

The Generative Dissensus of Reading the Feminist Novel, 1995-2020: A Computational Analysis of Interpretive Communities

Lisa Mendelman, Anna Mukamal

Lisa Mendelman, Menlo College

Anna Mukamal, Stanford University

Peer-Reviewers: Lauren Klein, Laura Mandell

Data Repository: <https://doi.org/10.7910/DVN/4ATJCF>

Article DOI: <https://doi.org/10.22148/001c.30009>

ABSTRACT

This article furthers ongoing work on the merits of the feminist novel's intrinsic variability by probing its dynamics in four publishing contexts: contemporary anglophone literary criticism, prestigious review publications, marketing materials, and online book reviews by social readers. We explore how these interpretive communities converge and diverge in their assessments of feminist fiction over the past twenty-five years by evaluating articles from the *MLA International Bibliography*, book reviews in *The New York Times*, *The New Yorker*, *Times Literary Supplement*, and other prominent periodicals, blurbs from Amazon, and *Goodreads* reviews. We trace the feminist novel's ambivalent fates—or rather, feminist novels' ambivalent fates—in and across these four domains. To do so, we engage computational methods of topic modeling, most distinctive word analysis, and named entity recognition. We synthesize these quantitative results with qualitative attention to provocative examples from our corpus. In so doing, we consider how literary scholars can develop more robust understandings of what feminism and feminist fiction mean to contemporary readers and what we stand to gain by bringing this diverse interpretive labor into our scholarly conversations.

Our synthetic interpretive approach reveals these communities' shared topical investments in feminist fiction, though the communities talk about these topics in importantly different ways. Together, their discourse converges on two organizing concerns: embodied subjectivity and temporality. Different configurations of these aspects of personhood in time inform the communities' vocabularies, their modes of self-address, the rationales they offer for reading feminist novels, and the forms of feminist subjectivity they promote. Our analysis thus demonstrates how novel reading can function as a mode of forging feminist knowledge and constructing feminist value systems: through imagined conversations with other readers, individuals and the interpretive communities in which they participate undertake this additive, pluralistic revisionary process with a connection to the past, a grounding in the present, and a commitment to ever-better futures.

What do you hear when you hear the word *feminism*? It is a word that fills me with hope, with energy. ... It brings to mind books written, tattered and worn, books that gave words to something, a feeling, a sense of an injustice, books that, in giving us words, gave us the strength to go on. Feminism: how we pick each other up. So much history in a word; so much it too has picked up.

—Sara Ahmed, *Living a Feminist Life* (2017)¹

Sara Ahmed's *Living a Feminist Life* (2017) asserts feminism's definitional instability as its strength and locates this asset in a literary genealogy that reciprocally influences feminism's capacious, shapeshifting project.² Participating in a nascent scholarly trend, Ahmed argues that feminism does not comprise a coherent set of beliefs, opinions, convictions, or political positions, but rather reflects a shared imperative of asking difficult, ethically-minded questions without holding out hope for a stable set of answers.³ This self-reflexive life project engages literature as a formative object in part because it provides models of feminist value systems—sometimes irreconcilable, sometimes harmonious, sometimes disappointing, but always operating as a spur to sharpen the reader's own sense of what it might mean to "live a feminist life."

This recent attention to feminism's definitional instability echoes a longer-standing critical ambivalence about defining feminist literature and the feminist novel in particular. Over the past four decades, scholars such as Nancy A. Walker and Rita Felski have articulated this ambivalence largely obliquely, skirting definitions of feminist fiction or avoiding ideological meanings by emphasizing literary elements (e.g., female protagonists, realist aesthetics) and only gesturing to sociopolitical implications.⁴ Ahmed participates in this tendency to elision, indirection, and even outright resistance when it comes to establishing exactly what constitutes the feminist novel.⁵ But perhaps these gymnastics are unsurprising: boundaries and distinctions seem in tension with many of these scholars' thesis that feminism is an inclusive, additive, evolving project that "creates endless eddies and currents that flow in often surprising and unexpected directions" in fiction.⁶ We extend this line of argument by surfacing contemporary readers' varied understandings of feminist novels—understandings that, as we detail below, ultimately circle around feminist

novels as works of fiction that somehow engage with a sociopolitical project of equity for all people, regardless of sex, gender, sexuality, and other categories of identity.

Our project furthers this ongoing work on the merits of the feminist novel's intrinsic variability by probing its dynamics in four publishing contexts: contemporary anglophone literary criticism, prestigious review publications, marketing materials, and online book reviews by social readers. We explore how these interpretive communities converge and diverge in their assessments of feminist fiction over the past twenty-five years by evaluating articles from the *MLA International Bibliography*, book reviews in *The New York Times*, *The New Yorker*, *Times Literary Supplement*, and other prominent periodicals, blurbs from Amazon, and *Goodreads* reviews. We trace the feminist novel's ambivalent fates—or rather, feminist novels' ambivalent fates—in and across these four domains. To do so, we build a corpus comprised of 335 documents and nearly 5 million words drawn from their digital platforms, in which the authors identify the text at hand as a feminist novel. We analyze this corpus using a series of computational methods—topic modeling, most distinctive word analysis, and named entity recognition—to illuminate the rhetorical patterns of our subcorpora and then close read our findings to explore their implications for the feminist novel. Specifically, we focus on which novels different interpretive communities choose to discuss and how they discuss them. In so doing, we consider how literary scholars can develop more robust understandings of what feminism and feminist fiction mean to contemporary readers and what we stand to gain by bringing this diverse interpretive labor into our scholarly conversations.

