

Modernism and Gender at the Limits of Stylometry

Sean Weidman <seangweidman_at_gmail_dot_com>, Pennsylvania State University
Aaren Pastor <aaren_dot_pastor_at_gmail_dot_com>, Pennsylvania State University

Abstract

Virginia Woolf writes in her novel *Orlando* that “it is clothes that wear us and not we them; we may make them take to mould of arm or breast, but they mould our hearts, our brains, our tongues to their liking.” Her observation remains vital to the author’s longstanding, feminist critique of essentializing discourses, but it also gives recourse to the ways our computational methods of studying literature and its history sometimes “mould [...] to their liking” their objects. When studying style and gender, digital humanists have tended to use computational methods to trace embedded, hidden linguistic structures, showing how they can contain, conceal, and condition the gendered lives of social groups and cultural milieus. In this essay, we present a stylometric case study — Woolf’s *Orlando* — that reminds us why, when dealing with gender in modern literature, computational critics must pay particular care when addressing and generalizing from modernism’s experimental styles. We outline the limitations of our own and prior approaches to questions of gender and literary style, and we ask whether flawed stylometric analyses can acquaint us with ways that modernism’s stylistic innovations productively haunt the conclusions of digital literary criticism.

1

Presumably

(once upon a time)

you thought MODERNIST STYLE was

worthy of your

ATTENTION and CURIOSITY.

“Manifesto of Modernist Digital Humanities” Alex Christie, Andrew Pilsch, Shawna Ross, Katie Tanigawa [Christie et al. 2014]

When Doug Mao and Rebecca Walkowitz described the innovative critical turns of the New Modernist Studies over a decade ago, the *digital turn* did not make the cut [Mao and Walkowitz 2008]. In fact, only very recently have accounts of modernism turned to the conceptual purview of the digital, often to reframe understandings of modernism and digital technology or to contextualize new digital technologies applied to modernism. Jessica Pressman’s *Digital Modernism* and Shawna Ross and James O’Sullivan’s edited collection *Reading Modernism with Machines* necessitate special note in this regard, as each study in its own way accounts for how a digital modernism, or a digital approach to modernism, is “aligned with strategies of the avant-garde: it challenges traditional expectations about what art is and does [and] illuminates and interrogates the cultural infrastructures, technological networks, and critical practices that support and enable these judgments” [Pressman 2014] [Ross and O’Sullivan 2016]. Stephen Ross and Jentery Sayers trace as much in “Modernism Meets Digital Humanities,” and Gabriel Hankins argues similarly that through the “weak program”

2

of digital methods applied to modernism, critics can activate a variety of subtler points of contact that better model how diffuse modernisms share aesthetic territory [Ross and Sayers 2014] [Hankins 2018 b, 582]. Elsewhere, Hankins goes so far as to say that the “new modernist studies is now always already digital,” and though Ama Bemba Adwetewa-Badu reminds us that “digital tools cannot do the critical and analytical work that we do,” she adds that “neither can we do the work that they do,” that is, the work necessitated by a messy and data-filled literary world [Hankins 2018 a] [Adwetewa 2020]. Look no further for corroboration than Laura McGrath et al.’s “Measuring of Modernist Novelty,” an essay that operationalizes modernist “newness” to measure evolutions of novelty in 20th-century fiction [McGrath et al. 2018]. Other digital forays in the past half-decade, as detailed expansively by J. Matthew Huculak, represent substantial collections, recoveries, expansions, and innovations of modernist archivism [Huculak 2018]. Cue in Amanda Golden and Cassandra Laity’s co-edited issue of *Feminist Modernist Studies* on modernist, feminist digital humanities [Golden and Laity 2018], and the digital turn of modernist criticism seems fully upon us, furnishing the New Modernist Studies with accounts that illustrate how the critical debates of modernism are newly accessible, and newly accessible to critique, through digital methods.

Even so, computational approaches to analyzing modernist texts (e.g. text analysis, stylometry, computational stylistics, network analysis, and topic modeling) so far account for a relatively scarce subset of modernism’s new digital surge. By far the most active scholars at the intersection of text analysis and modernist studies have been Hoyt Long and Richard Jean So, whose work on modernism at the Chicago Text Lab has spanned a wide range of modernist topics: they’ve analyzed the poetic publication networks of high modernism, provided a multifaceted analysis of the English-language haiku, and studied the movement of stream of consciousness as literary technique [Long and So 2013] [Long and So 2016 a] [Long and So 2016 b]. Other scattered exceptions exist too,^[1] but their work makes up the bulk of that subset of digital literary criticism that attempts to measure, model, and interpret the styles of modernism computationally.

Several practical reasons have shouldered much of the blame for this lack of critical diversity. Although new and compelling intellectual paradigms have attempted temporal and spatial refigurings of modernism through its complex global and planetary shifts [Hayot and Walkowitz 2016] [Wollaeger 2012] [Stanford 2015], the dominant critical views of modernist work still converge in the early 20th century.^[2] Computational literary criticism requires an accessible textual archive, and since the preferred periodization of modernism has gone largely unchanged, large portions of modernist corpora have long been under copyright until only very recently — and most works, at least widely, may remain so for the better part of the next decades.^[3] Unsurprisingly, easy access to reliable digital materials has been a strong determiner of the topic matter of computational literary study, which often seeks to analyze bodies of work macroanalytically or, in the now notorious Morettian variation, “from a distance.” These limits have held particular sway in the realm of stylometry, a broad approach to studying textual style that aims, usually, to analyze a large corpus of literature by counting certain words (or other linguistic elements like grammatical structures, punctuation, etc.), averaging the measurements out into a unique “signature,” and then comparing that stylistic composite to other texts, authors, periods, and so on. Applied to literature, these countings have helped scholars make claims about style as a measurable form as it exposes certain conditions, contexts, and forms of representation, both inside and outside of narratives; through style, that is, critics have argued that they can digitally identify, measure, and compare the flows and workings of social and cultural mechanisms — like productions and representations of gender, race, class, or nationality — as they slip into linguistic tendency.

If we put aside the material limitations of modernist literature’s computational study, we find that deeper conceptual concerns have also tended to restrict it, especially with regard to modernist stylometry and its relationship to gender (and other identity categories). Modernist social life was as much governed by the sociopolitical affordances of feminism’s first wave as anything else, and its consequences ripple powerfully through much of modernist literature. Coupled with the dicta of avant-garde modernist aesthetics — which, for the sake of space, we might vulgarly boil down to newness and experimentation — this access to new material conditions, sociopolitical spaces, forms of representation, and modes of being-in-the-everyday also released a well-documented *stylistic* first wave for many writers. Gertrude Stein, Virginia Woolf, H.D., Djuna Barnes, Mina Loy, and Jean Rhys, among a multitude of others, unwound the concertinaed styles of their literary predecessors to produce, if we might borrow from Cary Nelson, a literature “generically undecidable, sometimes feminist and lesbian, wholly unassignable to a humanizing persona, and

3

4

5

more purely and powerfully devoted to an exploration of how language works” than much of the literature that preceded it [Nelson 1986, 179].

Our aim in this essay is to consider the conceptual impediments to studying that undecidability of gender stylometrically, as an influencer (or correlative feature) of modernist literary style.^[4] We first locate where some of those difficulties are borne out in computational literary criticism broadly and, in particular, call attention to the way stylistic markers that have been taken as proxies for forms of gendered experience can prove unreliable in modernist contexts. Virginia Woolf’s famous meditation in *A Room of One’s Own* on androgyny and its creative potential might be invoked here: stylometric scholarship has not yet contended with how extensively the experiments of modernist style, full of gender explorations and polysemous difficulty, complicate the tidiness and differentiability sometimes required of text analysis. Such a claim is certainly not new in the realm of computational literary criticism generally, and as the next section details, many digital practitioners have already studied just how deeply our notions of gender undermine computational designs. It is simply to this mounting critical stack that we add a new stylometric case featuring a small corpus of modernist fiction, in an effort to show both the limits and potentials of its stylometric study.^[5]

Using *Orlando* as a featured example, we argue that the story of gender and modernism as seen through stylometry reveals a shared limit of its models — of identity and computation — to provide trustworthy exemplars. We then briefly consider some ways that the stylistic uniqueness of Woolf’s novel might help us understand modernism more completely. After all, if the full cultural stakes of modernism remain unearthed [James and Seshagiri 2014], if the legacies of modernism continue to haunt our aesthetic forms, and if gender is one site of particularly fruitful aesthetic experimentation, then stylometry may help further uncover the stylistic innovations and cultural remnants that constitute the topos of modernist influence. And since modernist texts, at least in the US, are beginning to shed their copyright shackles, the proto-framing for a modernist stylometry is perhaps already overdue.

The digital turn has provided an innovative set of tools for reading, encountering, and thinking through modernist texts, as singular materials to linger with but also to scrutinize through their forms of digital manufacture. We hope to show that stylometric analysis of someone like Woolf in particular, and perhaps many modernist writers in general, links critics to new ways of seeing the designs of gender, the stylistic transitions of which require stylometrists to contend carefully with the complexity and multiplicity of modernist literature and its shifting orientations.^[6] Such a view has the benefit of reiterating two important points: first that modernism’s collection of digital movements (*vis-à-vis* Pressman) occasioned an understanding of linguistic styles as flexible, experimental technologies that could hide and reveal, invoke and manufacture identity categories at the proverbial touch of a pen or click of a typewriter; and second that our computational methods themselves complicate the study of modernist representation by rewriting literary data into new and ongoing digital dialectics.

A Critical History of Gender, Literature, and Quantitative Formalism

Some framing is in order before we outline our modernist experiment, regarding first how we employ the concept of gender in relation to modernist literature, and second what digital humanists have tended to glean from the computational study of literary style and gender (using tools and methods at least adjacent to stylometry). After, we consider how treating Woolf’s *Orlando* as a case study of gender and modernist style may help scholars rethink their applications of stylometric critique.

