of the dataset itself. Who compiled these lists? How did they obtain the information about the newspapers? What information is not included? For instance, does the data include anything about the communities that various newspapers served? If we analyze the most popular words used in titles, what contours of the newspaper landscape in those years are we necessarily excluding? A humanistic approach to data analysis must foreground such conversations and teach students how to probe the intellectual, social, and political underpinnings of each dataset or corpus they encounter.

4. FOREGROUND MINDSET OVER METHOD

The courses I draw from in this paper largely teach humanities data analysis through coding in the programming language R. I would not argue that coding is the only way that effective humanities data analysis might be taught or practiced, in large part because this raises the barrier of entry too high. This is an area of inquiry that needs more voices, not fewer. There are many good tools out there for performing most common analysis tasks, and one can be thoughtful about all of the issues I have thus far outlined in graphical user interfaces (GUIs) as easily as in a programming language. And to be frank: my own work in this area began using GUIs and applications, which helped me cultivate approaches to questions that eventually required me to cultivate programming abilities. That early work was no less valid than the work I do now.

However, I will outline the reasons that I have moved more and more toward approaching humanities data analysis in my classes through programming rather than using common DH applications. In fact I should immediately modify that to say that I teach humanities data analysis through workbooks, which are in a way an intermediary form between an application and pure programming. In R such workshops are composed as RMD files,[14] and they weave together prose with executable blocks of code. In this form I can include relatively complex chunks of code, even in the first days that we work on programming, so we can move immediately to applications that will feel more consequential to students than printing "Hello, World" might. When looking at an RMD file like this, students might not understand each line. In contrast to a GUI, however, in this form each line is available for inspection, and good deal of the work we do together, at least initially, consists of running a block of code, looking at the results, and then working back through the code to understand what precisely it did step-by-step.

I follow this procedure for a few reasons. First, I largely concur with my colleague, historian Ben Schmidt (2016), that "digital humanists do not need to understand algorithms *at all*," insofar as understanding means grasping precisely what each Greek letter in an algorithm's equations stands for or being able to precisely reproduce the underlying math. What humanists do need, Schmidt continues, is "to understand the transformations that algorithms attempt to bring about." If students are to use computational techniques responsibly, they need to understand what happens to text or tabular data to move from a set of, say, novels to a vector space model. What as-

sumptions are baked into the method? What variables can be controlled, and what happens when they are changed? It is possible to do this work within a GUI? Gale's own Digital Scholar Lab is quite robust and good at foregrounding explanations of the methods it includes and the variables that researchers can control within it.

I do not wish to pretend that working in a programming language like R lays everything bare, given that many elements of particular algorithms are packaged within functions that can be run with little understanding. For me, however, working through the code helps students understand a little more fully the processes through which transformation happens, to see each major step in a given process, and to begin asking questions about how particular functions operate. My workbooks typically include chunks of prewritten code as well as instructions about lines that students can manipulate, as well as empty code blocks they can use to copy, paste, and modify the code I have given them. This kind of tinkering introduces a sense of code as a medium—it is literally text that can be copied, pasted, edited—and opens up to later workbooks in which students have more direct agency, as well as independent projects in which students ask questions about their own data.

Here's an important point to make about all of the pedagogy I have been describing: I cannot turn students into proficient, independent programmers in a four-week unit, or even in a full semester class, and that is not my goal. Nor is my goal to bring them to expertise in any particular method: e.g., classification or topic modeling. What I am trying to cultivate is a mindset for approaching data, exploring it, figuring out what questions computation might help answer about it, and then determining what methods might help answer those questions. Ultimately I want them to begin to understand how to outline a series of steps—an algorithm, in a way, though not expressed mathematically—that would allow them to transform the data from its original form into a form pertinent to the question at hand. To actually make that process happen would likely require more research and potentially collaboration. But I would argue the hardest skill for a humanities student to acquire is programmatic thinking, not programming.

A deeply humanistic, exploratory approach to data will not only result in students able to use methods developed for computer science or social science on humanities data, but also will result in the need for new methods required by our materials. This, I would argue, is an area largely unexplored in digital humanities or adjacent fields. We have seen quite sophisticated projects that apply methods such as vector space analysis or topic modeling to novels or historical newspapers, but we have not seen new methods emerging from and specifically for historical, literary, or related data. But this is the future I want to teach toward. One reason we need humanistic data analysis is that we cannot cede either the maintenance and algorithmic framing of digital cultural heritage entirely to engineers or, with all deference to our hosts today, to corporations. Certainly products such as the Gale Digital Scholar Lab (DSL)[15] can be useful starting points for research within their collections and to teach the affordances of the methods they incorporate. Gale's DSL is particularly powerful and flexible; I have been genuinely impressed with its capabilities.

Nevertheless it remains true that no standard suite of analytical tools, however thoughtfully developed, will capture the extent of what we might hope to learn from a historical, literary, or social dataset, and one of the great challenges of humanistic work is that the questions of interest are difficult to generalize or predict well enough in advance to develop a generic tool that will cover them. Teaching students using only out-of-the-box tools will limit their possibilities. In addition, I teach knowing that the majority of my students are not heading toward jobs at well-resourced universities that are likely to subscribe to the databases that make a product like the DSL useful. The majority of them are heading toward work in the much broader field of US universities and colleges or into work in the private sector. If these students are to continue data-rich humanistic work—and, essentially, train their own students to do the same—they will do so primarily using open access tools—including both GUIs and programming languages—and open access data. My students need to understand the basic underpinnings of such tools and be able to flexibly adapt to different institutional situations and, if necessary, know how to begin assembling their own suite of tools for the research and teaching they need to do.

NOTES

1. https://safiyaunoble.com/.
2. https://mathbabe.org/.
3. https://virginia-eubanks.com/.
4. https://bookbook.pubpub.org/data-feminism.
5. https://f18rwda.ryancordell.org/.
6. https://s19tot.ryancordell.org/.
7. http://s17hda.ryancordell.org/.
8. https://s19rm.ryancordell.org/.
9. http://www.dhsi.org/courses.php#DataAnalysis.
10. http://www.dear-data.com/theproject.
11. https://f18rwda.ryancordell.org/assignments/dear-my-data.html.
12. https://chroniclingamerica.loc.gov/search/titles/.
13. https://viraltexts.org/.
14. Examples of RMD files from my recent DHSI class are here: https://github.com/rccordell/DHSI-HDA/tree/master/exercises.
15. https://www.gale.com/intl/primary-sources/digital-scholar-lab.

REFERENCES

Bode, Katerine. 2018. *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor: University of Michigan Press.

Daston, Lorraine, and Peter Galison. 2017. *Objectivity*. New York: Zone Books.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5, no. 1. http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Fagan, Benjamin. 2016. "Chronicling White America." *American Periodicals: A Journal of History & Criticism* 26, no. 1: 10–13. https://muse.jhu.edu/article/613375/summary.

Klein, Lauren. 2015. "The Carework and Codework of the Digital Humanities." http://lklein.com/2015/06/the-carework-and-codework-of-the-digital-humanities/.

Library of Congress. *Chronicling America* (newspaper archive). https://chroniclingamerica.loc.gov/.

Library of Congress. *U.S. Newspaper Directory*. https://chroniclingamerica.loc.gov/search/titles/.

Lupi, Giorgia, and Stefanie Posavec. 2016. *Dear Data: A Friendship in 52 Weeks of Postcards*. New York: Princeton Architectural Press.

Ramsay, Stephen. 2011a. *Reading Machines: Toward an Algorithmic Criticism*. Urbana: University of Illinois Press.

Ramsay, Stephen. 2011b. "Who's In and Who's Out." Position paper delivered at the 2011 Modern Language Association (MLA) Conference. https://web.archive.org/web/20170426170232/http://stephenramsay.us/text/2011/01/08/whos-in-and-whos-out/.

Schmidt, Benjamin M. 2016. "Do Digital Humanists Need to Understand Algorithms?" In *Debates in the Digital Humanities*, edited by Lauren F. Klein and Matthew K. Gold. Minneapolis: University of Minnesota Press. http://dhdebates.gc.cuny.edu/debates/text/99.

Vee, Annette. 2017. *Coding Literacy: How Computer Programming Is Changing Writing*. Cambridge, MA: The MIT Press.

Witmore, Michael. 2012. "Text: A Massively Addressable Object" [2010]. In *Debates in the Digital Humanities*, edited by Matthew K. Gold. Minneapolis: University of Minnesota Press. http://dhdebates.gc.cuny.edu/debates/text/28.

