

Medieval Authorship and Canonicity in the Digital Age – an Introduction**

1 Two ‘Once New’ Philologies

More than three decades ago, in 1990, two different fields in medieval studies, literature and computational linguistics, witnessed the proclamation of a ‘New Philology.’ The launch of the first variant of such a New Philology, in a special issue of the American journal *Speculum*, is by now one of the most frequently recalled success stories in the recent history of literary studies (Bloch *et al.*). Under the impetus of Stephen G. Nichols, the issue’s contributors — all specialised in the study of vernacular literatures and historiography — argued in favour of a philology that, in the study of medieval texts, acknowledges the particularity of medieval manuscript culture (Nichols, “The New Philology”). Whereas traditional philology and literary history had always put forward the published edition of the text as the basis for any further analysis — according to either the ‘text-genealogical’ principles associated with Karl Lachmann (1793–1851) or the ‘best possible manuscript’ principle of Joseph Bédier (1864–1934) — the contributors to the *Speculum* issue argued that it was time for medievalists to privilege the dynamics of the ‘manuscript matrix’ as the object of study. This implied an awareness of the fact that medieval texts were constantly being rewritten in the age before the printing press, that a multitude of actors was involved in their redaction and transmission, and that the materiality of manuscripts and their paratexts were important indicators of how texts were compiled, read and appropriated in new contexts. The new orientation did not, however, come out of the blue. It was strongly inspired by insights from francophone scholarship and showed a clear affinity with Paul Zumthor’s (1915–95) recognition of the ‘mouvance’ in the medieval textual tradition (1972) and with Bernard Cerquiglini’s provocative essay *Éloge de la variante* (1989). The influence of this

* Ghent University, and Ghent University and Research Foundation Flanders (FWO).

** This special cluster contains a selection of contributions that grew out of the conference “The Medieval Literary Canon in the Digital Age,” held at Ghent University on 17–18 September 2018. The conference was organised by Mike Kestemont, Wim Verbaal and the authors of this introduction, in collaboration with the Henri Pirenne Institute for Medieval Studies at Ghent University, the international research network RELICS (Researchers of European Literary Identities, Cosmopolitanism and the Schools), and the research community “Digital Humanities Flanders” (DHu.f). We thank all participants at the conference for their generous exchange of ideas, as well as the editorial board of *Interfaces* for kindly welcoming this collection in its e-journal.

new approach to medieval textuality, which is known today under the label of ‘material philology’, came to be particularly strong.

Less well known is that in the same year, 1990, the Italian Jesuit Roberto Busa (1913–2011) also announced a New Philology (Busa, “Informatics;” see also “Half a Century Ago”). This godfather of twentieth-century computational linguistics had already been convinced of the possibilities of informatics since the late 1940s. It was his seemingly utopian plan to develop a lexical analysis of the entire oeuvre of Thomas Aquinas (1225–74) that led him to embrace computer science when it was still in its infancy. The most often recalled stage in Busa’s career is the moment when, in 1949, he managed to convince IBM’s founder, Thomas J. Watson (1874–1956), to join him in his project (Birnbaum *et al.* S1–S2). His *Index Thomisticus* resulted in fifty-six printed volumes, but was also launched on CD-ROM exactly forty years after his deal with Watson. This digital collection of about 180 texts and 11 million lemmatised words, which allowed concordances to be generated digitally, constituted the first machine-readable corpus of such a size. It was also this achievement that led Busa to argue for a new philology in which the main challenge was to advance artificial intelligence in the semantic processing and syntactic analysis of large quantities of texts. Busa’s new philology did not seek to reject traditional approaches (Busa, “Informatics” 343). It implied above all an awareness among philologists of the potential of computational research and a research agenda that aimed to help the further development of that potential. For Busa, this new philology was, in sum, about “a quality-leap and new dimensions” (Busa, “Informatics” 339).

