The Digital Middle Ages: An Introduction

By David J. Birnbaum, Sheila Bonde, and Mike Kestemont

Our aims in this supplement of *Speculum* are frankly immodest. In organizing a series of sessions devoted to the digital for the Medieval Academy annual meeting in 2016, we hoped, by bringing together a diversity of projects, to showcase for the Academy membership the wide range of exciting possibilities afforded by digital humanities (DH). The papers gathered here are drawn largely from those sessions, with several additions. We want to acknowledge the contributions of Sarah Spence and William Stoneman, coorganizers of the sessions, for their inspiration and help. This supplement is the first issue of *Speculum* devoted to digital medieval projects, and it is offered in an online, open-access format that reinforces the openness to which the digital aspires and which it encourages.

Busa

The advent of digital medieval studies is often attributed to the work of Roberto Busa (1913–2011). The Italian Jesuit priest was a philosopher and theologian who specialized in the lexical analysis of the works of Thomas Aquinas.[1] Because of the massive size of Aquinas's oeuvre, Busa quickly found himself in need of an indexing method to search the corpus, one that could surpass the labor-intensive system of handwritten fiche cards with which he began his work. Busa was quick to recognize the possibilities of the early computing systems that were developed in his lifetime, and in or around 1949 he reached out to Thomas J. Watson Sr., the founder of IBM. Watson and his staff at IBM were impressed with the aspirations of the Italian Jesuit (they called him "more American than the Americans"), but IBM was persuaded to participate in a joint research initiative only after the priest pointed out a flyer in the New York office of IBM that said, "The difficult we do right away, the impossible takes a little longer."[2]

Over the next thirty years, IBM and Busa created the *Index Thomisticus* project, the world's first sizable machine-readable corpus, containing, among an array of other related texts, an index verborum of all 118 works of Aquinas, totaling approximately 11 million words.[3] The project required an administrative and organizational staff that was unprecedented for a humanities research initiative at that time, as Aquinas's entire oeuvre had to be digitized onto punch cards (and later

---

[1] For an excellent *in memoriam* on the occasion of the centenary of his birth see Marco Passarotti, "One Hundred Years Ago: In Memory of Father Roberto Busa SJ," *Proceedings of the Third Workshop on Annotation of Corpora for Research in the Humanities (ACRH-3)*, ed. Francesco Mambrini, Marco Passarotti, and Caroline Sporleder (Sofia, 2013), 15–24.

[2] Passarotti, "One Hundred Years Ago," 17.

[3] The corpus can now be consulted online, http://www.corpusthomisticum.org/it/.

magnetic tapes), which were the primary data carriers in use in the middle of the twentieth century. Over the years, Busa would hire large numbers of young, local, female typists for this specialized task (see cover of this issue).

> Once, I was told by father Busa that he was used to choosing young women for punching cards on purpose, because they were more careful than men. Further, he chose women who did not know Latin, because the quality of their work was higher than that of those who knew it (the latter felt more secure while typing the texts of Thomas Aquinas and, so, less careful). These women were working on the Index Thomisticus, punching the texts on cards provided by IBM. Busa had created a kind of "school for punching cards" in Gallarate. That work experience gave these women a professionally transferable and documented skill attested to by Father Busa himself.[4]

In recent years, these aspects of the Index Thomisticus project have become the subject of research projects in the fields of oral history and gender studies, and the index helps us to realize that women played a foundational role in the early days of computer science and digital medieval studies.[5]

The open-minded priest never saw a conflict between the aims of his work and his religious calling, seeing the computer "as the son of man, and therefore grandson of God."[6] Busa praised both the speed and the enhanced accuracy of computer analyses. Today, one of the most esteemed awards in the field of digital humanities is named after the Italian Jesuit: the triennial Roberto Busa Prize issued by the Alliance of Digital Humanities Organizations (ADHO). Busa himself was the first recipient of the award in 1998 and he remained an active contributor in the community until his death.

Digital Humanities and Medievalists

In the decades following the onset of the Index Thomisticus project, medievalists were often early adopters of the digital, and continue to play an important role in the development of a broader field, which came to be called digital humanities.[7] This field took other forms and names during its emergence and subsequent development: humanities computing, humanist informatics, literary and linguistic computing, digital resources in the humanities, eHumanities, and others. These competing alternatives, among which "humanities computing" had long been dominant, have only recently made place for the newly canonical term "digital humanities," which today is rarely contested.[8] "Digital humanities" is generally meant to refer

---

[4] Quoted in Melissa Terras's "For Ada Lovelace Day—Father Busa's Female Punch Card Operatives" (blog), 15 October 2013, http://melissaterras.blogspot.be/2013/10/for-ada-lovelace-day-father-busas.html.

[5] See, for example, Julianne Nyhan and Andrew Flynn, *Computation and the Humanities: Towards an Oral History of Digital Humanities*, Springer Series on Cultural Computing (Cham, 2016), https://link.springer.com/book/10.1007%2F978-3-319-20170-2.

[6] Passarotti, "One Hundred Years Ago," 20.

[7] John Unsworth, "Medievalists as Early Adopters of Information Technology," *Digital Medievalist* 7 (2011), https://journal.digitalmedievalist.org/articles/10.16995/dm.34/.

[8] According to Kirschenbaum, the rise of the term "digital humanities" can be traced to "a set of surprisingly specific circumstances": (1) the publication of the 2006 Blackwell *Companion to Digital Humanities*, (2) the inauguration of the Alliance of Digital Humanities Organizations (ADHO), (3) the

to a broader field than "humanities computing." Whereas the latter is restricted to the application of computers in humanities scholarship and had narrower technical goals, the former also incorporates a "humanities of the digital," including the study (potentially via traditional means) of digitally created sources, such as art and literature.[9] DH is therefore profoundly multidisciplinary and attracts contributions from scholars and scientists both within and outside the humanities and the humanistic social sciences.[10]

Digital humanists have taken care to define themselves in an inclusive rather than exclusive manner. As a result, the term "digital humanities" connotes a greater sense of integration than the diversity of approaches that are sheltered within the "big tent" of DH and that are also reflected in the contents of this supplement.[11] Thus, while the definition of DH has been the subject of dedicated anthologies,[12] countless panel discussions, and even entire websites (http://whatisdigitalhumanities .com), a better question may be whether there still exist nondigital humanists today, since most scholars at least to some extent rely on computational aids, however basic, such as online search engines or word processors. Even the "original" objects of our research are most often mediated by the printed or online text or the slide or digital image.[13] The difference between the digital humanities and their less digital counterpart has become more a matter of degree than of kind.

It is clear that the digital humanities (and within it, digital medieval studies) are a practice-oriented community. It may be that it is a pragmatic methodological awareness that ties this community together, although theoretical self-reflection and meta-analysis have nonetheless become more prominent recently.[14] A number of theorists, including Willard McCarthy, recipient of the 2013 Busa award, and John Unsworth, have pointed to the necessary disjunction between the "object studied

---

inauguration of the NEH's Digital Humanities program. See Matthew Kirschenbaum, "What is Digital Humanites and What's It Doing in English Departments?," in *Defining Digital Humanities: A Reader*, ed. Melissa Terras, Julianne Nyhan, and Edward Vanhoutte (Farnham, 2013), 247–53, at 197–98 in particular.

[9] Alan Liu, "The Meaning of Digital Humanities," *PMLA* 128 (2013): 409–23.

[10] This multidisciplinary nature is treated as central in most general-purpose introductions to digital humanities, including *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Oxford, 2004); and *Digital_Humanities*, ed. Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp (Cambridge, MA, 2012).

[11] On the "big tent" discussion, see Matthew Jockers and Glen Worthey, "Introduction: Welcome to the Big Tent," in *Digital Humanities 2011*, ed. Alliance of Digitial Humanities Organizations (Stanford, 2011), vi–vii. Geoffrey Rockwell humorously noted, "Having wandered in the wilderness that was humanities computing since the late 1980s I find it ironic to be part of something that is suddenly 'popular' or perceived to be exclusive when for so many years we shared a rhetoric of exclusion." See the anthologized reprint, Geoffrey Rockwell, "Inclusion in the Digital Humanities," in Terras, Nyhan, and Vanhoutte, *Defining Digital Humanities*, 247–53, at 248.

[12] Consult Terras, Nyhan, and Vanhoutte, *Defining Digital Humanities*.

[13] See, for example, Matthew Battles and Michael Maizels, "Collections and/of Art and the Art Museum in the DH Mode," in *Debates in the Digital Humanities 2016*, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis, 2016), 325–44.

[14] The issue of "theory" in DH has been the subject of the first volume of the "Conversations" section in the online *Journal of Digital Humanities* in winter 2011 (http://journalofdigitalhumanities.org /1-1), in which a series of more spontaneous writings (e.g., from the blogosphere) about the topic have been collected.

and the representation of that object in digital analysis."[15] McCarthy has argued that the concept of "modeling" is a central characteristic of the digital humanities.[16] By a model, he means "a representation of something for purposes of study, or a design for realizing something new." Following Clifford Geertz, he distinguishes between models *of* things (for example, a grammar; a geographical map) and models *for* things (for example, an architectural plan). Depending on disciplinary traditions, scientific models are known under various names (representation, diagram, map, simulation, and so on). What such models typically have in common is that they offer a condensed, often simplified representation of things. Therefore, models are more easily manipulated than the things they represent, which allows for experimentation.

In McCarthy's view, "modeling," the heuristic process in which models are constructed and manipulated, is central to the digital humanities. Of course, models and modeling practices have long existed in humanities scholarship: the critical apparatus in printed editions of medieval works is but one classic example of a well-known edition model, which attempts to represent in condensed fashion the complex phenomenon of a medieval text tradition. What sets the digital humanities apart is an increased awareness of, and explicit interest in, modeling strategies, as a consequence of the field's intense interaction with computers. But computers can process only fully explicit and consistent models, which means that if computers are to analyze humanities data, our assumptions must be fully explicit and consistent. The need for explicitness and consistency can be alienating for scholars in humanities fields where the exceptional is often embraced. Scholars from poststructuralist paradigms might also mistake the need for explicitness for scientific positivism.

Digital Medieval Studies

Models and modeling provide a framework for presenting ongoing work in the field of medieval studies and for explaining the ways in which much of this work might deviate from what went before. First, much new groundwork is being done in digital medieval studies. High-resolution electronic manuscript facsimiles are produced in large quantities by heritage institutions in the GLAM (galleries, libraries, archives, and museums) sector around the globe. The British Library Digitised Manuscripts link and the Cathedral Library of Cologne (the Codices Electronici Ecclesiae Coloniensis) are two good examples.[17] In the future, initiatives like the IIIF (International Image Interoperability Framework) can be expected to enhance our capacity to inspect and compare primary sources on a scale—and with an immediacy—that would have been unimaginable for earlier generations of scholars. (We should remember that some libraries do not even allow visitors to inspect multiple physical items at the same time!) Institutions such as the Schoenberg Institute for Manuscript Studies lead the way in this respect, with its visionary director Will Noel

---

[15] See Julia Flanders and Fotis Jannidis, "Data Modelling," in Schreibman, Siemens, and Unsworth, *Companion to Digital Humanities*, 229–37.

[16] These views have been summarized in Willard McCarthy, "Modeling: A Study in Words and Meaning," in Schreibman, Siemens, and Unsworth, *Companion to Digital Humanities*, 254–70.

[17] http://www.bl.uk/manuscripts/ and http://www.ceec.uni-koeln.de/.

being honored as a White House Champion of Change in 2013 for his commitment to open science.[18]

Excellent examples of enriched digital libraries include the Online Froissart (where high-resolution facsimiles often are accompanied by transcriptions and historical information)[19] or Monasterium.net, a virtual archive that offers centralized access to over five hundred thousand primary diplomatic sources, such as charters, from more than one hundred European archives.[20] Equally representative is the database and interface supporting the DigiPal project, which won the Medieval Academy's first Digital Humanities Prize in the spring of 2017 (see the contribution to this supplement about this platform for the paleographic study of English manuscripts by the project's principal investigator, Peter A. Stokes).

Modeling choices present themselves at even the most basic research steps. With basic facsimile creation, for example, critical modeling choices must take into consideration bandwidth and memory limitations, which impose practical limits about the resolution at which manuscripts can be photographed, stored, and distributed. Starting from what resolution does a photograph offer a reliable representation of the immediate source? Can we reasonably expect that some users will ever need reproductions at 3,000 DPI or more? Johanna Drucker has therefore correctly stressed that what we might regard as raw *data* ("given") in the humanities are already the product of some form of modeling, however modest;[21] and she proposes that we use the term *capta* ("taken") to reflect the constructed nature of such data. Metadata yield similar concerns, since many collections are currently digitized at a more rapid pace than the GLAM institutions can manually annotate with metadata. Again, difficult choices have to be made: how can we responsibly (re)publish digital files for which no, incomplete, or only outdated metadata is available? Should we allow users from across the globe to crowdsource annotations for these newly digitized objects, or should this remain the domain of trained experts? Authority is a complex issue in this respect and presently under intense renegotiation. Much effort goes into federating access to heterogeneous data streams through the creation of information repositories that collect linked information. Descriptive metadata standards, such as Dublin Core or the Getty vocabularies, play an important role in this.[22] The contribution by Toby Burrows in this supplement sheds new light on how structured metadata can be leveraged in the field of manuscript studies.

