Renaissance and Reformation
Renaissance et Réforme

Augmented Criticism, Extensible Archives, and the Progress of Renaissance Studies

Michael Ullyot

Résumé de l'article

Durant les trois décennies qui ont suivi l'émergence de la nouvelle histoire, les études de la Renaissance ont développé grâce à un travail approfondi d'archives de nouvelles données à interpréter. Des développements similaires de plus grande ampleur nous attendent, tels que ceux que nous observons dans l'étude des archives imprimées anglaises. Des transcriptions pouvant être analysées par des logiciels permettent maintenant aux chercheurs d'utiliser des algorithmes révélant de nouveaux faits réels, et pourtant inaccessibles avant aujourd'hui. Il s'agit d'un argument non pas en faveur de la critique algorithmique, mais en faveur d'une critique plus vigilante, assurant que le jugement humain est bien au centre des hypothèses et des résultats des méthodes de recherche algorithmique. Cet article fait un portrait des méthodes émergentes qui ne sont possibles qu'en 2015, et qui pourraient avoir le même effet que le télescope pour l'astronomie.

Augmented Criticism, Extensible Archives, and the Progress of Renaissance Studies[1]

MICHAEL ULLYOT

University of Calgary

In the three decades since the rise of New Historicism, Renaissance studies has progressed through extensions of scholars' archival reach to new objects for new interpretations. The future will bring expansions on a larger scale, like those we now witness in English print archives. Machine-readable transcriptions of some fifty thousand texts now enable scholars to use algorithms that tell us things about them that are true, yet can only be known in the future. This is an argument not for an algorithmic criticism but for an augmented criticism, in which human judgments are the origin and outcome of algorithmic research methods. It sketches the emergent methods that are possible only in 2015, yet will do for the archival humanities what telescopes did for astronomy.

Durant les trois décennies qui ont suivi l'émergence de la nouvelle histoire, les études de la Renaissance ont développé grâce à un travail approfondi d'archives de nouvelles données à interpréter. Des développements similaires de plus grande ampleur nous attendent, tels que ceux que nous observons dans l'étude des archives imprimées anglaises. Des transcriptions pouvant être analysées par des logiciels permettent maintenant aux chercheurs d'utiliser des algorithmes révélant de nouveaux faits réels, et pourtant inaccessibles avant aujourd'hui. Il s'agit d'un argument non pas en faveur de la critique algorithmique, mais en faveur d'une critique plus vigilante, assurant que le jugement humain est bien au centre des hypothèses et des résultats des méthodes de recherche algorithmique. Cet article fait un portrait des méthodes émergentes qui ne sont possibles qu'en 2015, et qui pourraient avoir le même effet que le télescope pour l'astronomie.

In a 2010 essay for the *London Review of Books*, Keith Thomas describes the Oxford method of historical research he learned from his tutor, Christopher Hill. Thomas's subject is "the historical ethnography of early modern England," which encompasses "such diverse topics as literacy, numeracy, gestures, jokes, sexual morality, personal cleanliness or the treatment of animals."[2] Such catholic interests can appear anywhere, so Thomas reads a vast range of available texts

---

2. Keith Thomas, "Diary," *London Review of Books* 32.11 (2010): 36–37, 36. His most recent study is *The Ends of Life: Roads to Fulfilment in Early Modern England* (Oxford: Oxford University Press, 2009).

and documents to accumulate notes on slips of paper with various headings. He sorts and shuffles and files the slips away, to assemble them later into arguments.