If we start with an understanding of gender as a medley of sociopolitical conventions and laws discursively assigned to bodies and constructed and policed accordingly (à la Judith Butler), to imagine we might reliably evaluate gender computationally seems, conceptually, at least somewhat confused. Gayle Rubin’s sex-gender system, that “set of arrangements by which society transforms biological sexuality into products of human activity” [Rubin 1975, 159], has long-since clarified that gendered identities are sets of internalized, performative scripts reified through various legal, political, scientific, and cultural processes. This is an essential clarification for stylometric analysis, which purports to measure these discursive influences as they bleed through language. If it is only through socializing discourses that a body becomes vested with the idea of an essential sex/gender, stylometry only detects the un-isolated impressions of that social and cultural influence as they mix with a mess of other discursive powers.^[7]

In spite of these conceptual limitations, a variety of digital humanists have undertaken the intricate work of studying gender and style throughout literary history and at various scales. A first and landmark example is John Burrows's 1986 study of Jane Austen's novels, which simply used Austen's varied frequency of function words (pronouns, prepositions, articles, conjunctions, and other non-content words) to identify character dialogue patterns, narrative subtleties, thematic transformations, tonal ebbs and flows in characterization, directions and forces of speech, and so on, formal features that (he notes) regularly fall along typical 19th-century gender lines [Burrows 1986].^[8] Many have since built from Burrows's pioneering study and, among the more recent attempts to unravel gender's impact on literary style, Matthew Jockers's work on 19th-century fiction in *Macroanalysis* is likely the most well-known. Chief among Jockers's conclusions is that gender is the *least* influential determiner on literary style in the 19th century when compared to other categories of stylistic difference — e.g. author, genre, (decade-long) period, textual variations (punctuation, pronoun usage, etc.) [Jockers 2013, 96–8]. In a narrower study of 19th-century fiction, one framed more closely by the period's sociohistorical contexts, Jockers and Gabi Kirilloff looked at the depiction and action (via pronoun and verb combinations like *she walked* or *he dressed*) of male and female characters; they observed that men and women were narrated with different vocabularies according to 19th-century norms, and that 19th-century novelists had difficulty masking their own gender when they wrote characters of another [Jockers and Kirilloff 2016].^[9] In fact, Kirilloff and Jockers published an addendum to that study recently, pairing close reading of several of its outliers with those verb-usage lists, concluding that “[t]hrough the characters in these novels behave in ‘unconventional’ ways, all three struggle to mitigate traditional and emerging social values,” a dialectic they found “reflected not only at the level of plot but at the level of the sentence, and in fact, at the level of individual words” [Kirilloff et al. 2018, 842].

Although the above examples concern gender as it emerges in literature to reflect certain cultural norms and social circumstances, critics have also studied gender computationally in relation to its contemporary production, to its politics of reception, and to the literary marketplace; and each approach productively ties the stylistic into its socioeconomic imprints. Ted Underwood, David Bamman, and Sabrina Lee find that gender divisions between characters in fiction are becoming less distinguishable (i.e. men and women characters are being described and speaking/acting more similarly), but that the proportion of fiction written by women, and the proportion of women characters, dropped dramatically from 1800-1960. Hence “while gender roles were becoming more flexible,” Underwood et al. argue that the story of women writing literature is “a story of steady decline,” and that between 1850 and 1960, “fiction itself became more attentive to men,” such that women as characters were becoming less prominent, even in books by women [Underwood et al. 2018]. Jonathan Y. Cheng, meanwhile, takes physical descriptions as an entryway to studying gender and gendered embodiment over the last 150 years of English-language fiction; he uncovers that physical descriptions of characters has increased over time, but that for women's characterization, specifically, bodily descriptions remain curiously central, even as literary depictions of men and women become less binary [Cheng 2020, 28–9]. In a strictly contemporary account, after canvassing thousands of reviews from the *New York Times Book Review*, Andrew Piper and Richard Jean So discovered that reviewers still talk about women's books and men's books in stereotypically antithetical terms, “reproduc[ing] the public/private split bequeathed to us from the nineteenth century” [Piper 2016]. For Piper and So, their findings indicate that the damaging discourses around gender are still being subtly reproduced in the public apparatuses that surround literature.^[10] Their results were expanded upon by Eve Kraicer and Piper, who found similarly — and through a diverse and thorough assortment of computational models — that the representations of women in much of contemporary fiction mirrors their marginalized and decentralized positions in the literary marketplace [Kraicer and Piper 2019]. Matthew J. Lavin achieves something similar in his more located study of gender, reception, and the descriptive styles of *NYT* book reviews between 1905 and 1925; and, contrary to Underwood et al., Lavin finds that although women were in fact writing *more* fiction than men, that diverse work was still being discussed and reviewed within a reductive, domestic, private/public vocabulary [Lavin 2020].

The computational analyses we cover above largely discuss gender and literary production in the 19th-century or our contemporary period, and we draw attention to them not only to detail the range of current quantitative formalist work on gender and literature, but also to point to the critical gap that exists where a related digital modernist criticism could be, given that these studies have provided few insights about modernism.^[11] While the above is by no means a comprehensive list,^[12] as a largely current critical sampling, these studies showcase both the standard story of gender

and style leading up to or following the *fin de siècle* and how scholars have approached that story, which is most often by examining the stylistic indications of a period's culturally and socially produced gender binary, and then determining how those categorizations reveal new aspects of gendered subjectivity. In fact, this is probably a fair way to summarize the basic project of most computational studies of literary style and gender, the majority of which unearth, organize, and interpret linguistic traces through gender's various social and cultural histories, its modes of material production, and its narrativized representations. Yet as these scholars realize, utilizing digital methods to approach gender as a defining mechanism governing literary reception, critique, and re/production introduces as many problems as insights, especially when the results of digital literary analyses are placed within the same social and historical contexts from which the corpuses are pulled.

The first impulse of computational critics is often to test the already-reified categories around us, gender notwithstanding. But that approach can tend toward troublesome precarity — troublesome not merely from a computational standpoint either, per Johanna Drucker, who makes the following observation about the hazards of studying gender digitally in her meta-analysis of computational modeling:

The markers of gender identity may be not in the corpus but in the biographical details, expressed or suppressed, of authors. Looking at stylistic features and asking how they come to represent or construct conventions of gender as an effect of textual practice would lead in directions other than those that presume to sort according to gender as a given category. [Drucker 2017, 632]

So, as we harness computational methods for the textual analysis of gender, we cannot neglect the urgency of self-reflexivity, which must illuminate how the parameters of our textual data, as well as our own interpretive contexts, already construct the pseudo-essentialized elements of gender in our digital practices. Miriam Posner and Lauren Klein helpfully reimagine data in an analogous way, “as much an orientation toward one's sources as it is a primary category of knowledge” [Posner et al. 2017, 2]. Such a view, the scholars point out, yields DH methods more materially, socially, culturally oriented, attendant to the non-digital conditionings of the digital, the discourses of which are often grounded in dialectics of visibility and invisibility.

Take, though, a statement from some of the scholars participating in that digital praxis — for example, Underwood et al. in their *Cultural Analytics* piece — who often explain these shared limitations differently:

This essay considers both the gender positions ascribed to authors as biographical personages, and the signs of gender they used in producing characters. In both cases, we understand gender as a conventional role that people were expected to assume in order to become legible in a social context. Authors and characters have been coded according to a tripartite scheme (feminine / masculine / other or unknown), because that scheme organized most public representation of gender in the period we are studying. Gender can certainly be more complex than these categories suggest, and flexible ontologies can be designed to illuminate the complexity. But this essay is inquiring about the history of conventional roles, not about the truth of personal identity, or the underlying processes that produce gendered behavior. [Underwood et al. 2018, 2]

The message here begins to echo Drucker's, in that it is in the author's biographical details and situated experiences, which leech into the text as stylistic attributes, that the idiosyncrasies of the author's personal style become the clothing that outfits a novel's gendered position. But where Drucker implies that gender can operate as other than a given category, Underwood et al. must admit to retaining a tripartite structure for sorting gender, even though they make very clear that such a schema will fail to encapsulate gender's complex, messy, performative capacities. In most computational analyses of literary style and gender this is an inevitable rhetorical move that, as Kraicer and Piper explain, historicizes the results and “allows us to make explicit the otherwise latent ways in which gender is being mobilized and hierarchized within novels” [Kraicer and Piper 2019, 26]. After all, a first lesson of computational literary studies is that discourse predetermines data — depending on the tools we implement and the discourses we possess, data can always be made to appear a certain way. As they track some of those ways in their *Cultural Analytics* “Identity Issue” introduction, Susan Brown and Laura Mandell develop an exhaustive history of conceptualizing gender, outlining

the importance of such historicizing for digital projects, which must balance quantitative methodologies with that same reflexive process that asks how our methodological designs direct our analyses. “Historicizing helps to destabilize identities,” Brown and Mandell maintain, “and cultural analytics can make visible a kind of history that we have never seen before” [Brown and Mandell 2018, 13].

Perhaps, though, the kind of historicizing typical textual analysis does serves to stabilize more than destabilize. Underwood et al. say, for instance (to keep with our arbitrary, cherry-picked example), that their essay does not delve into “the underlying processes that produce gendered behavior.” But the assumption that such processes can be ignored when studying literary history seems suspect — because language, we know, is one such process, one that governs many others; and as Drucker reminds us, if you want to model gender you must produce (or reproduce) the underlying structures of gender. At some point, the current confines of stylometric analysis require that data be sorted into comparable categories, which means whether one gender, two gender, three gender, four, comparing categories or columns of gender identification remains much easier than comparing gender spectra. Does adding a third, non-binary category — or a fourth, or a fifth — unlock or ameliorate the hindered politics of representation at the heart of designing computational analyses around those historical notions of gender? As Pamela Caughie et al. emphasize in their work on feminist ontologies, the answer is: *of course not* [Caughie et al. 2018]. And yet, to label some identity as historically extant when it wasn’t *also* introduces a set of problems, and as Underwood and company note, to imagine there was more social space for expressions of gender fluidity than there actually was does us little good when analyzing textual traces that we aim to tie back into historically grounded claims about social realities.