Digital scholarly editing is one of the major stakeholders in the digital humanities community, and much of the activity in this area revolves around the Text Encoding Initiative (TEI, http://www.tei.org).[23] The TEI defines an influential set of guidelines for enriching texts with both interpretative and descriptive annotations using a

---

[18] http://dla.library.upenn.edu/dla/schoenberg/index.html.

[19] Peter Ainsworth and Godfried Croenen, ed., *The Online Froissart, Version 1.5* (Sheffield, 2013), http://www.hrionline.ac.uk/onlinefroissart.

[20] http://monasterium.net/mom/home.

[21] Johanna Drucker, "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5 (2011), http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

[22] Seth Van Hooland and Ruben Verborgh, *Linked Data for Libraries, Archives and Museums: How to Clean, Link, and Publish your Metadata* (London, 2014).

[23] Elena Pierazzo, *Digital Scholarly Editing: Theories, Models and Methods* (Farnham, 2015).

markup language called XML. The contribution by Franz Fischer offers a broad survey of the sort of digital editions and archives that currently live on the web.[24] Optical character recognition (OCR) has allowed us to turn (scans of) existing editions into machine-readable and searchable texts, which often then serve as the basis for new digital editions. The Electronic Beowulf project (http://ebeowulf.uky.edu) is an early seminal project that allowed greater access to an important medieval text. *Beowulf* is preserved in a single eleventh-century manuscript, which was damaged by fire in 1731. Transcriptions made in the late eighteenth century show that many letters then visible along the charred edges were subsequently lost. In 1845, each leaf was mounted into a paper frame. Scholarly discussion of the date, provenance, and creation of the poem continue around the world, and researchers regularly require access to the manuscript. Digitization of the entire manuscript provides a solution to problems of access and conservation. Immense corpora are today available to medieval linguists; well-known examples from the Anglo-Saxon world include the *Linguistic Atlas of Early Medieval English* (c. 650,000 words) or York-Toronto-Helsinki Parsed Corpus of Old English Prose (c. 1.4 million words), and resources on this scale have allowed scholars to verify long-standing questions in medieval studies using quantitative means.[25] The contributions by Maxim Romanov, Jeroen De Gussem, and David Joseph Wrisley in this supplement illustrate the sort of macro-analyses that large corpora enable.

Although much progress has been made on the level of OCR, the computational study and semiautomated transcription of handwritten materials remains a much more elusive application. Mike Kestemont, Vincent Christlein, and Dominique Stutzmann contribute an article to this supplement where the reader is introduced to the field of computer vision and its considerable potential for the study of medieval script. Many other interesting applications of digital script analysis have appeared in recent years, such as automatic writer identification for medieval documents.[26] Surely, we can expect much more progress in field of visual analyses for DH in the coming years.

In her history of humanities computing, Susan Hockey notes that the earliest work in the field of DH was strongly biased towards text.[27] For medievalists this is especially limiting, given the importance of manuscripts—including their illuminations or initials—in medieval culture. When compared to, for example, image or audio files, it is clear that plain text files come with much more relaxed computational demands in terms of storage, user interfaces, and processing power. This helps ex-

---

[24] Also see Greta Franzini, Melissa Terras, and Simon Mahony, "A Catalogue of Digital Editions," in *Digital Scholarly Editing: Theories and Practices*, ed. Elena Pierazzo and Matthew James Driscoll (Cambridge, UK, 2016).

[25] A relevant example is Jacob Thaisen, "Initial Position in the Middle English Verse Line," *English Studies* 95 (2014): 500–13. In this paper, Thaisen uses statistical language modeling to show that the beginning (and, to a lesser extent, the ending) of Middle English verse lines are relatively more stable in the transmission of texts.

[26] For example, for Chaucer manuscripts, Marius Bulacu and Lambert Schomaker, "Automatic Handwriting Identification on Medieval Documents," in *Proceedings of the 14th International Conference on Image Analysis and Processing* (Modena, 2007), 279–84.

[27] Susan Hockey, "A History of Humanities Computing," in Schreibman, Siemens, and Unsworth, *Companion to Digital Humanities*, 3–19.

plain why much of the early work has, for example, been lexicographic in nature. Word-level analyses, such as the Index Thomisticus, lent themselves well to computer-based indexing and quantification. Hockey gives special mention to medievalists like Roy Wisbey, who produced an index to Early Middle High German literature as early as the 1960s. But Hockey also emphasizes the serious limitations of both hardware and software (in regard to memory) with which early adopters struggled. (It is easy to forget that modern smartphones come with larger amounts of computer memory than the onboard computer of the Apollo 11 mission in 1969.) Even what are now regarded as relatively trivial issues, such as displaying basic medieval glyphs on a computer screen for the common Germanic characters thorn (þ) and eth (ð), remained a challenge until deep into the twentieth century. Today the Unicode Standard (http://unicode.org) seeks to provide support in operating systems and applications for all human writing systems, including those of the Middle Ages, and the Medieval Unicode Font Initiative (MUFI, http://folk.uib.no/hnooh/mufi/) promotes a microstandardization of the Unicode Private Use Area (PUA) specifically for Western medieval writing.

Stemmatology is another typically medievalist domain in which we find early adopters of computational methods. Collation software was able to align variant manuscript readings, which could serve as the input to the machine-assisted identification of a stemma codicum. Peter Robinson's Collate software was used to manage the variants in the Canterbury Tales Project and elsewhere and has now been succeeded by the open-source CollateX (http://collatex.net) for textual collation. An especially sophisticated exploration of machine-assisted stemmatology was Caroline Macé's Tree of Texts project at KU Leuven (Katholieke Universiteit, Leuven), which was the starting point for Tara Andrews's StemmaWeb Project: Tools and Techniques for Empirical Stemmatology (https://stemmaweb.net/).[28]

Within the domain of text analysis, computational stylistics (or "stylometry") also played an early role in the development of digital humanities, and the article by Jeroen De Gussem in this supplement describes an application of this technology to twelfth-century Latin literature. Stylistic phenomena belong to the realm of tangible poetics and have the advantage of being more amenable to quantification than hermeneutical features, that is, those relating to interpretation. This has led to advances in authorship attribution of medieval texts, such as the many medieval romances that (allegedly) resulted from collaborative forms of authorship. Results include the early study by John R. Allen on the authenticity of the Baligant episode in the *Chanson de Roland*[29] and the more recent investigation of the Middle Dutch *Walewein* by Karina van Dalen-Oskam and Joris van Zundert, which sheds new light on the complex interferences between authorial and scribal aspects of medieval texts.[30] Likewise, the well-known dual authorship of the *Roman de la Rose* is now

[28] Tara L. Andrews and Caroline Macé, "Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation through Graph Analysis of Texts and Stemmata," *Digital Scholarship in the Humanities* 28 (2013): 504–21.

[29] John R. Allen, "On the Authenticity of the Baligant Episode in the *Chanson de Roland*," in *Computers in the Humanities*, ed. John L. Mitchell (Edinburgh, 1974): 65–72.

[30] Karina van Dalen-Oskam and Joris van Zundert, "Delta for Middle Dutch: Author and Copyist Distinction in *Walewein*," *Literary and Linguistic Computing* 22 (2007): 345–62.

often used as a generic test case in the development of text analysis software.[31] In more focused contributions, the results of stylistic analyses have been linked to issues involving gender criticism.[32] Metrical analyses, such as the work by Friedrich Dimpel on Middle High German, are another domain where computational methods can be expected to break ground.[33]

Visualizations, Sound, and 3D Modeling

Although historically the digital humanities have been dominated by text-oriented paradigms, the community is increasingly engaging with multimodal research objects and methods.[34]

The Visual Turn

DH has adopted visualizations in many areas of research. Graphs, charts, diagrams, and other visual interpretations were common in pre-DH scholarship, but with DH has come the interest and ability to engage with large data sets and to represent them visually—see, for example, the varied visualizations in Maxim Romanov's contribution to this supplement.[35] Network visualizations are also frequently used, not only for textual exploration (De Gussem's paper), but also for geographic analyses, for instance in the papers by Romanov and Toby Burrows.[36] Another recent article points to the potential for visual analysis to produce results in the arena of image-feature analysis, taxonomy building, and clustering methods for medieval manuscripts;[37] see also Kestemont, Christlein, and Stuzmann's article on computational approaches to identifying scripts in this supplement. A number of recent projects have invested effort in virtual recreations of medieval libraries at Chartres, Lorsch, and elsewhere.[38] Manuscriptlink, a new digital humanities initiative, aims to reconstruct "virtual" medieval libraries by collaborating with collections around the

---

[31] Maciej Eder, Jan Rybicki, and Mike Kestemont, "Stylometry with R: A Package for Computational Text Analysis," *R Journal* 8 (2016): 107–21.

[32] Examples include Jan Ziolkowski, "Lost and Not Yet Found: Heloise, Abelard and the *Epistolae duorum amantium*," *Journal of Medieval Latin* 14 (2004): 171–202; Mike Kestemont, Sara Moens, and Jeroen Deploige, "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux," *Digital Scholarship in the Humanities* 30 (2015): 199–224.

[33] Friedrich M. Dimpel, *Computergestützte textstatistische Untersuchungen an mittelhochdeutschen Texten* (Tübingen, 2004).

[34] Susan Hockey, "History of Humanities Computing."

[35] S. Jänicke, G. Franzini, C. Faisal, and G. Scheuermann, "Visual Text Analysis in Digital Humanities," *Computer Graphics Forum* 35 (2016), doi:10.1111/cgf.12873.

[36] David Hadbawnick, "The Framing Narrative and the Host: Two Kinds of Anxiety in the *Canterbury Tales*," in *Open Access Companion to the Canterbury Tales*, http://www.opencanterburytales.com/open-review-home/the-framing-narrative-and-the-host/.

[37] Dominique Stutzmann, "Clustering of Medieval Scripts through Computer Image Analysis: Towards an Evaluation Protocol," *Digital Medievalist* 10 (2015), https://journal.digitalmedievalist.org/articles/10.16995/dm.61/.

[38] For Chartres, http://www.biblissima-condorcet.fr/fr/a-new-life-medieval-libraries-chartres; for Lorsch, https://www.uni-heidelberg.de/presse/news2012/pm20120323_lorsch_en.html.

world to reaggregate previously lost medieval volumes.[39] Burrows's contribution to the present supplement tackles related issues.

The Spatial Turn

The strategic use of digital mapping is an offshoot of visualization, one that is often directed toward graphic analysis of location, ownership, and distribution within geographic boundaries. Data sets providing greater access to larger spatial data sets have enhanced research in this area. For example, Harvard's *Digital Atlas of Roman and Medieval Civilizations* (DARMC) app provides GIS maps and geodatabases that are openly available and searchable online.[40] The Digitized Medieval Manuscripts app (DMMapp) provides original map resources online,[41] while the Digital Mappaemundi allows for searching between medieval maps and textual sources.[42]

Geographic Information System (GIS) technologies provide ways to map and compare spatial data.[43] For example, GIS has been used to investigate the history of medieval rural and urban landscapes. City Witness (http://www.medievalswansea .ac.uk/), a multidisciplinary research project, has created an online interactive map of Swansea, c. 1300, showing its principal topographical and landscape features, alongside an electronic edition of fourteenth-century texts. Together the map and texts provide multiple vantage points on the town and the significations attached to locations within the town by various social and ethnic groups (including Anglo-Norman and Welsh, lay and religious, male and female). The focus of the Mapping Medieval Chester project (http://www.medievalchester.ac.uk/index.html) is the identities that Chester's inhabitants formed between c. 1200 and 1500. Like City Witness, the project integrates geographical and literary mappings of the medieval city using cartographic and textual sources in order to understand how urban landscapes were interpreted and navigated by local inhabitants.

GIS has also been used to "map" individual objects like the manuscript page. Mapping texts through GIS is at the heart of David Wrisley's contribution to the supplement; and the Lancelot-Graal project (http://www.lancelot-project.pitt.edu /lancelot-project.html), featured in the article by Alison Stones in this supplement, is one of the leaders in this adaptation of GIS. In the Gough map project (http:// www.goughmap.org/), GIS was used to analyze the relational representation of space in medieval and contemporary maps, allowing us to understand that the fourteenth-century map was designed to be functional and demonstrated a high degree of spatial accuracy.[44] Mapping of places within charters or even hagiographic

---

[39] A description of the Manuscriptlink project is available on Vimeo (http://vimeo.com/98052555) and YouTube (http://youtu.be/B9r7F0PAeYQ).

[40] http://darmc.harvard.edu/.

[41] http://digitizedmedievalmanuscripts.org/.

[42] https://ihr.asu.edu/research/seed/digital-mappaemundi-resource-study-medieval-maps-and-geographic -texts.

[43] Ian Gregory, Christopher Donaldson, Patricia Murrieta-Flores, and Paul Rayson, "Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research," *International Journal of Humanities and Arts Computing* 9 (2015): 1–14.

