

Ryan Cordell*

Speculative Bibliography

<https://doi.org/10.1515/ang-2020-0041>

Abstract: This article proposes *speculative bibliography* as an experimental approach to the digitised archive, in which textual associations are constituted propositionally, iteratively, and (sometimes) temporarily, as the result of probabilistic computational models. Speculative bibliography is offered as a complement to digital scholarly editing, and as a direct response to the challenges of scale and labour that will make comprehensive editing of digital archives impossible. Rather than acting on specific, individual texts, a speculative bibliography enacts a scholarly theory of the text through a computational model, reorganising the archive to evidence a particular idea of textual relation or interaction. Such models, in which textual relationships are determined by formal, internal textual structures, constitute bibliographic arguments that can be verified, amended, extended, or contested on either humanistic or computational grounds.

Recurated and Reedited

Jerome McGann begins his 2014 book, *A New Republic of Letters*, channelling an imperative he first articulated at least thirteen years earlier: “Here is surely a truth now universally acknowledged: that the whole of our cultural inheritance has to be recurated and reedited in digital forms and institutional structures” (2014: 1). McGann despairs that current digital archives are, by and large, being created by for-profit entities, using errorful automated processes such as optical character recognition (OCR), and by workers trained in the technical aspects of digitisation, but not in the historical and cultural dimensions of the artefacts being digitised. What is required, McGann argues, is a sustained, collective scholarly intervention:

Digitizing the archive is not about replacing it. It’s about making it usable for the present and the future. To do that we have to understand, as best we can, how it functioned – how it made meanings – in the past. A major task lying before us – its technical difficulties are great – is to design a knowledge and information network that integrates, as seamlessly as possible, our paper-based inheritance with the emerging archive of born-digital materials. (2014: 21–22)

Corresponding author: Ryan Cordell, Northeastern University
E-Mail: r.cordell@northeastern.edu

What does it mean to make the archive “usable for the present and the future”, and how do notions of usability change due to digitisation? The massive scale and cost of the effort McGann advocates seem impossible in the current age of educational austerity – which is particularly felt in the humanities – and yet the imperative to design multiple paths into both “our paper-based” and “born digital” inheritances seems clear.

One common trope when writing about digital archives focuses on their scale: hence phrases such as ‘big data’ or ‘distant reading’. Because digital collections often aggregate material that was once separated in different libraries and archives, they are perceived as more comprehensive than their analogue forebears, and scholars imagine new questions that might be asked not of a collection of books a human might be able to read, but of entire corpora that a computer can process quickly. Such excitement is not misguided, as work in digital humanities, cultural analytics, and related fields over the past decade has shown through analyses that would have been nearly impossible prior to digitisation. Our work on the *Viral Texts* project at Northeastern University, for instance, uses text mining techniques to trace the reprinting of material in nineteenth-century newspapers across the U.S. and, more recently, across the globe.¹ This kind of pattern matching – identifying overlapping, matching strings of characters across tens of millions of newspaper papers – would be far beyond the capabilities of a human researcher or the operational capacities of an individual archive. The project relies on digital archives that federate newspapers physically distributed around the world, which means an analogue effort along these same lines would require more time than any researcher possesses and enormous geographic mobility.

While such research shows the potential for computational humanities research, we should not confuse the existing digitised archive with ‘the archive’ in an abstract sense. Even the most massive digitised historical archives comprise only a fraction of the physical archives from which they are derived.² The Stanford Literary Lab’s eleventh pamphlet describes this reality with a metaphor of nesting dolls: “The corpus is thus smaller than the archive, which is smaller than the published: like three Russian dolls, fitting neatly into one another” (Algee-Hewitt

¹ For more on the *Viral Texts* project, including links to our publications, see <https://viraltexts.org/>. To learn about our more recent international work, see the *Oceanic Exchanges* project at <https://oceanicexchanges.org/>.

