Text Minding: "A Response to Gender, Race, and Nationality in Black Drama, 1850-2000: Mining Differences in Language Use in Authors and their Characters"

Sean Ross Meehan <smeehan2_at_washcoll_dot_edu>, Washington College, Chesterton, MD

Abstract

A response to the Data Mining cluster, exploring the role of machine learning in textual study.

Head of the first division

Should he find his way to this special issue of *DHQ*, Sven Birkerts, author of *The Gutenberg Elegies*, would likely add machine learning to his cataract of digital distractions besetting literary reading in the electronic age. He would agree with the authors of "Gender, Race, and Nationality in Black Drama, 1850-2000," I presume, only when they conclude their discussion of the performance of algorithms in generating feature lists that successfully distinguish American and non-American playwrights in the Black Drama database, and anticipate the objection, "that the task itself is a trivial case, attempting to confirm a distinction that is all too obvious to be of significant literary interest." The depth of literary meaning, Birkerts would assert, cannot be mined by machines trained for data. I don't share Birkerts' vision that literary reading and new media technologies are mutually exclusive. He comes to mind upon reading these experiments in text mining, rather, as a way to illuminate what I think the significant literary interest of machine learning can and should be for humanists and literary critics. That interest, to put it simply before elaborating a bit further in this response, is a renewed and more robust understanding of the textuality, the principal object (if not subject) of our studies, that many critics (and print-centric critics like Birkerts among them) too frequently take for granted. [1]

The basis for such renewed critical potential is evident when the authors frame their experiments in machine learning as a complement, rather than supplement, to traditional text analysis. In summarizing the performance of text mining with regard to gender classification of characters within the drama database, the authors emphasize this complementarity in asserting, "The degree to which these lists reveal true differences among black American male and female authors is a matter for discussion. The important thing is that the mining algorithm gives fuel to the discussion and serves as a starting point for closer textual study." Machine learning, on this view, is a tool for closer textual study, a staple of literary criticism; this tool is particularly rich in its potential to identify meaningful patterns across increasingly greater amounts (and distances) of texts. The authors are justifiably circumspect in worrying about the ways in which such a computational tool may distort the meanings of literary texts or offer an analysis at too great a distance from such texts. Most interesting, to me, is the worry that the binary logic necessary for computation may reveal binary thinking in the texts, but with the result of unduly privileging the stereotypical force of such patterns. [2]

I offer to this concern a three-fold response. First, we should worry such things; but this concern, of course, is not limited to digital tools of analysis. As Jerome McGann reminds literary critics in *Radiant Textuality*, all of our critical tools of analysis are and always have been "prostheses for acting at a distance," the same "distance that makes reflection possible" [McGann 2001, 103]. Second, in view of this understanding of interpretation as a dynamic between tools and texts, always oscillating between method and meaning, I would suggest that the self-awareness of this critical problem made evident in "Mining Differences" is not just a credit to the article; this greater attention to the mediations of critical method is enhanced by the mediation of the machine. The point is that the texts we study, whatever the substrate, are [3]

already mediating technologies, compilations of linguistic and printing and picture-making machines. Textual technologies such as the machine learning and text mining software at issue here, can and should, if used thoughtfully, offer insight into the technologies of the texts we study. Katherine Hayles, emphasizing the need for literary critics to pursue "media specific analysis," argues that digital media give us the opportunity "to see print with new eyes" [Hayles 2002, 33]. Birkerts could certainly take note; despite his title and his defense of the book, there is very little Gutenberg in his de-materialized conception of reading.

This closer understanding of text technology, it seems to me, might begin to address the concern for binary opposition raised by the authors though not interrogated further in the article. That limitation of the machine may well be the source of literary insight. In his *Figures in Black*, the African American literary theorist and scholar Henry Louis Gates, Jr. offers a compelling analysis of how Frederick Douglass employs binary opposition in his first slave narrative as a complex rhetorical device aimed at undermining, by way of revealing, the binarity of language that slavery employs in its ontology of slaves versus men. The critical insight is that the arbitrary and binary mechanics of language as such can be appropriated and used differently by the author who knows how to operate and deconstruct the literacy machine. Douglass thus mines the language of slavery and race in his narratives in order to reveal a mind within the machine. The lesson is that how we read and write does indeed determine what we read. For Douglass, the interaction of medium and meaning drives the revisionary potential of his narrative; we see the black author in print with new eyes. I see a similar insight into the relation between ontology and textual technology suggested in "Mining Eighteenth Century Ontologies," where the knowledge discovery of machine learning locates its Enlightenment type in the knowledge discovery, ancestral hypertext of the *Encyclopedie*. We shouldn't be surprised, I suspect, to find similar insights regarding the ontologies of race and gender in the texts of black drama — a genre, after all, whose textuality is inherently and complexly multimedia. Surprised or not, the suggestive patterning of textuality that machine learning can reveal, combined with the critical self-awareness that the humanities seeks, represents a significant means for this critical discussion to move forward.

4

Works Cited

Birkerts 1994 Birkerts, Sven. *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. New York: Fawcett, 1994.

Gates 1987 Gates Jr., Henry Louis. *Figures in Black: Words, Signs, and the 'Racial' Self*. New York: Oxford University Press, 1987.

Hayles 2002 Hayles, N. Katherine. *Writing Machines*. Cambridge, MA: MIT Press, 2002.

McGann 2001 McGann, Jerome. *Radiant Textuality: Literature After the World Wide Web*. New York: Palgrave Macmillan, 2001.