Thanks to the further development and effective application of machine learning, computational text analysis has indeed made great qualitative progress since the early 1990s. The “new dimensions” promised by Busa have also manifestly unfolded. In the past few decades an increasing mass of texts from different times and regions and in multiple languages has become machine-readable and therefore suitable for large-scale analyses. This accessibility also, it is argued, finally offered unprecedented opportunity for studying authors and texts that had never made their way into the established literary canons. When in 2000 Franco Moretti presented for the first time his well-known concept of ‘distant reading,’ advocating the exposure of textual connections within enormous bodies of digitised texts, he explicitly stated that, contrary to traditional ‘close reading,’ his method allowed one to “look beyond the canon” (“Conjectures on World Lit-

erature” 57). In the same vein, Matthew Jockers, in his computationally-driven macro-analysis of nineteenth-century novels, reflected on how his new methodology had shown that the “the canonical greats” appeared to be “not even outliers; they are books that are similar to other books, similar to the many orphans of literary history that have been long forgotten in a continuum of stylistic and thematic change” (Jockers 168).

In presenting their respective views on what contemporary philology ought to do, both Stephen G. Nichols and Franco Moretti took a critical stance toward Ernst Robert Curtius’s (1886–1956) *Europäische Literatur und lateinisches Mittelalter* (1948) as an iconic expression of traditional philology. Nichols argued that Curtius’s classic, in its somewhat restrictive focus on the European ‘unity’ of poetic form in the Latin Middle Ages, had failed to take into account that the exact opposite is in fact far more characteristic of medieval literary production, namely its multiplicity and variance (Nichols, “The New Philology” 2). For Moretti, Curtius’s Latin Middle Ages and its *topoi*, which the latter presented as “die verwitterte Römerstraße von der antiken zur modernen Welt” (Curtius 29), offered too static a model to understand European literature (Moretti, “Modern European Literature” 86–88, 91, 98–99). However, if we compare the ways in which the two ‘once new’ philologies born in 1990 impact on today’s medieval studies, then a number of differences or at least apparent contradictions stand out as well.

First of all, it is evident that material philology succeeded early on in making its mark on the traditional field of research. The fact that Nichols’s first manifesto immediately appeared in *Speculum* certainly contributed to this rapid success. It is fair to say that within material philology interest in the digital humanities has grown rapidly; in particular in the digitisation of manuscripts and in new digital edition techniques that, in contrast to traditional printed critical editions, value the uniqueness of single manuscripts while making comparisons between manuscripts possible.¹ Yet it was not until 2017 that *Speculum* also devoted an (exclusively online) special issue to “The Digital Middle Ages,” offering fascinating samples of the most cutting-edge research in this field.

Secondly, one may wonder if the methods and principles of material and computational philology do not contradict each other. The fact that many computational analyses start from digital corpora based on editions, in which orthographic variation is often even filtered out in order to better reveal recurrent linguistic patterns in

1. For an excellent example, see the [Online Froissart Project](#) (Ainsworth and Croenen).

texts, is in a certain sense at odds with the appreciation of ‘variance’ in material philology.

Finally, the overtly post-structuralist agenda from which material philology emerged seems to have few obvious affinities with the research questions often found in computational linguistics. This is perhaps most apparent in the case of stylometry, or the study of style based on quantitative analysis, which also forms the central approach in the four case studies presented in this themed cluster of *Interfaces*. Indeed, much stylometric research is concerned with authorial attributions of disputed or anonymous texts. Such questions of attribution, of course, have little in common with the denial, within material philology, of the romantic concept of the ‘author’ as the unique and identifiable creative force that is supposed to have been at the basis of every ‘new’ text. Moreover, one can rightly ask whether their ultimately traditional fixation on authorship is not also canon-confirming, in spite of Moretti’s and Jockers’s ambition to break open canons via computational distant reading. In what follows we will dwell on these considerations by surveying stylometry’s origins and early history. Whereas this history is undeniably closely entwined with positivistic and romantic notions of individual authorship typical of nineteenth-century philology, the technical advancements and new scholarly insights of the past few decades are increasingly telling a much more nuanced story.