20

This is the uncertain impasse at which stylometric gender analysis has, so far, found a rather comfortable home. Underwood is right, as he aptly remarks elsewhere, that “nothing about this [binary] method compels us to stop at two perspectives” [Underwood 2020, 96]; but is the best we can really expect of such critique its regular revelation that “the very act of modelling carries with it the seeds of a constructionist recognition that a phenomenon could [or will soon] be modelled differently” [Brown and Mandell 2018, 17]?^[13] That leaves us with a set of untenably limited critical positions and several vital, unaddressed questions. Does the clumsiness of quantitative formalism actually force us to decide whether to (a) anachronistically apply new understandings of gender to contexts that cannot hold them, or (b) hand-wash our responsibility to witness and recover different kinds of gendered being, whenever they occurred? In the context of stylometry specifically, how can we responsibly do (a) while avoiding (b)? How can we avoid reproducing normative investigations of identity categories like gender through models of distant reading, given that those models function in systemic and epistemological hierarchies of power that were not designed with plurality in mind?^[14]

21

In short, we not only lack a spectrum-based computational model for categorizing and operationalizing gender stylistically, but — for scholars of modernism — we also lack a critical vocabulary to address how our digital methods at once allow and prevent us from analyzing those “thick textures of history” that the New Modernist Studies promised [Mao and Walkowitz 2008, 745]. This is a hinge, too, of Nan Z. Da’s argument against computational literary studies: that computational approaches to texts (stylometry among them) are frequently ill-applied to data not suited for its methods, and that its results are either incorrect, overstated, or unoriginal [Da 2019]. So, how might we develop a situated stylometry as “a digital strategy that *can* inform” our computational and interpretative methods [Christie et al. 2014]? How can we put at our service digital methods that, according to Drucker, only act as a kind of filter through which we pull existing social structures and hierarchies, like normative or binary ideas about gender? Can we avoid treating their computational products as purified or more extant forms of pre-constructed inequity, and instead allow our processes of modeling to deconstruct themselves, to produce findings resistant to the systems that arrange them?

22

In addition to underscoring the urgency of revising our methods to better fit gender’s actualities, we hope to demonstrate shortly — against recent critiques of computational literary scholarship — that even flawed stylometric models can teach us new things about modernism. In the following section, we posit that studying Woolf from within the binary she saw beyond is itself an approach that helps us to reconstruct her project of imaginative self-fashioning. Treating *Orlando* as an important edge-case for modernism’s digital-critical scaffolding, our investigation aims to recover one of Woolf’s stylistic innovations and, in doing so, to rethink a few of stylometry’s prevailing critical assumptions.

23

Anglophone Modernism and Woolf’s *Orlando*

As with any digital literary analysis, taking the stylometric baton into modernism began with an inevitably insufficient selection (and rejection) of texts to analyze, and the narrow variety we collected here is representative of three, semi-arbitrary governing features: first, each author needed to have written at least three works of fiction (or collections of short stories),^[15] second, those texts needed to be published between 1900 and 1940,^[16] and third, those works needed to be written in English. Bearing in mind the lessons of Jockers's *Macroanalysis*, by narrowing our scope so dramatically we tried our best to isolate gender and its stylistic pressures from other well-known influencers of style like genre, period, translation, and so on. We thus settled on 15 male and 15 female Anglo-modernists (Tables 1-2), and chose three to five works from each, for a total of 141 texts (68 female, 73 male).^[17]

Bowen, Elizabeth	Rhys, Jean
Compton-Burnett, Ivy	Richardson, Dorothy
Ferber, Edna	Stein, Gertrude
Hall, Radclyffe	Townsend Warner, Sylvia
Hurston, Zora Neale	West, Rebecca
Kavan, Anna	Wharton, Edith
Larsen, Nella	Woolf, Virginia
Mansfield, Katherine	

Chesnutt, Charles	Hemingway, Ernest
Conrad, Joseph	Huxley, Aldous
Dos Passos, John	Joyce, James
Dunbar, Paul Laurence	Lawrence, D.H.
Faulkner, William	Lewis, Wyndham
Fitzgerald, F. Scott	Steinbeck, John
Ford, Ford Madox	Waugh, Evelyn
Forster, E.M.	

If the process of forming a modernist corpus begets a state-of-the-field question about the current archive of modernism, this study's selection of authors succeeds only in capturing a small, Anglophone, transatlantic subsection of the diverse, global engagements with and reactions to the political, social, and cultural investments of modernity. It is representative, that is, of only one, precarious, incomplete variant of modernism — a small sample that allows us to make clear claims about one subsection of modernist work. Archives, digital or otherwise, summarize and sometimes disguise significant acts of scholarship, and our own corpus operates with an agenda that erases as much as it attempts to combat its erasures. The resulting selections are meant only to represent one specific stylistic case — a rather “traditional” canon of Anglophone modernist fiction^[18] — among the many stylistic cases of modernist literature, and we hope readers will forgive the indiscretions of our modernist assemblage, knowing that this list of authors could have taken any number of reasonable forms.^[19]

After just outlining some of the many problems of doing so, the irony of organizing our study around normative gender categories has not eluded us. Mandell announces that this very move is her topic of inquiry in the opening chapter of a recent *Debates in the DH* volume: “in some quantitative cultural analyses, the category of gender has been biologized, binarized, and essentialized in a trend that I call *stereotyping*”; and if literary data has “been collected, sorted, and even produced according to the categories M/F,” then it will find statistically significant differences whether or not those differences are significant [Mandell 2019]. An algorithm trained on constructed differences of data (i.e. assigned authorial gender) does not sort through the world it purports to represent, after all, but generates a new world through the measures of difference with which we provide it. Bearing that considerable problem in mind, and building on the foundations set by the many computational studies of 19th-century literature, our interest is not in asking if or where gender is marked in fiction, but whether and how it “gives-to-see” the literary styles and structures of modernist fiction. Because we rely on gender as a historical category, we can learn about its in/stability and trans/formation from the ways it is invoked, signaled, modified, or tailored textually, around specific cultural contexts and social meanings and in relation to other texts. As traditional, binary conceptions of gender began to be understood as culturally, socially, politically, and legally affixed to bodies, modernist authors might also have affixed them to — or unfixed them from — their texts at a macro level, and we want to track the unique structural styling of gender in those texts.

Like most, our methods for doing so stylometrically rely on the production of a table of word frequencies, from which stylistic similarity (based on statistical distance) can be measured. Traditionally, the most influential words in these

tables are those function words, English's grammatical glue (articles, prepositions, pronouns, etc.), because it turns out that each author — and, though to a lesser degree, each period, genre, and gender — uses each of those words at a remarkably consistent rate in their own work, and at a unique rate when compared to others.^[20] Stylometrists can then measure the statistical distances between the styles of individual texts, authors, or other groups (e.g. based on gender) to detect broad trends.^[21] This research has tended to take the form of sorting and examining diverse most-frequent-word (MFW) lists made up of other types of content words (nouns, verbs, adjectives, etc.) to locate stylistic quirks or generate topics for close-reading. For instance, one group of authors might prefer one term, set of terms, or set of topics over another — or, they might not — and by contextualizing and interpreting the divergences or convergences, critics can theorize what circumstances (generic, historical, social) might contribute to similarity or difference. However, in this study, we want to be particularly sensitive to the interpretive dangers in close-reading individualized wordlists (either for authors or their texts) because of how our digital design treats gender as a two-body problem. Fearing, that is, the equal likelihood of finding gender-specific stylistic trends and replicating essentializing discourses in our interpretations, we decided to take a less orthodox path by staying *macro* and running a series of broader, supervised machine-learning analyses on our modernist corpus.^[22]

So, we trained our machine on the two groups of modernist texts (female, male), asking it, in effect, to develop an average style for each group based on the frequencies with which its authors used common words.^[23] We then gave it sequences of random groupings of six different texts (three male, three female), which it had to correctly place in the corresponding author's associated sex. By random chance, the machine will assign any text to its correct group 50% of the time — with two possible groups in our analysis, the algorithm has the equivalent of a digital coin toss — which means that anything consistently or considerably more accurate suggests there may be measurable stylistic differences between the two groups. Among hundreds of classifying runs, misattributions occurred on average about 10% of the time; the classifier accurately determined the assigned gender of a text's author in around 90% of cases, which supports the notion that even given the experiments of modernism, the period's sex-gender binary seems to remain a strong determiner of style.^[24] We utilized a series of different test conditions to ensure our results weren't skewed with any one set of parameters,^[25] and though our tests turned up fleeting misattributions — among other examples, the machine assigned a random sample from Hemingway's *For Whom the Bell Tolls* to the female group, and it ascribed several samples from Radclyffe Hall's *Well of Loneliness* to the male set^[26] — we produced one misattribution that was both more frequent and more consistent than the rest: Virginia Woolf's 1928 novel, *Orlando: A Biography*.

Common stylometric practice says to zoom in here and discern why *Orlando* is an outlier — and options abound for doing so: we might, for example, compare individual wordlists of each gender grouping, or of the text itself, or of the words preferred or avoided by one gender grouping/text over another. Again, however, our interest is in staying macro. Having already separated the novels into male and female groups, the wordlists only and unsurprisingly provide the same stylistic snapshot that has been traced extensively elsewhere (as our previous section details). We know that even as standard uses sometimes lead to surprising ends, male and female authors normally reproduce socially, culturally, historically consistent accounts of gender in their writing, often through predictable and stereotypically gendered characterizations, vocabularies, narrative actions, identifiers, and locales.^[27] Relocating the gender difference we structured our study around is not so interesting to us. What we want to ascertain, separately, is whether we can deploy stylometric methods without double-reading gendered cultural meaning into a list of semantic differences like verb constructions or function words. What if we apply stylometry to zoom out to the level of a whole text rather than its list of words? What does gender play look like within our small modernist corpus if we treat it as, for instance, an element of narrative structuring?

Instead of studying how male and female authors diverged in word choice or grammatical preference, then, we decided to look at the overarching structure of that difference in our consistent outlier, *Orlando*. To do so, we ran a more specific classification Eder calls a *Rolling NSC*.^[28] This method also requires that we train the machine on our two gender groups, but rather than supply it with a random selection of texts to assign, we ask it to analyze slices of one particular text — in this case, *Orlando* — and then assign each slice to its stylistically closest group/gender. To make matters more interesting, we isolated our test author, so that none of Woolf's other works appeared in the training sample or

28

29

30

contributed to the averaged style of the female set; if she had a particularly strong author signal (and she does), this process would prevent her unique style from overpowering the average gendered style that results from the amalgam of her contemporaries' works.^[29] For sake of reference, we did the same analysis on a number of other randomly selected texts and, as is to be expected with a ~90% attribution rate, most slices of most novels were correctly assigned to their author's corresponding gender (Figures 1-2).^[30] In other words, most modernist authors in our corpus write more like the other modernists of their sex, and do so consistently and for the entirety of each novel, which appears to be the stylistic norm in literature regardless of period.