This approach extends ongoing work in reception studies by combining computational and close reading methods to evaluate prose content produced by literary critics, book reviewers for high-profile periodicals, Amazon's marketing team, and social readers. As Melanie Walsh and Maria Antoniak have recently proposed, “computational methods and internet data [from platforms like *Goodreads* and Amazon], when combined, can help literary critics capture the creative explosion of reader responses” and account for “readers' own critical voices”—particularly when the resulting findings are qualitatively analyzed as well.⁷ Indebted to Janice Radway's sociological work on gendered reading communities and genre fiction, our synthetic attention to the readerly discourse of “actual subjects in history” across

multiple interpretive communities allows us to examine the roles that sex and gender play in contemporary reading cultures and helps us to consider feminism's diverse social purchase in the twenty-first-century English-speaking world.⁸

Our corpus and our methodology iterate our argument that the feminist novel is multiple, collectively arbitrated, and open-ended—by definition and like feminism itself. Our corpus comprises contemporary readers' writing about novels rather than the text of these novels themselves, as noted above. As we explain in detail below, we recruit these platforms' internal algorithms to produce this corpus, first generating their respective lists of feminist novels and then populating the text associated with these lists on each of the platforms (journal article content, book reviews, marketing materials). This methodology allows us to examine how different interpretive communities themselves classify and discuss novels vis-à-vis feminism, rather than requires us to posit our own categorical assessments of what constitutes feminist fiction. We refer to these subcorpora as interpretive communities because they are comprised of published prose that addresses a network of other readers. As Stanley Fish defines this foundational concept of reader response theory: “interpretive communities are made up of those who share interpretive strategies not for reading but for writing texts, for constituting their properties.”⁹ We are especially interested in how the binding “interests, ... purposes and goals” that drive this collaborative meaning-making can be multiple, contentious, and unsettled—and productively so.¹⁰

In the first two sections of this article, we contextualize our inclusive method of corpus construction and its resulting hybrid dataset in terms of recent feminist work in the digital humanities (DH) and in literary criticism centered on gender and reception. In the third section, we gloss the metadata that situates our four interpretive communities as intersecting but importantly distinct. The subsequent four sections elucidate this generative dissensus by first employing a given computational tool and then qualitatively evaluating our findings, both as a whole and in terms of specific provocative examples. Our synthetic interpretive approach—which we detail at length in our methods section—reveals these communities' shared topical investments in feminist fiction, though we will also demonstrate that the communities talk about these topics in importantly different ways. Together, their discourse converges on two organizing concerns: embodied

subjectivity and temporality. These concerns align with Andrew Piper’s finding that novelistic discourse emphasizes “phenomenological encounter,” with a “linguistic investment [that . . .] falls overwhelmingly on sense perception and a sense of human embodiment.”¹¹ Piper argues that novelistic discourse focuses on “[n]ot the world itself, but a person’s encounter with and reflection upon that world—the world’s *feltness*.”¹² A key finding of our work is that discourse *about* the novel shares this focus. Moreover, we find that this nonfictional discourse takes up questions of embodied subjectivity in terms of the temporally-marked “social horizon”—“things ‘out there’”—that Piper identifies with nonfiction.¹³ Readers from the four interpretive communities that we analyze consistently bring this “combination of sense perception plus cognitive skepticism” to bear in discussions of feminist considerations and then extend this phenomenological orientation outward—linking the personal to the political, to invoke a clichéd feminist truism.¹⁴

More specifically, readers assess novels’ feminist qualities in terms of three aspects of temporally-oriented embodied subjectivity: 1) as the experience of living in our minded bodies, particularly in terms of thinking and feeling; 2) as socially constructed, especially in terms of intersectional categories of identity (e.g., by race, class, and ability as well as by sex and gender); and 3) as relays between individual, idiosyncratic being and collective life and community belonging. As we discuss in sequence below, the configurations of these aspects of personhood in time inform the communities’ vocabularies (“What We Talk About When We Talk About Feminist Novels”), their modes of self-address (“How We Talk *Distinctively* When We Talk About Feminist Novels”), the rationales they offer for reading feminist novels (“Why Read Feminist Novels?”), and the forms of feminist subjectivity they promote (“Who We Talk About When We Talk About Feminist Novels”). Our analysis thus suggests how novel reading can function as a mode of forging feminist knowledge and constructing feminist value systems: through imagined conversations with other readers, individuals and the interpretive communities in which they participate undertake this additive, pluralistic revisionary process with a connection to the past, a grounding in the present, and a commitment to ever-better futures. In so doing, contemporary readers reflexively redefine understandings of the feminist novel. Put differently: if these interpretive communities posit that there is no such thing as *the* feminist novel, we observe in turn that feminist novels are always in the making by their readership.

Feminist Scholarship in Digital Humanities and Literary Criticism

Our project takes its archival, methodological, and other theoretical and practical cues from ongoing work in feminist DH that foregrounds synthesis, inclusivity, and embodiment as crucial modes of knowledge production. In *Data Feminism* (2020) Catherine D’Ignazio and Lauren F. Klein ask, “What makes a project feminist? ... [A] project may be feminist in content, in that it challenges power by choice of subject matter; in form, in that it challenges power by shifting the aesthetic and/or sensory registers of data communication; and/or in process, in that it challenges power by building participatory, inclusive processes of knowledge production.”¹⁵ Similarly, Elizabeth Losh and Jacqueline Wernimont’s edited collection *Bodies of Information: Intersectional Feminism and Digital Humanities* (2018) pushes beyond “the representational politics of feminisms” to “assert that these feminisms function as sophisticated forms of critical theory.”¹⁶ In particular, Losh and Wernimont highlight “the material, situated, contingent, tacit, embodied, affective, labor-intensive and political characteristics of digital archives and their supporting infrastructures and practices” that feminist DH practitioners utilize to “express[] their concerns about present-day power relations and signify[] interest in collective and communal consciousness-raising efforts.”¹⁷ Our corpus gives equal attention to, and derives a majority of content from, non-scholarly critics, and, as we detail below, our computational and qualitative methods seek to gain purchase on readers’ reflections on their “material, situated, contingent” experiences of being in the world.