Introducing Digital Humanities in the Undergraduate Classroom: Strategies, Solutions, and Pedagogical Practices Using the Gale Digital Scholar Lab

Sarah L. Ketchley

Lecturer and Professor in Digital Humanities and Egyptology
University of Washington, USA
ketchley@uw.edu

Abstract: *This paper explores the practical aspects of teaching digital humanities skills to undergraduate students primarily from humanities and social science backgrounds. The results are presented in the form of a case study of a course titled Introduction to Digital Humanities taught at the University of Washington between 2015 and 2018. While many students who took the course had little to no previous experience in the field, the majority chose the course in order to become technologically proficient and remain competitive in the job market after graduation. The case study explores the lessons learned during the first three course sessions, and describes the strategies implemented during the fourth session to address many of these instructional and classroom management challenges. These strategies included the use of a cloud-based digital tool—the Gale Digital Scholar Lab—to create and manage corpora of primary source material related to the class research topic. Students worked to clean up OCR text in the platform, before using the integrated digital tools to conduct quantitative and qualitative text analysis. These analyses included topic modeling, n-grams, named entity recognition, and sentiment analysis. Students exported their results either as raw data (CSV/JSON) to explore further using external tools including Google Fusion Tables and Voyant, or as image files to include in their final Omeka exhibits.*

Keywords: digital pedagogy ■ undergraduate ■ digital tools ■ text mining ■ digital humanities ■ cloud-based digital humanities tools

The evolving field of digital humanities offers an opportunity for students to develop technical proficiency in the context of humanities content. The field has become known for core values that include openness, collaboration, diversity, experimentation, collegiality, and connectedness, an ethos that extends into the classroom environment and teaching practices (Spiro 2012). Digital humanities is not a subdiscipline of humanities; it doesn't belong under the roof of any one department. Its scope and potential are limited only by the imagination of those involved in it. The nature of digital humanities is collaborative—much work is project-based, with the ultimate goal of making content or research available in a digital format, generally online. This content could be a collection of primary and/or secondary sources, a database, or perhaps a geographical mapping site. Planning and developing a successful digital project involves effective project management and teamwork; consideration of best practices for project preservation and sustainability;

and development and implementation of topics as diverse as text modeling, working with images, maps, and statistics—all the while considering the aesthetics of online presentation and design. It is experiential education and technological training combined with traditional liberal arts skills. Digital humanities training and hands-on experience are attractive to many students because the skills learned are readily transferable to industry and beyond.

How best to bring the digital tools and practical skills we use as researchers into an undergraduate classroom? This paper will discuss strategies for introducing digital humanities to a diverse group of students from various university departments and with a wide range of skill sets. The experience is described from the viewpoint of a faculty member who is first and foremost a scholar in the humanities using digital tools to investigate historical research data, as opposed to a faculty member whose discipline is primarily technology focused.

CASE STUDY: INTRODUCTION TO DIGITAL HUMANITIES

Introduction to Digital Humanities was a course taught at the University of Washington between 2015 and 2018 to four separate cohorts of students. The three-to-five credit class had no prerequisites, and met semiweekly for two hours each session.

The initial three course sessions were very similar in nature, comprising mixed undergraduate and graduate groups from multiple departments working in vertically aligned teams, with the graduate students acting as project leads in their respective groups. The syllabus (see figure 1) was broad-based and wide-ranging, guiding students through the process of planning and building a digital exhibit based on a particular research question. In the first two sessions, this question was chosen by the lecturer based on her familiarity with the data—Middle Eastern travel journals—and her relative unfamiliarity with teaching digital humanities. By the third course session, students were able to choose their dataset from several options. The thematic flow of the class mirrored that of a successful digital humanities project, starting with developing team norms and work plans, identifying and implementing best practices for collecting and curating data, analyzing and visualizing the collected material, and finally building a digital exhibit to showcase the products of each student team's work.

The broad learning goals of the class included:

1. fostering core computing competencies for humanities and social science students;
2. promoting interdisciplinarity and collaboration;
3. using digital tools to investigate humanities data;
4. bringing a digital project through a complete lifecycle including planning, implementing, and then concluding, with an emphasis on building for sustainability; and
5. exposing students to a range of projects, digital tools, and research methodologies falling under the umbrella of digital humanities.

Unsurprisingly, these first three times the course was offered it presented a number of pedagogical and practical challenges (not least because they were somewhat experimental in nature), and naturally evolved and improved each time. By course number four, in Autumn 2018, all the solutions to the challenges described below were implemented in the classroom, and the experience became more manageable for the instructor and more streamlined for the students.

CHALLENGES PART I: CLASSROOM

- Initial promotion of the course was difficult, listed as it was in the Near Eastern Languages and Civilization (NELC) course catalogue, since students don't generally go looking for a digital humanities course in a NELC department. In an effort to boost enrollment numbers, the lecturer did a fair amount of promotion around campus, including making posters, pitching the course to a variety of student groups and classes, and posting to various social media channels.

FALL QUARTER 2018 SYLLABUS - INFO 498 C: AN INTRODUCTION TO DIGITAL HUMANITIES	
TOPIC	**THEME**
PLANNING	Welcome! Syllabus Review
	What is a Digital Project? Evaluating Digital Projects
	Copyright & Open Source Material
	Final Presentation Planning & Omeka Overview
	Digital Scholar Lab: Intro & Demo
	LAB: Project Planning and Collaborative Working
	Risk Assessment & Data Management Plans
COLLECTING & CURATING	Metadata Lab
	Creating a Digital Archive; What is OCR?
	Corpus Creation & Curation; Cleaning Data with OpenRefine
	Text Cleaning with RegEx, Lexos & Digital Scholar Lab
	Libraries & Digital Scholarship; Text cleaning in Hathi Trust
WORKING WITH DATA	LAB: Cleaning Content Sets
	Text & Corpus Analysis with Voyant & Digital Scholar Lab
	XML & TEI
	Geoparsing with Clavin, Mapping
	LAB WORK
	Named Entity Recognition, Sentiment Analysis & Ngrams
DIGITAL SHOWCASE	LAB WORK
	Final Group Presentations & Digital Showcase

Figure 1: Syllabus for Introduction to Digital Humanities.

- The amount of time available for in-person hands-on work was limited; the University of Washington operates on the quarter system, which is just 10 weeks long, and the digital humanities classes met twice a week for approximately two hours of in-person time each.
- The students were very diverse in terms of skills and background, which presented potential problems in pitching the course at a level which would offer a worthwhile educational experience for all.
- An appropriate setup of the physical classroom space was an important consideration. The traditional university classroom often presupposes a lecture-based, passive learning experience. The nature of digital humanities work is such that this type of environment would not be conducive to positive learning outcomes.
- The lecturer was a subject matter expert first and digital humanist second. It's often the case that humanities faculty are resistant to the notion of introducing a practical element to their classroom, not least because teaching tech to a group of students vs. more traditional humanities topics can present challenges to the subject matter specialist. In this case, the lecturer had spent eight years developing technical skills on her own time as a pathway to

using digital humanities methodologies to investigate her research data. During this time, collaboration with like-minded colleagues led to the development of an undergraduate internship program in digital humanities (at Newbook Digital Texts, a publishing house), and subsequently grant funding to develop and teach the initial three introduction to digital humanities courses. Such a lengthy apprenticeship period is often not viable or practical for faculty who are heavily invested in teaching and research.