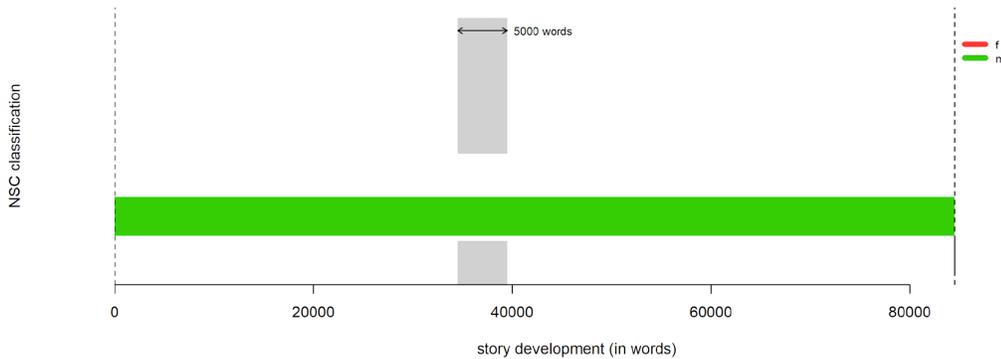


Figure 2. James Joyce's *Portrait of the Artist as a Young Man* as seen through the “rolling.classify” function of *Stylo*; the green line is actually made up of many thinner segments, 5000-word slices of the novel, each of which overlaps by 4500 words, and all of which were classified as “male” in style.

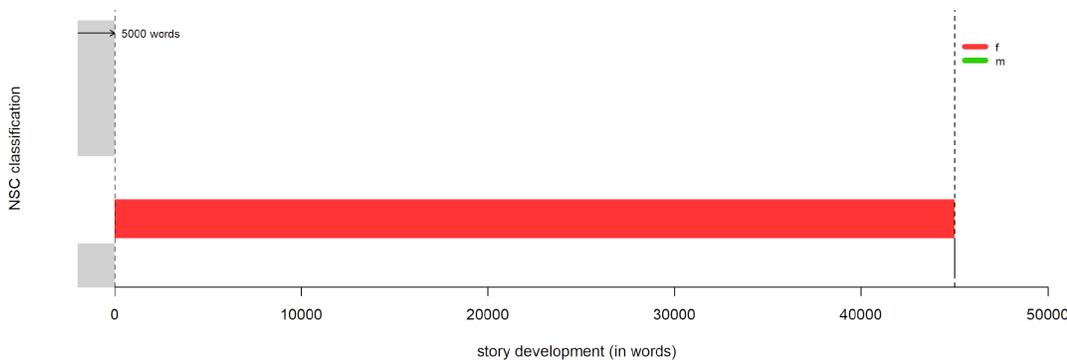


Figure 3. Jean Rhys's *Quartet* as seen through the “rolling.classify” function of *Stylo*; the red line is similarly made up of 5000-word slices of the novel, each of which overlaps by 4500 words, and all of which were classified as “female” in style.

But *Orlando* presents an atypical case — and to an extraordinary degree (Figure 3). About 32,000 words into the novel, the seemingly misclassified “male” style of Virginia Woolf abruptly changes, and the algorithm attributes the remainder of the novel’s style to the correctly assigned sex. More extraordinary, even, is what takes place narratively at the stylistic switch. The barely visible vertical dotted line marked ‘a’ in Figure 3, which lines up almost perfectly with the moment of stylistic gender reassignment, if you will, comes just after our protagonist Orlando is awakened in dramatic fashion from a weeklong slumber, and as those around him make a startling discovery:

31

We are, therefore, now left entirely alone in the room with the sleeping Orlando and the trumpeters. The trumpeters, ranging themselves side by side in order, blow one terrific blast —

32

‘THE TRUTH!’

at which Orlando woke.

He stretched himself. He rose. He stood upright in complete nakedness before us, and while the trumpets pealed Truth! Truth! Truth! we have no choice left but confess — he was a woman.

[Woolf 1992, 87]

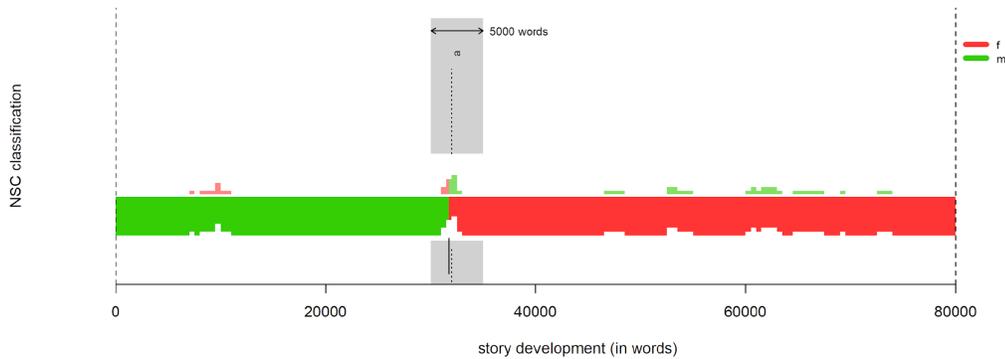


Figure 4. Virginia Woolf's *Orlando*. The slices (and 500-word incremental changes) are more visible here, and their overlap shows the sudden stylistic transition — from stylistically “male” to “female” — about a third of the way through the novel. The smaller peaks/valleys represent the model's less-certain detection of each gender's style.

As she narrates a sex change in her story, Woolf actively styles a gender signature to match that of her character, a creative feat no other author — of *any* period — has managed, as far as we know. It's difficult to explain just how unusual this is, in part because our subsection of modernist fiction doesn't speak for all of modernism (let alone all of fiction), but in larger part because we've just never seen anything like this before. If our analysis isolated a descriptive, content-based lexicon (nouns, verbs, adverbs, adjectives), something like this result would be expected, at least to a degree. Several of the studies we outlined in the last section confirm that regardless of an author's gender, talented writers are capable of imitating aspects of another gender in their writing (preferred behaviors, contexts, spaces, styles of speaking, etc.), a fact likely truer still for a writer as accomplished as Woolf. In our analyses, however, which again ranged from 50 to 500 MFW, the words we used to measure literary style are largely just those functional, non-descriptive terms void of explicit narrative context. This means that Woolf didn't simply change the *what*, the content her male-turned-female protagonist was thinking, saying, or doing — she fundamentally changed the *how*, the grammar and syntax that holds that content together.^[31] Perhaps how this came to pass, as *Orlando's* narrator relates, matters little: “It is enough for us to state the simple fact; Orlando was a man till the age of thirty; when he became a woman and has remained so ever since” [Woolf 1992, 88].

33

***Orlando* at the Limits of Modernist Stylometry**

If we turn to the literary criticism on *Orlando*, the impressiveness of this stylistic manipulation would seem inevitable given the theoretical projects the novel is said to carry out. “Long before university Gender Studies departments came into existence,” writes Anne Delaplace, “Woolf had already reflected on the dissonance between social and sexual identity” in *Orlando*, which retains echoes of the intermix[ing] of genders from the exploration of androgyny in *A Room of One's Own* [Delaplace 2011, 39]. Indeed, “in every human being,” says *Orlando's* narrator, “a vacillation from one sex to the other takes place, and often it is only the clothes that keep the male or female likeness, while underneath the sex is the very opposite of what it is above” [Woolf 1992, 121]. For Pamela Caughie, the novel provides nothing less than the archetypal “model of modernist life writing in the era of transsexualism,” a unique, modernist “transgenre” that “radically refigures the narrative of transsexualism” [Caughie 2013, 502–4].^[32] These literary critics identify clearly the power of Woolf's novel as it derives from close readings of context and narrative, producing feminist modes of representation for non-normative sexuality and genders through its play of literary genre — in Caughie's case, the biography, mock or not.

34

Elsa Högberg and Amy Bromley even consider *Orlando* the novel that inaugurated Woolf's theory of sentence morphology, the way Woolf playfully reshapes formal and generic sentence structures to mark a "modernist rewriting of biography and gender, and the orthodox ways in which they tend to shape life and the body" [Högberg and Bromley 2018, 2].

Keeping with these accounts, we simply want to make one vital addendum: that the novel's power — as proto-gender study or as transgenre — emerges from Woolf's style. Again, while not every novel in our analysis was always assigned so tidily to one gender or another, no text separated as consistently or in such a structured manner as *Orlando* — on cue, as it were, with the gender transition in its narrative. That level of stylistic control raises new questions about the scales of gender performativity, or rather focuses them down to the level of the line, the choice of one (seemingly insignificant) function word over the other, which complicates readings of *Orlando* as a tale of cross-dressing, in which clothes make the (wo)man without regard for the discursive power wielded in the creation of "men" and "women." Woolf, it seems, did not merely change the descriptions and clothes of her protagonist and so change the gender; she altered the very stylistic nuances that characterized Orlando when she was a woman from when he was a man. Hers isn't just a narrative exploration of transition either, in the vein of Dr. Matthew O'Connor in Djuna Barnes's *Nightwood* or Leopold Bloom in James Joyce's *Ulysses*.^[33] That her writing diverges so thoroughly from the stylistic doxa of her contemporaries suggests to us that, as Woolf celebrates the new possibilities of being gendered and representing genderedness, she delivers a linguistic paradigm of that futurity. In *Orlando*, Woolf not only cultivates a sustained, pluralist theory of gender transition, but she also practices the linguistic applications of its modern futures, a multifaceted engagement that perhaps no other author had actualized before.