The Oxford method will be familiar to all scholars who read archival texts and make arguments about them. Humanists, as a category, are consumers and digesters of texts; and those who study past cultures build arguments from archives by sifting through them for evidence. Our methods have precedents in the Renaissance humanists' habits of collecting, disassembling, and commonplacing that Ann Blair has described.[3]

But do they have a future? Thomas's essay closes with a turn to the techniques of digital search and catalogue and retrieval systems, and their implications for his system. In melancholic tones, he claims that computer-aided research is obviating his method:

> I have become something of a dinosaur. Nowadays, researchers don't need to read early printed books laboriously from cover to cover. They have only to type a chosen word into the appropriate database to discover all the references to the topic they are pursuing. I try to console myself with the reflection that they will be less sensitive to the context of what they find and that they will certainly not make the unexpected discoveries which come from serendipity. But the sad truth is that much of what it has taken me a lifetime to build up by painful accumulation can now be achieved by a moderately diligent student in the course of a morning.[4]

Thomas concludes that query tools will expedite scholarly inquiries, making them faster and more reliable. His word choices are important. When "a chosen word" can refer to "the topic they are pursuing," in the fullness of its "context," then our proverbial student will be able to find the same topics as a Fellow of All Souls.

Thomas demurs that the student's lack of expertise, the kind accumulated through a lifetime of reading, may leave him or her ignorant of this context.

---

3. See Ann M. Blair, *Too Much to Know: Managing Scholarly Information before the Modern Age* (New Haven and London: Yale University Press, 2010).

4. Thomas, "Diary," 37. Daniel Shore's forthcoming book, *Cyberformalism: The History of Syntactic Forms in the Digital Archive*, promises to address "how search engines transform literary and philological inquiry," with specific reference to the EEBO-TCP database I address in this article. See Shore, "Extensions of the Book," *The Collation*, 13 March 2014, collation.folger.edu/2014/03/extensions-of-the-book/.

Consider a topic like "personal cleanliness," which Thomas mentions. It's not difficult to imagine it including words that pertain to soap and water, to good and bad odours, probably to attractiveness or morality; but it would take an expert in the history of medicine to know that it is unlikely to include words pertaining to disease, which was not yet associated with hygiene. Thomas's "moderately diligent student" might search in vain for those words, lacking the historical expertise of the scholar. He or she will use the search tools as an instrument of narrow depth, not exhaustive breadth. That is what distinguishes experts from novices: the qualitative knowledge of a field that provokes quantitative inquiry and argument, the judgment that no search engine will provide.

Words are not topics, because quantities are not qualities. That is to say, a search engine can retrieve every instance of a given word, along with its immediate context. But it is little more than an electronic concordance; it cannot tell the scholar, novice or expert, what the word's qualitative, topical meaning is. For that, we need to combine human expertise with a more robust computational process like topic modelling. This process "identifies words that tend to co-occur together in multiple places in multiple documents," writes Matthew Jockers.[5] A topic is "a collection of words that have different probabilities of appearance in passages discussing the topic," in Ted Underwood's definition.[6]

Topic modelling divides texts into these collections of words that tend to appear together. If the word "bed" appears near other words like "chamber" or "soft," it has a different topic (domestic furniture) than the word "bed" that appears near "violets" or "ground-plot" (a tract of land; also figurative for the source of a scheme or plan). The experienced reader, familiar with phrases like "a bed of roses" (Marlowe) or "an imaginative ground-plot" (Sidney), will recognize the latter topic from this cluster of words.

That is to say, the Oxford method's future will be augmented, not replaced, by computational processes. My purpose is to argue that Thomas is right to suggest that the future methodologies of humanist inquiry will be disrupted by computational processes. But those processes will have human expertise and judgment—not topic modelling algorithms—at their centre. Their methodologies will start with what Jockers and others call unsupervised pattern

5. Matthew Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana: University of Illinois Press, 2013), 124.

6. Ted Underwood, "Topic Modeling Made Just Simple Enough," *The Stone and the Shell*, 27 April 2012, tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/.

recognition, rather than with Thomas's preconceived "topic" of a search. To explain, computers begin their algorithms in ignorance, so to speak, of the topics they will find. They follow human instructions ("compare the proximity of every word in a text to every other word") to generate lists of proximate words for human experts to analyze. Topic modelling is unsupervised search, because it "creates topical categories without a priori subject definitions," writes Sharon Block.[7] "With no human input into what constitutes a theme, a motif, a topic, the model collects distributions of co-occurring words and then returns them in a manner that allows us to examine, assess, interpret, and intuit what they all have in common, that is, their shared 'theme.'"[8]