CLASSROOM SOLUTIONS

- The Informatics department proved to be a more logical place to list the introductory digital humanities course. Offering both majors and minors in the field, its classes attract students from varied departments, and without promotion or publicity, the intro class filled within a day of being open for registration.
- To maximize the amount of time available for hands-on work with the teacher and TA present, the lecturer opted to switch from spending time giving overview lectures to a flipped classroom setup. (Stommel 2013) The expectation was set that students carry out preliminary preparation before meeting in class. This involved watching introductory videos created for each class session, supplemented by course readings and online participatory discussion boards. In-class sessions began with Q&A based on work already completed, and students came to class ready to put some of the theories they had learned about into practice.
- Students completed a pre-class survey which included questions about previous experience, interests, skills, and goals so that they could be placed in well-balanced teams from day one. They were able to interact with and learn from each other immediately, while also building invaluable social skills. The survey also provided the lecturer with guidance about the level of detail necessary for each class topic, and presented opportunities to promote peer learning and engagement. Students took a second survey mid-way through the course to identify any lingering issues or concerns, as well as to solicit requests for discussion topics. The opportunity to interact with students one-on-one in each class session was invaluable in enabling the lecturer to customize the learning experience.
- The University of Washington has a number of Active Learning Classrooms (ALC) which are ideal for the type of hands-on team-based work which characterizes digital humanities, centered on discussion, one-on-one interactions, and immediate feedback and troubleshooting during in-class activities (Hornby 2017). Practical features include easily movable furniture to create hubs for teams to work in, with easy options to bring the students together in a larger group setting. Each of these hubs had a larger monitor screen at each table for collaborative work, as well as writable wall surfaces for planning and brainstorming activities, and—essentially—multiple appropriately situated electrical outlets. This classroom setup facilitated regular interaction between the teacher, TA, and students, which in turn enabled swift feedback and responses to technical issues or student questions. Students found this non-traditional setup both dynamic and practical, as it enabled them to engage not only in the material they were working with, but also in the learning process more broadly since they had a measure of control over their own experience. The team-based setup also facilitated ready collaboration and the continued development of social skills.
- In recognition of the fact that she was not an expert in every topic to be covered in class, the lecturer framed the introductory course as an opportunity to collaborate with colleagues in the library and in other university departments, in order to offer students a comprehensive and well-rounded syllabus. For instance, a lecturer in the Library and Information Science program covered topics like copyright and working with open source material, and the Metadata Librarian discussed Dublin Core standards and the importance of clear and consistent records as a matter of project sustainability. Faculty from Human Centered Design and Engineering spoke about effective project planning and design as students began the process of building their digital exhibits.

In the Autumn 2018 course, the lecturer was able to collaborate with a commercial vendor, Gale, in a trial first classroom use case of its recently released Gale Digital Scholar Lab.

CHALLENGES PART II: TECHNOLOGY

The first three times the course was taught involved experimentation with many digital tools and platforms, which, in retrospect, presented a somewhat ambitious undertaking and often created rather than solved problems. A selection of these challenges is described below.

- Students showed up with a wide variety of tablets and laptops with diverse operating systems and even system languages. Since the class relied heavily on software that students either downloaded to their local machines or used online, at least one whole session was dedicated to downloads and installs, and troubleshooting continued over the course of the quarter. Under these circumstances, it was often difficult to offer an engaging educational experience.
- The range of analyses that students were conducting in the first three classes necessitated the use of multiple platforms over the course of the quarter. Each platform had its own quirks, and each had a learning curve (sometimes a steep one). It was often the case that workflow was neither clear nor intuitive, and learning how to access and navigate the platforms took significant time. (Christie 2017).
- On a broader scale, use of technology in the course was hampered by institutional limitations, particularly in terms of IT and software support, and dedicated support for digital scholarship.

TECHNOLOGY SOLUTIONS

Collaboration with Gale, and use of the Gale Digital Scholar Lab, enabled student use of a single solution for much of the process of corpus building, data curation, analysis, and visualization. Since the Gale Digital Scholar Lab is cloud-based, the types of computers students brought to class (or not) were no longer an issue. Similarly, by limiting the number of secondary tools students installed and used, this aspect of classroom management became less of a headache and the focus shifted towards developing research questions and ultimately presenting the results in a digital exhibit.

AUTUMN 2018 INTRODUCTION TO DIGITAL HUMANITIES

The remainder of this paper will focus on the specifics of the Autumn 2018 course, discussing the syllabus and the digital tools used by students in class. The demographics of the 2018 course were much broader than earlier iterations, since Information School classes are heavily sought-after by the student population at large (see figure 2). The class filled within hours, with a cohort of 35 students. Twenty-one university departments were represented in class, and while most students were university juniors and seniors (third/fourth year), 90 percent of the group had no previous experience in using digital tools to investigate humanities data—indeed most were confused about what digital humanities represented as a discipline. They joined the class either to develop digital literacy, or to apply existing technical or discipline-specific tools to a humanities dataset.

The broad course learning objectives have been described above, and in order to meet these goals, the lecturer worked to create an environment and class culture that encouraged autodidacticism and peer learning, as well as knowledge-sharing, participation, and collaboration (see also Savonic and Tagliaferri 2017). The lecturer modeled this ethos by inviting guest presenters to share expertise, highlighting the benefits of cross-disciplinary collaboration. Flipping the classroom helped de-emphasize the traditional classroom model of the professor as the authority figure and shift the in-class focus towards building relationships with individual students, while keeping an eye on team dynamics at the same time. Early class sessions focused on

Autumn 2018 Student Demographics

Percent	University Department
24%	Geography
15%	Economics
12%	Informatics
6%	Library & Info Science
6%	English
6%	Mathematics
3%	Philosophy
3%	Environmental Science and Resource Management
3%	Business Administration
3%	Communication
3%	History
3%	Art
3%	French
3%	Business Administration (Finance)
3%	Earth & Space Sciences
3%	Near & Middle Eastern Studies
3%	Political Science
3%	Pre-Architecture
3%	Material Science & Engineering
3%	Psychology
3%	Sociology

Figure 2. Autumn 2018 Student Demographics.

discussing and developing strategies for professionalizing roles within teams, which included establishing and maintaining expectations for working in a team environment, delivering on time, and reporting progress and final outcomes. While each student had a designated role within their team (for example, project manager, content specialist, data wrangler, visualization specialist, etc.), it was important to foster an environment where equal weight was accorded to what might otherwise be labeled as faculty vs. technician roles.

In order to encourage a spirit of inquiry and curiosity, and since students were exploring datasets in topics which were new to them, a goal was to encourage students to experiment without fear of failure. To this end, one of the bedrock assignments was a weekly worklog documenting class notes, reading, and discussions, as well as full step-by-step guides to tools used and analyses run. Provided that students kept a complete record of their work, they were not penalized if they ran an analysis that returned lackluster results, or experimented with a tool that, ultimately, did not work out as expected.

The research theme of the course was historic menus, which was chosen because it was broad yet inclusive enough to engage everyone in the class, with plenty of scope for student teams to personalize their research question. The theme was inspired by New York Public Library's (NYPL) "What's on the Menu" project, and students mostly used content sourced from Gale Primary Source archives as well as some additional material from the NYPL project and

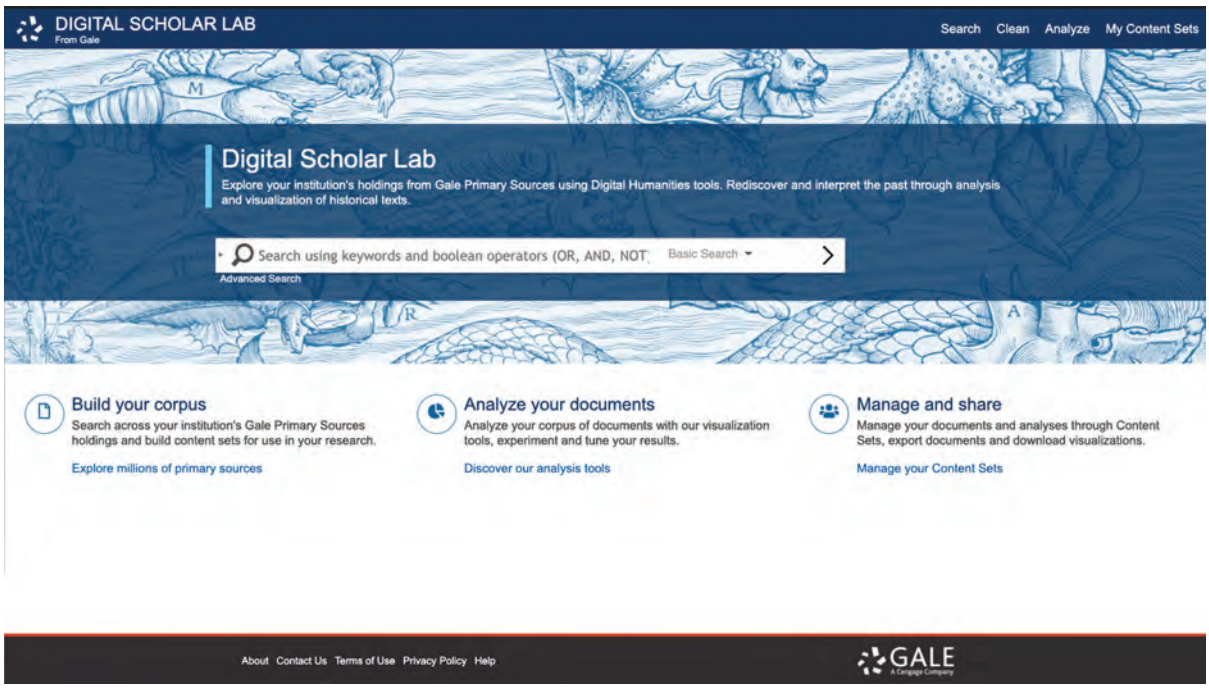


Figure 3. Screenshot of Gale Digital Scholar Lab Landing Page.

other open source repositories. It's often the case that students and faculty may be unaware of the extent and range of digital collections available to them via their library website. A significant benefit of Gale Digital Scholar Lab usage is that it promotes awareness and drives usage of an institution's existing database collections, including Gale Primary Sources, which are readily available to support research, teaching, and digital scholarship.