In this our study deviates from the findings of several aforementioned accounts of gender and literary style (e.g. the work of Jockers and Kirilloff (2016) on gender in 19th-century fiction), which tend to suggest gender-disciplinary social norms had significant and, by extension, inextricable effects on literature's stylistic structures. Most stylometric studies that isolate gender within a corpus of literature treat it similarly, as a governing stylistic force that manifests in formal semantic preferences or syntactic characteristics. In the case of Anglo-modernist fiction, though, staying out of the wordlists forces a change in perspective.^[34] Perhaps the period's history of gender and literary style is more than a sum of linguistic tendencies, or more than gender's socially, culturally, politically enforced written marking. In her innovative styling of gender transition, Woolf shows that this history is also an account of gender's *unmarking* and *undisciplining*. Perhaps all that Woolf intuited is that gender needn't have been as influential on literary style as we think it must have been; or, perhaps Woolf saw that our ideas about gender do not so much influence literary style as literary style influences our ideas about gender. If *Orlando* declares gender as merely the newest formal territory on which writers could experiment with identity, then it uncovers for us Woolf's broader way of modern seeing, which could understand gender not as an imminent, immutable, uniform stylistic determiner, but as a flexible literary landscape — or, what comes to the same thing, as a social text, a style.^[35]

Taking a step back, and stealing from the punchy clarity of Stephen Ramsay, we want to repeat a key clarification: "We are not trying to solve Woolf" [Ramsay 2011, 15]. We instead want to use *Orlando* to wonder whether the current registers of stylometry and computational stylistics stymie some of the radical potentials of modernist style. Because we constructed this stylometric study around a constricting binary gender logic, it can only tell us so much about that binary or what constitutes its differences, especially with tools applied as narrowly as we apply them here (e.g. on a relatively small modernist corpus, and then on a single modernist text relative to that corpus). Stylometric methods, like other forms of quantitative formalism, can't exactly be employed in pursuit of such a task, particularly when our assumptions need to be levied in service of cultural and social conclusions to make them critically relevant. Most studies of gender and literary style have tended to assume that if a linguistic difference can be computationally tracked along gender lines, that difference must divulge something about gender's lived differences or material history. But we are not convinced this is always true. Sometimes the only thing stylistic difference informs is our understanding of failed historical representation, or a history of *unlived* difference, or (in our case) gender's otherwise hidden potentials. Despite writing from within a descriptive system that inhibits gender identity's expression and possible futures — and despite being studied from within the same — Woolf still teaches us about the ways gender/ing could be deployed, inverted, and resisted, a didactic function that materializes even through, even as, the digital recording of a broken and insufficient

35

36

37

gender binary.

Our brief analysis of *Orlando* should suggest that stylometry even applied narrowly can still teach us a great deal about, say, the depth of modernism's *avant garde* or about the limitations of studying style as a consistent and formal indicator of real-world messiness. After all, it's through our study's significant structural limitations that we see the extent to which modernists could make their styles fit (or an ill fit for) its categories — see, that is, how Woolf fashioned gender in, out, and around the binary trying to contain it. So what other realignments might Woolf's modernist experiment afford within the genealogy of gender and computational literary studies? Two observations seem essential: the first, a minor point about exemplarity, will lead us to the second, a broader claim about the limits of studying modern(ist) sociocultural forms through stylometric analysis.

38

Christy Burns asserts that “in *Orlando* Woolf has already [...] anticipated our attempts to clothe her writings in our own desires,” for the very reason that “the curious residue of language [...] both invites and resists our insistent refigurations, our attempts to make Woolf conform to our societal demands” [Burns 1994, 359]. In this unique anticipatory capacity Woolf's novel comes to represent an exceptional case in our corpus of modernist fiction, both through its multi-century generic scope and through its stylistic flexibility. Its uniqueness thus poses an interpretive problem: How can we draw general conclusions about Anglophone modernism, or about gender and modernist literature, from a text so different from the rest, a novel that singularly rejects what we thought we knew about the relationship between gender identities and literary style?^[36]

39

According to Ben Etherington and Sean Pryor, by focusing on such exceptional cases scholars “can actively resist the falsifications and exclusions performed by the abstract generalisations of history, theory, and society at large” [Etherington and Pryor 2019, 7]. Stylometry notwithstanding, computational methods of analyzing literature attempt to discern a set of generalities that can cater to the macroanalytic at the cost of particulars, outliers or exceptions that sometimes require culling or discarding. But if Woolf's unprecedented literary experimentalism is what makes her an exemplary modernist — which is the case made by Burns and many other modernist critics — perhaps *Orlando*'s exceptionalism is what makes it an ideal text from which to generalize. If, in numerous cases of Anglophone modernism, the experimentally exceptional *is* the general,^[37] then the discovery of new forms of difference in that literature, like *Orlando*'s stylistic play, requires that the very category of “general” be productively expanded to include even its most remarkable outliers. And when the general example produces an updated interpretive scale or model, then it also produces an updated politics of historical interpretation consistent with that model, a politics that scholars like Burns would insist Woolf not only refuses but remakes.^[38] Another way to say this is that, as Woolf invents a new means by which gender could be individually styled and proven as style, she shows that our general, normative understanding of the conditions and capabilities of fiction for representing gendered life is incomplete and in need of expansion. Literary critics have known about this capacity of Woolf for a long time, of course, but this stylometric view of *Orlando* demonstrates just how acutely she played with gender as itself a type of literary style. There is no solving Woolf here, as Ramsay says, because Woolf fundamentally changes the stylistic equation, widening modernism's scope of the representatively possible.

40

To our second point, then, and contrary to what most digital humanists have found by studying the literature of the 19th century, Woolf proves again that modernist literature generated denaturalizing and atypical movements of gender *as* and *through*, and not merely *in*, its styles. Modernists, we know, were among the first generations of writers fundamentally influenced by the reality that unconscious forces construct us,^[39] and it's essential to remember that their words are thus not simply markers of hidden sociocultural powers and discourses, but that they instead represent — for one of the first times in modern literary history — active, ongoing, deliberate engagements with those discourses as they knew they were appearing in their literature. Hopefully it's clear how our point about exemplarity is a stepping-stone to this. When we arrive in modernism, a period in which authors, figures, writers, and thinkers were quite explicitly styling gender and its cultural markers, tracing a gender signal in writing and attaching its features to latent cultural realities is no longer quite as defensible by the logic used in stylometric studies of 19th-century literature. As Woolf herself says elsewhere, theirs was already “an age clearly when we are not fast anchored where we are; things are moving round us; we are moving ourselves” [Woolf 2009, 74].

41

Simply put, what we imagine to be the latent, stylistic territory of stylometry (and other computational methods invested in written style) cannot be trusted to have been latent at all in modernism. Woolf's experiment reminds us that the fundamental assumption of traditional stylometry is notoriously unreliable: namely, that authors and specified groups have distinct, independent, hidden markers of style we can compare and from which we can glean social and cultural insight — about, for instance, the intentional or accidental written production of gender. Earlier we argued that modernism hasn't been studied digitally for two reasons, but this would offer a third: treating modern style as a proxy for (or reflection of) the linguistic features of modern identity can neither be done dependably nor without presuming to already know, and thus arrange a study around, where identity begins and ends for literary stylists. Scholars leveraging stylometric methods to analyze modernist literature must contend with modernism's unique self-awareness in this regard — the ways stylistic experimentation may limit the conclusions that can be drawn from modernist language about the conditions of modern life. *Orlando's* remarkable stylistic play gives us reason to pause, in other words, reason to question stylometry's ability to contextualize gender and modernism in the same breath.

We do not mean to suggest that stylometry is fundamentally an ill-fitted tool for studying modernism and gender, or even modern identity writ large. We simply want to point out that modernism, perhaps more so than other aesthetic movements or literary periods, necessitates a critical mutuality that requires from its scholars an upfront and explicit negotiation between object and method, accident and intention, style and styling. As interpretive constraints, these tensions might even address why there isn't much serious stylometric work on modernism — and perhaps why there should be. If we want the future of modernity's digital literary study to have any kind of educational end, and if we indeed think that stylometry can spill forth new insights about (or models of) the representational histories of social experience (vis-à-vis linguistic style), then a crucial part of such projects must be to gauge how the limitations of our tools might help us to study differently the sociocultural traces of modern identity and its stylistic waves. Maybe Woolf isn't the first or only author to make the necessity of this reorientation so evident, but *Orlando* certainly illustrates the extents to which, when applied to and in modernism, both our literary data and our digital-critical methods must become more than "mere containers" of language's curious residues [Posner et al. 2017, 4] [Burns 1994, 359].

Notes

[1] One might just as easily scan this essay's bibliography for other examples, but among them: McGrath et al. (2018); Ross and O'Sullivan (2016); Jeffrey Drouin, "Close- And Distant-Reading Modernism: Network Analysis, Text Mining, and Teaching *The Little Review*" (2014); Sean Weidman and James O'Sullivan, "The limits of distinctive words: Re-evaluating literature's gender marker debate" (2018).

[2] In fact, David James and Urmila Seshagiri recently bolstered that organizing logic, arguing how "[r]etaining modernism across deep time can dehistoricize it as a movement but repoliticize it as a global practice, a practice that serves instrumental ends in the context of cultural circumstances with which modernist writing has yet to be associated" [James and Seshagiri 2014, 90].

[3] Peter B. Hirtle of Cornell University Library hosts a "Copyright Information Center" page that's kept updated for the complex copyright terms of public domain in the US. See Hirtle's article, "When Is 1923 Going to Arrive and Other Complications of the U.S. Public Domain" (2012), for an explanation of why published works after 1923 have remained out of the public domain. As of 1 January, 2019, a variety of US-published works from 1923 have entered the public domain, and each subsequent year will see a correlative year's worth of once-copyrighted works do the same. An exception must be made, however, for Matthew Huculak and Claire Battershill's *Open Modernisms* project, which as of this writing represents an online archive of nearly 500 modernist works in various genres (see also Claire Battershill et al.'s *Making The Modernist Archives Publishing Project* (2017)). Otherwise, in the past half-decade, the HathiTrust Research Center is the only digital collection that offers access to sets of post-1923 texts for scholars doing computational research, and it comes with several restrictions.

[4] We should clarify that our use of the term *style* diverges from other variants used by modernist literary scholars (e.g. Rebecca Walkowitz), which tend to invoke the term loosely and synonymously with *form*, insofar as each denotes a literary pattern traceable through (for example) close reading. In the realm of stylometry, a notion of literary style is technical and tied to word use, and the term stands in for the aggregate of a set of formal, observable textual features. Our underlying assumption is that word preference can be measured by first counting words, both within individual texts and comparatively within a group of texts, and that those counts will be unique along different axes (e.g. different from writer to writer; genre to genre; etc.). By extension, those counts tell us something about the content and *style* of a text/writer relative to other texts. For a terrific account of the interdisciplinary use of the term among textual studies fields, see Herrmann et al. (2015).