Is this disruption inevitable, or will it remain on the fringes of Renaissance studies? Because the field relies on human interpretations of archival texts, it is subject to a methodological shift that is already underway in the fields of eighteenth- and nineteenth-century studies, particularly due to the computational text-analyses and visualizations of digital humanists like Jockers and Franco Moretti.[9] Scholars in those fields have had longer access to machine-readable transcriptions of archival texts than Renaissance scholars have had. Recent and future expansions of access to transcriptions of texts in Renaissance archives—particularly English printed text archives—will enable more powerful queries as our text-analysis algorithms grow increasingly capable of addressing their linguistic features.

The question remains whether this expansion of access will change our field's future research methods. To address this question, I turn to the last major expansion in this field, the rise of New Historicism three decades ago. Thanks to New Historicism, the archival reach of Renaissance studies over those decades has expanded beyond literary and canonical texts to the fullest possible array of archival texts. But that word, "possible," raises a recurrent problem. It has been possible for scholars to read only a limited number of texts in their lifetimes, so there always remain books and manuscripts outside an ever-widening circle. The reach of New Historicists always exceeds their grasp; their new constellations of archival evidence that reveal social energies still rely on the contingent

7. Sharon Block, "Doing More with Digitization: An Introduction to Topic Modeling of Early American Sources," *Common-place: The Interactive Journal of Early American Life* 6.2: n.p.; cited in Jockers, 124.

8. Jockers, 123.

9. Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso, 2007); Franco Moretti, *Distant Reading* (London: Verso, 2013).

discoveries they have made with limited time and contingent access. This problem endures today, let it be said: of the seventy thousand English books printed between 1473 and 1700 a little over fifty thousand have been transcribed, and the proportion of transcribed manuscripts is far lower. New technologies of archival access bring us closer to realizing the ambitions of past generations of historians and critics, but our arguments will never rest on the irrefutable grounds of comprehensive archival evidence.

There are interpretive gains and losses in topic modelling, which like other computational processes must reduce multivalent words to linguistic units in order to view them synoptically. So while we enjoy expanded access to growing collections of words, our machines also must treat those words with narrower criteria, such as parts of speech or frequency of recurrence. I conclude that those quantifiable features of a text do not limit critics to quantifiable (statistical) interpretations, but inform qualitative (nuanced) interpretations of features like generic qualities and the micro-rhetorical acts that constitute genre, in the recent work of Jonathan Hope and Michael Witmore.[10] Criticism, like historiography, is augmented by our tools' ability to gather lists of textual evidence for us, based on our criteria. Then begins the real work, of restoring depth to their flattened surfaces.

The Oxford method's textual sifting and cataloguing exemplifies the scholar's instrumental preconceptions about archival texts. It relies on the practised human expert to recognize significance where it lies, to have preconceived ideas about which qualities (Thomas's "topics") in an archival text are worth transcribing onto a slip of paper. Those transcriptions are always instrumental, gathered for use in future arguments—even if those arguments are presently unknown. The aim is to integrate texts into new systems of thought, to temper preconceived ideas with new discoveries.

My emphasis on "newness" is not an argument for novelty, for inventive departures from the archival evidence—but rather for the opposite, for plausible formulations and explanations of that evidence. Plausibility is the first measure of the value of a new discursive framework: Is it acceptable to the community of scholars? The second measure is more active: Is it capable of

---

10. J. Hope, and M. Witmore, "The Hundredth Psalm to the Tune of 'Green Sleeves': Digital Approaches to Shakespeare's Language of Genre," *Shakespeare Quarterly* 61.3 (2010): 357–90.

provoking new debates and new frameworks? In other words, our proposed exegeses aspire only to an incremental authority, an ability to provoke and facilitate new exegeses.