Over the course of the quarter students learned best practices for planning their projects, creating and curating a dataset, and using digital tools to analyze the material they had collected, before building a digital exhibit to display the results of their work. We used Omeka for display purposes, and Gale Digital Scholar Lab for the process of collaborative dataset creation, curation, and analysis.

The workflow we followed in the course is summarized here on the landing page of the Gale Digital Scholar Lab (see figure 3). Students began by building a corpus of material (described as "content sets" in the Gale Digital Scholar Lab), going through the process of curating it and then cleaning the optical character recognition (OCR) data, before taking their content sets through the various forms of text mining analysis and generating visualizations of analysis results. The importance of workflow and process was highlighted by the lecturer in the form of recorded tutorial videos, by an in-class demo, and by directing students to the contextual Help documentation on each webpage of the Gale Digital Scholar Lab—which, importantly, includes a glossary of digital humanities terms, in recognition of the fact that vocabulary is often the biggest barrier to understanding in a new field.

Each student team had a group Gmail login to facilitate collaboration on the same material without interfering with the work of any of the other teams in the class. When they logged in, students accessed a personalized workbench area that stored the content sets they created and the analyses they ran, along with descriptive metadata about the documents in each content set (see figure 4). Students also had the option to export a variety of information including individual primary source images, OCR text, visualizations, and raw tabular data, which they could then use in external tools or, in the case of the intro class, include in a digital exhibit.

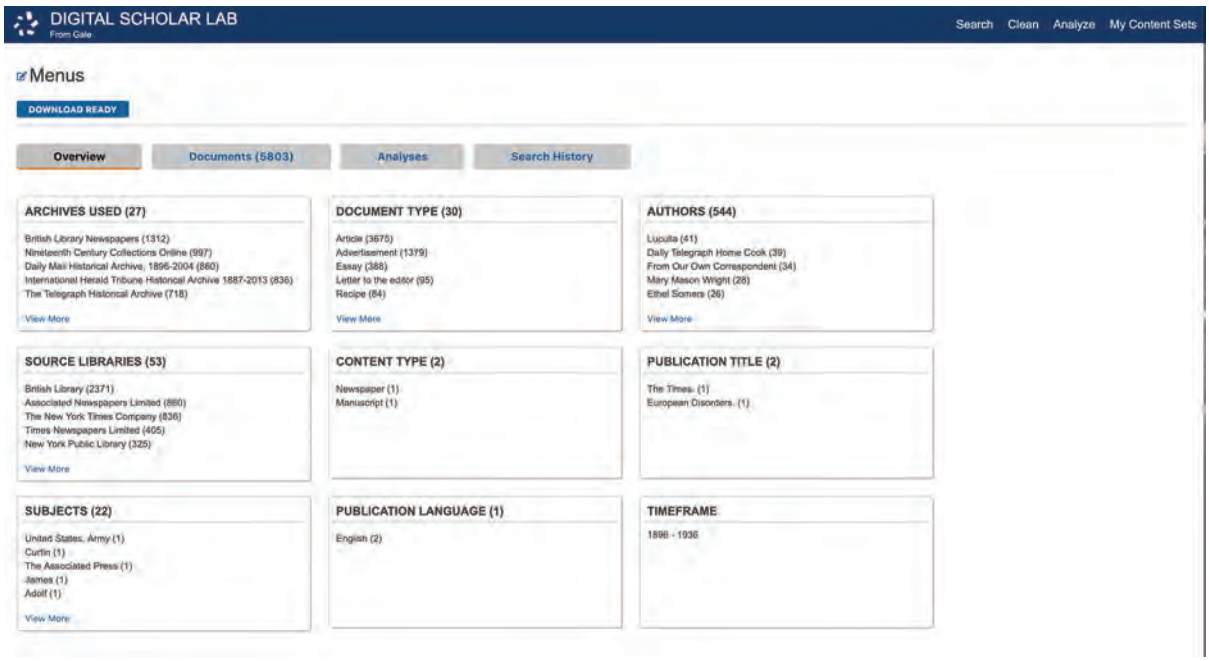


Figure 4. Screenshot of Gale Digital Scholar Lab's My Content Sets.

1. Creating and Curating Content Sets

The starting point for student work was the creation of content sets related to their team topic. Searching for the word "menus" in its broadest sense returned thousands of results, so students worked with Advanced Search filtering to curate their results into manageable content sets. The teams brainstormed menu-related topics of interest and then iterated through the process of searching and curating to hone and develop their research question, and to source sufficient information to analyze.

As students decided which documents were most relevant for research, they used the metadata as a guideline, as well as the snippet of OCR text for each document on the Search Results page. Content sets can contain up to 10,000 documents, which was more than adequate for class purposes.

Data curation takes a long time if it is done well, yet it is an essential building block if analysis results are going to be meaningful. Students were able to navigate into specific documents to take a closer look at a side-by-side view of the relevant original article and its OCR text, to determine its suitability for inclusion (see figure 5).

In going through this process, they used the OCR confidence level that is called out for each document as a guideline for curation, not necessarily discounting material with a low confidence rating because it could have been image heavy, or scanned from a poor quality original or microfiche, or even an older version of the OCR scanning software.

The results of this iterative process of data curation were that each of the seven student teams developed research questions and built a collection of digital objects in the Gale Digital Scholar Lab. There was great diversity in the topics chosen, ranging from "what did vegetarianism mean before World War I?," "evolving attitudes towards Chinese food in the US," and "food patterns in boom and bust," to "the evolution of the Thanksgiving menu since the 1800s," "the history of the famous Terrace Room in the Hotel New Yorker," "the history of eggs Benedict as a breakfast dish," and lastly, "By the Glass"—a project mapping wine on menus to their contextual history.

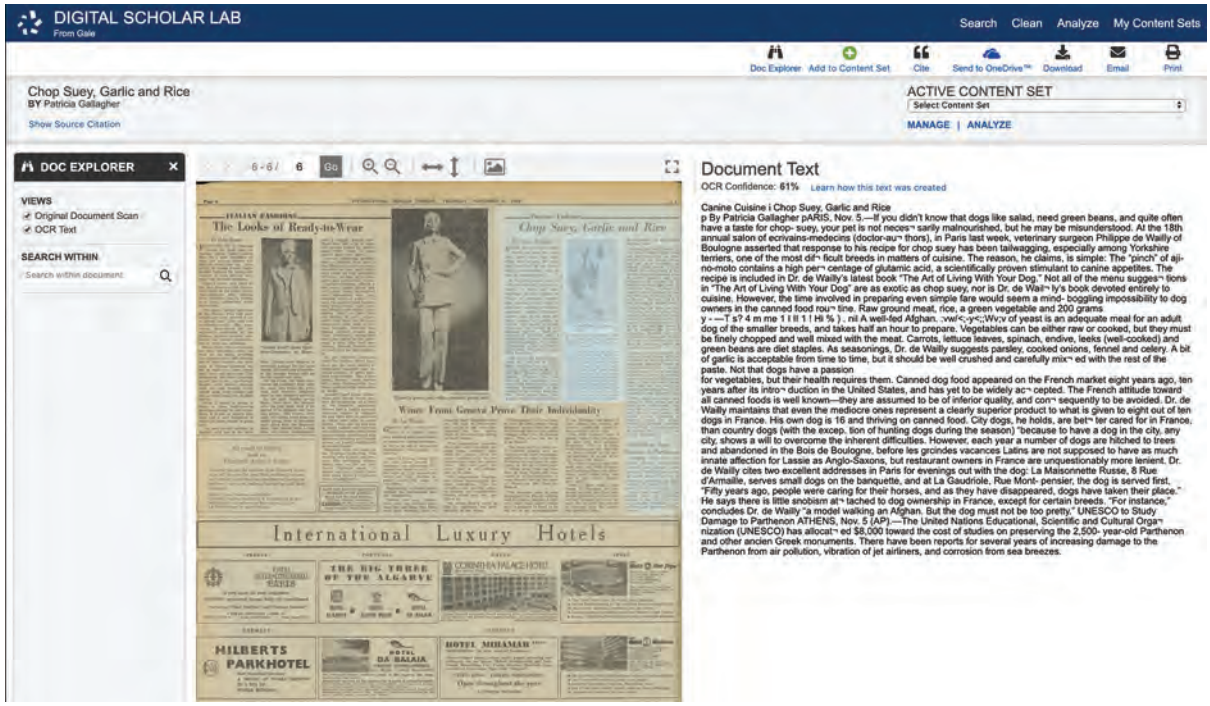


Figure 5. Screenshot of Gale Digital Scholar's Document Explorer.