² I am careful here to refer to the “digitised archive” rather than the “digital archive” to keep my focus squarely on the mass of historical materials that have been remediated into digital forms. The digital archive is much larger, including a wide variety of born-digital materials such as websites, computer programs, video games, and the like. This latter archive presents a host of additional challenges to those I outline in this introduction, which I touch on later in the article.

et al. 2016: 2). The Literary Lab team sees a profound change underway in the digital era, when “the relationship between the three layers has changed: The corpus of a project can now easily be (almost) as large as the archive, while the archive is itself becoming – at least for modern times – (almost) as large as all of published literature” (Algee-Hewitt et al. 2016: 2). This may be true of some digital archives, if we slice a given domain finely enough, but this ideal of corpus-archiving convergence is not true for many media, periods, or genres.

As of January 2020, for example, The Library of Congress’s *Chronicling America* collection of digitised historical newspapers includes nearly 16 million pages published between 1789–1963. Those 16 million pages, however, represent only a small fraction of the newspapers published in those years. For context, *Chronicling America* publishes fewer than ten titles published in New York City out of the hundreds or even thousands of newspaper titles that appeared in that city over the timeline of the archive’s holdings. I write this not to chide the Library of Congress – time and money for digitisation are likewise scarce, and choices must be made – but to dispel the notion that scale is a particular challenge of the digital age. Our analogue archive remains more extensive than our digitised historical archive, though it is not necessarily addressable in the same ways. Scale, then, is a problem – perhaps *the* problem – of the physical and digital archive alike, and not one simply solved by digitisation and computational research methods.

This fact brings us back to McGann’s charge and the sobering realisation that the vast majority of the analogue archive was never curated or edited in the first place, perhaps rendering the mandate to recurate or reedit premature. The promises of digitisation first emerged even as the field of literary study was undergoing a (relatively) new process of canon reformation and archival reclamation. As a result, there have never been certainties about what materials merit careful remediation into digital form. Moreover, there is neither the time nor the labour that would be required to digitise everything to the standards scholars such as McGann might wish, particularly in an age of austerity toward humanistic research. In such an environment, digital scholarly editing remains an essential endeavour, but it must proceed in parallel with other experiments that meet the mass digitised archive where it is, leverage the unique affordances of digital media to explore its contours, and identify areas of scholarly interest from a well of largely undifferentiated content.

Toward Speculative Bibliography

This article proposes *speculative bibliography* as a complementary, experimental approach to the digitised archive, in which textual associations are constituted

propositionally, iteratively, and (sometimes) temporarily, as the result of probabilistic computational models. A speculative bibliography enacts a scholarly theory of the text, reorganising the archive to model a particular idea of textual relation or interaction. Many computational processes already create data we might identify as speculative bibliographies: algorithms that detect the relative prevalence of “topics” across documents, identify sequences of duplicate text in a corpus, or more simply list texts that share search terms.

Our work on the *Viral Texts* project at Northeastern University, for instance, uses text mining techniques to trace the reprinting of material in nineteenth-century newspapers across the globe.³ To simplify things just a bit, our method posits that the editorial practice of reprinting can be modelled in this way: “If passages of text from separate pages contain at least five matching phrases of five-words length and their words overlap each other by at least 80 %, they should be considered ‘the same’ and clustered together”. Essentially, we have developed a textual model that is agnostic on questions of author, title, genre, or similar categories that are largely absent from either the nineteenth-century newspaper page or the metadata of twenty-first-century digitised newspaper archives. To write that another way, it is a method that does, I believe, account how the nineteenth-century newspaper “made its meanings”, as well as the ways that current digital newspaper archives mask those meanings. Speculative bibliography recognises that a unique affordance of digital media is the ability to rapidly reorganise or reconfigure its contents and seeks to identify meaningful patterns for exploration within collections that are often messy and unevenly described. In the *Viral Texts* model, textual relationships are determined by the formal structures internal to the texts themselves, but our algorithm is nonetheless a bibliographic argument.

Why “Bibliography”?