[44] Christopher D. Lloyd and Keith Lilley, "Cartographic Veracity in Medieval Mapping: Analyzing Geographical Variation in the Gough Map of Great Britain," *Annals of the Association of American Geographers* 99 (2009): 27–48.

texts can allow for a deeper understanding of the construction of a sociopolitical landscape. The mapping of the locations and types of miracles within the *Life of Sainte Foy of Conques* provides evidence for the spatial extent of the monastery's influence and for differences within it.[45]

Three-Dimensional Reconstructions

Seminal three-dimensional reconstructions of past buildings and spaces have included the reconstructions of the church at Cluny by the laboratory at Darmstadt University;[46] the Amiens Cathedral website directed by Stephen Murray at Columbia University;[47] and the MonArch website, with its three-dimensional reconstructions, time slider, and linked textual sources for Saint-Jean-des-Vignes, Soissons, produced by Sheila Bonde and Clark Maines.[48] See also the contribution by Sheila Bonde, Alexis Coir, and Clark Maines on the abbey of Ourscamp in the current supplement. An ambitious recent project harnesses the results of archaeological survey and historical sources to create a complete three-dimensional reconstruction of the architecture of the entire medieval town of Montieri, Italy. This 3D reconstruction has aided researchers in their analysis of the architecture and layout of the town and will also make contributions to heritage and tourism.[49]

The Sonic Turn

Digital advances that allow us to recreate medieval manuscripts or to see three-dimensional recreations of medieval structures have made important contributions to the understanding of the medieval past. Having a full understanding of how people experienced these objects and buildings carries this understanding still further.

Sound studies have been strongly linked to heritage and conservation, often focusing on the capture of songs, music, and sounds of our cultural environment.[50] For medievalists, the recreation of past music and soundscapes links these efforts to the three-dimensional architectural reconstructions. One digital resource is provided by DIAMM (the Digital Image Archive of Medieval Music) at Oxford University, which presents information on thousands of manuscripts, as well as nearly fifteen thousand images and associated metadata. The online forum *Sounding Out!*

---

[45] Faye Taylor, "Mapping Miracles: Early Medieval Hagiography and the Potential of GIS," in *History and GIS: Epistemologies, Considerations and Reflections*, ed. Alexander von Lünen and Charles Travis (Heidelberg, 2012), 111–25.

[46] Manfred Koon and Horst Cramer, *Cluny: Architektur als Vision* (Heidelberg, 1993).

[47] http://www.learn.columbia.edu/Mcahweb/Amiens.html.

[48] http://monarch.brown.edu/monarch/index.html.

[49] Daniele Ferdani and Giovanna Bianchi, "3D Reconstruction in Archaeological Analysis of Medieval Settlements," in *Archaeology in the Digital Era*, vol. 2, *e-Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, UK, 26–29 March, 2012*, ed. Philip Verhagen (Amsterdam, 2013), 156–64.

[50] See Tanya Clement, "When Texts of Study Are Audio Files: Digital Tools for Sound Studies in Digital Humanities," in Schreibman, Siemens, and Unsworth, *Companion to Digital Humanities*, 348–57.

provides space for publication, posts, discussion, and recordings.[51] The Stanford Center for Computer Research in Music and Acoustics (CCRMA) is a multidisciplinary facility where digital technology is used as an artistic medium and research tool.[52] One recent recreation of the medieval soundscape for the cathedral of Santiago da Compostela, led by Rafael Suárez from the Universidad de Sevilla, found that the acoustic conditions for pilgrims in the nave were compromised, while the acoustic conditions in the choir were ideal for both plainchant and polyphony.[53] See also the contribution to this supplement by Bissera Pentcheva and Jonathan Abel, which explores the acoustics of Hagia Sophia; and the article by Spyridon Antonopoulos, Sharon Gerstel, Chris Kyriakakis, Konstantinos T. Raptis, and James Donahue describing the acoustic aspects of Byzantine churches in Thessaloniki.

Immersive Environments and Heritage

The ability to make a virtual visit to medieval sites is one offshoot of digital work with a heritage application, and Google and UNESCO have collaborated to offer virtual visits to several important locations.[54] IIVE (Interactive Immersive Virtual Environments) provide an interactive engagement for the "viewer" as part of a museum or heritage display. Second Life and its open-access counterpart, OpenSimulator; Myo; GoogleGlass; and Oculus VP are all potential applications. These virtual worlds, where users are represented by avatars, allow interaction between users and the environment and are thus appropriate for simulating (past) environments in real time. They may (though they do not always) include senses beyond the visual, especially harnessing sound. One such site, focused on the cathedral of Saint Andrews in Scotland, combines three-dimensional reconstruction of the cathedral buildings, the movement of processions, music, and other sounds, experienced through an avatar.[55] While brick and mortar museums are costly to build and maintain, and travel to an archaeological site may not be practicable (especially after an excavation has ceased), a simulated experience of a site visit can be created through digital technology. Two archaeological projects from the Roman world, the Rome Reborn and Portus projects, have provided these technologies for virtual visitors.[56] One recent medieval application has been realized for a tenth- to twelfth-century Muslim suburb

[51] https://soundstudiesblog.com/2016/04/04/17060/.

[52] https://ccrma.stanford.edu/about.

[53] Rafael Suárez, Alicia Alonso, and Juan J. Sendra, "Intangible Cultural Heritage: The Sound of the Romanesque Cathedral of Santiago de Compostela," *Journal of Cultural Heritage* 16 (2014): 239–43, http://www.sciencedirect.com/science/article/pii/S1296207414000788.

[54] http://whc.unesco.org/en/news/570/.

[55] S. Kennedy et al., "Exploring Canons and Cathedrals with Open Virtual Worlds: The Recreation of St Andrews Cathedral, Saint Andrews Day, 1318," *Digital Heritage* (2013), https://risweb.st-andrews.ac.uk/portal/files/75971074/digitalheritage2013_submission_536.pdf.

[56] Kimberly Dylla et al., "Rome Reborn 2.0: A Case Study of Virtual City Reconstruction Using Procedural Modeling Techniques," in *CAA 2009: Making History Interactive; 37th Proceedings of the CAA Conference March 22–26, 2009, Williamsburg, Virginia* (Oxford, 2010), 62–66; S. Keay et al., "The Role of Integrated Geophysical Survey Methods in the Assessment of Archaeological Landscapes: The Case of Portus," *Archaeological Prospection* 16 (2009): 154–66. On the use of Second Life in archaeology, see Luis Miguel Siquiera and Leonel Morgado, "Virtual Archaeology in Second Life and Open Simulator," *Journal of Virtual Worlds Research* 6 (2013): 1–16.

of Sinhaya, outside Zaragoza. The visualization of Sinhaya was based on the archaeological evidence of excavations as well as archival material. Photorealistic lighting algorithms were developed by Grupo de Informática Gráfica Avanzada (GIGA), and the virtual animation can be viewed in a low-cost CAVE-like system.[57]

Open Medieval Studies

In the digital humanities, traditional print publication forms have not ceased to exist, but they are often complemented and supported by electronic formats. Many digital humanists are attracted by the low threshold and immediacy of electronic communication platforms, so that scholarly communications increasingly happen in less formal blog posts, comments sections, and online or micromessaging platforms, such as Twitter. Many digital humanists are inspired by the open-science movement, which advocates the widest possible electronic distribution via open-access repositories and journals, not only of research results, but also of primary research data (for example, editions) and any home-brewed software that enabled the research. Some scholars even wonder whether it would not be desirable to open up the entire research cycle to the wider public—which is still often the primary source of humanistic scholarship—from funding proposal, to software development, to peer review.

Such ambitious proposals are sometimes contrasted to more conventional scholarship in the humanities, where scholars are imagined as sitting alone and brooding over their work for a prolonged period, until the research is finally perfected and released in a format that will, they hope, last for ages. Such conventional longer-term projects—typically undertaken by individuals instead of teams—are today under pressure from digital humanists, who argue that more traditional forms of scholarship and the associated publication culture lead to less sustainable research, although the reverse is also true from some perspectives.

Creating and publishing a traditional print edition of medieval documents does not easily allow future generations to refine this work and add layers of annotation and analysis, especially if the original source data is not released with the print volumes. With electronic publication, supplying the primary data also means that future scholars will not need to go through a cumbersome and error-prone digitization process. The use of version-management platforms, such as GitHub (http://github .com), are helpful in this respect, because they allow scholars—and their peers—to keep track of, comment on, and distinguish among versions of the work in real time. Thus prepublication feedback can be taken into account by scholars, and minor postpublication corrections need not wait until the next print edition to be integrated.

The ease with which digital scholarly work in medieval studies can be modified over time makes it more fluid than print formats. Umberto Eco was but one among many contemporary thinkers to link the instability of modern electronic resources to the medieval transmission culture of texts. The internet is a young and still-fragile

---

[57] Diego Gutierez et al., "Archaeological and Cultural Heritage: Bringing Life to an Unearthed Muslim Suburb in an Immersive Environment," *Journal of Cultural Heritage* 5 (2004): 63–74.

medium that struggles with the well-known phenomenon of "dead links" that naturally compromises the citability of sources. The citability of both scholarship and data sources is an obvious "teething issue" of DH and therefore an important concern in many projects.[58] And so is the durability: books are not copied and distributed as easily as digital files, but we have books that are still usable after centuries, while computer files can become unreadable in a decade as operating systems, storage formats, encodings, and application software go out of fashion. In many cases, the fragility of a digital edition lies not only in the data, but also in the application we use to interact with the data. For example, a digital concordance may have practical advantages over a paper one, but only as long as it is not locked into a hardware or software environment that has gone out of fashion.

Intellectual property rights, economic costs, and privacy issues also stand in the way of the naïve realization of the ambitious goal of a completely open medieval studies.[59] The entire Patrologia Latina, for instance, can today be found in digitized versions online. While the quality of such freely available texts is generally lower—they abound in OCR errors—than what can be found in established subscription databases such as Brepols's Library of Latin Texts (http://www.brepolis.net/), it is an interesting, but also worrying, development that the mere "availability" of a particular text version is rapidly becoming a selection criterion that rivals the age-old importance attached to the criterion of quality.[60] Many digital humanities venues already require scholars to submit their source data together with their papers, which is impossible in the case of copyrighted editions. Whereas for many applications in data mining, the minute differences among editions of the same text will not matter very much, it is frustrating that the high-quality materials produced by previous generations of scholars are sometimes severely underused in DH because of accessibility and "shareability" issues.

When it comes to the economic side, open-access journals, such as the *Digital Medievalist Journal* (https://journal.digitalmedievalist.org/), are courageous initiatives because they cannot count on a steady income flow in the form of subscription fees to guarantee their future sustainability. Many open-access journals will in fact charge the author a substantial "article processing charge" (APC), which raises concerns about the financial independence of scholars without institutional backup, especially retired scholars or those in alternative academic careers (#alt-ac). One major advantage of traditional print journals is therefore that they are largely free of charge to authors. The open-access community is currently working to overcome this challenge: national academies will probably take up new responsibilities, and initiatives such as the Open Library of Humanities (https://www.openlibhums.org/)—which aims to cover fully the APCs for the journals in their collections—can be expected to play a more prominent role in the near future.

---

[58] Jonathan Blaney and Judith Siefring, "A Culture of Non-citation: Assessing the Digital Impact of British History Online and the Early English Books Online Text Creation Partnership," *Digital Humanities Quarterly* 11 (2017), http://www.digitalhumanities.org/dhq/vol/11/1/000282/000282.html.

[59] Such issues lie at the heart of the current work by Walter Scholger, University of Graz.

[60] This issue is, for instance, raised in David Bamman and David Smith, "Extracting Two Thousand Years of Latin from a Million Book Library," *Journal on Computing and Cultural Heritage* 5 (2012): 2–13.

All in all, copyright remains a largely unsettled matter in the humanities today and, arguably, too few medievalists, digital and nondigital alike, are properly informed about the various licensing possibilities. Open-access licenses, such as the Creative Commons (CC), allow authors to enforce an attribution to the original creator of a work (e.g., CC-BY) or add restrictions with respect to the commercial usage of their work (e.g., CC-NC) or subsequent reuse (CC-SA, CC-ND). The fact that this digital supplement to *Speculum* is published in open access, under a liberal license that encourages wide dissemination, reflects these concerns in the DH community. We are thankful to the University of Chicago Press and to the Medieval Academy of America for their openness to this project as well as for their support.

A Panoramic Reading of Speculum

From a methodological perspective, it is vital that new approaches to medieval culture not lose touch with traditional and more conventional scholarly methods. Nevertheless, thought-provoking tensions have emerged between the digital and traditional humanities. In 2010, for example, Google collaborated with a large number of scientists to publish an influential *Science* paper on the well-known Google Books project.[61] In this project, the California technology giant claims to have digitized roughly 4 percent of all books ever printed—and the expansion of the data set is still ongoing. Because this data set is easily searchable online,[62] it offers a convenient resource, which today is probably queried by scholars more often than they care to admit. In this paper, Jean-Baptiste Michel et al. discuss an emerging research field called "culturomics," the study of high-throughput cultural data through lexical analysis, and they focus on the diachronic analysis of word frequencies in English books (1800–2000). Although their word-counting strategy was simple, their research demonstrated that word usage in large corpora correlates with cultural developments. The relative frequency of the word "slavery," for instance, peaked in their data during the U.S. Civil War and later during the civil rights movement. In addition to a large array of linguistic analyses, their word counts even demonstrated the active censorship of Jewish artists, such as Marc Chagall, in Nazi Germany.