2. Preparation for Analysis

An important yet often overlooked part of the text mining and analysis process is the cleaning of OCR text. At the time the class started, text cleaning wasn't an inbuilt feature in the Gale Digital Scholar Lab; it was added to the platform in November 2018. Therefore, initially at least, students had to export their OCR texts and clean them outside the Gale Digital Scholar Lab using Lexos and OpenRefine. A number of students experienced some of the familiar download and install challenges and, particularly with OpenRefine, a steep learning curve to get up and running. Once cleaning was introduced to the Gale Digital Scholar Lab, students were able to move immediately from their content sets into the cleaning process in a more streamlined and intuitive fashion.

An important starting point for this type of work is familiarity with the dataset, which enables the researcher to make considered choices about how and what to clean. Students had gained a measure of this insight through the process of careful content set curation. Text cleaning is another slow and iterative process, and in the case of the Gale Digital Scholar Lab, involved downloading a sample set of ten cleaned and ten uncleaned documents for side-by-side comparison, to evaluate the effect that text cleaning choices had on the dataset. Students would inevitably see errors that they had missed, then return to their cleaning configuration to tweak it accordingly. The value of doing this work was underscored by comparing analyses run on uncleaned datasets vs. those that students had taken the time to standardize and clean. Results were invariably more significant and comprehensible in the latter.

3. Analyzing Texts

One of the most useful features of the Gale Digital Scholar Lab is that primary source content and text analysis tools are integrated in a single streamlined work environment. There were six tools for text mining and analysis integrated in the platform, four of which were open source (topic modeling, named entity recognition, parts of speech tagging, and clustering analysis), and two that were custom implementations by Gale (sentiment analysis and n-grams). The Analysis tool

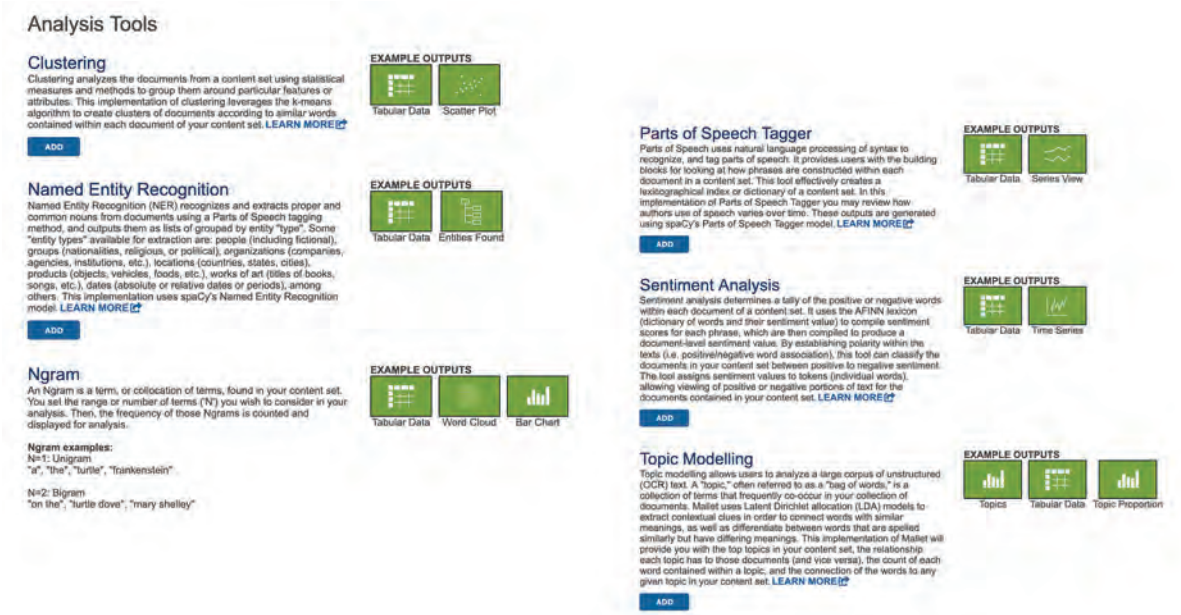


Figure 6. Screenshot of Gale Digital Scholar Lab's Analysis Tools.

page gives a high-level definition of each tool, with links out to developer documentation and algorithms that are important for reproducibility and citation (see figure 6). The platform's Help documentation goes into more detail about the uses and limitations of each tool.

Students focused on four of these tools, largely due to the constraints of time. They began by running a topic modeling analysis to discern if there were themes or topics not immediately visible in their content sets that may merit further investigation. The tool is built on MALLET, which is open source, and widely used by digital humanists. Until recently it was run from the command line, which was out of scope in the context and constraints of the introductory class. In using the topic modeling tool built into the Gale Digital Scholar Lab, students were able to choose the number of topics they wanted the algorithm to return as well as the number of words per topic. In most cases we stuck to the defaults (ten in each case), which returned the most prevalent themes in each team's dataset, many of which were surprising to students and helped shape the direction of further research.

Students ran n-gram analyses to identify the most prevalent words in their datasets, and learned through the process of text cleaning to move beyond the common stop words into more meaningful analysis and visualization. This tool was also a useful starting point for students as they began to explore the material they had gathered; switching between n-grams of various lengths (unigrams, bigrams, trigrams, etc.) enabled them to look at specific words in their broader context.

One of the students noted the significance of the sentiment analysis they ran on their dataset, commenting "Sentiment analysis was by far the most useful because it really captured what happened in 1934. However, we learn that as people got used to the conditions of the Great Depression, they start being more positive again even in the midst of the time period." A few of the conclusions that students were able to draw from their analyses included noting the increasing secularization of material related to Thanksgiving in modern menus vs. those from the nineteenth century; charting the rise of so-called Chinese restaurant syndrome as an extension of anti-Chinese sentiment, as well as tracking the invention of chop suey and the Americanization of Chinese food; and mapping the origins of eggs Benedict as a breakfast staple and its rise in popularity over time.

The final tool students worked with was named entity recognition, which extracts and lists entities such as place names, people's names, organization names, and so on from a given content set. Students used this process as the foundation for mapping their datasets; for example, extracting place names from a Chinese menu dataset related to President Nixon's visit to China in 1972, downloading this information as a CSV or JSON file, and then visualizing the data using Google Fusion Tables.

4. Exporting Data and External Analysis

The Gale Digital Scholar Lab export functionality enabled students to combine datasets created in the platform with those gathered from other sources, and then to analyze all the material using external tools. A CSV download of a topic modeling analysis includes all the data seasoned users of MALLET would expect to see, with the addition of listings for individual Gale document ID numbers.

Some teams also work with material sourced from other open repositories, which they included in their final digital exhibit. Additional external platforms students used were also cloud-based, and included Knightlab's StorymapJS for building project narratives, and Voyant for text mining.