2 Stylometry and Authorship

Although ‘stylometry’ as a term was coined in the nineteenth century, it has become commonplace to associate the method with earlier philological approaches dating back to at least the Italian humanists of the fifteenth century. Often considered as one of its forefathers is Lorenzo Valla (1407–57), whose unmasking of the controversial *Donatio Constantini* as a Carolingian forgery was primarily based on stylistic arguments (Eder 63–64). Although Valla’s approach was indeed formalistic and focused on matters of style, he did not, however, apply statistical analysis. In that regard, it was rather his contemporary Leon Battista Alberti (1404–72) who was in the vanguard (Ycart). In 1466, Alberti composed a mathematical treatise on cryptography called *De componendis cyfris*. One could argue that by finding out statistically informed characteristics of language, namely the frequency patterns of vowels in Latin, Alberti was already practicing

2. A manual or computer-assisted way (e.g. through machine-driven retranslation or paraphrasing) to obfuscate the writing style of a text and circumvent stylometry's potential to recognise authorship.

an early kind of 'adversarial stylometry.'²² He explored ways to obfuscate the style and content of a text through encryption with the aim of concealing an author's identity or message.

Regardless of the intriguing parallels with such distant ancestors, the cradle of stylometry is clearly to be found in the positivist spirit, formalism and empiricism of the late nineteenth and early twentieth century. It is telling that one of the earliest scholarly articles for computational approaches to style appeared in *Science*. It was written not by a philologist but by the American physicist and meteorologist Thomas Corwin Mendenhall (1841–1924). Mendenhall took up the novels of Charles Dickens (1812–70) to verify if frequency distributions apply to style as well. He manually counted word lengths for small segments of text, and by plotting these lengths he stumbled upon what he called 'characteristic curves' that appeared to be consistently the same for texts of the same authorship (Mendenhall). Another notable figure active in these same decades was the Polish philosopher and philologist Wincenty Lutosławski (1863–1954), who wrote *Principes de stylométrie* in 1890, thereby establishing the eponymous method (Lutosławski). Lutosławski was able to establish the chronology of Plato's writings by focusing on what he himself called 'stylèmes,' which he understood to comprise rare words used in a conspicuously high number, word frequencies, word position in the sentence, and proportional frequency of the parts of speech. Around the same time, the British statistician (George) Udny Yule (1871–1951) introduced vocabulary richness as a stylometric feature, a technique which is still used today. Armed with this and other methods, Yule verified suspicions that the *De imitatione Christi*, the influential and intensively translated devotional treatise of the Modern Devotion movement, was written by the Augustinian canon Thomas of Kempen (1380–1471) (Yule). A final figure of achievement in the early field of stylometry is the American linguist and philologist George Kingsley Zipf (1902–50), especially known for his controversial and still much-debated 'Zipf's law.' Zipf pointed out that about half the words human beings use in writing and conversation correspond to the 150 most frequent words, a phenomenon which he explained in his 'principle of least effort' (Zipf). He argued that human beings tend to minimise the number of letters — or words — necessary to bring a message across, which is why (generally) half of any language consists of the same words over and over. These are grammatical or syntactical words which despite their omnipresence are often overlooked, such as conjunctions, pronouns, prepositions, adverbs and particles.

The breakthrough of stylometry came in the early 1960s, when the revolutionary advent of early computing advanced the evidence that these ‘function words’ — whose ‘silent’ omnipresence Zipf had already pointed out — convey significant information about the writer using them. The book *Inference and Disputed Authorship: The Federalist*, published in 1964 by the two American statisticians Frederick Mosteller (1916–2006) and David Lee Wallace (1928–2017), intended to formulate an answer to the long-standing authorship controversy around the pseudonymous late eighteenth-century Federalist papers. Mosteller and Wallace were able to show that the statistical analysis of function words was extremely efficient for distinguishing works of different authorship, and their book became the foundational scholarly work of non-traditional authorship attribution. All the contributions in the current cluster of *Interfaces* discuss function words in much detail and with further evidence to prove their effectiveness, which demonstrates the lasting significance of Mosteller and Wallace’s revolutionary discovery of a ‘stylistic DNA’ or ‘stylistic fingerprint’ sixty years on.