Our work also dialogues with recent scholarship that looks beyond literary content to consider how sex and gender shape the broader field of anglophone cultural production. This work participates in a tradition of sociological criticism focused on gender and reception, such as Janice Radway’s ethnographic studies of the readership of romance novels and Joan Shelley Rubin’s recovery of the tastemaking institutions and individuals engaged with so-called middlebrow culture.¹⁸ Recent efforts to rethink literary history by centering marginalized interpretive communities analyze “mass-produced literary texts ... [as] used by real people” in conversation with more traditionally authoritative cultural sources as well as with one another and

engage in a multi-modal practice to do so.¹⁹ For instance, Rita Felski's *Hooked: Art and Attachment* (2020) "looks closely at how people connect to novels" and other media by combining ethnography, biography, and book history, among other modes of inquiry, and does so in order to draw out "similarities that are often overlooked" between "ordinary and academic interpretation."²⁰ Timothy Aubry differently elucidates how "readers creatively appropriate and reimagine" contemporary American novels "in order to make them serve a variety of personal and practical functions" by merging literary critical readings of these texts with analyses of *The Oprah Winfrey Show's* Book Club discussions, Amazon Customer Reviews, and more.²¹ In a third example of work along these lines, Elizabeth Long harnesses longitudinal ethnographic interviews with participants in women's book clubs along with cultural history and close reading to probe the intellectual and interpersonal import of discussing private reading in social collectives.²² Long and Aubry situate themselves in Radway's lineage in a manner akin to Felski's opening gambit: "How, asks Janice Radway, can we more fully engage the vitality and ongoingness, the forward trajectory and the unsecured nature, of social processes" and do so "without pitching aesthetic experience outside the social world"?²³ Our project likewise draws its critical and creative energy from this provocation.

DH scholars participate in this ongoing line of critical inquiry by combining computational and traditional approaches to explore textual objects beyond canonical literature and readers beyond the academy. Matthew J. Lavin examines how the "crucial categorical norm" of gender influences critical reception and cultural capital by using machine learning and traditional book historical approaches to evaluate early twentieth-century *New York Times* book reviews.²⁴ Finding "gendered patterns in both subject matter and structural vocabularies of [these] book reviews," Lavin identifies a "division between 'what men write about' and 'what women write about' that has not been observed when primary texts such as novels were analyzed using large scale, computational methods."²⁵ Karen Bourrier and Mike Thelwall differently explore how gendered cultural expectations drive "the social lives of books" by comparing the incidence of women authors and female characters in Victorian literature on *Goodreads*, in the *MLA International Bibliography*, and on syllabi from the *Open Syllabus Project*.²⁶ Bourrier and Thelwall find "a strong correlation between" academic and popular selections of Victorian literature but observe importantly distinct modes of discussing these

works: *Goodreads* reviews manifest supposedly “unacademic forms of reading” such as “readerly identification, reading for character, and reading for plot” that are no less sophisticated and politically engaged in their commitments to strong female protagonists.²⁷ Melanie Walsh and Maria Antoniak similarly seek “to hear nonacademic readers’ voices” at a scale “difficult if not impossible before the internet” by analyzing 120,000 reviews of 144 “classics” as most shelved and read by *Goodreads* users.²⁸ While demonstrating that *Goodreads* reviewers participate in a robust, cross-media “classics industry,” Walsh and Antoniak emphasize that the platform nonetheless compensates these users by offering them “a chance to reflect on their lives and relationships to power, a conduit for connecting to others, and an opportunity to enter a critical conversation that has long excluded them.”²⁹ Thus, in addition to these projects’ topical interests in sex and gender, their praxis also aligns with feminist investments as they seek to leverage DH methods to interpellate additional interpretive communities into academic discourse and, in so doing, to democratize the intellectual labor of literary criticism.

Our project also participates in ongoing work in feminist literary criticism by scholars focused on anglophone prose production. Recent work in this field tends to employ a both/and strategy in response to a lineage of frequently polarized approaches to literature by and about woman-identified individuals. Monographs by Suzanne Bost, Shelley Fisher Fishkin, Sharon Patricia Holland, Tanya Ann Kennedy, and Emily Westkaemper and anthologies such as *Anthropocene Feminism* (2017) and *The Bloomsbury Handbook of 21st-Century Feminist Theory* (2019) discuss literary objects in terms of their aesthetic qualities and their cultural work, read canonical and avant-garde works together with mass-market/popular/middlebrow/“genre” fiction, consider the overlapping audiences and plural reception histories of this media, and foreground the intersectional identities—and especially the formative work of race, ethnicity, sexuality, class, ability, and nationality—that shape individuals’ actually lived lives as well as textual representations of embodied experience.³⁰ Like Bourrier and Thelwall’s project, much of this work seeks to revalue supposedly naïve reading practices including a focus on plot and narrative realism and an interest in character and the workings of attachment and identification.³¹