5. Research Results and Digital Exhibits

The final stages of the process involved each team building an Omeka digital exhibit, with the following rubric:

> *Explore creating pages and sub-pages within your Omeka exhibit as a way of displaying the relevant information. Be thoughtful about design, aesthetics, and usability. (Note: the theme you choose will dictate certain layouts, for example, whether the menu is a sidebar or header menu.)*
>
> *1.* ***LANDING/HOME PAGE*** *giving details about your group's project, with appropriate navigation. This should be linked from the main class webpage. Think of a creative title for your work! (e.g., not "Group 1's Project" but "Gastronomy at the Turn of the Century," or something like that).*
>
> *2.* ***ABOUT PAGE***:
> - *Team Members—with a photo and brief bio.*
> - *Each team member should compile their worklogs for the quarter into a single document and store it as a file in Omeka. Link to the file at the end of Team Member bio.*
> - *Completed Project Charter.*
> - *Project One pager.*
> - *Draft Data Management Plan (you don't need to worry about funding issues, but give some consideration to backup and storage).*
>
> *3.* ***A COLLECTION OF ITEMS*** *in Omeka*
>
> *Each person should collect at least five and up to ten items, creating a collection of around twenty to fifty items based on your group size. Each person should have appropriately completed metadata fields for their items, showing consistency across the entire group's collection, and completeness (no empty fields).*
>
> *4.* ***DATA VISUALIZATIONS***
>
> *Include a page that links to at least one detailed data visualization, comprising narrative, analysis, and images from your Omeka collection, which you have worked on collaboratively as a group. (Storymap is an example of a platform that will fulfill these criteria.) Choose from the categories that mirror what we work on in class:*
> - *text analysis (e.g., word frequency analysis, sentiment analysis, topic modeling, theme analysis, etc.)*

- *mapping*
- *data visualization (e.g., network analysis)*

AND include examples (in the form of a link, image, or screenshot) from each of the remaining categories based on your work in class during the hands-on sessions. These might include:

- *word clouds based on the Voyant/N-Grams/Topic Modeling/Text Analysis classes*
- *one or more maps from the Mapping class*
- *term frequency chart or topic modeling sunburst wheel*

The final exhibits and student presentations were impressive, and testament to the students' enthusiasm for their project work and commitment to their team's effort. See figures 7a through 7c for the project home page, the opening screen of one student's exhibit, and an example of a link from that student exhibit's home page.

LEARNING OUTCOMES: MAPPING SKILLS TO JOBS

One class session was dedicated to a discussion of identifying and documenting the transferable skills that students were learning and using, primarily in this digital humanities course but also in their broader undergraduate education. This conversation took place in week seven, once students had had time to work with their teams and make considerable headway in their digital research project. It was prompted largely by lecturer interactions with students during team check-ins each week, when it became apparent that there was general concern about employment prospects particularly for humanities students beyond college. While identifying that all undergraduate students will have learned how to think critically, reason analytically, and communicate effectively, beyond this, the group was able to clearly identify a range of specific skills they were developing and using in class as being relevant and marketable to employers. From data wrangling and management, through project planning and design, these skills also included writing in various formats (for an online audience, for documenting technical decisions, etc.),

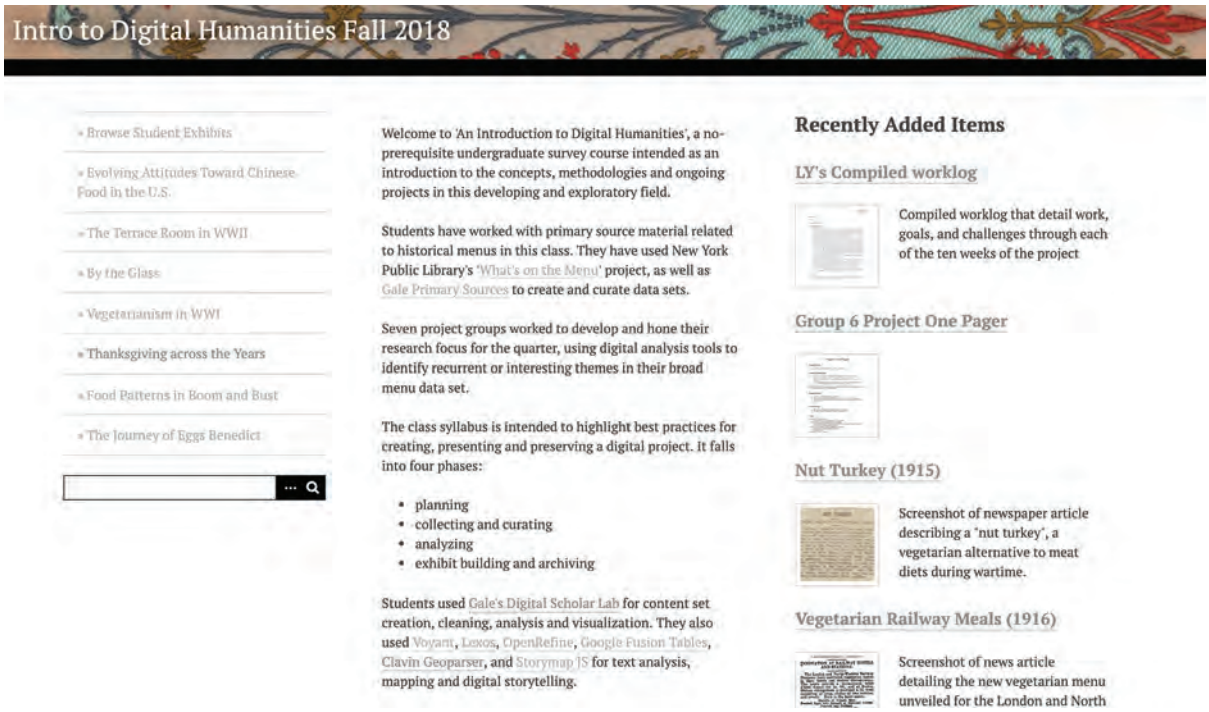


Figure 7a. Home page for the Intro to Digital Humanities Fall 2018 projects.

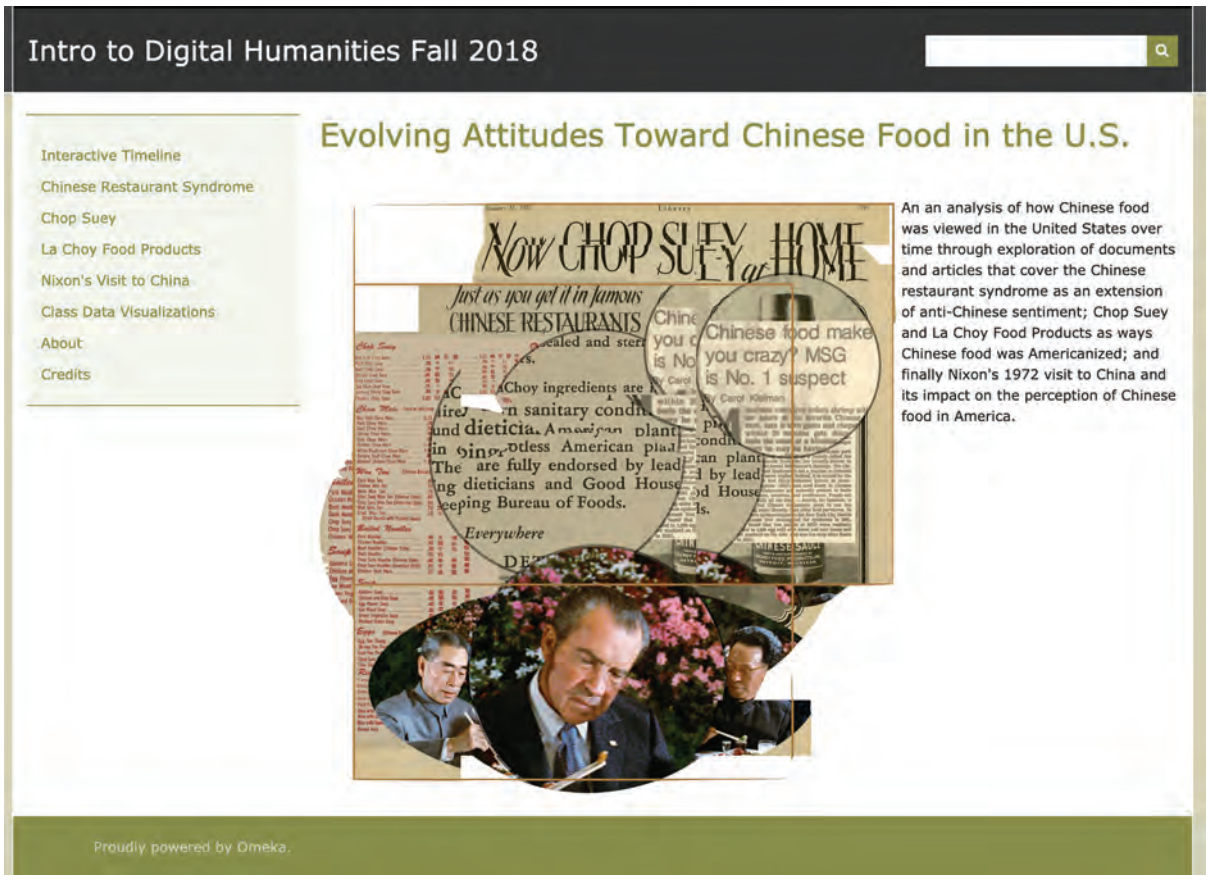


Figure 7b. Student Exhibit: Evolving Attitudes toward Chinese Food in the U.S.

interpreting and explaining data analysis and visualizations, and carrying out research as part of a team, working collaboratively and responsibly. In short, students saw great value in the type of work they were doing in class, and the dots it connected as they transitioned towards life beyond university.