To oversimplify just a bit, we might describe bibliography as a system for modelling textual relationships. The bibliographer decides that these texts belong together because they share certain metadata (e.g. author, genre, era of publication) while these others might belong together because they share formal material features (e.g. octavo format, dos-à-dos binding). In many bibliographic traditions,

³ The fullest exposition of our methods can be found in the book chapter “Textual Criticism as Language Modeling”, published in draft on the University of Minnesota Press’s Manifold Scholarship platform.

these relationships are mapped out quite methodically and procedurally – dare I write algorithmically? –, which is perhaps why bibliography can seem “dry as dust” to outside observers. However, bibliographers share a conviction Jonathan Senchyne summarises beautifully in his new book that “[m]aterial textuality means [...] the material presence of something is itself figurative and demands close reading” alongside a text’s linguistic content (2019: 6).

The constellation I want to gather under the sign of speculative bibliography comprises computational and probabilistic methods that map relationships among documents, that sort and organise the digital archive into related sets. I employ *bibliography* to insist that such methods belong to the textual systems they transform, and should themselves be objects of research and scrutiny, described with the same rigor with which bibliographers and book historians describe historical technologies of knowledge production.

Earlier in the digital age, bibliographer Thomas Tanselle proposed a definition for what he called an “electronic edition” of a text, as “all copies resulting from a single job of typographical composition”:

[I]n order to include modern methods of book production which do not involve actual type setting, an edition should be defined as all copies resulting from a single job of typographical composition. Thus whether printed from type (set by hand or by machine), or plates, or by means of a photographic or electronic process, all copies that derive from the same initial act of assembling the letterforms belong to the same edition. (1975: 18)

Tanselle was working to expand the intellectual boundaries of bibliography for a publication environment in which texts could proliferate almost instantly and almost infinitely, to think about word processing from a materialist perspective. Where previous bibliographers had focused on composition as the setting of metal type (whether moveable characters of cold type or a line of hot type), Tanselle recognised that typing at a computer keyboard was also an act of inscription, committing a particular arrangement of typographic characters – “assembling the letterforms” – to memory.

In a recent article in *Book History* (Cordell 2017), I pivot from Tanselle to argue that humanities scholars need to take optical character recognition (OCR) seriously as a material and cultural artefact – to grapple more fully with how the technical, social, and political structures through which mass digitised archives are created, and to apprehend the OCR layers that are typically covered by the interfaces of digital archives. OCR software scans digitised page images, attempting to recognise the letterforms on the images and transcribe them into a text file. I argue that we might consider OCR a species of compositor, setting type in a language it can see but not comprehend, and thus that OCR data derived from a historical text is a new edition – a copy “resulting from a *single job of typographical composition*” – of that

text. It is a kind of offset composition, in which the programmer sets the rules for recognition that the program will follow to create many editions.

In naming OCR-derived text as a new edition, I argue against the language of surrogacy for describing digitised historical texts. Thinking of digital objects as surrogates masks the material and cultural circumstances of their creation, obscures the affordances of the electronic media created through digitisation, and can encourage the devaluing (or even deaccessioning) of material holdings. Myths of surrogacy can short-circuit media-specific engagements with digital archives, leading us to *use* digitised materials just as we would their analogue forebears when we could be leveraging their digitality toward new kinds of associations and discoveries.

This article expands the frame of that argument further to consider code – such as that underlying reprint detection or classification – as another job of typographical composition that inscribes a theory of textual relationship, at least when its objects are bibliographic or textual. That last caveat is important, because just as we would not claim all written analysis as bibliographic, we should not claim all code as bibliographic. Annette Vee points to the double valence of code when she writes:

Programming has a complex relationship with writing; it is writing [...] because it is symbols inscribed on a surface and designed to be read. But programming is also *not* writing, or rather, it is something more than writing. The symbols that constitute computer code are designed not only to be read but also to be executed, or *run*, by the computer. (2017: 20)

Vee continues in a line that echoes (though I suspect accidentally) Tanselle, stating that “programming is the authoring of processes to be carried out by the computer” (2017: 20). Denis Tenen argues something similar when he writes: “Unlike figurative description, machine control languages function in the imperative. They do not stand for action; they are action” (2017: 94). A program becomes speculative bibliography when its action operationalises a theory of textual relationship. If OCR is a species of compositor, such algorithms might be species of editor, set loose with one unwavering principle of selection apiece. Speculative bibliography is simultaneously action and documentation.