Humanist disciplines that focus on archival texts, and literary criticism in particular, are not an empirical enterprise. These disciplines are not testing hypotheses by questioning whether they fit the "facts" of a text, but testing their ability to provoke new hypotheses. Literary criticism, for one, "operates within a hermeneutical framework in which the specifically scientific meaning of fact, metric, verification and evidence simply do not apply," writes Stephen Ramsay. "'Verification' occurs in a social community of scholars whose agreement or disagreement is almost never put forth without qualification."[11]

These categories do not apply to archival humanism because facts (for instance) obtain everywhere, not just in the local environment of a particular text or set of texts. Thomas gathers quotations on a topic to build a plausible, persuasive exegesis about the phenomenon he describes ("personal cleanliness") in the culture he studies (early modern England). At a given moment he stops to write up his results, to present us with a compellingly cross-referenced pattern of evidence. The basis of the historian's arguments is his descriptions and citations of texts' features and excerpts, in relation to other texts' features and excerpts. Similarly, Ramsay writes, "[t]he [literary] critic who endeavors to put forth a 'reading' puts forth not the text, but a new text in which the data [its language] has been paraphrased, elaborated, selected, truncated, and transduced."[12] The aim of the historian or the critic who makes arguments about texts is to suggest that the pattern extends beyond his or her circle of evidence; that his or her choice of evidence is not arbitrary but exemplary.

The assumption that our claims about a chosen cluster of evidence are extensible to other texts is necessary, given the limits of human expertise and lifespans. It is also productive: it drives the incremental progress of our understanding of historical cultures through their textual remains. But my premise, like Ramsay's, is that archival humanists should adopt a more robust sense of progress than human readings afford. By augmenting human readings with computational processes like topic modelling, our arguments will have three

---

11. Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Urbana: University of Illinois Press, 2011), 7.

12. Ramsay, 16.

advantages over human readings alone: we will treat (1) more evidence (2) more instrumentally (3) with fewer preconceptions.

Human readings are implicitly exemplary, making arguments that are extensible from some texts to a wider circle of texts. They are therefore provisional in two senses: they provide the basis on which other scholars will build new arguments; and they are subject to others' correction and revision. Consider how many arguments about archival subjects begin by positioning themselves in dependent or corrective relationships to other arguments. They claim that those arguments have done influential work, but "have failed to confront the complementary evidence"[13] or are, worse, ignorant of conflicting evidence.[14]

This corrective or, more generously, cumulative process gives archival work a sense of progress, as measured by the novelty or complementarity of new discoveries relative to older ones. When scholars make archival discoveries and then make new arguments about them, the novelty of those arguments owes to the novelty of the discovery—not of the artifacts themselves, which are obscure but hardly new. Newly discovered artifacts can threaten settled orthodoxies, like the canon of texts attributed to Shakespeare. To simplify, this sense of progress falsely presumes that everything to be known about the old texts is already known, and novelty depends only on obscurity.

This anxiety that we will miss some crucial piece of evidence is the rationale behind various projects to expand our archival reach. We use different terms for these projects, depending on our training. If we are historians like Thomas, we might be retrospective ethnographers. If anthropologists, we might trace our origins to Clifford Geertz's methods of "thick description."[15] If literary critics, we might practise "archival poetics," or critical readings of archival materials for their poetic or subtextual qualities.[16]

---

13. Anthony Grafton and Lisa Jardine, *From Humanism to the Humanities: Education and the Liberal Arts in Fifteenth- and Sixteenth-Century Europe* (London: Duckworth, 1986), xii.

14. In *"Counterfeiting" Shakespeare: Evidence, Authorship, and John Ford's Funerall Elegye* (Cambridge: Cambridge University Press, 2002), Brian Vickers dismantles Don Foster's and others' evidence for attributing two poems to William Shakespeare. In a book of 16,895 unique words, "evidence" appears 262 times, starting of course with the title. Variations on the words "ignorant" or "ignore" appear sixty-two times, mostly directed at Foster (although "ignorantly" appears in one of the contested poems).