This work variously responds to the questions Felski poses at the outset of *Literature*

After Feminism (2003): “How ... does feminist criticism change the way we think about readers? Does plot have a gender? How have feminists talked about female authors? What is the role of value in feminist scholarship?”³² We, too, take up these questions in exploring how different readers answer her second, third, and fourth questions. As we detail following the discussion of our methodology below, our analysis indicates that individuals identify and imagine different plot structures as feminist. For example, a three-star *Goodreads* review of Zora Neale Hurston’s *Their Eyes Were Watching God* (1937) asserts, “Janie was badass up until she met Tea Cake. Then it was ‘Feminism? What feminism?’” Yet another *Goodreads* reviewer locates Hurston’s feminism in this romantic plot, claiming that “many of us ... don’t recognize the power within ourselves to be strong and capable until we’ve been screwed over by a man.” Within and across interpretive communities, readers differently evaluate the same narrative in terms of progressive values and commitments to equity and prosocial change. Our answer to Felski’s first question, then, includes a pluralization of reading practices and communities, with a particular interest in the experiential as well as ideological values these practices prioritize and these communities assert.

Methods & Corpus

Our corpus consists of a total of 4,945,592 words comprised of literary criticism (LC), prestige review publications (PR), Amazon marketing materials (AM), and *Goodreads* reviews (GR), all published over the past quarter century. We built this corpus as follows (see Figure 1 for a tabular overview):

1. We began with LC. We searched in the full text of articles indexed by the *MLA International Bibliography* for “feminist novel” and “feminist fiction,” filtering for peer reviewed journal articles published on or after January 1, 1995 written in English about anglophone novels or novels translated to English of all periods. We seek to surface readers’ understandings of the feminist novel, but we needed to define some search terms to identify these objects. We therefore determined to use the most explicit identification available: “feminist novel.” This initial search term yielded 34 results. We wanted a larger corpus and more novels identified as feminist, so we expanded to “feminist fiction,” which yielded 21 results. When we excluded articles that

did not identify specific novels as feminist (e.g. short book reviews of monographs) or that were about non-anglophone novels, the results yielded a total corpus of 40 literary critical articles that identified 58 discrete novels as feminist. We downloaded these articles using our institutional permissions and created plain text files of the full text, including titles, abstracts (if published in the downloaded article), and footnotes, producing a total subcorpus of 314,975 words.³³

2. For PR, we searched for “feminist novel” in *The New York Times Book Review*, *The New Yorker*, *The Los Angeles Review of Books*, *TIME*, *Times Literary Supplement*, *The Washington Post*, *Slate*, *Jezebel*, and *London Review of Books*’ respective search functions, filtering for articles published on or after January 1, 1995 about anglophone novels or novels translated to English in all periods.³⁴ This search returned 97 prestige review articles that identified 87 discrete novels as feminist. We downloaded these articles using our institutional permissions and created plain text files of the full text, including titles, producing a total subcorpus of 144,131 words.
3. For AM, we searched Amazon Books for “feminist novel.” Amazon’s search function does not include a filter for the date of materials release or web posting, but as the site did not begin selling media until 1998, all materials were produced after January 1, 1995. This search returned 240 results. In order to build a subcorpus with a word count in the same order of magnitude as our other subcorpora, we determined to use the top 128 novels written in or translated to English across all time periods. For these 128 novels, we created plain text files of blurbs, editorial reviews, and first chapter “preview” excerpts (if published on Amazon’s site), producing a subcorpus totaling 152,587 words.³⁵
4. For GR, we identified the top 70 novels written in or translated to English shelved—meaning identified by readers—as “feminist,” as of May 2020. *Goodreads*’ search function does not include a filter for the date of materials release or web posting, but as the site did not launch until January 2007, all materials were produced after January 1, 1995. We included novels from all time periods. The “feminist” shelf contains many texts that are not novels, such as memoirs and cultural histories, which we excluded. We also excluded any reviews not written in English. We created plain text files of the reviews of each novel, including roughly equivalent word counts for each of the five

star categories (1 to 5 stars), producing a subcorpus totaling 4,333,899 words across 350 documents (5 documents per each of the 70 novels). As we explain in relevant sections below, for certain methods we downsampled this data by including reviews from across all star categories in one plain text file for each novel to produce a representative sample of 358,551 words across 70 documents, yielding a more comparable number of words across subcorpora.

For each of these novels, we populated metadata including publication year and author gender.

Subcorpus	Number of distinct documents	Number of distinct feminist novels	Total number of words
Literary criticism (LC)	40	58	314,975
Prestige reviews (PR)	97	87	144,131
Amazon marketing (AM)	128	128	152,587
Goodreads (GR)	350 total, downsampled to 70	70	4,333,899 total, downsampled to 358,551

Figure 1. Summary at a glance of the contents and size of each subcorpus; we detail the overlaps in titles (column three here) in the following section.

These communities overlap in their content ownership and production as well as their interpretive interests. Most notably, Amazon owns at least half of our subcorpora, as the company acquired *Goodreads* in 2013 for \$150 million.³⁶ Many readers also participate in multiple interpretive communities. For instance, young adult (YA) novelist Jennifer Mathieu, author of the 2017 YA novel *Moxie* (identified as feminist by GR), recently reviewed a children’s book about middle school dress codes in *The New York Times*.³⁷ Similarly, *Washington Post* book critic Ron Charles often self-markets by posting snippets from his reviews on *Goodreads*, just as many

social readers direct readers to their personal book review blogs hosted off *Goodreads*. We attend to these exchanges by comparing computational models of individual interpretive communities to those of our corpus as a whole and elaborating these models through close reading of each community’s published prose.