Consider a classification algorithm. We use these in the *Viral Texts* project to sort millions of individual reprints into generic categories: fiction, news, poetry, science writing, domestic writing, etc.⁴ Classification is typically a supervised

⁴ The fullest exposition of our classification experiments can be found in the book chapter “Classifying Vignettes, Modeling Hybridity”, published in draft on the University of Minnesota Press’s Manifold Scholarship platform.

task, in which researchers tag a set of documents – the training dataset – as belonging to the various genres they hope to identify. From this training data, different classification algorithms can be used to compare unknown texts against known. Some classification algorithms use words to determine belongingness: e.g. domestic fiction will use words like “eye”, “heart”, “mother”, or “tear” in much higher proportion than we would expect from random chance, while news articles will disproportionately use words like “ship”, “president”, “bill”, or “debate”. I am making these lists up, because in reality they depend entirely on a specific research corpus and researchers’ initial genre classifications, but word-based classification works roughly in this way. There are also other classification methods that use topic or vector space models to establish relationships among the words in different texts.

I would argue that these processes are displaced forms of editing that operate with less precision, but with greater speed and scale, than solely human endeavour; they are a kind of offset editing. Editors create models of textual relationship through tagging their training data and then operationalise those models across a wider textual field than they could edit alone. In our case, we spend a few weeks manually tagging several hundred texts per genre in order to classify millions of unknown texts in an afternoon. Importantly, these methods are not binary, but probabilistic across all genres for which we create a training set, so that one text might be classified as 79 % likely to be poetry and 65 % likely to be religious. If we later seek out popular newspaper poetry or religion (or religious poetry), we would find such a text. Of course, with such a method there are many false positives or false negatives: texts that a human reader would recognise as an account of a cricket match, to cite a genuine example, but that the classifier identifies as poetry, or texts that a human observer would recognise as poetry but that a classifier fails to identify as likely to be such.

As scholars such as Richard Jean So and Hoyt Long have demonstrated, even textual groupings that might appear errorful can be formally and interpretively rich. In their case, poems that seemed mistakenly identified by a classifier as haiku pointed them to a subtle, “ontologically distinct” relationship among modernist poems, a diffuse “haiku style” that “appear[s] to saturate a much broader array of poems, adding up to a kind of Orientalist milieu that is related to the haiku style but is also part of something larger”. For Long and So, “the ontology of machine learning proves invaluable despite its relatively impoverished notion of the poetic text” because “[i]t not only extends our capacity to find textual patterns that extend to lesser-known and marginal poets but also to cultural-historical contexts that might otherwise remain beyond our purview” (2016: 265–266). If classification methods are bibliographic because they posit textual relationships and paths through mass digital archives, they are speculative because the paths

they posit are probabilistic, experimental, and iterative, and because they might force fundamental reconsideration of the categories we bring to textual sorting and organisation.

Why “Speculative”?

With speculative bibliography, I seek an intellectual frame that recognises the practical, theoretical, and historiographical potential of exploratory computation without resorting to dehistoricised, idealised notions of ‘big data’, to negotiate a middle ground between strong theories of ‘distant reading’ or ‘cultural analytics’ on the one hand and the “scholarly edition of a literary system” more recently advocated by Katherine Bode. I am fully convinced by Bode’s argument that, “[c]ontrary to prevailing opinion, distant reading and close reading are not opposites. These approaches are united by a common neglect of textual scholarship: the bibliographic and editorial approaches that scholars have long depended on to negotiate the documentary record” (2018: 19). Bode rightly points out that most “data-rich literary history” projects fail to fully delineate “the broader relationship between the literature of the past and the disciplinary infrastructure used to investigate it” (2018: 52). She advocates instead for the “scholarly edition of a literary system” in which