15. Clifford Geertz, *The Interpretation of Cultures* (New York: Basic Books, 1973).

16. Cyndia Susan Clegg, "Archival Poetics and the Politics of Literature: Essex and Hayward Revisited," *Studies in the Literary Imagination* 32 (1999): 115–32. No doubt the reader can add other schools of

I submit that a more productive enterprise would preserve this attention to human expertise, but would also urge scholars toward a more agnostic openness to what a collection of texts can reveal through computational processes. The trouble with human critics—and I speak from experience—is our reliance on interpretive heuristics, given our limited time and attention. We also read, and often think, in linear terms: from act 1 to act 5; from cradle to grave; from prompt to heuristic. So Jonathan Hope and Michael Witmore use text-analysis algorithms on Shakespeare's plays because they are "productively indifferent to linear reading and the powerful directionality of human attention."[17]

Computational methods are unbound by our conventional habits of interpretation; they take it in unexpected directions, because they have no expectations to begin with. They also suggest that novelty can arise from new techniques of analysis, rather than (only) new objects for analysis. As Witmore claims below, they can show us what was always there in the old texts, only hidden from view. But before I conclude with these methods, I will address their preconditions: the expansion of our dataset through machine-readable transcriptions of Renaissance texts, particularly in English; and the features and motives this expansion shares with the last time it happened in Renaissance studies, thirty years ago.

New Historicists have dominated Renaissance studies since at least 1986, when Arthur Kinney introduced a special issue of *English Literary Renaissance* on "cultural poetics"—his preferred term—with these prescient words: "[M]any recent thinkers […] see literary texts as one of many signs, signs of no greater or lesser importance than political events, social customs, or public gestures. Thus while they keep both history and literature in play, they deny the primacy of literature (and the isolation of it)."[18] The critics Kinney introduced became household names, at least if your household had graduate students: Stephen Greenblatt, Jean Howard, Louis Montrose. They resisted the old-historicist division of contextual and textual, setting obscure texts alongside better-known ones. Even if they reinforced older ideas of canonical literature—using unknown texts as entry-points to better-known texts—they still introduced

---

archival practice to this brief list.

17. Hope and Witmore, 359.

18. Arthur Kinney, "Preface," *English Literary Renaissance* 16.1 (1986): 3.

readers to lesser-known authors like Thomas Churchyard or Barnaby Googe, sitting alongside Edmund Spenser and Philip Sidney. Similarly, New Historicists dismantled the hierarchical divisions between literary and extra-literary texts, and between familiar and unfamiliar authors; but they did not quite equalize their mutual influence and set them in mutual conversation.

In the same journal as Kinney's essay, Jean Howard praised the New Historicists' flattening of social and literary texts, but challenged her colleagues to "be more overtly self-conscious of [their] methods and [their] theoretical assumptions." She charged them to expose these assumptions and write "polemical" essays:

> The best criticism performs two tasks at once: the practical business of reading another text and the critical business of explaining the terms of that reading. […] Essays which explain how and why one does and should read in a particular way are both more generous and more risky since they […] expose what is difficult and what is at stake in "making knowledge" at this historical moment.[19]

My own purpose is not to dissect New Historicist habits of thought (a kind of meta-criticism I have indulged in elsewhere[20]) but to expose the assumptions and parameters of Renaissance studies on the verge of its next great expansion of archival data, and its potential integration of the algorithms that can process that data.

The archival scope of English Renaissance studies, in which many of the New Historicists originated, has expanded considerably since 1986. Many archives have become databases of language-based inquiry, as more printed books gain more accurate digital surrogates. In English, some fifty thousand of the seventy thousand books printed between 1473 and 1700, now in the Early English Books Online (EEBO) database, have been painstakingly transcribed by human readers into machine-readable (read: searchable) text files by the Text Creation Partnership (TCP). The Early Modern OCR Project (eMOP) is actively working to accelerate these transcriptions using Optical Character

19. Jean Howard, "The New Historicism in Renaissance Studies," *English Literary Renaissance* 16.1 (1986): 42, 31.

20. Michael Ullyot, "The Rhetoric of Anecdotes in New Historicism," *Clio: A Journal of Literature, History, and the Philosophy of History* 40.3 (2011): 307–29.