Our methodology recruits the communities’ internal logics to generate their respective lists of novels, via their proprietary platforms, search algorithms, and organizing functions (see Appendix for complete lists). This approach bears out Katherine Bode’s call for centering the “real epistemic, ethical, and political challenges” of constructing DH data sets and, more broadly, of “building a field in which humanities and statistical ways of knowing coexist and enhance each other.”³⁸ It also iterates D’Ignazio and Klein’s argument for “participatory, inclusive processes of knowledge production” in DH work.³⁹ This approach further performs intersectional feminist commitments to theorizing individual experience in dialogue with collective understanding, as it allows us to observe textual patterns across many readers’ prose artifacts and then synthesize this larger-scale cultural analysis with attention to individual readers’ responses.

We describe these methods in greater detail at the beginning of each section that follows. We start with a discussion of our metadata about the novels in our corpus, which we uploaded into and analyzed using Tableau. We then turn to the computational methods we pursued in sequence.

Community Consensus

As Figure 2 demonstrates, only one author appears in all four subcorpora, Marilyn French. Nine authors appear in three of the four subcorpora: Naomi Alderman, Margaret Atwood, Kate Chopin, Erica Jong, Chris Kraus, Doris Lessing, Toni Morrison, Alice Walker, and Virginia Woolf.

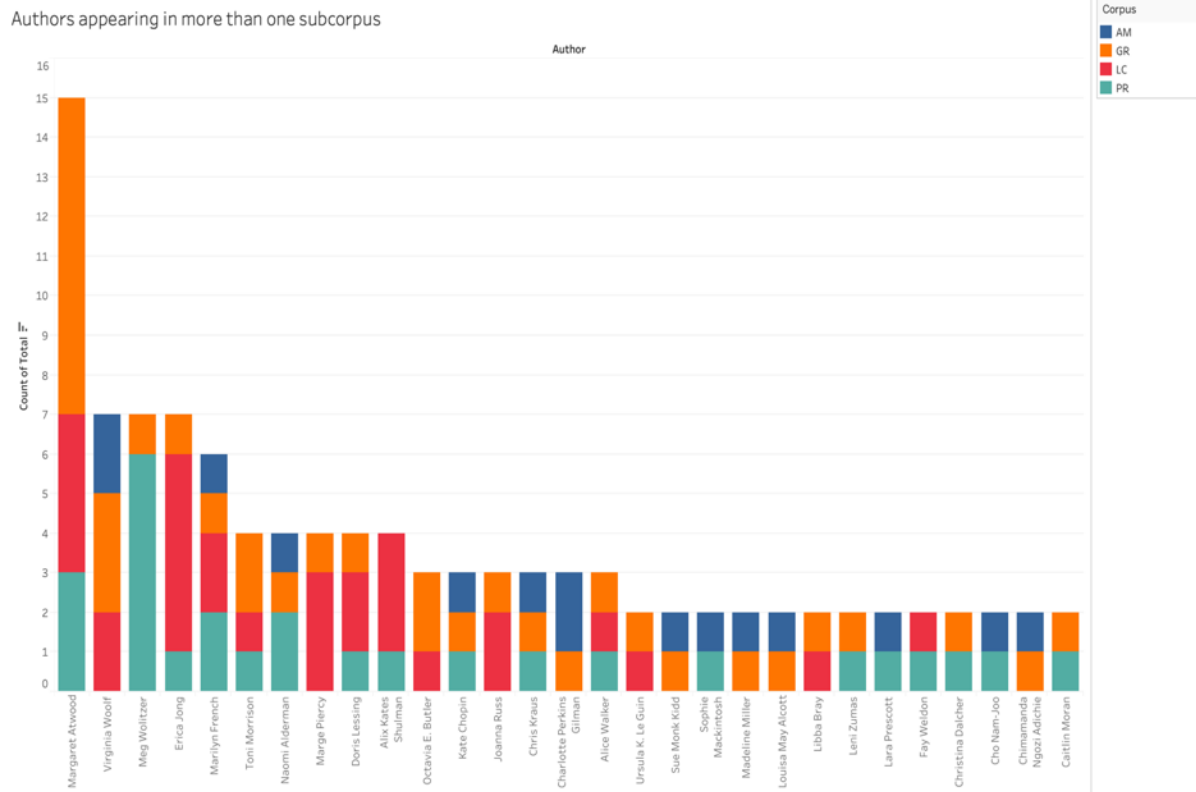


Figure 2. Authors appearing in more than one subcorpus, sorted in descending order by frequency of appearance, colored by subcorpus. For complete lists of each author appearing in each subcorpus, please see Appendix.

Figure 3 shows that there is no single novel that appears in all four subcorpora, but four novels appear in three of four subcorpora: Atwood’s *The Handmaid’s Tale* (1985), Chopin’s *The Awakening* (1899), Kraus’s *I Love Dick* (1997), and Alderman’s *The Power* (2016). There are no novels in common between AM and LC, suggesting the divergence of literary critics and consumer-driven marketers as well as reflecting the feminist novel’s multiplicity and collective arbitration in the contemporary literary field.