[a] curated dataset replaces the curated text. In the form of bibliographical and textual data, it manifests – demonstrates and, specifically, publishes – the outcome of the sequence of production and reception, including the current moment of interpretation, described in the critical apparatus. The model it provides is stable: it is published and accessible for all to use, whether for conventional or computational literary history. But that stability does not belie or extinguish its hypothetical character. Rather than showing a literary system, it presents an argument about the existence of literary works in the past based on the editor’s interpretation of the multiple transactions by which documentary evidence of the past is transmitted for the present. Its suitability and reliability for literary-historical research is established by a relationship between the historical phenomena and the data model that is explicitly interpretive and contingent rather than supposedly direct or natural. (2018: 53)

Bode models this in the carefully-curated datasets she compiles about Australian newspaper fiction, which I would point to as an exemplar for computational work going forward. Bode’s carefully assembled metadata allows her to write more definitively about the objects of her analysis than is typically possible in large-scale computational work and to write an alternative account of Australian fiction, rooted in periodical culture, that takes seriously a rural, popular readership ignored in previous scholarly work.

However, I do want to carve out space for approaches to the digital archive that are bibliographically informed while being experimental, exploratory, even playful. The term ‘speculation’ has a long history in the digital humanities, as a term that pairs the technical and ludic. In an essay from the 2004 volume that named the field of digital humanities, Johanna Drucker and Bethany Nowviskie write about the tensions inherent in the term ‘speculative computing’:

Speculative computing is a technical term, fully compatible with the mechanistic reason of technological operations. It refers to the anticipation of probable outcomes along possible forward branches in the processing of data. Speculation is used to maximize efficient performance. By calculating the most likely next steps, it speeds up processing. Unused paths are discarded as new possibilities are calculated. Speculation doesn’t eliminate options, but, as in any instance of gambling, the process weights the likelihood of one path over another in advance of its occurrence. Logic-based, and quantitative, the process is pure *techne*, applied knowledge, highly crafted, and utterly remote from any notion of *poiesis* or aesthetic expression. Metaphorically, speculation invokes notions of possible worlds spiraling outward from every node in the processing chain, vivid as the rings of radio signals in the old RKO studios film logo. To a narratologist, the process suggests the garden of forking paths, a way to read computing as a tale structured by nodes and branches. (2004: 441)

For Drucker and Nowviskie, speculative computing is evocative almost despite itself, “conjuring images of unlikely outcomes and surprise events, imaginative leaps across the circuits that comprise the electronic synapses of digital technology” (2004: 441–442). Prediction is interpretation in this framework: a model of thought instantiated in code. For the digital humanities, this idea is important because “[s]peculative approaches make it possible for subjective interpretation to have a role in shaping the *processes*, not just the *structures*, of digital humanities” (2004: 442).

More recently, Nowviskie draws inspiration from Afrofuturism and other speculative practices to name a “design problem” in the construction of “digital humanities collections – archival and otherwise” that she argues are “more likely to be taken by their users as memorializing, conservative, limited, and suggestive of a linear view of history than as problem-solving, branching, generative, non-teleological”. Nowviskie challenges this memorialising impulse, asking:

Are we designing libraries that activate imaginations – both their users’ imaginations and those of the expert practitioners who craft and maintain them? Are we designing libraries emancipated from what I’ll shortly demonstrate is often experienced as an externally-imposed, linear and fatalistic conception of time? Are we at least designing libraries that dare to try, despite the fundamental paradox of the Anthropocene era we live in – which asks us to hold unpredictability and planetary-scale inevitability simultaneously in mind? How can we design digital libraries that admit alternate futures – that recognize that people require the freedom to construct their own, independent philosophical infrastructure, to escape

time's arrow and subvert, if they wish, the unidirectional and neoliberal temporal constructs that have so often been tools of injustice? (2016: n. pag.)

Nowviskie argues that librarians should “take seriously the Afrofuturist notion of cultural heritage not as content to be received but technology to be used” and thus to “position digital collections and digital scholarly projects more plainly not as statements about what was, but as toolsets and resources for what could be” (2016: n. pag.).

