

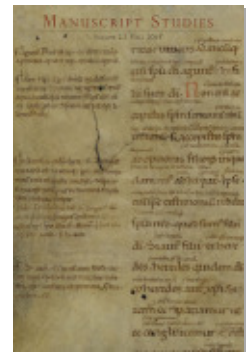


PROJECT MUSE®

*Analysis of Ancient and Medieval Texts and Manuscripts:
Digital Approaches* ed. by T. L. Andrews and C. Macé
(review)

Alexandra Gillespie

Manuscript Studies: A Journal of the Schoenberg Institute for
Manuscript Studies, Volume 2, Number 2, Fall 2017, pp. 577-580
(Review)



Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/mns.2017.0029>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/676184>

Chinese literacy primer, the *Explanations on the Meaning of the Three-Syllable Classic* (Item 30, *San zi jing jiangyi*, 三字經講義). Hand copying also enabled the circulation of texts whose printing had been proscribed, as is the case with the *New Edition of the Lawyer's Thunder to Scare Heaven* (Item 31, *Xinke falü jingtian lei*, 新刻法律驚天雷), a manual for success in legal cases, which was banned in the eighteenth century for its disruptive social effects.

It is clear from the images in Pedersen's catalogue that many of the manuscripts were already in a very fragile state when they were acquired, and that many cannot be repeatedly handled without further deterioration. Open-access digitization would liberate their fascinating contents to a broad readership.

Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches. T. L. Andrews and C. Macé, eds. Turnhout: Brepols, 2014. 338 pp., 27 b/w illustrations, 51 color illustrations, 26 b/w tables. €97. ISBN: 978-2-503-55268-2.

ALEXANDRA GILLESPIE
University of Toronto

IN 2010, STANFORD UNIVERSITY Library estimated that less than 5 percent of ancient and medieval documents and manuscripts had been made available online. In 2017, that number is probably still below 10 percent. The vast majority of digitized manuscripts and documents await transcription; many classical and medieval texts are available online only in outdated or unreliable editions. It is still not clear whether emerging specifications—for example, the World Wide Web Consortium's Web Annotation Data Model or the International Image Interoperability Framework—will deliver on their promise that cultural heritage data sets can be released from institutional “silos,” aggregated, shared, linked, and opened to new forms of

computational inquiry. And even if they can be, what sort of inquiries can and should be undertaken? What new results can researchers expect?

This book provides answers to those questions, although the editors and authors regularly acknowledge that, while the work of imaging, encoding, and exposing texts and manuscript images continues, such answers will be tentative and somewhat premature. Joris J. van Zundert notes in his conclusion to the collection that “most contributions in this volume in some way refer to scholarly edited texts.” Essays on the *Vita et miracula s. Symeonis Treverensis* by Tuomas Heikkilä, Petrus Alfonsi’s *Dialogus* by Philipp Roelli, and Richart de Fournival’s *Bestiaires* by Jean-Baptiste Camps and Florian Cafiero describe advances in the use—since the 1990s—of phylogenetic and other “tree”-building algorithms for textual stemmatics. The principles that lie behind this work are nineteenth-century and Lachmannian: agreements in error suppose a genetic relationship between the witnesses attesting the error. Computation improves rather than changes method in these cases. It improves the range, scale, accuracy, evidentiary basis, and verifiability of scholarly findings. It allows inquiry to proceed when in an earlier era it would have been stalled by practicalities or undermined by cumulative human error.

The ease and speed with which a large and highly varied data set can be processed by a machine enables Alberto Cantera to take on the daunting tradition of Avestan manuscripts of the Zoroastrian liturgies, which were transmitted orally for approximately eight hundred years before the first written copy. Karina van Dalen-Oskam uses similar processing power to isolate, by cluster and principal component analysis, the distinctive diction and vocabulary of scribes of the Middle Dutch *Rijmbijbel*. Francesco Stella, Luca Verticchio, and Stefania Pennasilico extend these methods even further beyond their usual domain of authorship attribution, to apply them to the literary genre of the epistle. In these essays, computation produces statistical support for that which scholars usually intuit. Petrarch did devise a new vocabulary for letter writing; Alcuin was inclined to imitate his Augustinian models. Scribes did innovate as they copied the *Rijmbijbel*, and their innovations are clustered by region and chronology.