[5] And this line probably owes a great critical debt to those recent critiques of computational literary study proffered by Katherine Bode (2017)

and Nan Z. Da (2019) — though, our study tends toward slightly more optimism: in the end, while we suggest stylistic critique needs to be wary of drawing firm conclusions about modern life via modernist fiction, we nevertheless think stylometry has much to offer the study of modernism.

[6] In this we uphold Losh et al.'s call for a “genuinely messy, heterogeneous, and contentious pluralism” as the underlying ethic of our digital methods, a critical approach that may also productively join — or productively digitize — the political investments of our data, its structures, and our own methods of analyzing and contextualizing modernism [Losh et al. 2016, 98].

[7] For Bode, this manner of viewing texts often ends in “dismiss[ing] the documentary record's multiplicity” [Bode 2017, 92].

[8] Burrows popularized the field of literary computational stylistics with his book, and his was one of the first to recognize the scales of semantic meaning within patterns of function word usage; although he doesn't really study gender in Austen explicitly, literary stylometrists continue to build on his principles.

[9] Refer also to Jan Rybicki (2016) and Mark Algee-Hewitt (2015) for two more essays utilizing similar methods to reach similar ends.

[10] Although not limited to modern literature, this claim is made even clearer and more forcefully by Earhart et al. (2020) in their recent account of gender and scholarly citational practices.

[11] In fact, aside from Lavin's essay about modern reviewing, we can think of only two studies that employ some form of quantitative formalist approach to modernist work and even vaguely relate it to gender. Stephen Ramsay's *Reading Machines* opens with a chapter that analyzes Virginia Woolf's *The Waves* in relation to its feminist criticism, and David Hoover's later “Argument, Evidence, and the Limits of Digital Literary Studies” positions itself directly opposite Ramsay's earlier study by rereading *The Waves* with different computational methods [Ramsay 2011] [Hoover 2016]. Among other thematically adjacent studies: see González et al. (2019) for an account of gender, stylometry, and *modernismo*; and outside of modernism's computational literary study, see Churchill et al. (2018) regarding their work on Mina Loy, style, and UX design and their interactive digital project of feminist modernist design (which builds on D'Ignazio and Klein's (2016) foundational feminist visualizations essay).

[12] Among other, longer accounts, Katherine Bode's *A World of Fiction: Digital Collections and the Future of Literary History* (2018) and Ted Underwood's *Distant Horizons: Digital Evidence and Literary Change* (2019) also each contain a section on gender and (mainly) 19th-century literature.

[13] Richard Jean So contends similarly that errors help us realign models to the unseen peripheries of data, and echoes Brown and Mandell's sentiment through the oft-quoted adage of famed statistician George E.P. Box: “All models are wrong, but some are useful” [So 2017, 669]. Andrew Piper makes a comparable remark when he considers that “[m]odeling puts computation not on the outside of what is known but as part of the process itself,” a reflexive process toward the contingency of knowledge he terms *the New Recursivity* [Piper 2016]. Piper also argues convincingly that every aspect of the modeling process, especially those required by its implementation on particular data selections, necessitates reduction — and that in its ubiquity reductiveness can actually be generative [Piper 2017, 654].

[14] A case in point about this own study, which began many years ago (and before the resources of HathiTrust were widely accessible): our corpus of women authors is almost entirely hand-scanned and OCRed, because the continued gender inequality of the literary marketplace, which Piper and So have studied, also occupies the realm of text digitization. Riddell and Bassett (2020) have measured this gender inequity, finding (in a corpus from the 1830s to the present day) that novels by women have been digitized at substantially lower rates than novels by men. For a more detailed account on the many levels of infrastructural relevance women writers require to be studied by digital methods, see Laura Mandell's “Gendering Digital Literary History: What Counts for Digital Humanities” (2016), then see Roopika Risam's “Navigating the Global Digital Humanities: Insights from Black Feminism” (2016) for a take on the complexity of foregrounding racial and multicultural diversity while doing so.

[15] While our analyses included the default standardization of all texts with z-scores (such that variations in a term like “novel-length” no longer come into play) and might have analyzed works of intentionally disparate lengths, we wanted to ensure our corpus maintained a cohesive genre; the phrase “novel-length fiction” is one common compromise.

[16] Give or take a year or two — we cheated, for example, to fit in Conrad's *Heart of Darkness* (1899).

[17] For the sake of space, the full text list isn't included here, but it is of course available upon request.

[18] We would be remiss if we didn't mention Evans and Wilkens (2018) as a recent computational study that adds to the mounting rationales

against such a canon. The authors argue convincingly that, when modeling British fiction as a whole (and not just its canonical works), the modernist period produced narrative attentions to international locales that greatly outnumbered national ones. Most modern British fiction, that is, spent more time discussing international milieus than not, which reaffirms concerns about the representative validity of something like an orthodox canon of British modern fiction.

[19] Hence conventional caveats apply: our corpus may indeed produce a narrative driven by ease-of-access or proximity over actual representativeness; a larger, more diverse, more global (in short, a “different”) corpus may have provided different results; and the claims we hope to make about gender and modernist style thus can’t reliably be extrapolated to modernist literature’s other flavors without further analysis. We admit this is a substantial, but so far largely unavoidable, limitation of studying the literary canon of a period still heavily under material and economic wraps. It’s reason, too, to be skeptical of the midrange scale of our study, which ends by looking at one text and one author and is thus not nearly as macroanalytic as most stylometric studies of literature. (It should be said that Marks Algee-Hewitt and McGurl explore this and other rationales of corpus-making in the Stanford Literary Lab’s *Pamphlet 8*, “Between Canon and Corpus: Six Perspectives on 20th-Century Novels,” 2-8.)

[20] Though, there are indeed more nuanced measures of stylistic distance that weight function and content words in different ways. Regardless, this general technique is often called the *bag of words* approach, and while it is popular in text analysis it also has its drawbacks. Its most basic model treats every word in each text in the same way, regardless of that word’s (a) syntactic or semantic contexts and (b) relation to the narrative or literary forms — thus, every novel merely becomes a countable bag of words. The approach makes measuring stylistic difference simple and effective, but in doing so it erases the context of all other literary elements in a work, which makes “reading” the resulting wordlists a precarious task. Take the exchange in *Orlando* between the newly acquainted Shelmerdine and Orlando as a brief example — and take first a jumbled approximation of what our machine sees, in the spirit of *big-data* experimentalism: “a” (2), “cried” (2), “[a]re” (2), “you” (2), “he” (1), “man” (1), “Orlando” (1), “she” (1), “Shel” (1), “woman” (1). What’s happened in this sequence, now that we’ve merely counted words and glossed over all context? Maybe a man and a woman were crying together, but who really knows? The flattening perils of decontextualized computational analysis are indeed laid bare. Now, here’s the moment as it appears in the actual narrative: “‘You’re a woman, Shel!’ she cried. ‘You’re a man, Orlando!’ he cried.” Even in a short, two-sentence example, Woolf’s tremendous gender-/word-play and its meaning for the speakers, who have fallen for one another (and soon marry) in part because of their gender non-conforming androgyny, is entirely lost if one merely rushed to count one’s computational chickens.

[21] Although we don’t engage with it here, Underwood and So have raised a few conceptual concerns with this approach, asking recently whether statistical distance and stylistic distance are comparable measures of relation at all [Underwood and So 2021].

[22] A basic note may be necessary regarding the difference between supervised and unsupervised machine learning techniques (though, there are other techniques that borrow from both types). Supervised techniques tend to require the input of pre-classified data to “learn from,” so that the algorithm can track and then predict future patterns from similarly classified data (e.g. one might train a road sign classification machine on thousands of different stop sign images, and then feed it other random images and ask it to output whether or not they feature stop signs). Unsupervised machine learning, conversely, tends to model the distribution or configuration of input data (e.g. providing a large, unsorted data set of road sign images and some metric of sorting them might provide correlations, groupings, or trends to help identify or separate those images). For further reading on this difference, and for the other variations of these methods, see Shalev-Shwartz and Ben-David (2014), 4-6.

[23] Our stylistic analysis was done entirely through Eder et al.’s *Stylo* package in the statistical computing program, R [Eder et al. 2016]. Although we’ve chosen a supervised analysis, its methodological limitations are significant. The benefit of unsupervised learning (e.g. PCA) is that the machine doesn’t know how many groups of data we think we’re studying, so we can’t privilege a group split just because we think there is one; the difficulty becomes identifying what features, exactly, constitute the groupings output by the machine. Inversely, the downside of a supervised analysis is that we organize the data in two pre-conceived groups and ask that each text be assigned to one group or the other. The benefit is that we at least think we have an idea of where machine-located differences are coming from — i.e. in this case, we isolate our groups based on gender. Both methods put us in an at least somewhat compromised position when drawing conclusions about gender and its stylistic features.

[24] One digital interpretation of the stylistic distinctiveness between the genders in modernism has been advanced elsewhere [Weidman and O’Sullivan 2018]. Again, as Mandell notes, this measure of “textual gender” alone is not terribly interesting; traditional social and cultural gender signaling is perfectly expected in literary work and, in this case, we also constructed our data to produce a split along that line of pre-sorted difference [Mandell 2019].

[25] We forced the machine, for example, to cull words that didn’t appear in most or all of the texts (to prevent particularly uncommon themes, narrative locations, or character names to artificially amplify differences); we used different MFW counts (50, 100, 200, ... 1000) to see if a more robust set of word frequencies would change our findings; we employed a variety of sampling methods, from analyzing each text in its entirety to

including only one or two random, bag-of-word samples; and we even swapped between statistical distance measures, seeing similar results from Burrows's Delta and the Wurzburg Cosine — though we talk more in footnotes 28 and 34 about why we limited this particular adjustment.

[26] This finding about Hall is especially intriguing; though we haven't space to discuss the novel here, Hall's protagonist, the "sexual invert" Stephen Gordon, is narrated through much of the same language as Woolf's in *Orlando*, a comparison that certainly warrants further study.

[27] Many of the studies from Jockers and Underwood, for example, discern as much in their 19th-century and contemporary corpuses, and alongside at least one other aforementioned piece [Cheng 2020], one of our prior essays confirms those findings in modernist fiction [Weidman and O'Sullivan 2018].