REFERENCES

Álvarez Sánchez, Adriana, and Miriam Peña Pimentel. 2017. "DH for History Students: A Case Study at the Facultad de Filosofía y Letras (National Autonomous University of Mexico)." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Bretz, Andrew. 2017. "The New Itinerancy: Digital Pedagogy and the Adjunct Instructor in the Modern Academy." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Brown, Susan. 2016. "Tensions and Tenets of Socialized Scholarship." *Digital Scholarship in the Humanities* 31, no. 2 (June 1, 2016): 283–300. https://doi.org/10.1093/llc/fqu063.

Christian-Lamb, Caitlin, and Anelise Hanson Shrout. 2017. "Starting from Scratch? Workshopping New Directions in Undergraduate Digital Humanities." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Christie, Alex. 2017. "Building a Toolkit for Digital Pedagogy." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Clement, Tanya. 2012. "Multiliteracies in the Undergraduate Digital Humanities Curriculum: Skills, Principles and Habits of Mind." In *Digital Humanities Pedagogy: Practices, Principles and Politics*, edited by Brett D. Hirsch, 365–88. Cambridge, UK: Open Book Publishers.

Dunn, Stewart. 2016. "Quantitative, Qualitative, Digital. Research Methods and DH." *Stuart Dunn* (blog), September 21, 2016. https://stuartdunn.blog/2016/09/21/quantitative-qualitative-digital-research-methods-and-dh/.

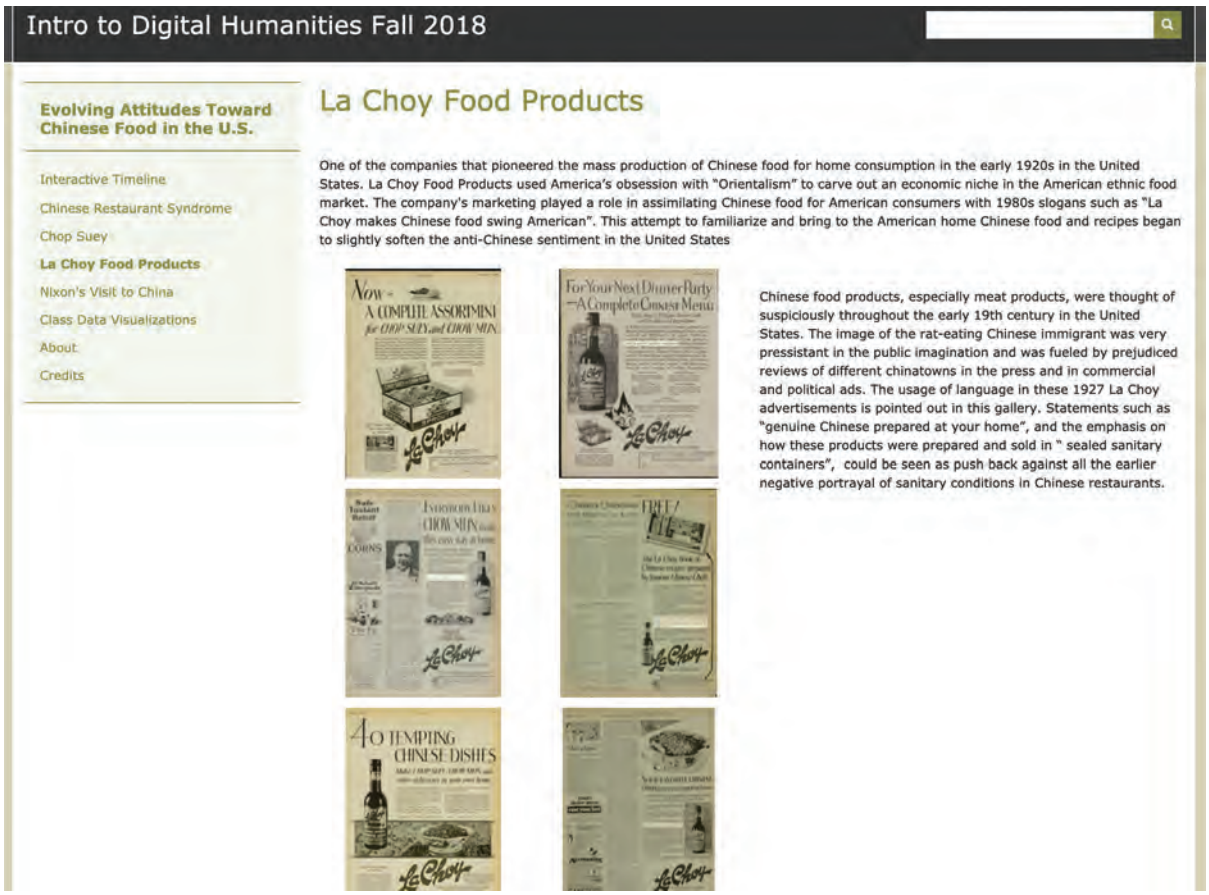


Figure 7c. Evolving Attitudes toward Chinese Food in the U.S.: Link for La Choy Products.

Estill, Laura. 2017. "Collaborative Knowledge Creation and Student-Led Assignment Design: Wikipedia in the University Literature Class." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Faull, Katherine M., and Diane K. Jakacki. 2015. "Digital Learning in an Undergraduate Context: Promoting Long-Term Student–Faculty Place-Based Collaboration." *Digital Scholarship in the Humanities* 30, no. suppl_1 (December 1, 2015): i76–82. https://doi.org/10.1093/llc/fqv050.

Gale. n.d. "Gale Digital Scholar Lab: Open New Research Pathways." Accessed March 5, 2019. https://www.gale.com/intl/primary-sources/digital-scholar-lab.

Gold, Matthew K., ed. 2012. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press.

Google. n.d. "Google Fusion Tables." Accessed March 30, 2019. https://fusiontables.google.com/DataSource?dsrcid=implicit.

Hirsch, Brett D., ed. 2012. *Digital Humanities Pedagogy: Practices, Principles and Politics*. Cambridge, UK: Open Book Publishers. https://doi.org/10.11647/OBP.0024.

Hswe, Patricia, Tara LaLonde, Kate Miffitt, James O'Sullivan, Sarah Pickle, Nathan Piekielek, Heather Ross, and Albert Rozo. 2017. "A Tale of Two Internships: Developing Digital Skills through Engaged Scholarship." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

HumTech. 2015. "A Student Collaborators' Bill of Rights." University of California, Los Angeles, June 8, 2015. https://humtech.ucla.edu/news/a-student-collaborators-bill-of-rights/.

Jenstad, Janelle, Kim McLean-Fiander, and Kathryn R. McPherson. 2017. "The *MoEML* Pedagogical Partnership Program." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Kelley, Shannon. "Getting on the Map: A Case Study in Digital Pedagogy and Undergraduate Crowdsourcing." 2017. *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Kennedy, Kara. 2017. "A Long-Belated Welcome: Accepting Digital Humanities Methods into Non-DH Classrooms." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Koh, Adeline. 2014. "Introducing Digital Humanities Work to Undergraduates: An Overview." Hybrid

Pedagogy, August 14, 2014. http://hybridpedagogy.org/introducing-digital-humanities-work-undergraduates-overview/.

Ledolter, Johannes, and Lea VenderVelde. 2019. "A Case Study in Text Mining: Textual Analysis of the *Territorial Papers*." *Digital Scholarship in the Humanities*.

Lexos (website). Wheaton College (Norton, MA). Accessed March 30, 2019. http://lexos.wheatoncollege.edu/upload.

Locke, Brandon T. 2017. "Digital Humanities Pedagogy as Essential Liberal Education: A Framework for Curriculum Development." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Mahony, Simon, and Elena Pierazzo. 2012. "Teaching Skills or Teaching Methodology?" In *Digital Humanities Pedagogy: Practices, Principles and Politics*, edited by Brett D. Hirsch. Cambridge, UK: Open Book Publishers, 2012.