Recognition, which (simply put) enables machines to recognize individual letters: even those long s's, even in old microfilms. While there are promising signs that projects like Early Modern Manuscripts Online will integrate transcriptions of non-printed material to this dataset, most of the expansion of the past decades has been in printed books.

The corpus that the TCP is publicly releasing over time (2015–2020) is a considerable trove of data. Martin Mueller's term for it is the "Book of English."[21] The TCP's transcriptions of these texts are imperfect, but improving; in 2010, Mueller began supervising a collaborative curation of these texts using a tool called AnnoLex, offering "a social and technical space of collaborative curation, where individual texts live as curatable objects continually subject to correction, refinement, or enrichment by many hands, and coexisting at different levels of (im)perfection."[22]

In 2013, the same project corrected some six hundred plays printed before 1660. Its immediate goal is to produce texts that reliably and thoroughly represent the contents of their original sources. But its more distant goal is what Mueller calls "the *algorithmic amenability* of the digital surrogate, its capacity for being variously divided or manipulated, combined with other texts for the purposes of cross-corpus analyses, having data derivatives extracted from, or levels or metadata added to it."[23] By "surrogate" Mueller means the machine-readable transcriptions of archival documents, the source for these derivatives and base for these superstructures. Hence the importance of making it reliable and trustworthy: like producing an edition, it is the work that makes all other work possible.

Does the digital surrogate not also circumscribe and overdetermine the work that can derive from it? If its goal is algorithmic amenability, does that mean it can only provide the evidence for linguistic arguments? Algorithmic

21. Martin Mueller, "Towards a Book of English: A Linguistically Annotated Corpus of the EEBO-TCP Texts," presented at *"Revolutionizing Early Modern Studies"? The Early English Books Online Text Creation Partnership in 2012*, Conference: EEBO-TCP 2012, University of Oxford, 17–18 September 2012, http://j.mp/eebo-tcp2012.

22. Martin Mueller, "EEBO-TCP 2012: The Future of the TCP as a Public Domain and Collaboratively Curated Corpus of Early Modern English," *Scalable Reading*, 26 September 2012, http://scalablereading. northwestern.edu/2012/09/26/eebo-tcp-2012-the-future-of-the-tcp-as-a-public-domain-and-collaboratively-curated-corpus-of-early-modern-english/. For AnnoLex, see http://annolex.at.northwestern.edu.

23. Martin Mueller, "The Great Digital Migration," *Scalable Reading*, 13 August 2012, http://scalablereading.northwestern.edu/2012/08/13/the-great-digital-migration/. (My emphasis.)

text-analysis tools enable a new kind of criticism, one that relies on the linguistic qualities of texts that Renaissance literary critics have tended to ignore in the past three decades. This criticism may extend the project of New Historicism by expanding the range of archival texts under scrutiny, but that scrutiny will necessarily be of their linguistic qualities.

Literary critics tend to consider linguistic features only as a means to higher-order interpretation. Lisa Samuels and Jerome McGann call these features "the operating system of language," adding that literary critics tend to treat linguistics as "pre-interpretive and pre-critical."[24] Yet they also enable a new kind of literary criticism, what Samuels and McGann call deformative criticism—which disrupts our usual progress through a text, revealing relationships between words that may be widely dispersed.

Algorithmic criticism, in Ramsay's more recent formulation, bridges the gap between linguistic features and literary criticism. Ramsay's project is to move text analysis from the explicit limits of "low-level linguistic phenomena" that computational linguists study, capable only of identifying "a relatively narrow range of propositions," toward a seed-bed of propositions, provocations, debates, and discussions: the standard measure of a literary argument's value.[25]

So Willard McCarty has studied personification in Ovid's *Metamorphoses* by following Morton W. Bloomfield's reframing of the trope's location: "the momentary, ontological dimension of the trope *as a linguistic event* rather than the temporal, narratological one *as a technique of story-telling*." Bloomfield wrote that "personification should be studied at the microscopic level, 'grammatically,' as a phenomenon brought about by discernible operations of language."[26]