[28] For a detailed explanation of the *nearest shrunken centroid* (NSC) method of classification, see Jockers and Witten (2010). For a specific description of the technical features of NSC as applied in *Stylo*, see Eder (2016). As a stylometric classifier that finds an averaged stylistic profile, NSC provides certain benefits over a standard variant like Burrows's Delta in Eder's rolling method, which tries to pinpoint moments of stylistic "takeover" in a single text after being trained on one more groups of texts [Eder 2016, 459–63]. With Delta, each text in each training group is treated distinctly and is not consolidated into a stylistic profile (i.e. *Orlando* is measured against *each* training sample/text, ranking the "styles" that are closest to the test text); NSC, however, produces composites against which we can compare our test text (i.e. *Orlando* against the averaged style of all female texts or all male texts). For our study, where we want to isolate the gender binary as a two-class problem, NSC is a natural fit.

[29] This control turned out to matter little — when we reintroduced her remaining texts to the training set and re-ran the analysis, the results were nearly identical.

[30] Although we analyzed each text at 50-1000 MFWS with culling from 0-50% and received remarkably negligible variation in our results, for exact reproducibility's sake, each NSC classification image we included here was produced at 500 MFWS and 0% culling.

[31] Burrows's aforementioned 1986 study, among other contemporary versions — e.g. Pennebaker's *The Secret Life of Pronouns* (2011) — was among the first to explore just how important function words were to the close analysis of literary style. Remarkable, then, that Woolf seems to have divined and manipulated this reality half-a-century earlier and without the benefit of computational analysis.

[32] Several related essays warrant a brief mention here: Brenda Helt's "Passionate Debates on 'Odious Subjects': Bisexuality and Woolf's Opposition to Theories of Androgyny and Sexual Identity" (2010) provides an argument for using bisexuality to describe Woolf's depictions of desire; Jessica Berman's "Is the Trans in Transnational the Trans in Transgender" (2017) offers a discussion of how Orlando's transnational roaming contributes to Woolf's critique of imperial masculinity via Orlando's seamless gender transition; and Madelyn Detloff's "Camp Orlando (or) Orlando" (2016) provides an account of the camp sensibilities and reparative work of *Orlando*.

[33] We steal both examples directly from Emma Heaney's magnificent book, *The New Woman: Literary Modernism, Queer Theory, and the Trans Feminine Allegory* (2017). In fact, Heaney's exploration of the history and production of trans femininity provides an important clarification to our study's finding, detaching Woolf from a legacy of modernist trans feminism [Heaney 2017, 302 (note 11)].

[34] We understand, of course, that we still depend on averaged literary styles (via a bag-of-word list of most frequent terms) in this study, and we do not call that essential stylometric practice into question. What we're after here, rather, is a reimagining of how eschewing the inexact close-readings that tend to follow computational measures of style can actually help us accomplish something critically important. Deciphering averaged stylistic differences — or, what's more probable, incidentally lodging our pre-held assumptions into stylistic peculiarities — does not expand or contextualize our conclusions as much as it forecloses their messiness in almost-assuredly biased explainers. By restraining that stylometric impulse, we hope to let the text's relation to the corpus (rather than specific terminological connections/disparities) do the critical work for us. We want to venture an observation here, one helped along by Adam Kilgarriff's influential essay, "Language is never, ever, ever, random" (2005) and its adjacent, clarifying distinction between "randomness" and "arbitrariness" in interpretations of linguistic phenomena in literary corpora. That linguistic structures appear nonrandom or predictable in their relation does not make that relation meaningful — just as syntax can colonize meaning, narrative structures can demand certain forms, patterns, and distributions of linguistic content, which detaches judgments about their meaning from their possible literary-historical or sociocultural arbitrariness. Having used a specific corpus organization and specific modeling tools, all of which were designed to treat gender as a bimodal stylistic question, we think finding in *Orlando* a nonrandom and narratively nonarbitrary gender flip is more than a linguistic oddity.

[35] Following Rachael Scarborough King's delineation of form and genre, we might even call the literary style of gender a *genre*, an organizing metanarrative, "a collection whose members are assembled and whose boundaries are always permeable" [King 2021, 262].

[36] This is a problem Piper (2020) has recently aimed to tackle at much greater length and with much greater care than we do here.

[37] As Pryor notes elsewhere, “the category of modernism was developed through attention to exemplars” [Pryor 2011, 37].

[38] Etherington and Pryor also clarify this point: “because exemplarity conditions the production of knowledge, helping to construct the very object of inquiry, it is also a political problem” [Etherington and Pryor 2019, 5].

[39] This is a claim that has long since entered modernist criticism's common vocabulary, but see, among other examples: Maud Ellman, *Nets of Modernism: Henry James, Virginia Woolf, James Joyce, and Sigmund Freud* (2010); Matt Ffytch, “The Modernist Road to the Unconscious” (2012); and, for a classic account of this impact on modern culture generally, Michael North, *Reading 1922: A Return to the Scene of the Modern* (1999).

Works Cited

- Adwetewa 2020** Adwetewa-Badu, Ama Bemma. “Poetry from Afar: Distant Reading, Global Poetics, and the Digital Humanities.” *Modernism/modernity Print Plus* 5, no. 1 (2020).
- Algee-Hewitt 2015** Algee-Hewitt, Mark. “The Performance of Character: Digital Models for Gendered Speech in Romantic period Literature.” Lecture at Simon Fraser University, 2015.
- Algee-Hewitt and McGurl 2015** Algee-Hewitt, Mark and Mark McGurl. “Between Canon and Corpus: Six Perspectives on 20th-Century Novels.” *Literary Lab: Pamphlet* 8 (2015): 1-27.
- Battershill et al. 2017** Battershill, Claire, Helen Southworth, Alice Staveley, Michael Widner, Elizabeth Willson Gordon, and Nicola Wilson. *Scholarly Adventures in Digital Humanities: Making The Modernist Archives Publishing Project*. Routledge, 2017.
- Bode 2017** Bode, Katherine. “The Equivalence of ‘Close’ and ‘Distant’ Reading; or, Toward a New Object for Data-Rich Literary History.” *Modern Language Quarterly* 78, no. 1 (2017): 77-106.
- Bode 2018** Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. U of Michigan P, 2018.
- Brown and Mandell 2018** Brown, Susan, and Laura Mandell. “The Identity Issue: An Introduction.” *Cultural Analytics* (Feb. 2018).
- Burns 1994** Burns, Christy L. “Re-Dressing Feminist Identities: Tensions between Essential and Constructed Selves in Virginia Woolf’s *Orlando*.” *Twentieth Century Literature* 40, no. 3 (1994): 342- 64.
- Burrows 1986** Burrows, John F. *Computation into Criticism: A Study of Jane Austen’s Novels and an Experiment in Method*. Oxford: Clarendon Press, 1986.
- Butler 1990** Butler, Judith. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge, 1990.
- Butler 1993** Butler, Judith. *Bodies That Matter: On the Discursive Limits of “Sex*. Routledge, 1993.
- Caughie 2013** Caughie, Pamela L. “The Temporality of Modernist Life Writing in the Era of Transsexualism: Virginia Woolf’s *Orlando* and Einar Wegener’s *Man Into Woman*.” *Modern Fiction Studies* 59, no. 3 (2013): 501-25.
- Caughie et al. 2018** Caughie, Pamela L., Emily Datskou, & Rebecca Parker. “Storm clouds on the horizon: feminist ontologies and the problem of gender.” *Feminist Modernist Studies* 1, no. 3 (2018): 230-42.
- Cheng 2020** Cheng, Jonathan Y. “Fleshing Out Models of Gender in English-Language Novels (1850-2000).” *Cultural Analytics* (Jan. 2020).
- Christie et al. 2014** Christie, Alex, et al. “Manifesto of Modernist Digital Humanities.” Humanities Commons, 2014.
- Churchill et al. 2018** Churchill, Suzanne W., Linda A. Kinnahan, and Susan Rosenbaum. “Feminist designs: modernist digital humanities and *Mina Loy: Navigating the Avant-Garde*.” *Feminist Modernist Studies* 1, no. 3 (2018): 243-56.
- D’Ignazio and Klein 2016** D’Ignazio, Catherine, and Lauren F. Klein. “Feminist Data Visualization.” *IEEE VIS4DH Conference*, 2016.
- Da 2019** Da, Nan Z. “The Computational Case against Computational Literary Studies.” *Critical Inquiry* 45, no. 3 (2019): 601-39.
- Delaplace 2011** Delaplace, Anne. “Transcending Gender: Virginia Woolf and Marcel Proust, *Orlando* and *Albertine*.” *Virginia Woolf Bulletin* 37 (2011): 37-41.