This last observation, however, should not give the false impression that the progress of stylometry has stagnated since Mosteller and Wallace. The tide in technical advancements since the 1960s announced the arrival of the digital age and has brought methodological improvement and progress to Mosteller and Wallace’s initial discovery, whose computer, after all, was still approximately the size of a car. Especially since the 1980s, the field of stylometry has been able to benefit from the improvements in computing performance. Worthy of note in this regard is John Burrows’s (1928–2019) introduction of multivariate analysis of style with Principal Components Analysis or PCA (Burrows), which had by 2000 become “the standard first port-of-call for attributional problems in stylometry” (Holmes 114).

Around the turn of the millennium, the field gradually witnessed the impact, as we noted above, of artificial intelligence and machine learning, combined with a larger arsenal of stylistic techniques and feature types (Stamatatos 539). The advancement of these techniques allowed stylometrists to work not only with the traditional bag-of-words approach,³ but also with *n*-grams,⁴ rhythmic and auditive aspects of style, lemmatised, grammatical and syntactic features, and even word embeddings for capturing words’ semantics through context (Mikolov *et al.*). The simultaneous arrival of machine-learning frameworks has moreover allowed for a better-informed assessment of the accuracy and reliability of this variety of stylometric

3. The bag-of-words approach represents a document as a ‘bag’ or ‘multiset’ of words. It exclusively takes into account word frequencies, disregarding context, word order or any other orderly principle of grammar or syntax.

4. *N*-grams are sequences of *n* (variable number of) characters/ words/parts-of-speech from a given sample of text or speech.

methods. Stylometry is also increasingly being made more accessible to non-experts in user-friendly packages with graphical user interfaces such as the Lexomics group's 'Lexos' (Kleinman and LeBlanc) or the Computational Stylistics Group's 'Stylo with R' (Eder, Rybicki and Kestemont), and has gradually become more transparent in its mode of operation.

This increase in the precision and accuracy of stylometric methods is not merely promising from a computer-scientific point of view, but also from a literary-historical one. Thanks to its technical advancements, stylometry is becoming increasingly attuned to challenging simplistic notions of individual authorship and can help scholars sharpen their understanding of literary writings as the result of layered, complex authorial roles. Stylometry's focus has in the past years been able to shift beyond attribution for the sake of attribution. We find stylometric scholarship exploring the implications of multi-authored or collaborative contexts, posterior redaction and editorial amendments of texts, stylistic influence and apprenticeship, intertextuality and shared linguistic communities, cross-linguistic authorship or authorship filtered through translation, stylistic development within authors' texts or entire oeuvres, or of different style registers for characters in works of fiction. The realisation and exploration of such complex models of authorship instantly draws attention to the contributions of the anonymous, marginal or suppressed voices of literary history that we have lost track of or forgotten. In other words, stylometry is becoming better equipped to explore (and even confirm) those aspects of textual instability which Nichols had presented as an essential characteristic of medieval literary production. As such, digital methods have developed at least one significant way of questioning the medieval canon, precisely at a juncture where the interests of the two 'once new' philologies of 1990 converge.

3 Questioning Canonicity in the Digital Age

A repeated promise in the wake of the 'digital turn' in literary scholarship is, as we have seen, that the growing availability and accessibility of digitised historical texts will enable scholars to transcend the limitations of traditional literary canons. However, much of the digital scholarship within medieval studies still seems to hinge primarily on well-conserved texts and often studied authors that continue to attract academic interest (*Roman de la Rose*, Christine de Pisan etc.