Linguistic features are what Michael Witmore calls the "addressable" components of a text; it means that "one can query a position within the text at a certain level of abstraction": genre, individual lines of print, parts of speech.[27] These categories are familiar enough, like parts of speech—but when you regularize them across large text corpora, and write algorithms that can recognize and isolate them, you can begin saying definitive things about (say) the average

24. L. Samuels and J. J. McGann, "Deformance and Interpretation," *New Literary History* 30.1 (1999): 25–56; 35.

25. Ramsay, 8–9.

26. Willard McCarty, *Humanities Computing* (Houndmills: Palgrave Macmillan, 2014), 55, 54.

27. Michael Witmore, "Text: A Massively Addressable Object," *Wine Dark Sea*, 31 December 2010, http://winedarksea.org/?p=926.

number of speakers in every scene by Middleton, or the distribution of proper nouns in historical drama, or the total number of lines of verse in every text printed in 1589.

Statistical text-analysis is simply an automated mode of address: recognizing what the philosopher Wientin Meillassoux, whom Witmore quotes, calls "those aspects of the object that can be formulated in mathematical terms." These mathematical aspects, Witmore continues, "can be meaningfully conceived as properties of the object in itself." Pairing massive addressability with future statistical procedures, Witmore concludes that "something that is arguably true now about a collection of texts can only be known in the future."[28] The key word is "known": what is true can only be demonstrated through the arguments of critics, wielding the right subsets of data ("aspects of the object") that make a compelling case for given "properties of the object."

For instance, Hope and Witmore use a program called DocuScope to run the statistical procedure of principal component analysis (PCA) to identify 101 language action types (or LATs) in Shakespeare's plays: uncertainty, disclosure, fear, sadness, reassurance, confrontation, question, denial, aside, and so on. (Mueller describes DocuScope as "a very large dictionary of short phrases or grammatical patterns that are mapped to a taxonomy of about 100 microrhetorical acts.")[29] These 101 LATs identify both significant and consistent differences between Shakespeare's comedies and histories. This technique makes "genre visible on the level of the sentence," and leads them to conclude that *Othello* is—linguistically, at least—more comic than tragic.[30]

What gives a text like *Othello* its genre? Is it the narrative that unfolds, and its resemblance to other narratives we assign with the labels "tragedy" or "comedy"? Resemblance is essential to categorization. But resemblance to what? Hope and Witmore offer quantitative evidence that these 101 linguistic categories add at least another dimension to a text beyond its narrative.

28. Michael Witmore, "The Ancestral Text," *Wine Dark Sea*, 9 May 2011, http://winedarksea.org/?p=979.

29. Mueller, "Future of the TCP." These 101 LATs are in seventeen clusters comprising fifty-one dimensions. So (in Docuscope's documentation) "the First Person [Options] cluster conveys the perspective of a unique entity looking out on the world from the inside." Its four LATs are First Person, Self-Disclosure, Self-Reluctance, and Autobiography; it includes words and phrases like "I am," "my," "me," "I'll," and "myself."

30. Hope and Witmore, 368n18.

"[A]lgorithmic transformation can provide the alternative visions that give rise to such readings," writes Ramsay: "[T]he narrowing constraints of computational logic—the irreducible tendency of the computer toward enumeration, measurement, and verification—is fully compatible with the goals of criticism […] because critical reading practices already contain elements of the algorithmic."[31] Those elements are the critic's transformation of a text into a "reading" of that text, based on his or her preconceived topics or themes. Using a machine to gather every instance of the text features and patterns that a critic uses to read a text "is the same thing at a different scale and with expanded powers of observation," he adds.[32]

The distinction is, I suggest, somewhat stronger than a matter of scale. As I have argued, the machine is a powerful augmentation of the human critic's anterior judgment because of its indifference to our habitual classifications, our authorial attributions and hierarchical rankings of genres and forms, our intentions, preconceptions, and value judgments. This indifference, combined with its singular purpose of following our instructions, makes it the most promising means to realize the New Historicist dream of an extensible archive.