In December 2010, the paper's two lead authors presented their thought-provoking work at the annual meeting of the American Historical Association. The association's President, Anthony Grafton, would later offer a fascinating account of this event: "For all their panache—and all the fun their tool permits—Lieberman-Aiden and Michel also inspired a little worry, as well as some hard thinking about the status of our discipline."[63] Grafton regretted that the paper's list of authors, though sizable, did not include a single historian, and that this lack of historical expertise occasionally showed in their presentation. He stated, in disappointment, "[A]pparently, historians have not established, in the eyes of many of their colleagues in the natural sciences, that they possess expert knowledge that might be valu-

[61] Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331 (2011): 176–82.
[62] https://books.google.com/ and https://books.google.com/ngrams.
[63] https://www.historians.org/publications-and-directories/perspectives-on-history/march-2011/loneliness-and-freedom.

able, or even crucial." Lieberman-Aiden and Michel would later counter this view, stressing that they did receive input from academic historians, although their names were not included in the final author list.[64] They stated that, "while 'expert knowledge' is important, shared paradigms, a shared language, and common intellectual values are a big part of what makes a successful team come together. This suggests that history departments have to grapple with several emerging responsibilities: to encourage familiarity with quantitative methods, with computational techniques, and—as you so eloquently wrote—with large-scale collaboration."[65]

The research initiative behind the culturomics paper is a typical instantiation of what is today commonly called "distant reading" in the digital humanities, a loosely defined notion seminally introduced by Franco Moretti in a momentous series of essays.[66] At present, distant reading (sometimes also known as macroanalysis, algorithmic criticism, panoramic reading, and other terms) plays a role in a variety of approaches to text analysis in the humanities where large bodies of texts are queried and analyzed using a combination of techniques from language technology, information retrieval, and data science.[67] Common to all these approaches is the strategy that an important part of the conventional reading process is in fact deliberately outsourced to a machine; human intervention is largely postponed to the interpretation of the simplified model that the algorithms yield. As Moretti noted, the reader's distance from the original text as such becomes a function of the increased scope of the reading effort.

Grafton's mixture of fascination and worry is probably representative of the attitude that many scholars today entertain towards such forms of distant reading. The fact that a crucial part of the reading process is outsourced to a machine calls into question the quality of the textual model that state-of-the-art computational methods can deliver. As a way of interrogating these issues, it will be illustrative to discuss a small, yet representative and critical, distant-reading exercise.[68] For this, the University of Chicago Press granted us access to a digital version of the *Speculum* archive covering the entire seventy-year period between the journal's inaugural issue in 1926 and December 2016. As in any sizable corpus nowadays, the quality of the digital text varies enormously: for the part up to volume 84, we have to work from the uncorrected output of optical character recognition, whereas we can work with clean, born-digital data from volume 85 onwards.

As can be gleaned from the bar plot on Fig. 1, where the token counts have been aggregated on a yearly basis, the full data set amounts to over 65 million tokens

---

[64] http://www.culturomics.org/Resources/faq/thoughts-clarifications-on-grafton-s-loneliness-and -freedom.

[65] Ibid.

[66] For example, Franco Moretti, "Conjectures on World Literature," *New Left Review* 1 (2000): 54–68. These essays were later brought together and commented on in Franco Moretti, *Distant Reading* (London, 2013).

[67] Relevant references for these terms include Matthew Jockers, *Macroanalysis: Digital Methods and Literary History* (Lincoln, 2013); Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Lincoln, 2011); Thomas Crombez, "Het onbehagen in de digitale cultuur: De opkomst van Digital Humanities," *Meta: Het Vlaamse tijdschrift voor bibliotheek en archief* 4 (2013): 8–13.

[68] All software used for this exercise is made available on https://github.com/mikekestemont /panorama.

Speculum archive (1926-2016)
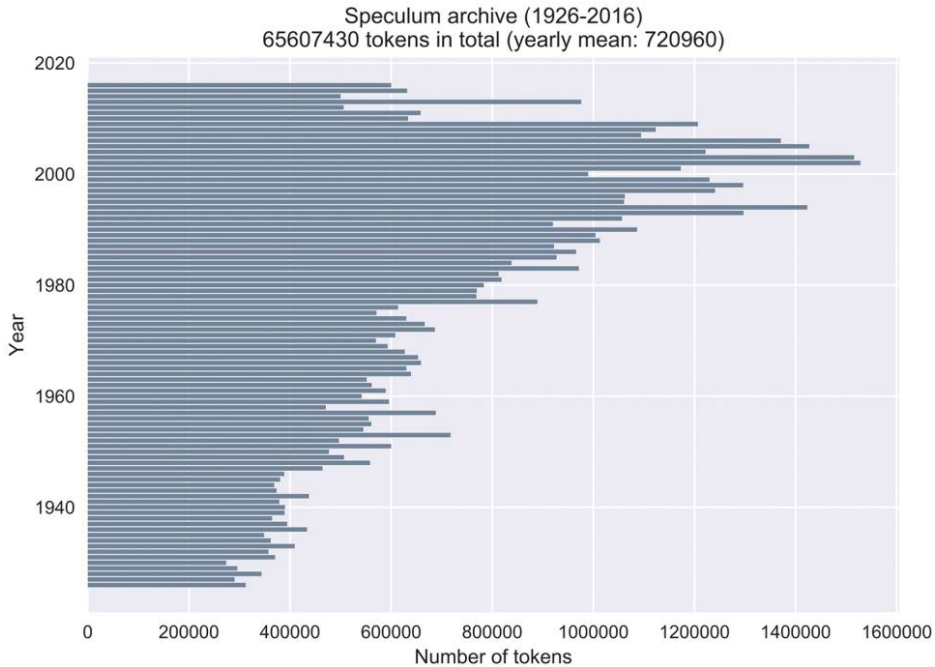65607430 tokens in total (yearly mean: 720960)

Fig. 1. Word counts for the data in the *Speculum* archive, aggregated at the year level (1926–2016).

(that is, words, but also punctuation marks and other symbols), although the numbers show severe fluctuations over the individual years. Nevertheless, the impressive size of the archive raises the intriguing question whether valuable patterns could be extracted from this data, which might yield a "panoramic view" of the journal's contents and thematic biases as well as its development throughout the years. Which medieval authors and texts rank highest, for instance, in *Speculum*'s popularity hit list; and which scholarly approaches have grown into or out of fashion over the years? For this effort, we made use of a range of computational techniques, which are representative of the state of the art in textual modeling strategies in the digital humanities nowadays. Hopefully, this array of methods will allow us to showcase, in a nontechnical language, the opportunities and, perhaps more importantly, the challenges that arise from such a "vanilla" application of distant reading.

   One common preprocessing step in textual analysis is to apply a so-called tagger to the material, an established procedure in natural language processing. In this exercise, we segmented the original raw stream of characters in a *Speculum* article into meaningful token units—for example, clitic "don't" will be restored to "do" and "n't".[69] We applied the Stanford CoreNLP software suite to the archive, which offers a host of basic procedures and which is maintained by one of the leading re-

---

[69] For an introduction to basic methods in NLP, consult, for instance, Dan Jurafsky and James Martin, *Speech and Language Processing*, 2nd ed. (Upper Saddle River, 2009).

Speculum 92/S1 (October 2017)

Table 1

An Example of a Sentence (Randomly Drawn from a 1955 *Speculum* Issue)
as Tagged by the Stanford CoreNLP suite

Index	Token	Lemma	Part of speech	Named entity
1	This	this	DT	O
2	Raimond	raimond	NN	PERSON
3	took	take	VBD	O
4	the	the	DT	O
5	title	title	NN	O
6	of	of	IN	O
7	Baron	Baron	NNP	PERSON
8	de	de	IN	PERSON
9	Saint-Gilles	Saint-Gilles	NNP	PERSON
10	in	in	IN	O
11	1066	1066	CD	DATE
12	.	.	.	O

search groups in the field of language technology.[70] Above, we show an example in Table 1 for the suite's output for a randomly selected sentence from a 1955 *Speculum* contribution. As can be seen in this example, the software will attempt to determine for each token its lemma (or uninflected dictionary headword; that is, past *took* becomes *take*), its part of speech (*title*, as it is used in this example, is an NN, or singular noun) and an indication of whether the token is a named entity (*1066* is a date, but *Raimond* is categorized as a person). The software makes these decisions on the basis of a statistical assessment of a token's appearance (for example, is the token capitalized?) and the lexical context in which it appears (for example, is the token preceded by an adjective?).

Because of the ambiguous nature of human language, such an automatic enrichment of the material will naturally yield many errors, especially in the case of the OCR-entered data with its unstable spellings, but it nevertheless already offers many possibilities for creatively querying the corpus. One interesting question might be which medieval dates have been most frequently mentioned in *Speculum* over the years. For this, we traced the cumulative frequency of all numbers in the data set that had been marked as a date in the named entity column and that fell in the "medieval" range of 500–1500. In the scatter plot in Fig. 2, we plot the twenty-five dates with the highest cumulative frequency in the corpus as a function of their coefficient of variation over the documents in the corpus, to keep track of their dispersion over the material.

The higher up in the plot a date can be found, the more frequent the date is; the more leftwards it is positioned (and the larger its font size), the better is it distributed over the individual documents in the corpus. The top of the list is clearly dominated

[70] Christopher Manning et al., "The Stanford CoreNLP Natural Language Processing Toolkit," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (Baltimore, 2014), 55–60.
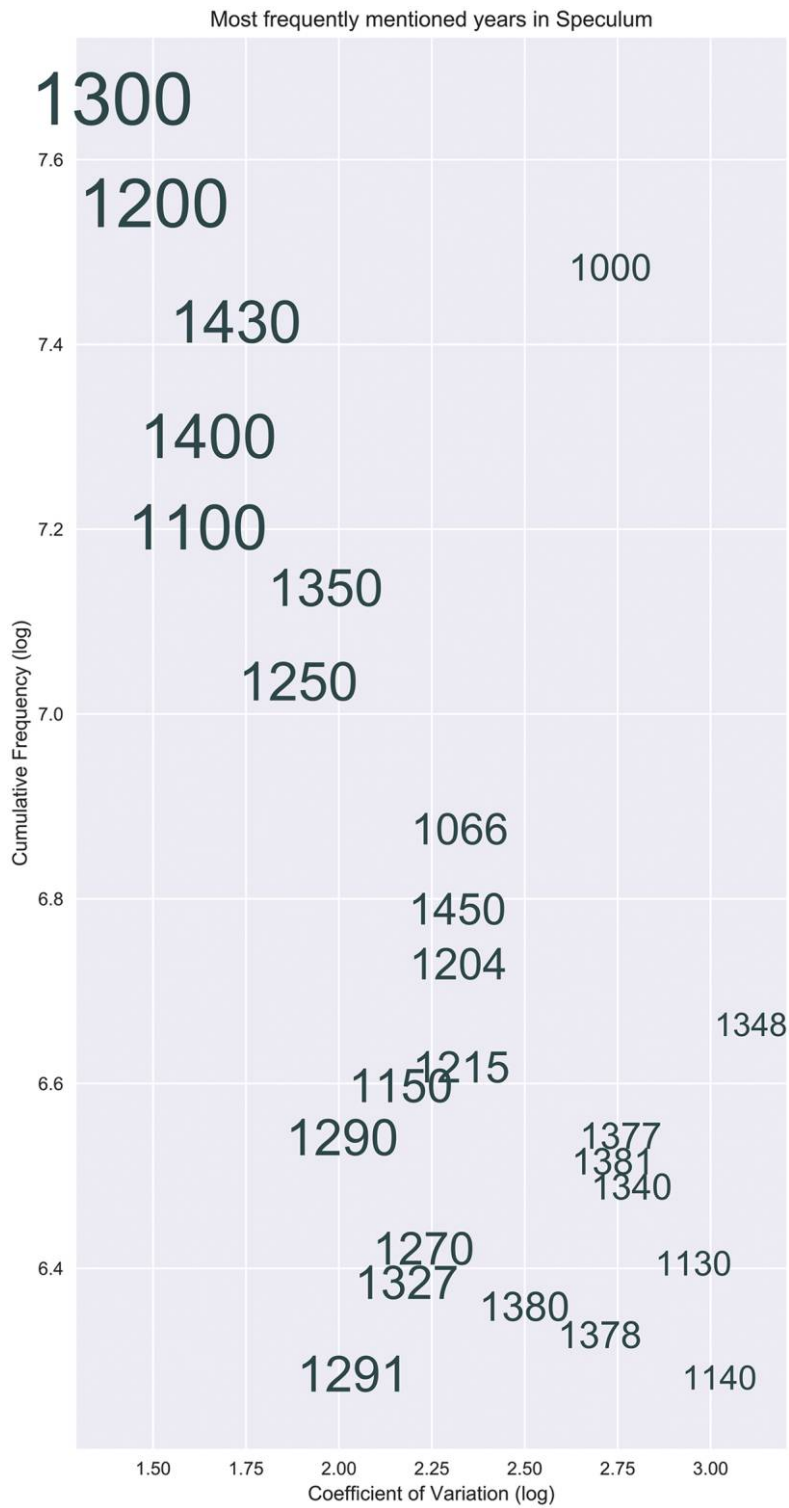
Fig. 2. Scatter plot showing the most commonly mentioned years (500–1500) in *Speculum*.

by round dates (1300, 1200, 1000). This reflects the fact that medievalists have generally preferred to think along conventional decennial, centennial, and millennial boundaries. Nevertheless, it is tempting to link a number of dates higher up in this hit list to well-known events in the medieval period, including 1066 (the Battle of Hastings) or 1204 (the sacking of Constantinople in the Fourth Crusade). For an iconic date such as 1430, it is interesting that one might be tempted to link it to several events, which helps explain its prominence: one can think of the siege of Compiègne and Joan of Arc's capture, but also Philip the Good's marriage and the installment of the Order of the Golden Fleece. Interestingly, the years 1000 and 1348 (outbreak of the Black Plague) have a lower dispersion than their elevated frequency might make us suspect.