— see e.g. Nichols, *From Parchment; Digital Library*). It may be relevant, therefore, to question the criteria that define our textual canons and the ways in which the rise of digital analyses may impact on them.

In a thoughtful article on this subject, Lars Boje Mortensen recently proposed a fine-grained model to assess medieval literary canons by analysing the forces that hold them in place (“The Canons”). The model distinguishes between four levels of canonicity. It draws its inspiration from Aleida Assmann’s conceptual distinction between ‘Canon’ and ‘Archive’ in the construction and maintenance of cultural memories. In Assmann’s theory, the Canon designates the ‘working memory’ that supports collective identities and that is built on a selective number of normative and formative texts and other cultural products, while the Archive denotes the cultural ‘reference memory’ that is passively maintained and stockpiled for potential future reframing and reinterpretation (Assmann; see also McGann 47–48). The first level in Mortensen’s four-tiered approach is that of the High Canon, encompassing texts and authors that are globally appreciated. They enjoy a multimedial presence in popular culture and dominate scholarship. The names belonging to the second level, the Broad Canon, are well-known within medieval studies but hardly visible in popular culture. Here we find both representatives of the learned culture of the Middle Ages, that are of transnational significance, and texts and authors that can be considered as foundational within national cultures and historiographies. All are often studied, edited, anthologised, translated etc. The third level, that of the Open Archive, contains texts that are generally well accessible in decent editions and listed in repertories and literary histories, but that are the object of only limited and specialised study. Finally, the Closed Archive comprises all kinds of texts hidden in manuscripts that are less known or studied, remain poorly or even unedited or are only known through reconstruction on the basis of other texts.

The four case studies collected in this cluster of *Interfaces* fit in with, and flesh out, Mortensen’s four-tiered model in a particularly appropriate way. While each individual article addresses and questions issues of authorship and scribal roles from its own specific angle, they collectively offer an original perspective on how computational methods in dialogue with traditional hermeneutics can also lead to new approaches to the four different levels of medieval canonicity.

Jeroen De Gussem’s article hones in on the joint authorship of the *Vita* of the twelfth-century visionary Hildegard of Bingen (1098–1179), an author who by now may be said to have secured her place

in the High Canon of medieval literature (Mortensen 58). However, her *Vita* contains rare and disputed autobiographical fragments which have often raised suspicions that these were heavily revised by consecutive hagiographers. Armed with computational stylistics, De Gussem establishes in considerable detail the layered character of the text, thereby bringing to light its collaborative authorship. By illustrating the involvement of Hildegard and a team of biographers in the *Vita*, De Gussem highlights the importance the visionary and her community attached to her constructed *persona*, her remembrance by posterity and her possible ‘canonicity,’ or even saintly ‘canonisation.’

Mary Dockray-Miller, Michael D.C. Drout, Sarah Kinkade and Jillian Valerio continue on De Gussem’s trail of hagiography and composite authorship, but in relation to a text that can be considered as one of the eleventh-century classics from the Broad Canon of England’s literary history. By making use of Lexomic technology developed at Wheaton College (Massachusetts), their piece explores the authorship of the contested prosimetric *Vita* of Edward the Confessor (1003–66) written around the time of the Norman Conquest of 1066. The candidates conventionally proposed in this authorship debate are the itinerant continental monks Goscelin (d. after 1107) and Folcard (fl. 1060s) of Saint-Bertin, who were from the mid-eleventh century onward recruited by a number of notable monastic houses in England for their hagiographical skill. In making a case for a composite authorship of the *Vita Æwardi*, Dockray-Miller *et al.* break new ground by challenging the ‘individual attribution’ of the text to a single author. In taking us through the complex and composite stylistic fabric of the *Vita*, they not only shift the focus from a single individual author to an entire school of writing, but also attach central importance to Queen Edith of Wessex (1029–75), King Edward’s widow who commissioned the work. They finally argue that if there is one authorial voice that may have overseen the composition of the *vita* in its entirety, it must be that of the well-educated Edith.