In an environment where library collections are inevitably incomplete and shaped by the biases of generations of curators, Nowviskie advocates for a new imagination that is not content with what the archive is, but insists on what it might be. Lauren Klein takes up a similar mandate for scholars of early American literary history, suggesting a “speculative aesthetics of early American literature [that] might be said to promise a new method for eliciting knowledge about a people, community, or culture, while also helping to conjure a sense – in the full meaning of that word – of what our distance from the past will forever preclude from view” (2016: 439).

What I am calling speculative bibliography is speculative in both senses Nowviskie and Drucker elicit. It is an anticipatory processing of bibliographic data in order to maximise possible paths of discovery – not all operations produce meaningful literary-historical insights, but some do – and it is also an imaginative act that asks how the archive might make meanings if differently arranged and juxtaposed.

Speculative Editions

In the *Viral Texts* project, we reorder the historical newspaper archive to see not individual newspapers over time, but the tendrils of textual repetition, quotation, circulation, and theft that linked papers and readers together. Since the beginning of the project, we have wrestled with how best to make our data available to other scholars to argue with or against. On the one hand, when we publish an article, its arguments are based on particular texts: the hundreds of reprints of the poem “Beautiful Snow”, for instance, or of a listicle outlining the habits of successful young businessmen. The reprints from which we developed our argument were identified by a particular run of our algorithm and exist as data in a spreadsheet, itself a historically specific textual artefact. We recognise that scholars reading our 2016 *American Periodicals* article should be able to refer to the 276 reprints of “Beautiful Snow” we consulted when writing it, and we dutifully published a spreadsheet alongside that article that includes all of the texts we cite in it (Cordell

and Mullen 2017). Ultimately, however, our argument is not about any particular reprint of “Beautiful Snow” but about the event of that poem across the country and the world: the many reprints, the readers who loved it, the poets who parodied it and parodied its parodies, the editors who debated its authorship. And our picture of that event continually shifts as we refine our reprint detection methods and add new historical newspaper data to our inquiries. I want readers to find those 276 reprints from 2016, of course, but I also want them to find the 291 reprints we know of in 2020, or the 500 (I write optimistically) we will know of in 2025.

As we develop the *Viral Text* project book *Going the Rounds* with the University of Minnesota Press, we are experimenting with an approach that weaves together textual editing and computational speculation.⁵ For each work that we write about – by which I mean a set of witnesses that we would identify as reprints of the same work – we create a clean transcription from one witness we specifically reference. These transcriptions, which we refer to as “anchor texts”, become stable points of reference incorporated into each new iteration of our algorithm: a seed around which reprints of that particular text will be clustered in subsequent analyses.⁶ In each new dataset we create, these transcriptions are the reference points that allow researchers to quickly home in on familiar texts, while allowing textual clusters to shift as we experiment with the parameters of our algorithm or expand the source data we analyse. Thus when the book is published, readers will be able to find the texts on which its arguments are based in our public database, but also to see how our picture of nineteenth-century newspaper reprinting is evolving in real time.

By providing a stable bibliographic reference point within a speculative computing environment, these transcriptions also enable us to better understand and critique the effects made by changes to our algorithm. From the computer science perspective of our project, we have always wrestled with the lack of “ground truth” – a well-known and described dataset in which to test whether a method is returning reliable results. No index or hand-tagged archive of nineteenth-century newspaper reprinting exists – even at a relatively small scale – that we could use to ensure we are finding the reprints we should be finding before applying our methods across a larger, unknown collection. Even from the CS side, then, our work is speculative, and we have relied on estimates of recall due to the fundamental incompleteness and uncertainty of historical data. Anchor text transcriptions allow us to directly compare textual clusters

⁵ For more on our book in progress, see <https://manifold.umn.edu/projects/going-the-rounds>.

⁶ Our transcriptions are available at <<https://www.dropbox.com/sh/74nl5ook9540xbu/AACBicW5ppS8YZ5lteUKMika?dl=0>> and will be incorporated into a database, which will be released concurrently to the book project.

from experiment to experiment and see precisely how changed parameters affect our results.