Such corpus-level aggregations of frequencies are an interesting toy to help us characterize medieval studies from a panoramic viewpoint, but the tagging of our material also allows us to query *Speculum* in a more specific fashion. For the line charts in Figs. 3a–b, for instance, we have calculated the relative frequency of all nouns (whether plural or singular) in the material throughout the period 1926–2016. Using a common statistical test (Kendall's *tau*) we have queried the results for the five nouns that have shown the steadiest decrease (a) or increase (b) in usage throughout *Speculum*'s history. The results in Fig. 3a are not particularly exciting, and show merely that traditional (Latinate) citation styles (op. cit., ff., loc.) are growing out of fashion among *Speculum* authors. Fig. 3b, on the other hand, suggests a clean and surprisingly linear frequency increase of the words "role" and "context": this phenomenon strongly suggests that medieval studies, as represented by *Speculum* articles, have been marked in the twentieth century by a transition towards a more functionalist and contextualized approach to the Middle Ages, something that has already been often observed in literary studies. The shift in the use of the words "overview," "focus," and "potential" seems on the other hand to be of a metascholarly nature and might signal a trend towards greater scholarly professionalization and specialization in the broader field of medieval studies.

Our analyses so far have been purely lexical or carried out at the level of individual words. The problem with such a brute surface-counting approach is that it conceals the actual context in which words are used. If a word has been frequently used in *Speculum*, that would indeed seem to attest to the cultural salience of the word in the world of medievalists, but this context-free approach cannot tell us whether the term has primarily positive or negative connotations, nor does it indicate the scholarly context in which it is typically used. To remedy this situation, the digital humanities increasingly make use of methods borrowed from distributional semantics, an exciting research domain in natural language processing (or computational linguistics). In this field of study, researchers build upon the general idea that words derive meaning primarily from the lexical context in which they appear.[71] For example, in a sentence such as "I made the *blarf* fetch the stick" or

---

[71] A classic reference (among many others) is Zellig S. Harris, "Distributional structure," *Word* 10 (1954): 146–62. An interesting recent opinion piece about distributional methods in language technology is Chistopher D. Manning, "Computational Linguistics and Deep Learning," *Computational Linguistics* 41 (2015): 701–7.
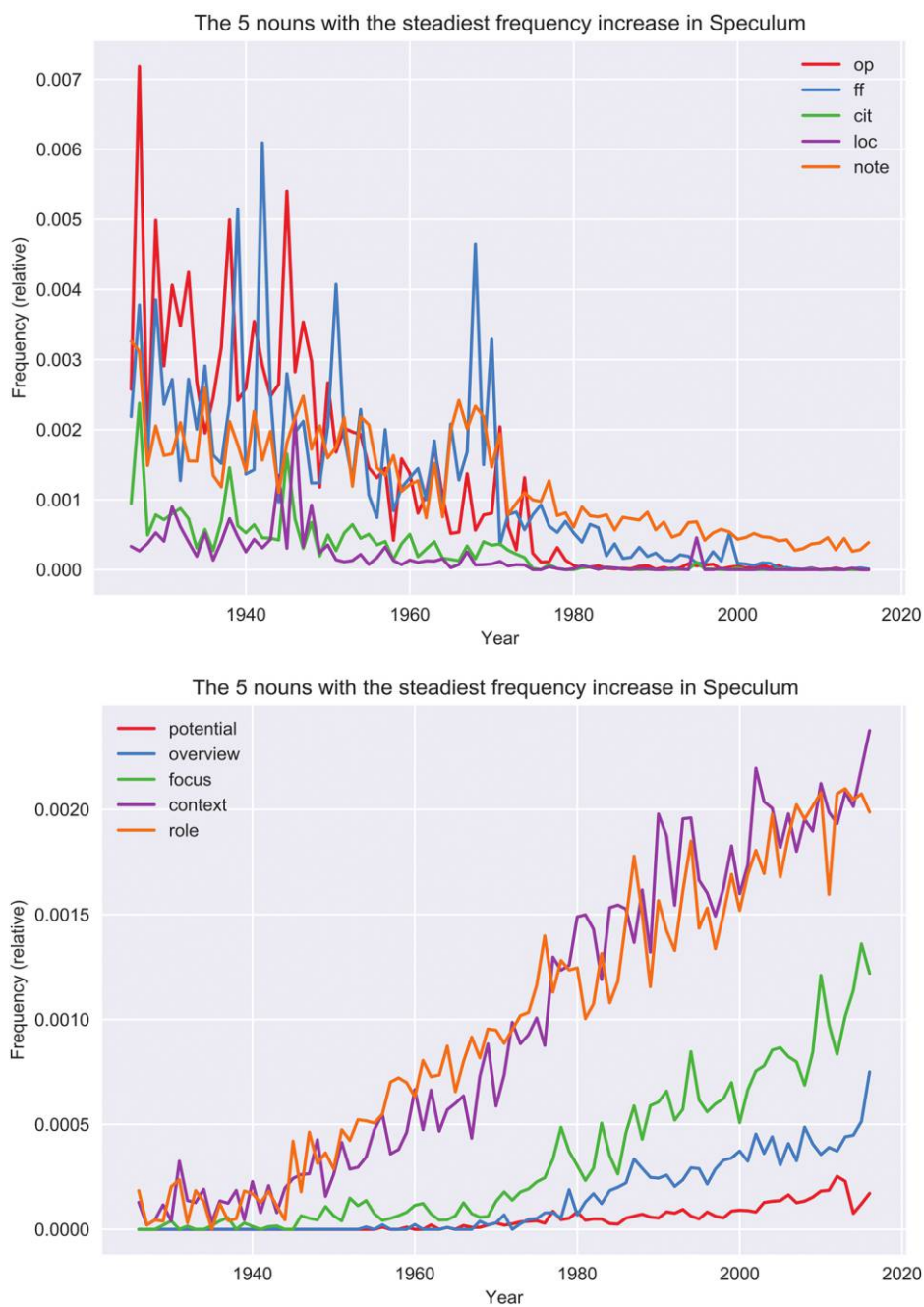
Fig. 3. (a) The relative frequency of nouns with the most linear drop in frequency. (b) The relative frequency of nouns with the most linear increase in frequency.

"I took the **blarf* for its evening walk," the context in which the nonexistent term **blarf* appears strongly hints towards a domestic animal—perhaps a dog.

In distributional semantics, researchers attempt to model the distributional patterns in word co-occurrences found in large corpora, such as the *Speculum* archive. The underlying assumption is that the vocabulary can be modeled into a set of semantic fields or topics; these topics consist of clusters of words that typically co-occur in documents or paragraphs and that are therefore more likely to belong to the same topic than words that never appear in the same context.[72] Each of the topics in such a "topic model" can be assumed to bear a certain weight, or topic score, on each document in a corpus: a newspaper article about a famous soccer player's transfer to Real Madrid, for instance, might be characterized as being 92 percent about "sports," 6 percent about "finance," and 2 percent about "Spanish lifestyle." We have subjected the *Speculum* archive to a topic-modeling exercise using the well-known method NMF (non-negative matrix factorization). We have asked the method to extract the 250 most salient topics from consecutive segments of 500 words, which did not include any so-called stopwords (such as articles, punctuation marks, or prepositions).

We have cherry-picked a representative selection of these topics and visualized them as a series of word clouds in Fig. 4. This selection clearly demonstrates the international and thematic variety of *Speculum* contributions over the history of the journal. In these clouds, the font size of the individual words reflects their relative importance to the topic. Note that the topic model itself does not produce a neat "label" for a topic, but its most significant words typically give a solid indication as to the semantic scope of a particular theme. These topics form relatively neat word clusters, even though this analysis does not depend on any external, handcrafted resources such as dictionaries: the model derives its semantic knowledge in a completely data-driven fashion solely from word usage statistics in a large corpus.

This overview of topics does attest to the dominance of insular topics, including those capturing the thematic fields surrounding the *Canterbury Tales*, *Beowulf*, Monmouth's Arthuriana, *Piers Plowman* and its alliterative colleagues, or the dominance of Cluniac monasticism and cathedral architecture. Nevertheless, the topical diversity is rich enough to include twelfth-century Latin literature from France, such as the Cistercian cluster of literature around Bernard, and also the world of Scandinavian sagas and of Dante's *Commedia*. A number of topics also clearly reflect higher-level thematic interests, such as courtly love, as well as themes within medieval architecture, Islamic studies, and gender studies. Many topics also seem to pick up on the major cultural clashes that characterized the medieval period, including the confrontation of Christian with Arabic culture in medieval Spain or the tension between Christianity and Judaism—note the presence of high-polarity terms such as accusation, violence, and murder in the latter topic.

Interestingly, this topic model also enables us to study the *Speculum* archive in a more diachronic fashion. If we were to calculate the average presence of a specific topic in all the *Speculum* issues that were published in a given year, plotting these scores on a timeline might provide insight into thematic evolution. In Figs. 5a–d,

---

[72] On topic modeling, consult for instance David M. Blei, "Probabilistic Topic Models," *Communications of the ACM* 55 (2012): 77–85.

Fig. 4. Word clouds representing a cherry-picked selection of topics from our thematic model (250 topics in total). Only the most salient words are plotted for each topic; the font size of the individual words gives an indication of their relative importance to the topic at hand.

we have plotted a number of trend lines for a selection of topics that seem to reveal interesting evolutionary patterns. The gender-related topic 156 (*women, female, male*), for instance, seems to have gained prominence only in the eighties, and the same goes for the sociocultural, functionalist approach to literature (*social, cultural, culture*), which seems to be captured in topic 231. One of the more obvious "downward" trends is the declining use of Latin throughout the journal's history (topic 3)—our analyses also suggested similar trends for other languages, such as French and German—suggesting that *Speculum* is becoming a more monolingual journal. Other topics are characterized by more local peaks, such as topic 48, which reflects a elevated number of citations in the field of medieval Aristotelianism (*averrois, aristotelem, commentariorum*) in the period 1950–70.
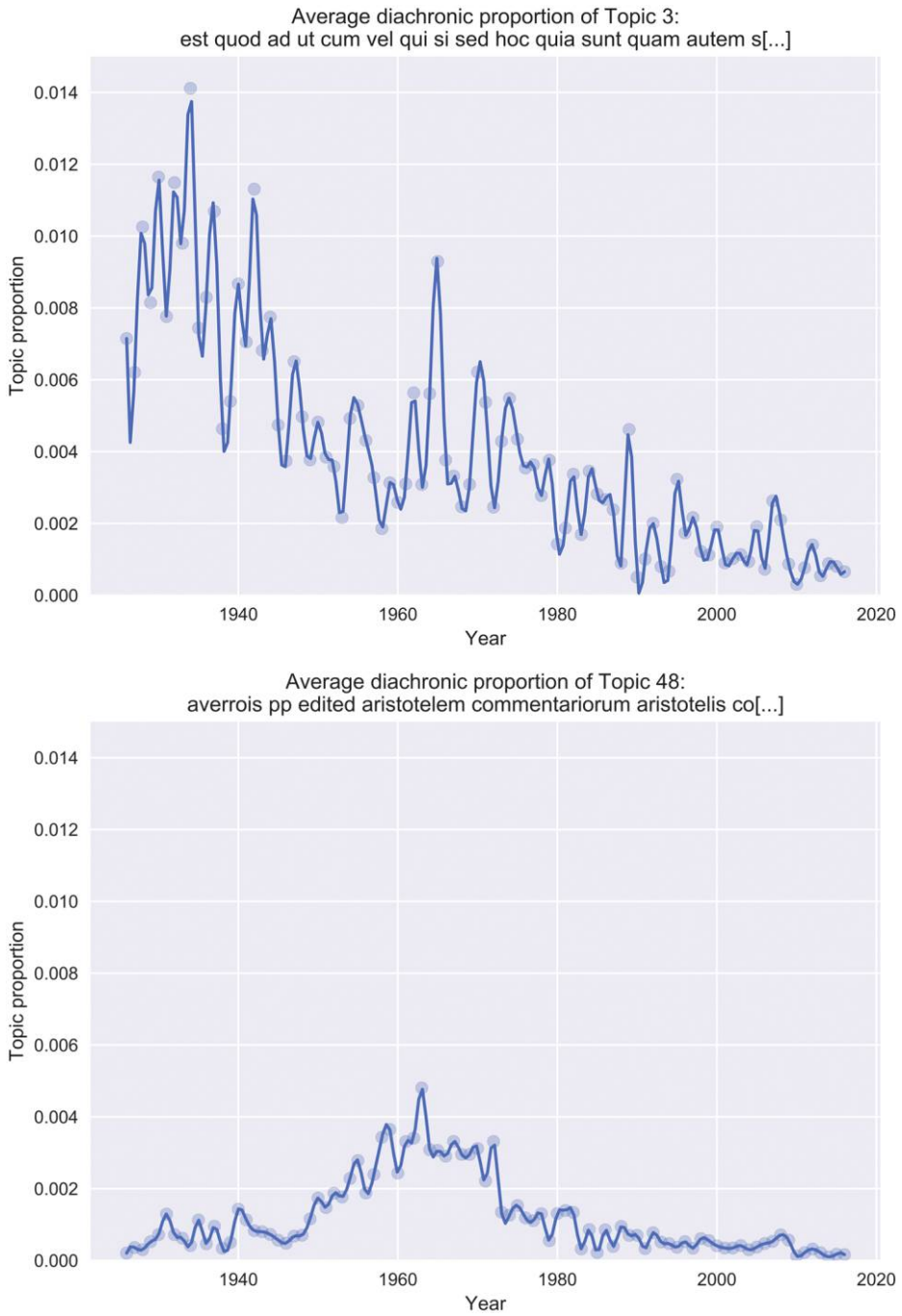
Fig. 4 (*Continued*)

   While topic modeling thus offers insights not available from simpler word-count approaches, it also raises new issues. Does the word *Bernard*, for instance, refer to the medieval author Bernard of Clairvaux or the present-day scholar Bernard McGinn (or both)? The problem that arises here is that even named entities can be ambiguous, and to achieve a more holistic approach to autonomous machine reading, such entities must be disambiguated. "Wikification" is a term that is used colloquially to denote the process of cross-document named-entity disambiguation in natural language processing.[73] Many software tools, such as the Stanford CoreNLP suite used above, are available today to tag automatically the named entities in free-running text, such as the names of individuals or places. While this process of named-entity recognition is already a crucial step towards knowledge extraction, the ambiguity of named entities presents a major obstacle on the road towards a machine's autonomous text understanding. In a sentence like "Clinton took the stage," it is unclear whether the named entity refers to Hillary Clinton, Bill Clinton, or the eponymous funk musician, George Clinton.