With the article of Eveline Leclercq and Mike Kestemont, we temporarily leave the realm of purely literary texts to further widen Mortensen’s idea of the Open Archive to documentary sources. With acknowledgements to the literally ‘open archives’ in the form of open-access databases such as *Diplomata Belgica* and *Chartae Galliae*, the authors pair distant reading with conventional diplomatic approaches to the formulaic language of charters. They present their double method as ‘distant diplomatics,’ and engage in disentangling

the multiple authorial strata (issuer, dictator, scribe, etc.) in charters and in detecting traces of the local preferences and compositional habits of the chanceries which the charters' scribes depended on. Leclercq and Kestemont present a thorough analysis of the development of a specific *dictamen* in a corpus of twelfth-century Latin charters from the Cambrai episcopal chancery. But more importantly, their article offers a promising methodological exploration of the potential of stylometry in the field of diplomatics.

As the only contributor in this cluster focusing on vernacular medieval texts, Gustavo Riva statistically analyses the rubrics to a corpus of short *Reimpaargedichte* in miscellany manuscripts from the twelfth to the sixteenth century. In doing so he draws attention to what is commonly called the 'paratext,' denoting the structural and marginal components of texts that until now remain hidden in the stratum of what Mortensen designates as the Closed Archive. It is in rubrics, Riva argues, that one can find the traces of the anonymous scribes responsible for preserving, copying and transmitting medieval texts, who by their rubrication "named and renamed" them, and who both literally and figuratively coloured these texts' reception for posterity. By statistically aggregating information about their lengths, their lexical variability, their most common lexical properties and their authorship, Riva's distant reading of rubrics permits the conclusion that they are rarely uniform and are dependent upon time- and place-bound conventions.

One final thought, before letting the articles speak for themselves: it is clear that the individual case studies presented here, despite focusing on different levels of canonicity, do not really question this hierarchy as such. Does this mean that the influence of the digital turn in medieval studies leaves traditional canons untouched? That is doubtful. As Mortensen has also noticed, the canons of medieval literature looked completely different in the past, especially in the centuries before the rise of romanticism and nation-states. To understand the "ups and downs in the long afterlife of medieval texts," Mortensen argues, it is not enough to look only at the influence of "ideology, political and educational context or shifts in literary taste" (Mortensen 47). Since the early Middle Ages, the accessibility of texts, dependent as it was on means of material transmission and the milieus in which these texts were collected and read, has also been an essential parameter in determining their popularity. It is therefore inevitable that the growing digital availability of texts and manuscripts, the facilitation of new research questions and the in-

creasing globalisation of education and learning will lead again to rearrangements of the scales of canonicity.