Algorithms are extensions, as Ramsay suggests, of the practices of literary criticism. They originate with critics, who notice something remarkable about a text, who want to recognize and name that effect. Once we identify the formal features, they are transferrable to other instances in other texts. If the features are quantifiable, and if the other texts are machine-readable, then the algorithm simply automates their detection.

This argument began with Thomas's use of the term "topic," and took up "topic modelling" as the ground-plot for a broader argument about algorithmic text-analyses of machine-readable archival texts. It is fitting, then, to conclude by examining the idea of a "model." It has a long history in the observational and computational sciences, but will (I think) have a long future in the archival humanities. McCarty has done more than anyone else to theorize the power and limitations of models, particularly in a chapter devoted to this topic in his seminal *Humanities Computing* (2005). He reminds us of "the fundamental dependence of any computing system on an explicit, delimited conception of

31. Ramsay, 17, 16.

32. Ramsay, 17.

the world or 'model' of it."[33] A model, he writes, is "an attempt to capture the dynamic, experiential aspects of a phenomenon" that are always provisional, always "*temporary states in a process of coming to know* rather than fixed structures of knowledge."[34]

For this reason, models are not conclusive in themselves; they are a means to satisfying conclusions. A topic model of an archival text only attempts to represent that topic through linguistic categories. The topics that algorithms are capable of isolating are from a dataset of words that can be numbered, listed, visualized, and otherwise rearranged or "deformed" by quantitative criteria: linguistic categories, frequencies, relationships/contexts, tenses, speakers, and so on.[35]

But models also extend beyond singular texts, allowing Mueller's "cross-corpus analyses." So Jockers argues that topic models allow scholars to widen their scope from a more restrictive examination of seminal texts to a study of the "aggregated ecosystem" or "economy of texts."[36] That is, they make us consider texts within the landscape of other, contemporary texts, and identify what topics they have in common. McCarty's metaphor for all computational algorithms is "telescopes for the mind," because they "extend our perceptual scope and reach […] [and] transform our conception of the human world just as in the seventeenth century the optical telescope set in motion a fundamental rethink of our relation to the physical one."[37]

To Thomas's concern about searches leading to topics, McCarty would reply that computers are "*modelling machines, not knowledge jukeboxes.*"[38]

And yet models, whether human or computational, are necessarily simplifications of their phenomena, like maps of a territory. To come full circle, let's return to Thomas in the *London Review of Books*—this time in Eamon Duffy's 2009 review of his synoptic study *The Ends of Life: Roads to Fulfilment in Early Modern England*. Duffy puts his finger on a problem with the Oxford method,

---

33. McCarty, *Humanities Computing*, 21.

34. McCarty, *Humanities Computing*, 23, 27, emphasis in original.

35. Samuels and McGann, "Deformance and Interpretation."

36. Jockers, 32.

37. Willard McCarty, "A Telescope for the Mind?" in *Debates in the Digital Humanities*, ed. Matthew Gold (Minneapolis, MN: University of Minnesota Press, 2012), 113.

38. McCarty, *Humanities Computing*, 27, emphasis in original.

whose wide-ranging quotations are "all presented as equivalent testimony to the mindset of an age," leaving it vulnerable "to the charge that it falsifies the material selected by decontextualising it."[39]

Thomas's essay the following year concedes as much, and asserts that he exercises his judgment about writers' and speakers' motives and contexts. Our scraps of paper need to emerge from deep immersions in archival materials, before we make them extensible to database reports and output files. Otherwise those alluring surfaces will be like our knowledge, forever superficial. Like Thomas, I see new technologies of access and manipulation changing the way archival scholars are engaging and will engage with old texts. But I put those technologies in a more tenuous position, proving their worth to human readers when they augment our judgment.

39. Eamon Duffy, "Common Thoughts," *London Review of Books* 31.14 (2009): 18.