   In wikification studies, researchers attempt to extract clues from the semantic context in which a named entity occurs to help disambiguate these mentions. If the sentence reads "Secretary Clinton took the stage," the apposition "Secretary" would strongly suggest that the sentence refers to Hillary, since she is the only disambiguation

---

[73] Xiao Cheng and Dan Roth, "Relational Inference for Wikification," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (Seattle, 2013): 1787–96.

Fig. 5. Four plots showing the diachronic presence of selected topics in *Speculum* issues on an annual basis.

Average diachronic proportion of Topic 156:
women female male men woman gender feminist sex sexual nuns [...]

Average diachronic proportion of Topic 231:
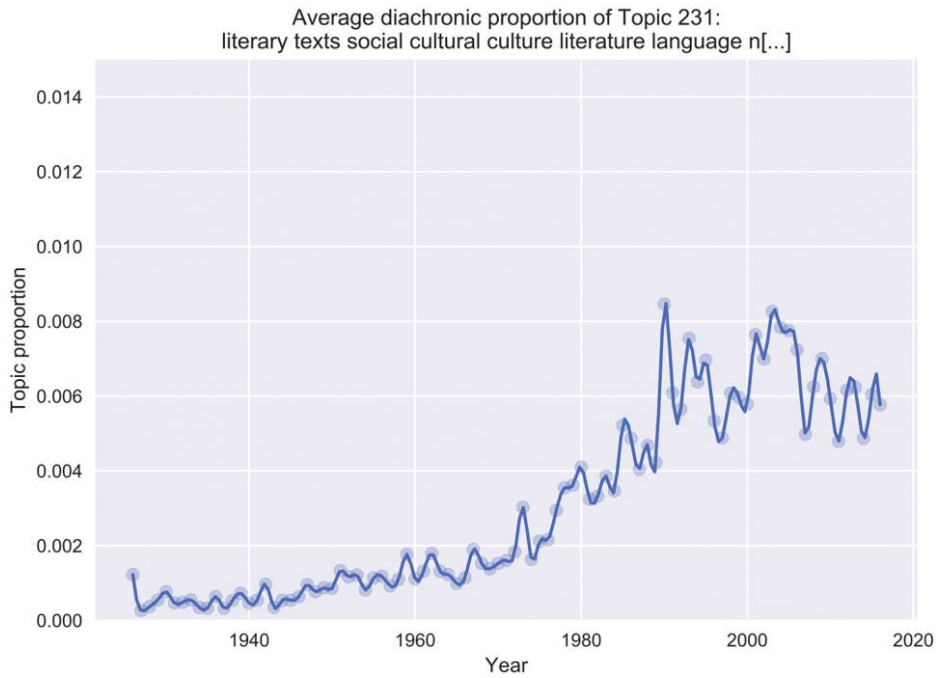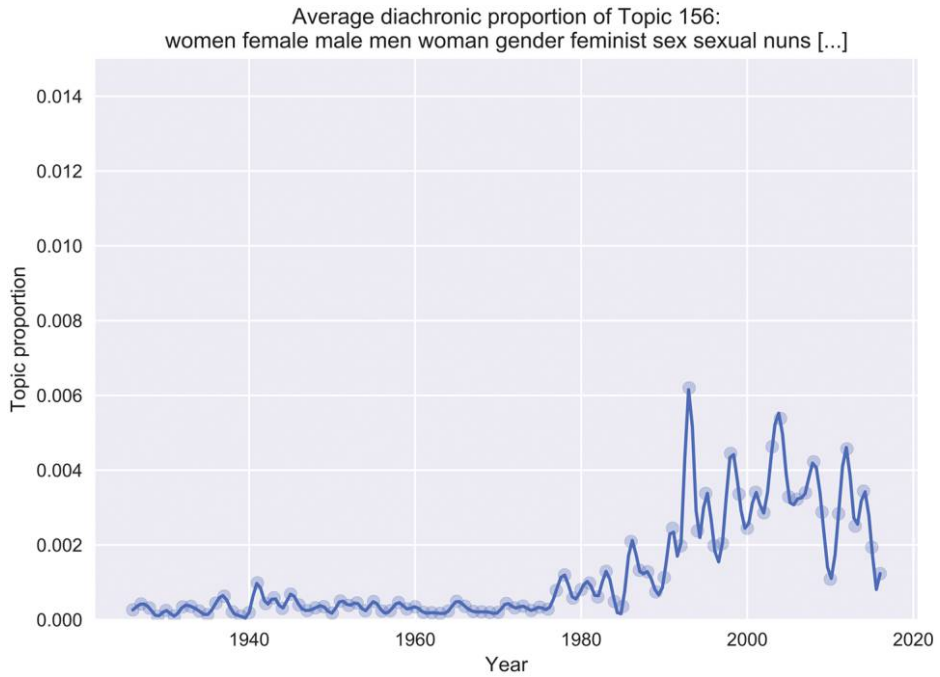### literary texts social cultural culture literature language n[...]

Fig. 5 (*Continued*)

candidate to have held this specific office. Additionally, wikification systems can exploit the fact that the named entities that are mentioned in a text typically form a semantically coherent set: in the sentence "Clinton took the stage with Bob Marley," the relatively unambiguous identification of the musician Bob Marley would suggest that the Clinton in this sentence is the artist George Clinton.

Scholars often turn to Wikipedia as a resource for mining fixed, unique identifiers for named entities. Through linking the named entities to the single, relevant entry for a named entity in the well-known encyclopedia, the algorithm effectively performs cross-document named-entity resolution. Additionally, Wikipedia is built on top of a rich ontological structure, so that various sorts of metadata can be harvested for each entity, in the form of descriptive labels indicating whether an individual was, for example, a philosopher or a king. Wikipedia has an impressive scope, but at the same time the use of a wikifier introduces strong biases. Uncommon named entities that have not yet received an identifiable Wikipedia page will be ignored by necessity. Likewise, the fact that we use a wikifier for the English language might bias our analysis towards entities that are relatively more salient, culturally speaking, in the Anglo-Saxon part of the world. When we apply the Illinois wikifier[74] to *Speculum*'s plain-text archive, a superficial reading of the wikifier's output anecdotally suggests that the wikifier struggles with the poor OCR quality of the earliest volumes, but nevertheless is able to output interesting annotations:

> Long ago Sir John Rhys offered a solar interpretation of Arthurian lore , but , according to Loomis , he did not work out the Celtic mythological system from the evidence of the Irish and Welsh legends themselves.

Note that the wikifier deals well in this example with abstracting over superficial synonyms: in texts that mention *Bernard*, *Bernardine*, or *Bernard of Clairvaux*, the entities will be mapped to the same unique identifier as their Latin equivalents, such as *Bernardus Clarevallensis*. Therefore, such a tool offers much more powerful search capabilities than raw text data. Nevertheless, the "disambiguations" are certainly not flawless, and at times are tragically hilarious—all sorts of American celebrities, including famous wrestlers and pop artists, would appear to have made a much larger contribution to medieval scholarship than we might have anticipated. Nevertheless, when aggregated at a higher level, the wikifier's output is accurate enough to draw up even more insightful hit lists than the ones we have shown so far. In Fig. 6, we have exploited the metadata that the wikifier attaches to each entity to draw up a list of the most salient—or, at least, most frequently mentioned—authors (a), poems (b), and saints (c) in *Speculum*.
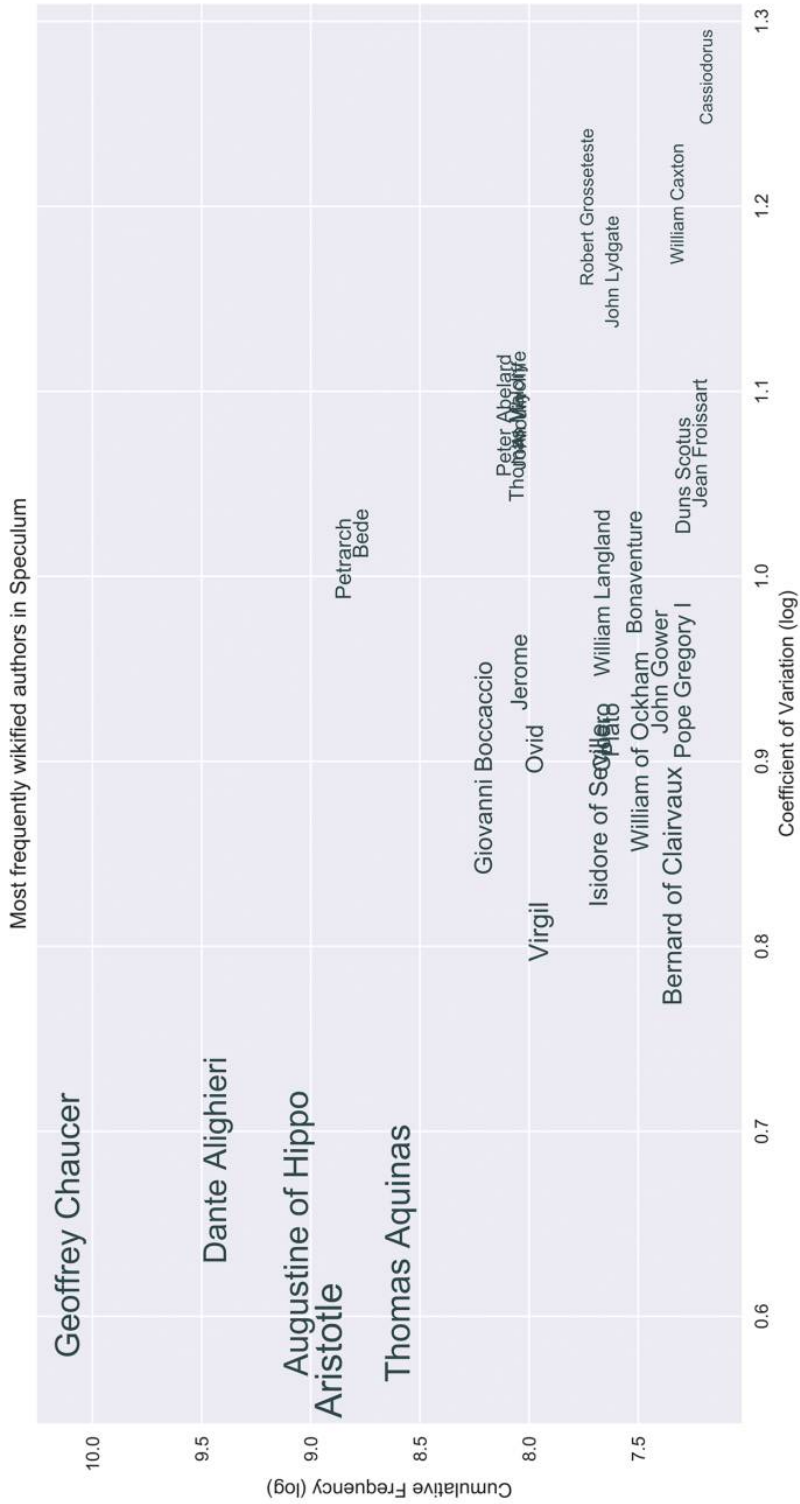