Bibliography

- Ainsworth, Peter and Godfried Croenen. [The Online Froissart](#) (2013). Accessed 19 November 2021.
- Assmann, Aleida. "Canon and Archive." *Cultural Memory Studies: An International and Interdisciplinary Handbook*. Ed. Astrid Erll and Ansgar Nünning. Berlin: De Gruyter, 2008. 97–107.
- Birnbaum, David J., Sheila Bonde and Mike Kestemont. "[The Digital Middle Ages: An Introduction](#)." *Speculum* 92.S1 (2017): S1–S38. Accessed 19 November 2021.
- Bloch, R. Howard, Alison Calhoun, Jacqueline Cerquiglini-Toulet, Joachim Küpper, and Jeanette Patterson (eds.). *Rethinking the New Medievalism*. Baltimore: Johns Hopkins University Press, 2014.
- Burrows, John F. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford, UK: Clarendon Press, 1987.
- Busa, Roberto. "Informatics and New Philology." *Computers and the Humanities* 24.5 (1990): 339–43.
- . "Half a Century of Literary Computing: Towards A 'New' Philology." *Historical Social Research* 17.2 (1992): 124–33.
- Busa, Roberto et al. *Index Thomisticus*. [Web edition](#) by Eduardo Bernot and Enrique Alarcón (2005). Accessed 19 November 2021.
- Cerquiglini, Bernard. *Éloge de la variante: Histoire critique de la philologie*. Paris: Seuil, 1989.
- Curtius, Ernst Robert. *Europäische Literatur und lateinisches Mittelalter*. Bern: A. Francke AG, 1948.
- [Digital Library of Medieval Manuscripts](#). Baltimore, MD: Johns Hopkins Sheridan Libraries. Accessed 19 November 2021.
- Eder, Maciej. "A Bird's Eye View of Early Modern Latin: Distant Reading, Network Analysis and Style Variation." *Early Modern Studies after the Digital Turn*. Ed. Laura Estill, Diane K. Jakacki and Michael Ullyot. *Medieval and Renaissance Texts and Studies* 502. *New Technologies in Medieval and Renaissance Studies* 6. Toronto: Iter Press, 2016. 61–88.
- Eder, Maciej, Jan Rybicki and Mike Kestemont. "Stylometry with R: A Package for Computational Text Analysis." *R Journal* 8.1 (2016): 107–21.
- Holmes, David I. "The Evolution of Stylometry in Humanities Scholarship." *Literary and linguistic computing* 13.3 (1998): 111–17.
- Jockers, Matthew Lee. *Macroanalysis. Digital Methods and Literary History*. Champaign, IL: University of Illinois Press, 2013.
- Kleinman, Scott and Marc LeBlanc. "Lexos: An Integrated Lexomics Workflow." *Digital Humanities 2017: Conference Abstracts* (McGill University & Université de Montréal, August 2017): 734–35.
- Lutosławski, Wincenty. *Principes de stylométrie*. Paris: Ernets Leroux, 1890.
- McGann, Jerome. *A New Republic of Letters. Memory and Scholarship in the Age of Digital Reproduction*. Cambridge, MA: Harvard University Press, 2014.
- Mendenhall, Thomas Corwin. "The Characteristic Curves of Composition." *Science* 9.214 (1887): 237–49.

- Mikolov, Thomas, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Dean Jeffrey. "Distributed Representations of Words and Phrases and their Compositionality." *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*. 2013. 3111–19.
- Moretti, Franco. "Modern European Literature: A Geographical Sketch." *New Left Review* 206 (1994): 86–109.
- . "Conjectures on World Literature." *New Left Review* 1 (2000): 54–68.
- Mortensen, Lars Boje. "The Canons of Medieval Literature from the Middle Ages to the Twenty-First Century." *Analecta Romana Instituti Danici* 42 (2017): 47–63.
- Mosteller, Frederick and David L. Wallace. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. Springer Series in Statistics. New York: Springer Verlag, 1984. 2nd ed. of *Inference and Disputed Authorship: The Federalist*. Reading, Mass.: Addison-Wesley, 1964.
- Nichols, Stephen G. "Introduction: Philology in a Manuscript Culture." *Speculum* 65.1 (1990): 1–10.
- . *From Parchment to Cyberspace. Medieval Literature in the Digital Age*. Medieval Interventions. New Light on Traditional Thinking 2. New York: Peter Lang, 2016.
- Stamatatos, Efstathios. "A Survey of Modern Authorship Attribution Methods." *Journal of the American Society for Information Science and Technology* 60.3 (2009): 538–56.
- Ycart, Bernard. "Alberti's Letter Counts." *Literary and Linguistic Computing* 29.2 (2014): 255–65.
- Yule, George Udny. "On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship." *Biometrika* 30.3/4 (1939) 363–90.
- Zipf, George Kingsley. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press, Inc., 1949.
- Zumthor, Paul. *Essai de poétique médiévale*. Paris: Éditions du Seuil, 1972.