- Drouin 2014** Drouin, Jeffrey. "Close- And Distant-Reading Modernism: Network Analysis, Text Mining, and Teaching *The Little Review*." *The Journal of Modern Periodical Studies* 5, no. 1 (2014): 110-35.
- Drucker 2017** Drucker, Johanna. "Why Distant Reading Isn't." *PMLA* 132, no. 3 (2017): 628-35.
- Earhart et al. 2020** Earhart, Amy E., Roopika Risam, and Matthew Bruno. "Citational politics: Quantifying the influence of gender on citation in *Digital Scholarship in the Humanities*." *Digital Scholarship in the Humanities*. Preprint, 3 August 2020.
- Eder 2016** Eder, Maciej. "Rolling stylometry." *Digital Scholarship in the Humanities* 31, no. 3 (2016): 457- 69.
- Eder et al. 2016** Eder, Maciej, Jan Rybicki, and Mike Kestemont. "Stylometry with R: A Package for Computational Text Analysis." *R Journal* 8, no. 1 (2016): 107-21.
- Etherington and Pryor 2019** Etherington, Ben, and Sean Pryor. "Historical poetics and the problem of exemplarity." *Critical Quarterly* 61, no. 1 (2019): 3-17.
- Evan and Wilkens 2018** Evans, Elizabeth F., and Matthew Wilkens. "Nation, Ethnicity, and the Geography of British Fiction, 1880-1940." *Cultural Analytics* (Jul. 2018).
- Golden and Laity 2018** Golden, Amanda, and Cassandra Laity, "Feminist Modernist Digital Humanities", *Feminist Modernist Studies* 1, no. 3 (2018), 205-210.
- Gonzalez et al. 2019** González, José Eduardo, Montserrat-Fuente Camacho, and Marcus Barbosa. "Detecting Modernismo's Fingerprint: A Digital Humanities Approach to the Turn of the Century Spanish American Novel." *Review: Literature and Arts of the Americas* 51, no. 2 (2019): 195-204.
- Hankins 2018 a** Hankins, Gabriel. "The Weak Powers of Digital Modernist Studies." *Modernism/Modernity* 25, no. 3 (2018): 569-585.
- Hankins 2018 b** Hankins, Gabriel. "We Are All Digital Modernists Now." *Modernism/Modernity Print Plus* 3, no. 2 (2018).
- Hayot and Walkowitz 2016** Hayot, Eric, and Rebecca L. Walkowitz, eds. "A New Vocabulary for Global Modernism". Columbia UP, 2016.
- Heaney 2017** Heaney, Emma. *The New Woman: Literary Modernism, Queer Theory, and the Trans Feminine Allegory*. Northwestern UP, 2017.
- Herrmann et al. 2015** Herrmann, J. Berenike, Karina van Dalen-Oskam, Christof Schöch. "Revisiting Style, a Key Concept in Literary Studies," *Journal of Literary Theory* 9, no. 1 (2015): 25-52.
- Hirtle 2012** Hirtle, Peter B. "When Is 1923 Going to Arrive and Other Complications of the U.S. Public Domain." *Searcher* 20, no. 6 (2012): 22-8.
- Hirtle 2018** Hirtle, Peter B. "Coypright Information Center." *Cornell University Library*. Jan. 10, 2018.
- Hoover 2016** Hoover, David L. "Argument, Evidence, and the Limits of Digital Literary Studies." *In Debates in the Digital Humanities* 2016, edited by Lauren F. Klein and Matthew K. Gold. U of Minnesota P, 2016, 230-50.
- Huculak 2018** Huculak, J. Matthew. "What Is a Modernist Archive?," *Modernism/Modernity Print Plus* (2018).
- Högberg and Bromley 2018** Högberg, Elsa and Amy Bromley. "Introduction: Sentencing Orlando." *Sentencing Orlando: Virginia Woolf and the Morphology of the Modernist Sentence*, edited by Elsa Högberg and Amy Bromley. Edinburgh UP, 2018, 1-14.
- James and Seshagiri 2014** James, David, and Urmila Seshagiri. "Metamodernism: Narratives of Continuity and Revolution," *PMLA* 129, no. 1 (2014): 87-100.
- Jockers 2013** Jockers, Matthew. *Macroanalysis: Digital Methods and Literary History*. U of Illinois Press, 2013.
- Jockers and Kirilloff 2016** Jockers, Matthew, and Gabi Kirilloff. "Understanding Gender and Character Agency in the 19th Century Novel." *Cultural Analytics* (Dec. 2016).
- Jockers and Witten 2010** Jockers, Matthew L. and Daniela M. Witten. "A Comparative Study of Machine Learning Methods for Authorship Attribution." *Literary and Linguistic Computing* 25, no. 2 (2010): 215-23.
- Kilgarriff 2005** Kilgarriff, Adam. "Language is never, ever, ever, random." *Corpus Linguistics and Linguistic Theory* 1, no. 2 (2005): 263-76.

- King 2021** King, Rachael Scarborough. "The Scale of Genre." *New Literary History*, vol. 52, no. 2 (2021): 261-84.
- Kirilloff et al. 2018** Kirilloff, Gabi, Peter J. Capuano, Julius Fredrick, and Matthew L. Jockers. "From a distance 'You might mistake her for a man': A closer reading of gender and character action in *Jane Eyre*, *The Law and the Lady*, and *A Brilliant Woman*." *Digital Scholarship in the Humanities* 33, no. 4 (2018): 821-44.
- Kraicer and Piper 2019** Kraicer, Eve, and Andrew Piper. "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction," *Cultural Analytics* (Jan. 2019).
- Lavin 2020** Lavin, Matthew J. "Gender Dynamics and Critical Reception: A Study of Early 20th-century Book Reviews from *The New York Times*." *Cultural Analytics* (Jan. 2020).
- Long and So 2013** Long, Hoyt, and Richard Jean So. "Network Analysis and the Sociology of Modernism," *boundary 2* 40, no. 2 (2013).
- Long and So 2016 a** Long, Hoyt, and Richard Jean So. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning." *Critical Inquiry* 42, no. 2 (2016): 235-67.
- Long and So 2016 b** Long, Hoyt, and Richard Jean So. "Turbulent Flow: A Computational Model of World Literature." *Modern Language Quarterly* 77, no. 3 (2016): 345-67.
- Losh et al. 2016** Losh, Elizabeth, Jacqueline Wernimont, Laura Wexler, and Hong-An Wu. "Putting the Human Back into the Digital Humanities: Feminism, Generosity, and Mess." In *Debates in the Digital Humanities 2016*, edited by Lauren F. Klein and Matthew K. Gold. U of Minnesota P, 2016, 92-103.
- Mandell 2016** Mandell, Laura C. "Gendering Digital Literary History: What Counts for Digital Humanities." In *A New Companion to Digital Humanities*, 2nd ed, edited by Susan Schreibman, Ray Siemens, and John Unsworth. Wiley-Blackwell, 2016, 511-23.
- Mandell 2019** Mandell, Laura C. "Gender and Cultural Analytics: Finding or Making Stereotypes?" *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein. UP of Minnesota, 2019.
- Mao and Walkowitz 2008** Mao, Douglas, and Rebecca Walkowitz. "The New Modernist Studies," *PMLA* 123, no. 3 (2008): 737-48.
- McGrath et al. 2018** McGrath, Laura B., Devin Higgins, and Arend Hintze. "Measuring Modernist Novelty." *Cultural Analytics* (Nov. 2018).
- Nelson 1986** Nelson, Cary. *Repression and Recovery: Modern American Poetry and the Politics of Cultural Memory, 1910-1945*. U of Wisconsin P, 1989.
- North 1999** North, Michael. *Reading 1922: A Return to the Scene of the Modern*. Oxford UP, 1999.
- Piper 2016** Piper, Andrew. "There Will Be Numbers." *Cultural Analytics* (May 2016).
- Piper 2017** Piper, Andrew. "Think Small: On Literary Modeling." *PMLA* 132, no. 3 (2017): 651-8.
- Piper 2020** Piper, Andrew. *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Elements in Digital Literary Studies, edited by Katherine Bode, Adam Hammond, and Gabriel Hankins. Cambridge UP, 2020.
- Piper and So 2016** Piper, Andrew, and Richard Jean So. "Women Write About Family, Men Write About War." *New Republic*, Apr. 8, 2016.
- Posner et al. 2017** Posner, Miriam and Lauren F. Klein. "Editor's Introduction: Data as Media." *Feminist Media Histories* 3, no. 3 (2017): 1-8.
- Pressman 2014** Pressman, Jessica. *Digital Modernism: Making It New in New Media*. Oxford UP, 2014.
- Pryor 2011** Pryor, Sean. "A poetics of occasion in Hope Mirrlees's *Paris*." *Critical Quarterly* 61, no. 1 (2019): 37-53.
- Ramsay 2011** Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. U of Illinois P, 2011.
- Riddell and Barrett 2020** Riddell, Allen and Troy J. Barrett. "What Library Digitization Leaves Out: Predicting the Availability of Digital Surrogates of English Novels." *arXiv:2009.00513 [cs.DL]*. Preprint, 1 September 2020.
- Risam 2016** Risam, Roopika. "Navigating the Global Digital Humanities: Insights from Black Feminism." In *Debates in the Digital Humanities 2016*, edited by Lauren F. Klein and Matthew K. Gold. U of Minnesota P, 2016, 359-67.

- Ross and O'Sullivan 2016** Ross, Shawna, and James O'Sullivan, eds. *Reading Modernism with Machines: Digital Humanities and Modernist Literature*. Palgrave Macmillan, 2016.
- Ross and Sayers 2014** Ross, Stephen, and Jentery Sayers. "Modernism Meets Digital Humanities." *Literature Compass* 11, no. 9 (2014): 625-633.
- Rubin 1975** Rubin, Gayle. "The Traffic in Women: Notes on the 'Political Economy' of Sex." In *Toward an Anthropology of Women*, edited by Rayna R. Reiter. Monthly Review Press, 1975, 157-210.
- Rybicki 2016** Rybicki, Jan. "Vive la différence: Tracing the (authorial) gender signal by multivariate analysis of word frequencies." *Digital Scholarship in the Humanities* 31, no. 4 (2016): 746-61.
- Shalev-Schwartz and Ben-David 2014** Shalev-Shwartz, Shai, and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge UP, 2014.
- So 2017** So, Richard Jean. "All Models Are Wrong." *PMLA* 132, no. 3 (2017): 668-73.
- Stanford 2015** Stanford Friedman, Susan. *Planetary Modernisms: Provocations on Modernity Across Time*. Columbia UP, 2015.
- Underwood 2019** Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. U Chicago P, 2019.
- Underwood 2020** Underwood, Ted. "Machine Learning and Human Perspective." *PMLA* 135, no. 1 (2020): 92-109.
- Underwood and So 2021** Underwood, Ted and Richard Jean So, "Can We Map Culture?" *Cultural Analytics* (June 2021).
- Underwood et al. 2018** Underwood, Ted, David Bamman, and Sabrina Lee. "The Transformation of Gender in English-Language Fiction." *Cultural Analytics* (Feb. 2018).
- Weidman and O'Sullivan 2018** Weidman, Sean G., and James O'Sullivan. "The limits of distinctive words: Re-evaluating literature's gender marker debate," *Digital Scholarship in the Humanities* 33, no. 2 (2018): 374-90.
- Wollaeger 2012** Wollaeger, Mark, and Matt Eatough, eds. *The Oxford Handbook of Global Modernisms*. Oxford UP, 2012.
- Woolf 1992** Woolf, Virginia. *Orlando: A Biography*. (1928). Vintage, 1992.
- Woolf 2009** Woolf, Virginia. "Poetry, Fiction and the Future." *Selected Essays*. Oxford UP, 2009, 74-84.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.