Armin Hoenen describes some experiments that supplement van Dalen-Oskam's findings on the habits of scribes. He sets out to test the hypothesis that types of scribal error might be reproducible automatically. In theory, an algorithm that mimics single-letter substitution could produce something that resembles a medieval scribal corpus. Hoenen's results are negative, and revealing. The algorithms he uses produce more variation over a few copies than scribes do over dozens. Scribes, that is, did not copy "automatically": they used phonographic, orthographic, metrical, and semantic cues to maintain the integrity of their message.

Other contributions here are less concerned with the results of computational analysis than with the promise and the structure of data sets. Samuel Rubenson describes a relational database of the *Apophthegmata patrum* that allows literary comparison across all languages and manuscripts. Maxim Romanov argues for text-mining techniques as a way to approach hundreds of thousands of surviving medieval Arabic biographical texts. Eugenio R. Luján and Eduardo Orduña describe the Hesperia database of Paleohispanic languages, which covers over two thousand inscriptions from 500 BC to AD 200, in five Iberian languages, two as yet unidentified. Charlotte Tupman and Anna Jordanous consider the opportunities offered by linked data and the semantic web for aggregating data about wisdom literature in Greek, Arabic, Spanish, and other languages. Linda Spinazzè gives an account of *Musisque Deoque*, a digital archive of Latin poetry that makes room for searchable variants for the texts it archives, and, where possible, for online digitized manuscripts of the witnesses to those variants.

As the volume draws to a close, three contributors consider what computation has to offer paleographers and codicologists. Digitization makes manuscripts widely available as sets of two-dimensional images. How does this change study of the documents and books that bear witness to texts? Patrick Andrist's essay notices that digital manuscript scholarship depends, first of all, on an automated search of metadata about ancient and medieval documents and books—but descriptions of these objects have never been standardized and are often inaccurate. Andrist proposes some sensible solutions for the structuring of data about composite manuscripts—which are

especially problematic for cataloguers—using XML and SQL databases. The problems that Ainoa Castro Correa identifies with paleographical method also pre-date digitization. Paleographers aim to identify, localize, and date scripts and the work of individual scribes. They do so using inexact verbal descriptions of visual data, so that the reliability of any finding “depends on the authority of the author and the faith of the reader” (quoting Albert Derolez). Castro Correa argues persuasively that digitization can mitigate but cannot yet solve this problem. Digital archives supply visual evidence to supplement published descriptions of scribal hands, but only experts are qualified to judge that evidence, and their judgments remain subjective. A number of researchers have devised algorithms that turn the task of identification over to a machine. But such algorithms “learn” on the basis of sample data sets, whose characteristics have been described and defined by paleographers in entirely traditional—and so traditionally limited—ways.

Castro Correa argues that ink analysis may be a more fruitful approach for researchers interested in computational approaches to digitized books. At least in theory, such analysis could provide measurements that do not depend on a scholar’s perception. Ira Rabin proves that point nicely in his essay on the use of multispectral imaging to distinguish tannin, soot, and iron gall inks. Rabin makes a good case for refining codicological descriptions with this evidence, which can be collected by any researcher equipped with a handheld microscope and an NIR (near-infrared reflectance) light source. *Analysis of Ancient and Medieval Texts and Manuscripts* thus ends on a note of both caution and optimism. Computational methods will not magically solve the most challenging problems in the fields of manuscript and textual studies. Critical judgment and informed historical reasoning remain vital to these fields. That is a good thing. Digital approaches to culture and history reveal the damaging artificiality of strict distinctions between humanist, qualitative insights and “scientific,” quantitative ones. There is still much to learn from old books and texts, and we need all sorts of traditional, new, and as yet unimagined ways of learning it.