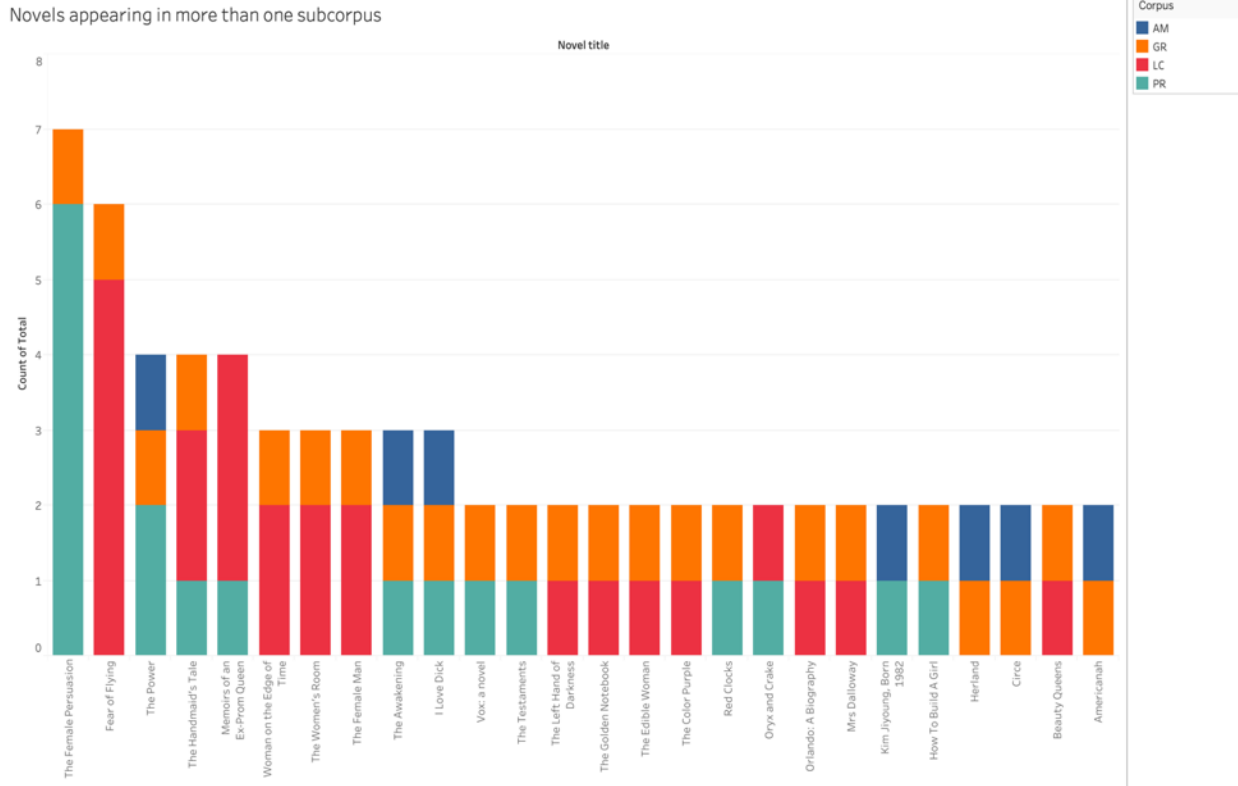


Figure 3. Novels appearing in more than one subcorpus, sorted in descending order by frequency of appearance, colored by subcorpus. For complete lists of each novel appearing in each subcorpus, please see Appendix.

Individual interpretive communities also express more internal consensus about authors than novels. The following authors and texts appear more than twice in the same domain: Erica Jong’s *Fear of Flying* (1973) appears in five literary critical articles, while Alix Kates Shulman’s *Memoirs of an Ex-Prom Queen* (1972) appears in three articles. Meg Wolitzer’s *The Female Persuasion* (2018) appears in six prestige review articles. Atwood has eight different texts shelved “feminist” in *Goodreads*, spanning five decades of literary production: *The Edible Woman* (1969), *The Handmaid’s Tale* (1985), *Cat’s Eye* (1988), *The Robber Bride* (1993), *Alias Grace* (1996), *The Blind Assassin* (2000), *The Penelopiad* (2005), and *The Testaments* (2019). Suzanne Collins and Virginia Woolf each have three texts shelved in *Goodreads* (Collins: *The Hunger Games* (2008), *Catching Fire* (2009), *Mockingjay* (2010); Woolf: *Mrs. Dalloway* (1925), *To The Lighthouse* (1927), *Orlando* (1928)). As these findings suggest, these communities identify feminist novels as largely written by women: the entire corpus only includes thirty discrete novels by male-identified authors out of 313 unique novels overall. However, there are novels written by male-identified authors in all four subcorpora, with AM

featuring the most male-identified (thirteen).⁴⁰ These findings suggest that, in the cross-section of our corpus, marketers have been more invested than other interpretive communities in the idea that non-female-identifying authors could contribute to the body of feminist novels.

These findings also indicate that feminist fiction has a long, unevenly distributed history that tends to serve the present. Figure 4 shows subcorpora by novel publication year. We see a density of novels post-1920s and particularly post-1960s. These temporal dynamics align with the overall increase and specific documented trends in novel publishing over the course of the twentieth and early twenty-first centuries.⁴¹ However, the timeline also rehearses familiar date markers in the traditional wave model of historical feminisms. Figure 5 focuses in on this contemporary density. These temporal maps may suggest that current readers recruit feminist fiction that serves their immediate needs and thus look to pasts that serve these transformative future-minded aims. We imagine, for example, that searching LC and PR outlets with a time range of 1970 to 1995 would have resulted in novels with more temporal diversity, as professional critics sought to recuperate a genealogy of female authors and craft a feminist canon (perhaps obviously, however, and as mentioned in the corpus construction section above, this earlier timeframe would have been impossible for GR and AM). The contemporary interpretive communities' diverse timeframes suggest that each cohort seeks different resources to sustain its present-day work. In particular, the findings suggest that LC invests more heavily in the past, while AM invests more heavily in the present.