Fig. 6. Subplots showing the most frequently mentioned authors (a), poems (b), and saints (c) in *Speculum* on the basis of the wikifier's named entity disambiguation.
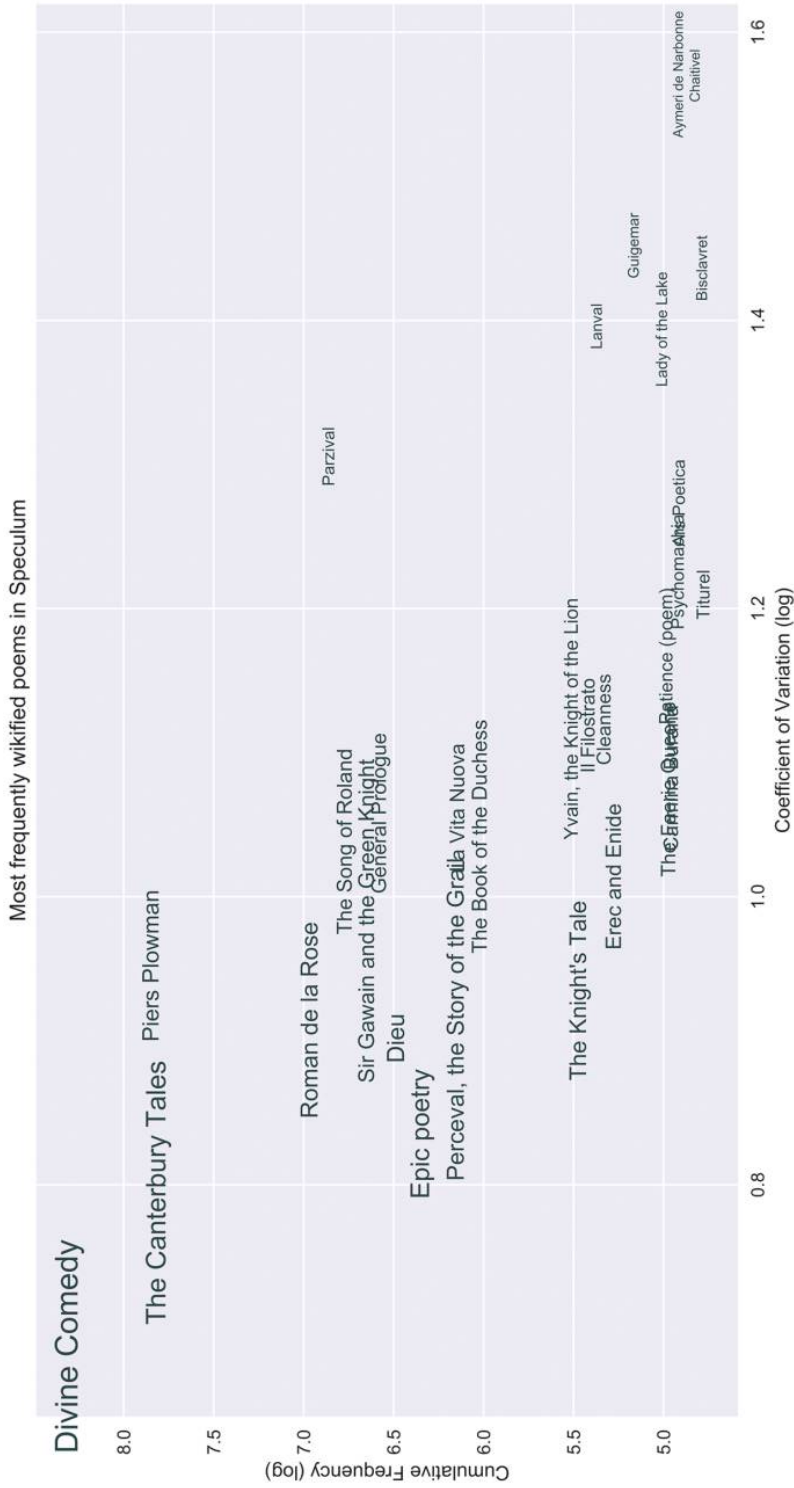
Most frequently wikified poems in Speculum

Divine Comedy

The Canterbury Tales   Piers Plowman

Roman de la Rose

The Song of Roland
Sir Gawain and the Green Knight
Dieu   General Prologue

Epic poetry
Perceval, the Story of the Grail   Vita Nuova
The Book of the Duchess

The Knight's Tale

Yvain, the Knight of the Lion
Il Filostrato
Cleanness

Erec and Enide

The Faerie Queene   Lanval
Patience (poem)   Ars Poetica
Psychomachia   Guigemar
Titurel   Lady of the Lake
Bisclavret

Parzival

Aymeri de Narbonne
Chaitivel

Cumulative Frequency (log)
8.0   7.5   7.0   6.5   6.0   5.5   5.0

Coefficient of Variation (log)
0.8   1.0   1.2   1.4   1.6

Fig. 6 (*Continued*)

Fig. 6 (*Continued*)

S29

While such hit lists are interesting in their own right, looking at mere frequency does not reveal the complex relationships that might exist among them and with other words with which these entities are typically associated. To study and visualize these, we turn to one final technique, from the sphere of distributional embeddings: word embeddings. Just like topic modeling techniques, word embeddings build upon the distributional hypothesis that words with a similar meaning will have the tendency to appear in similar contexts. However, whereas topic modeling techniques are geared to finding good representations for topics and documents, word embedding can yield much more fine-grained representations for individual words. Word embeddings will represent the items in a vocabulary using a numerical vector, or a list of numbers that aim to characterize the word meaning. The advantage of such a word-level model is that we can apply straightforward arithmetic to these vector representations and ask the model, for instance, to return the five words that it deems most similar to a certain query term. If we apply a popular word-embeddings model (word2vec) to our wikified corpus, we can inspect the immediate semantic neighborhood of the following terms listed in Table 2.[75] Using the vector representation that we can extract for our wikified authors, we can also use these embeddings to visualize the relationships between our authors in a dendrogram, or tree diagram. In Fig. 7, the wikified links take the form of leaves in a tree, which are eventually joined into new nodes in a branch structure. The branches reflect the distances between the representations that we obtained for these authors. Note how the structure that arises from this tree makes sense (monarchs cluster with monarchs, philosophers with philosophers, and so on) but also offers some surprising results: Ovid and Virgil, for instance, cluster with Boccaccio, Petrarch, and Dante, instead of with other authors from antiquity, such as Cicero or Plato. Note, also, how the tree realizes at the top level what seems to be a fairly neat split between vernacular authors and nonvernacular authorities.

Such word embeddings have attracted a good deal of attention recently, mostly because it has been shown that these models are able to solve independently an interesting form of analogical reasoning problem. For example, when asked which word is to "woman" as "king" is to "man," a model trained on English Wikipedia text will output the word "queen."[76] The task is simply solved through the following equation: $king - man + woman$. The idea is that the model takes its vector representation for the word *king*, "subtracts," or removes, all abstract properties that it associates with the word *man*, and then adds all the properties that it associates with the word *woman*. The model subsequently returns the word that is closest to the result of the operation. Other, culturally intriguing outputs of the original model were: $japan - sushi + new\_york \rightarrow pizza$ and $belgium - brussels + france \rightarrow paris$. As an interesting *Spielerei*, note that such a model is able to solve thought-provoking questions such as "Who is the Chaucer of the French?" through simply modeling it in the form of the equation $Geoffrey\_Chaucer - English\_language + French$

[75] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig, "Linguistic Regularities in Continuous Space Word Representations," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2013), 746–51.
    [76] Ibid.

Table 2
The Nearest Neighbors for a Selection of Canonical Entities
Using a Word-Embeddings Model

King_Arthur	Chrétien_de_Troyes	Geoffrey_Chaucer	Charlemagne
Matter_of_Britain	Perceval,_the_Story_ of_the_Grail	The_Canterbury_Tales	Louis_the_Pious
Round_Table	Yvain,_the_Knight_ of_the_Lion	General_Prologue	Pepin_the_Short
Gawain	Cligès	The_House_of_Fame	Charles_the_Bald
Mordred	Erec_and_Enide	Troilus_and_Criseyde	Clovis_I
Round_Table_ (Camelot)	Erec	Troilus	Carolingian_dynasty

_language. A number of actual results from drawing such a deliberately provocative analogy from our *Speculum* model are given below:

Geoffrey_Chaucer – English_language + French_language → Jean_de_Meun

Geoffrey_Chaucer – English_language + Latin → Ovid

Geoffrey_Chaucer – English_language + Italy → Giovanni_Boccaccio

While the output from such a naïve model naturally should be taken with a grain of salt, such exercises are nevertheless useful because the models are built in a purely data-driven way, and researchers have noted that these models generally tend to reproduce the cultural biases that are present in the material on which they have been based.

Conclusion: The "Canon" of Medieval Studies

Proponents of distant reading have often praised the ability of computer techniques to broaden our reading scope beyond the obligatory canon of Chaucers, Dantes, and Chrétiens. Moretti, for instance, famously suggested that computer techniques would finally allow us to tackle what Margaret Cohen has called the "Great Unread," the *oubliëtte* of historic literature.[77] So far, however, the results in this respect have been limited, and many digitization projects still center around the comfortable and well-known pantheon of canonized authors—the disproportional attention for a figure like Chaucer in traditional medieval studies, for instance, has been remarkably closely paralleled in the digital universe so far. This is but one case where digital medieval studies can probably do a better job of living up to its promise and lure our attention away from an already overexposed medieval canon towards the lesser-known peripheries of medieval culture.

Nevertheless, it is troubling that much new digital medieval work responds more closely to the questions and concerns of nineteenth-century medieval scholarship than those of the twentieth or twenty-first centuries. In the field of text analysis, for instance, practitioners have so far shown little interest in modern literary theory,

[77] Some of his seminal essays on the matter have been reproduced in Moretti, *Distant Reading*.
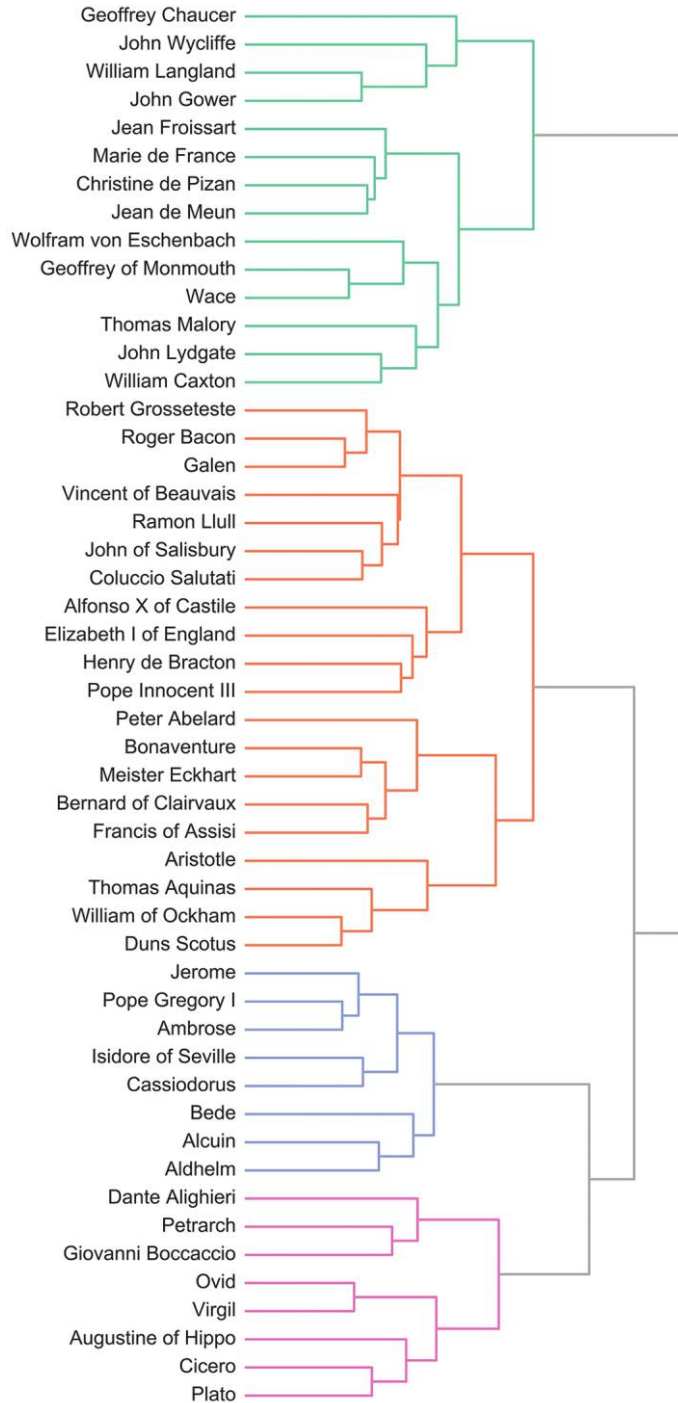
Fig. 7. A dendrogram representing the outcome of a cluster analysis, where the (dis)similarities between writers are visualized as a tree structure. The dissimilarities here are based on the embeddings we obtained for these writers and which capture the semantic context in which these writers are typically mentioned in *Speculum*.

and especially poststructuralist approaches. The postmodern dismissal of—and lack of interest in—the authorship of texts may also explain why digital scholars might keep their distance from a field that does not value issues central to much of digital medieval studies. Influential digital humanists, such as Geoffrey Rockwell or Stephen Ramsay, might interpret this observation in the light of their—as they themselves admit—rather polemic view of digital humanities as a community of "builders": a community that "does" instead of "talks," one that "makes" instead of "writes"[78]— and, we could add, perhaps also a community where scholarship is often so experimental that it is more like "playing" than "working."[79]

The Brothers Grimm rediscovered medieval literature in nineteenth-century Germany and took pains to initiate the scholarly study of a strange cultural phenomenon from a distant past, still fundamentally new to them at the time. They found themselves confronted with the need to catalog, describe, and edit an unstructured mass of new sources, and they struggled to apply the existing scholarly models that they had inherited from their humanist predecessors. Because of the European dimensions of many medieval phenomena, they were also involved in constant negotiations through their international scholarly correspondence, for example, about the authenticity of particular text versions or the directions of cultural exchange in medieval Europe. It would not be far-fetched to liken the condition of present-day digital humanists to their nineteenth-century precursors. Modern digital humanists, too, are confronted with the scholarly study of a medieval heritage that they often have to digitize from scratch, even as they define a scholarly, digital practice without a tradition of existing models that can be applied easily to the computational study and dissemination of these artifacts and new insights about them. Working as a community, many digital humanists are currently reinventing important aspects of medieval studies in that process, through fundamental discussions about the purpose and meaning of the field.