Martin, Allan. 2008. "Digital Literacy and the 'Digital Society.'" In *Digital Literacies: Concepts, Policies and Practices*, edited by Colin Lankshear and Michele Knobel, 151–76. New York: Peter Lang.

Mauro, Aaron, Daniel Powell, Sarah Potvin, Jacob Heil, Eric Dye, Bridget Jenkins, and Dene Grigar. 2017. "Towards a Seamful Design of Networked Knowledge: Practical Pedagogies in Collaborative Teams." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

McCallum, Andrew Kachites. 2002. "MALLET: A Machine Learning for Language Toolkit." University of Massachusetts, Amherst. http://mallet.cs.umass.edu.

MLA Commons: Digital Pedagogy in the Humanities (website). Accessed February 26, 2019. https://digitalpedagogy.mla.hcommons.org/.

Modern Language Association. n.d. "Text Analysis." MLA Commons: Digital Pedagogy in the Humanities. Accessed March 26, 2019. https://digitalpedagogy.mla.hcommons.org/keywords/text-analysis/.

Morris, Sean Michael. 2013. "Decoding Digital Pedagogy, Pt. 1: Beyond the LMS." Hybrid Pedagogy, March 5, 2013. http://hybridpedagogy.org/decoding-digital-pedagogy-pt-1-beyond-the-lms/.

Murphy, Emily Christina, and Shannon R. Smith. 2017. "Introduction." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Murphy, Emily Christina, and Shannon R. Smith. 2017. "Undergraduate Students and Digital Humanities Belonging: Metaphors and Methods for Including Undergraduate Research in DH Communities." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

New York Public Library. n.d. "What's on the Menu?" Accessed March 12, 2019. https://www.nypl.org/node/163851.

Newbook Digital Texts (website). Accessed March 30, 2019. http://www.newbookdigitaltexts.org/.

Omeka (website). Accessed March 30, 2019. https://omeka.org/.

OpenRefine (website). Accessed March 30, 2019. http://openrefine.org/.

Saum-Pascual, Alex. 2017. "Teaching Electronic Literature as Digital Humanities: A Proposal." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Savonick, Danica, and Lisa Tagliaferri. 2017. "Building a Student-Centered (Digital) Learning Community with Undergraduates." *Digital Humanities Quarterly* 11, no. 3 (October 2, 2017).

Spiro, Lisa. 2012. "'This Is Why We Fight': Defining the Values of the Digital Humanities." In *Debates in the Digital Humanities*, edited by Matthew K. Gold. Minneapolis: University of Minnesota Press, 2012.

Stommel, Jesse. 2013. "Decoding Digital Pedagogy, Pt. 2: (Un)Mapping the Terrain." Hybrid Pedagogy, March 5, 2013. http://hybridpedagogy.org/decoding-digital-pedagogy-pt-2-unmapping-the-terrain/.

StoryMap (website). Accessed March 30, 2019. https://storymap.knightlab.com.

University of Washington Libraries. 2017. "Active Learning Classroom Resources: ALC Research & Pedagogy." Updated June 26, 2017. http://guides.lib.uw.edu/c.php?g=342298&p=2838500.

Uprichard, Emma. 2014. "Big-Data Doubts." *The Chronicle of Higher Education*, October 13, 2014. https://www.chronicle.com/article/Big-Doubts-About-Big-Data-/149267.

Voyant Tools (website). Accessed March 30, 2019. https://voyant-tools.org/.

Wiersma, Ashley. 2013. "Talking about Digital Pedagogy." *Cultural Heritage Informatics Initiative* (blog), March 12, 2013. http://chi.anthropology.msu.edu/2013/03/talking-about-digital-pedagogy/.

Authors

Ryan Cordell is Associate Professor of English at Northeastern University and a Core Founding Faculty Member in the NULab for Texts, Maps, and Networks (https://northeastern.edu/nulab/). His scholarship seeks to illuminate how technologies of production, reception, and remediation shape the meanings of texts within communities. He primarily studies circulation and reprinting in nineteenth-century American newspapers, but his interests extend to the influence of computation and digitization on contemporary reading, writing, and research. Cordell collaborates on numerous DH projects including the NEH- and ACLS-funded Viral Texts project (http://viraltexts.org), and as primary investigator for the Digging into Data project Oceanic Exchanges (http://oceanicexchanges.org). At Northeastern, he serves as graduate program director for the English Department and directs the Huskiana Experiential Letterpress Studio (https://cssh.northeastern.edu/huskiana/). He is also a senior fellow in the Andrew W. Mellon Society of Critical Bibliography at the Rare Book School (http://rarebookschool.org/admissions-awards/fellowships/sofcb/) and serves on the executive committee of the MLA's Forum on Bibliography and Scholarly Editing.

Joris van Eijnatten is Professor of Cultural History at Utrecht University, the Netherlands. As a cultural historian, he works on various interrelated fields, including the history of ideas, religion, media, and communication. He is specialized in particular in digital historical research (digital history, digital humanities). His current projects are geared towards the digital analysis of cultural memes (continuously recurring patterns of words and concepts) in the nineteenth and twentieth centuries, based on newspapers, periodicals, parliamentary records, and other digitized texts. Despite his passion for Python, he is writing an analogue book on the history of the Dutch Republic in the eighteenth century. His teaching covers various areas, ranging from Big History and the history of cultural history to the conceptual history of modern "isms." Joris van Eijnatten is a co-founder and editor of the open-access journal *International Journal for History, Culture and Modernity*.

Tessa Hauswedell is a cultural historian who has worked extensively with digitised historical newspapers on two large-scale digital humanities projects. As part of the Oceanic Exchanges project (funded by the Digging into Data Challenges and the EHRC), she has conducted text-mining on *The Times Digital Archive* (TDA) to research debates about female emigration from Britain in the nineteenth century. She has also worked on the HERA-funded Asymmetrical Encounters project, where she researched debates about European cities in multilingual European newspaper collections. Most recently, she is co-editor of *Re-mapping Centre and Periphery: Asymmetrical Encounters in European and Global Context* (UCL Press, 2019).

Sarah Ketchley is a Faculty Affiliate Instructor in the Department of Near Eastern Languages and Civilization at the University of Washington, and Adjunct Associate Professor at the University of Maryland University College. Her current research and teaching interests include the history of the Golden Age of Egyptology at the end of the nineteenth and beginning of the twentieth centuries. She has developed and taught introductory-level courses in digital humanities to multiple cohorts of undergraduate and graduate students. In 2012, she co-founded Newbook Digital Texts, an innovative digital humanities publishing house re-imagining and restructuring traditional academic research, publication, and education. Newbook's flourishing internship program offers students the unique opportunity to work with primary source material, to transcribe and encode texts, and to create digital maps and visualizations. Sarah joined Gale in 2018 as a Digital Humanities Specialist, working to promote and support digital scholarship using the Gale Digital Scholar Lab.

Julianne Nyhan is Associate Professor of Digital Information Studies at UCL (University College London), where she leads the Digital Humanities MA/MSc programme. She is also Deputy Director of the UCL Centre for Digital Humanities. Julianne has published widely on the history of digital humanities, most recently (with Andrew Flinn) *Computation and the Humanities: Towards an Oral History of Digital Humanities* (Springer 2016). She is a co-investigator of a Leverhulme-funded collaboration with the British Museum on the manuscript catalogues of Sir Hans Sloane (https://tinyurl.com/y7zvrthm); a UK principal investigator of a Digging into Data challenge "Oceanic Exchanges: tracing global information networks in historical newspapers" (http://oceanicexchanges.org/); and a co-investigator of a Marie Curie action "Critical Heritage Studies and the Future of Europe" (http://cheurope-project.eu/).

Ulrich Tiedau is a historian and digital humanist. He has worked on and across the borders of humanities and communication technologies for most of his professional life and published widely on Belgian, Dutch, and German history as well as on distance education and digital scholarship. He is an Associate Director of the UCL Centre for Digital Humanities and also serves as principal editor of *Dutch Crossing: Journal of Low Countries Studies*, which in December 2009 was awarded an honourable mention in the *Phoenix Prize for Significant Editorial Achievement* by the Council of Editors of Learned Journals. He is a co-convenor of the Low Countries history research seminars at the Institute of Historical Research (IHR).

Index

Note: Page numbers in *italics* refer to figures. Page numbers followed by *t* refer to tables.

D

E

F

G

H

O

P

Q

R

S

T

U

V

W

X