By taking a speculative approach to building bibliographies, the *Viral Texts* project puts into practice a method for identifying sets of formally related texts worthy of study, by virtue of their duplication, from the massive archive of nineteenth-century newspapers. To expand our scope a bit, we might imagine other algorithms trained to recognise particular historical fonts, or to identify woodcuts within a collection of historical books. As a complement to Bode's "scholarly edition of a literary system", I would propose the *speculative edition*: all texts associated through a single computational model. We should not take the probabilities underlying speculative bibliography as *the truth* about literary history, but as propositions that demand testing and argumentation. Given the scope of our digitised collections, however, I would argue that many branching, speculative bibliographies will be necessary to identify fruitful paths of scholarly inquiry, and must proceed in dialogue with the careful editing and curation undertaken by textual scholars. Too often we separate our *data* from our *analysis of the data*, as if the one exists simply to illuminate the other. To resist this impulse, we need to integrate our computational experiments into our archival interfaces, provided as alternative paths into and through material.

Works Cited

- Algee-Hewitt et al. 2016. *Canon/Archive: Large-Scale Dynamics in the Literary Field*. Stanford Literary Lab, Pamphlet 11. <<https://litlab.stanford.edu/pamphlets/>> [accessed 11 June 2020].
- Bode, Katherine. 2018. *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor, MI: University of Michigan Press.
- Cordell, Ryan. 2017. "'Q i-Jtb the Raven': Taking Dirty OCR Seriously". *Book History* 20: 188–225.
- Cordell, Ryan. 2019a. "Textual Criticism as Language Modeling", published in draft on the University of Minnesota Press's Manifold Scholarship platform. *Manifold.umn.edu*. <<https://manifold.umn.edu/read/untitled-883630b9-c054-44e1-91db-d053a7106ecb/section/ea1f849a-bac1-4e9d-85f4-149d0083a6a4>> [accessed 15 June 2020].
- Cordell, Ryan. 2019b. "Classifying Vignettes, Modeling Hybridity", published in draft on the University of Minnesota Press's Manifold Scholarship platform. *Manifold.umn.edu*. <<https://manifold.umn.edu/read/untitled-bd3eb0af-fdad-4dd6-9c94-3fd15d522ab6/section/06899e82-8f06-43d2-9fc9-ea04dffef886>> [accessed 15 June 2020].
- Cordell, Ryan and Abby Mullen. 2017. "'Fugitive Verses': The Circulation of Poems in Nineteenth-Century American Newspapers". *American Periodicals: A Journal of History & Criticism* 27.1: 29–52.
- Drucker, Johanna and Bethany Nowviskie. 2004. "Speculative Computing: Aesthetic Provocations in Humanities Computing". In: Susan Schreibman, Ray Siemens and John Unsworth (eds.). *A Companion to Digital Humanities*. Oxford: Wiley Blackwell.

- Klein, Lauren F. 2016. "Speculative Aesthetics". *Early American Literature* 51.2: 437–445.
- Long, Hoyt and Richard Jean So. 2016. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning". *Critical Inquiry* 42.2: 235–267.
- McGann, Jerome. *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction*. Cambridge, MA: Harvard University Press, 2014.
- Nowviskie, Bethany. 2016. "Speculative Collections". *Nowviskie.org*. <<http://nowviskie.org/2016/speculative-collections/>> [accessed 16 June 2020].
- Senchyne, Jonathan. 2019. *The Intimacy of Paper in Early and Nineteenth-Century American Literature*. Amherst, MA: University of Massachusetts Press.
- Tanselle, G. Thomas. 1975. "The Bibliographical Concepts of 'Issue' and 'State'". *The Papers of the Bibliographical Society of America* 69.1: 17–66.
- Tenen, Dennis. 2017. *Plain Text: The Poetics of Computation*. Palo Alto, CA: Stanford University Press.
- Vee, Annette. 2017. *Coding Literacy*. Cambridge, MA: MIT Press.