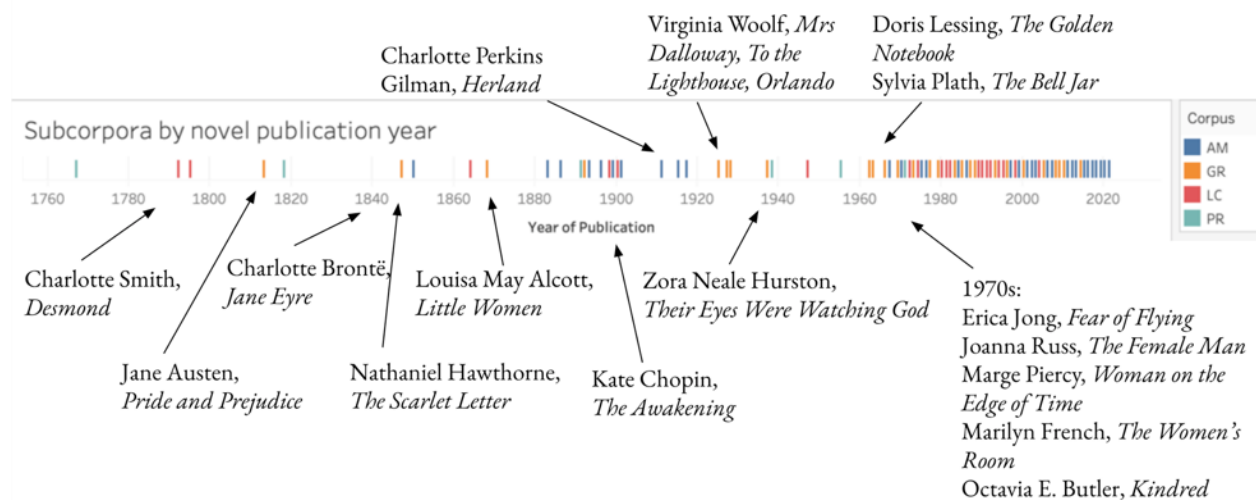


Figure 4. Subcorpora by novel publication year, all inclusive.

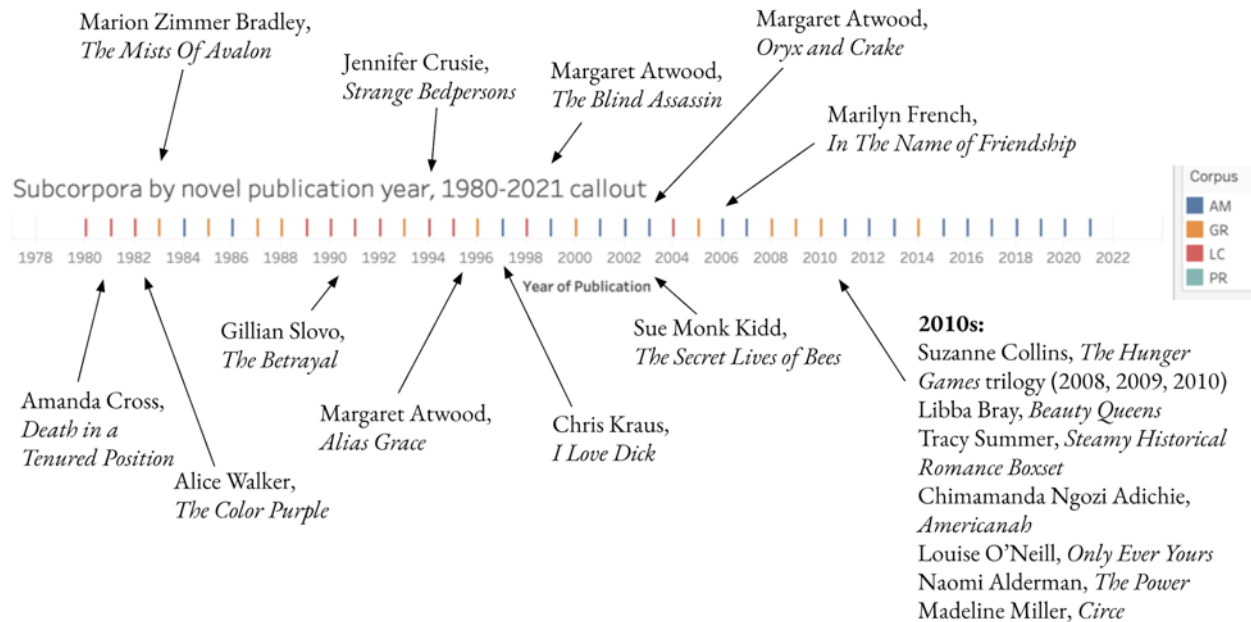


Figure 5. Subcorpora by novel publication year, 1980-present. Atwood appears multiple times, but we also see a more generically diverse landscape, with Suzanne Collins’s bestselling *Hunger Games* trilogy appearing alongside other YA fiction, including Libba Bray’s *Beauty Queens*. Similarly, Tracy Summer’s Amazon.com Services LLC-published *Steamy Historical Romance Boxset* is identified alongside Chimamanda Ngozi Adichie’s National Book Critics Circle Award-winning *Americanah*.