This situation leads to a complex, opaque, and fascinating relationship between digital medieval studies and their conventional counterpart. On an anecdotal level, digital humanists are inspired by the relative freedom they enjoy in the experimental playground that is DH, where scholars can operate largely outside the gaze and criticism of the conventional humanities. According to some, DH can be viewed as a deliberately "undertheorized" field,[80] where young scholars are not hampered by the mechanisms of intimidation and exclusion that are often related to the concept of "theory."[81] Others have claimed that DH is in fact much more theoretical than the traditional humanities, because of the central place that is assigned to fundamental methodological debates about modeling in the humanities. In a famous blog

[78] See, for example, their polemic pieces reprinted in Terras, Nyhan, and Vanhoutte, *Defining Digital Humanities*: Stephen Ramsay, "On Building," 243–45; and Geoffrey Rockwell, "Inclusion in the Digital Humanities," 247–53.

[79] On less targeted research methods facilitated by the digital, see Stephen Ramsay, "The Hermeneutics of Screwing Around," http://quod.lib.umich.edu/d/dh/12544152.0001.001/1:5/--pastplay-teaching-and-learning-history-with-technology?g=dculture;rgn=div1;view=fulltext;xc=1#5.1.

[80] Rockwell, "Inclusion in the Digital Humanities."

[81] For "theory" as a source of intimidation, see Jonathan Culler, *Literary Theory: A Very Short Introduction* (Oxford, 1997), 15.

post, "Who You Calling Untheoretical?," Jean Bauer quoted Susan Smulyan, who shouted on one occasion, "The database *is* the theory!"[82]

While the presence of higher-level theoretical and methodological debates is not open to question, the relationship between traditional and nontraditional schools in medieval studies merits a closer look here. Scholars in digital humanities typically justify their existence through an active affiliation with older humanities disciplines[83]—in fact, one could say that it is primarily this affiliation that separates the digital humanities from computer science. In medieval studies, too, the link between traditional and digital practitioners is crucial if the medieval field is to advance as a whole. Importantly, this requires a mutual interest from both parties and a fundamental willingness to learn from one another, while not neglecting the rich tradition of medievalist scholarship.

While we expect digital medieval studies to become more mainstream in the future, it will remain important to maintain dedicated outlets for digital medievalists to reflect on the more technical aspects of their work. A number of more recently inaugurated specialized journals, such as the *Digital Medievalist Journal* (https://journal.digitalmedievalist.org/) and *Digital Philology: A Journal of Medieval Cultures* (Johns Hopkins University Press) merit watching, in addition to the more established, multidisciplinary journals in DH, such as *LLC: Digital Scholarship in the Humanities* (Oxford University Press; formerly known as *Literary and Linguistic Computing*) and *Digital Humanities Quarterly*, both published on behalf of ADHO. Likewise, the book of abstracts of the annual global conference in DH organized by ADHO (http://adho.org/) helps keep track of current developments in the field.

Equally important for the further development of the field are platforms of a more pedagogical nature, where practical tutorials are offered that can help novice practitioners of the DH to acquire digital skills that may not yet be part of curricular training programs in higher education. Websites such as the Programming Historian (http://programminghistorian.org), for example, offer a wide range of peer-reviewed tutorials on technical skills. Other popular pedagogical resources for novice scholars are the many long-standing training events that are annually organized in the DH community, such as the Digital Humanities Summer Institute at the University of Victoria, the European Summer University in Digital Humanities at the University of Leipzig, and the Digital Humanities at Oxford Summer School (DHOxSS) at the University of Oxford. The THATcamp (The Humanities and Technology Camp) "un-conferences" held in various locations have also spread the word about digital methods and approaches to a broad audience.[84] Apart from a longer exposure to digital humanities practices, such events have an important social dimension by allowing newcomers to build up a network in DH.

---

[82] Jean Bauer, "Who You Calling Untheoretical?," *Journal of Digital Humanities* 1/1 (2011), http://journalofdigitalhumanities.org/1-1/who-you-calling-untheoretical-by-jean-bauer/.

[83] Cf. Liu, "The Meaning of Digital Humanities."

[84] http://thatcamp.org/.

THIS DIGITAL SUPPLEMENT

This supplement is divided into four sections that aim to represent many of the trends we have traced above. In the first section, "Manuscripts and Images," four papers engage with approaches to manuscript analysis. Toby Burrows introduces a project that collates the manuscripts formerly in the collection of Sir Thomas Phillipps and explores the challenges of analyzing large corpora. The enormous manuscript collection assembled by Phillipps in the nineteenth century was subsequently dispersed to institutions and private collectors around the world. Because the evidence relating to the provenance and history of these manuscripts is extensive and varied, developing a coherent framework for analysis required implementing a new data model for manuscript provenance. As well as examining the technical processes involved in this work, Burrows presents the results of applying this approach to two specific research questions: the histories of the group of manuscripts that were owned by both Thomas Phillipps and Alfred Chester Beatty, and the combined histories of the former Phillipps manuscripts that are now in institutional collections in North America.

Although it is well known that many scribes had several scripts and even alphabets available to them, there has been little discussion of the phenomenon from a paleographical point of view, and even less of the methods to address it. In his contribution about multigraphism in late Anglo-Saxon manuscripts, Peter A. Stokes examines the work of two multigraphic scribes in detail, drawing on the DigiPal framework and exploring the capabilities that it gives for communication and analysis of script and the insights it provides about late Anglo-Saxon scribal practice and multigraphic script in general.

Mike Kestemont, Vincent Christlein, and Dominique Stutzmann propose what they call "artificial paleography," based on the adaptation of technology from the field of computer vision and artificial intelligence to the paleographic study of medieval manuscripts. Their paper focuses on the automatic identification of script types in medieval manuscripts, which is an important step on the road to the fully automated "machine reading" of these documents. The work is presented in the context of a recently organized competition, or "shared task," on this subject, which is an increasingly common scientific format in the world of digital scholarship. In addition to a high-level introduction to the computer models they use, the paper focuses on the interpretation of these complex systems against the background of traditional paleography.

Murray McGillivray and Christina Duffy shine the new light of spectrometry to see beneath the illuminations of the well-known Gawain manuscript. Their article engages with the techniques of multispectral imaging to examine the illustrations in London, British Library, MS Cotton Nero A.x., the unique manuscript of *Sir Gawain and the Green Knight* and three other important Middle English poems. Imaging reveals the ink drawings under the later paint and detects differences from the illustrative goals, damaged and faded portions of images that were restored, and the intentional deployment of chemically different pigments that have come to look similar with the passing of time.

The second section, on "Mapping," includes two articles that illustrate the use of Geographic Information Systems (GIS) in medieval studies. David Joseph Wrisley explores digital mapping for medieval studies at multiple scales for both close and distant readings. His article distinguishes mapping geographical information from historical GIS, and it presents several findings of the Visualizing Medieval Places (VMP) project for the study of medieval French texts. Wrisley argues for the need to expand the project into a research architecture that allows social cocreation of data and explores the affordances of linked open data. M. Alison Stones describes the evolution of the web-based Lancelot-Graal project, which adapts GIS to the geography of the manuscript page, using it as part of a comparative examination of differences in the choice, placement, and treatment of subjects in manuscript illustrations.

The third section, "Texts and Editions," brings together four articles. Jeroen De Gussem traces the "secretarial trail" of Bernard of Clairvaux by using the techniques of stylometry. The literary style of Bernard of Clairvaux (c. 1090–1153) was of such grandeur that it was imitated by the greatest theologians of his time, providing an "architecture" for a Cistercian way of writing. Bernard's best imitators were, in fact, found by his side, in the scriptorium of Clairvaux. These scribes were trained to mimic their abbot's preferred wording and his mastery of rhetorical twists, and although Bernard made a habit of rereading, correcting, and repolishing his works, it is often unclear how we should estimate his secretaries' part in the ultimate constitution of his oeuvre. The focal figure in Bernard's scriptorium was Nicholas of Montiéramey, who served the abbot from c. 1138–41 to c. 1151–52, and in this article, the dynamics of kinship between Bernard's and Nicholas's oeuvres are laid bare through stylometric methods. The stylistic familiarity between their texts can teach us more about the nature of collaboration in the scriptorium of Clairvaux as well as allowing for a better close reading of Bernard's more dubiously attributed texts.

Maxim Romanov presents an algorithmic analysis of medieval Arabic biographical collections, a unique data collection whose sheer size has hindered a holistic scholarly treatment so far. His paper illustrates the sort of macroanalyses that large and understudied corpora enable, with an emphasis on the geographic and temporal distribution of the entities in his data. Romanov discusses the complexities of tagging, structuring, and sustaining these data and offers valuable pointers to practical tools and realistic methodologies.

Mark Cruse performs a quantitative analysis of toponyms in a manuscript of Marco Polo's *Devisement du monde* (London, British Library, MS Royal 19 D 1). Scholars have long noted that Marco Polo's account presents many textual problems, and not only to modern scholars. The text's toponyms also posed a particularly great challenge to the scribes who copied the early manuscripts because so many were unknown, and quantitative analysis of the toponyms in the oldest Old French copy of the account (Royal 19 D 1) confirms the scribal uncertainty that attended the copying of these words. By distinguishing between familiar and unfamiliar toponyms, by assigning the occurrences to specific scribes, and by quantifying the number of variants and the degree of orthographic and phonetic variance for each toponym, the article argues that we can identify the words and contexts that proved difficult to scribes. Rather than regarding these variants as errors, Cruse ar-

gues, we should analyze them as forms of reader response. An analysis of these toponyms in their manuscript context as semantic markers devoid of modern annotation enables us to encounter Polo's text as its earliest readers did—as the description of an as yet unknown world teeming with exotic places rich in significance. Ultimately, the ways in which scribes responded to the toponyms in Polo's account reflect not only scribal practice, but also the processes by which new geographical information was absorbed by medieval readers.

Franz Fischer's article surveys a series of digital scholarly editions with a focus on the options and requirements for developing digital textual corpora. On the one hand, textual—or, rather, editorial—plurality seems to be one of the main characteristics of digital editions; on the other, the usefulness of a corpus depends substantially on the uniformity and representativeness of the texts that it includes. Based on a clear yet flexible definition of digital critical editions, Fischer makes several proposals to resolve the conflict between a variety of editorial approaches and a desirable homogeneity within a corpus. Through the inclusion of editions that are digital in a wide sense and critical in a narrow sense, a focus on works rather than documents, and linkage to, or integration of, external resources, he argues that it is possible to create a valuable and truly digital corpus of critical editions. The usefulness of its features and the technical framework of such a corpus would be based on an elementary data model for metadata, text, annotation, and paratexts.

The fourth section, on "Multimediality: Space and Sound," presents three articles that explore reconstructions of medieval architectural space and of the sounds within medieval buildings. Sheila Bonde, Alexis Coir, and Clark Maines use computer-aided design (CAD) technology to reconstruct, represent, and study architectural process at the Cistercian church at Notre-Dame d'Ourscamp, concentrating on the late thirteenth century, when workers dismantled the church's Romanesque east end and replaced it with a new Gothic choir. They argue that digital representation has the potential to encourage viewers to engage with the fuller life cycle of a building, and that it encourages researchers to analyze the three-dimensional application of their interpretations of building change. The goal of their digital project has been to promote a fuller understanding of the process by which medieval builders dismantled parts of earlier buildings to attach newer extensions. The article and CAD project present an extended examination of the construction sequence and engage with issues of uncertainty in virtual representation.

The remaining two articles in this section examine the sounds of Byzantium. The international team of Spyridon Antonopoulos, Sharon Gerstel, Chris Kyriakakis, Konstantinos T. Raptis, and James Donahue investigates the acoustic aspects of Byzantine liturgical spaces in Thessaloniki's churches. Their project unites scientific analysis of acoustics with consideration of the architectural frame and imagery of choral performance. Their project aims to identify and preserve the acoustic signatures of the churches under study and to capture the multisensory experience of the Byzantine worshipper.

Bissera Pentcheva and Jonathan Abel present the method and the results of the Stanford University multidisciplinary Icons of Sound project. They argue that digital technology allows us to transcend a text-based encounter with Byzantine liturgical music and restores the performative aspects of the sung rite, and their focus is on Hagia Sophia: its acoustics, aesthetics, and music. The article details the effects

of the domed structure on the experience of sung chant within it: the amplification of sounds together with overlapping of notes and an "acoustic waterfall" produced both an aural and an optical brightness. Using digital technology, Icons of Sound has successfully imprinted the acoustic signature of the building on the live performance of Byzantine cathedral chant.

The articles in this supplement thus combine to offer a window into the wealth of approaches and experiences that medievalists have brought to the field of digital humanities. It is hoped that this contribution to *Speculum* incites (even more) new interest and fresh activity in this promising field.

David J. Birnbaum, University of Pittsburgh (djbpitt@pitt.edu)
Sheila Bonde, Brown University (sheila_bonde@brown.edu)
Mike Kestemont, University of Antwerp (mike.kestemont@uantwerp.be)