Our metadata thus reveals a productive discord across interpretive communities about 1) who writes feminist novels, 2) what these novels are, and 3) the timeframes in which these novels appear. The metadata indicates more consensus about identifying authors as feminists, rather than novels as such, both across interpretive communities and within them. The metadata also demonstrates the ways in which a sense of when feminist literature proliferates changes to serve different readers. These findings suggest a potential distancing from the seeming essentialism, or at least cohesion, of “feminist” as an adjectival descriptor, as opposed to the open-ended nature of feminism as an encompassing concept. For example, *Goodreads*’ “feminism” shelf is much more robust than its “feminist” shelf, both in the total number of texts shelved and the number of times readers have shelved its top seventy novels as “feminism.” This disparity perhaps reflects feminism’s definitional instability, as individuals appear to be more comfortable identifying people with the noun (a feminist as a type of person who is committed to women’s rights and gender equity), rather than applying the overdetermined adjective to a given novel that is then imagined to comprehensively promote a specific, coherently articulated political project.

What We Talk About When We Talk About Feminist Novels

Topic models can help us probe these interpretive communities' shared interests in feminist fiction. We produced our models using Mallet with default settings (20 top terms for each topic) and removing stopwords.⁴² We then looked at the posterior probabilities of the words in each topic and the topics in each document. We were interested in the relationships between the topical investments of corpus documents, both within each subcorpus and across subcorpora, so we endeavored to see what could be surfaced by "reading" the models themselves (rather than, for instance, analyzing individual corpus documents for their proportion of topics). We then produced a principal component analysis (PCA), transforming the topics' high-dimensional data into principal components that we could visualize in two dimensions on a PCA biplot.

We are sensitive to the concerns of interpreting topic models that Benjamin M. Schmidt has highlighted and Mark Algee-Hewitt has recently elaborated.⁴³ As Schmidt has demonstrated, topics are not as "coherent" and "stable" as one might assume; we therefore adopt his recommendation of "ground[ing] the analysis of topic models in the words they are built from" by analyzing the terms of each topic in the context in which they appear across the documents.⁴⁴ Algee-Hewitt emphasizes that "lists of top terms may overrepresent extremely high or low frequency terms at the expense of more meaningful words; labeling topics often overgeneralizes their specificity; and networks of topics are still based on a limited list of top terms."⁴⁵ We assess topical coherence and seek out patterns across a given topic's top terms, thereby attempting to avoid a facile reading of lists of top terms and to refrain from overly reading into individual words. Finally, we employ PCA to understand the topical distribution in the larger picture of our heterogeneous corpus. We remain cautious about the pitfalls of topic modeling as an explanatory method and therefore employ topic modeling as an exploratory metric and engage other computational methods in tandem.⁴⁶

Table 6 presents a 10-topic model of the full corpus. Overall, the contents of our model suggest how discussions of the feminist novel center on questions of personhood and temporality, with topics often drawing more heavily on one or the other concern but consistently addressing them both.⁴⁷ Questions of personhood

drive topics 0, 1, 3, 5, 6, and 7; questions of time drive topics 2, 8, and 9. Topic 0, for example, addresses maternal corporeality (“mother,” “desire,” “body”); topic 3 addresses anglophone nationality (“social,” “political,” “english,” “british”); topic 5 addresses gender and race (“unnatural,” “gender,” “black”); and topic 6 addresses biopolitics (“gender,” “war,” “oil,” “body”). Meanwhile, topic 2 addresses domestic temporality (“years,” “day,” “time,” “home”); topic 8 addresses an extended historical arc of feminist sexuality (“sex,” “radical,” “movement,” “liberation”); and topic 9 addresses the building of social worlds in time (“lives,” “society,” “change,” “past”).

Topic	Top terms
0	material mother desire world beauty body joanna girls female matter identity agency jimmy space bodies experience crake meaning ethical motherhood
1	author story book times review york love family books reviews historical life debut read sisters girl fiction war tale woman
2	back years long thought day made room head don't looked time house mother father it's home face put night dark
3	woman life women social marriage political english fiction writing mother children kate british husband south writers story family press love
4	book read story it's don't love i'm rev reading people feel characters character didn't make time things good books felt
5	narrative feminist unnatural gender page historical end woman black fiction narratives family female utopian life form ways social utopia relationships
6	gender women female war translation oil body gilead social male reta housewife handmaid's shields art violence herland bodies novels aunt
7	love novels life husband woman art wife write literary man it's heti writes young work writer writing sex heroine reading
8	feminist sexual women feminism women's women's york sex room novels radical fear movement female liberation romance political feminists sexuality flying

9	women world men time people power lives work life real man place society live change human characters back past children
---	---

Figure 6: A 10-topic model of the full corpus.

At the same time, using a PCA biplot we can see how different communities engage these topics differently. Figure 7 shows a PCA biplot for this model, with the principal components (PCs) accounting for 54.96 percent of the cumulative proportion of variance. The red arrows represent the topic loadings for PC1 and PC2 and function to spatially distribute all corpus documents according to their topical investments. Thus, this biplot indicates a strong differentiation between the four interpretive communities, with the most topical overlap between the LC and PR communities. The biplot also indicates that AM and GR are largely separated by two individual topics: topic 1 works along the PC1 axis, particularly determining the clustering of AM documents (in coral) at the top left; topic 4 particularly determines the clustering of GR documents (in green) at the top right. Meanwhile, PR documents (in purple) operate in a space between AM and GR.

This distribution suggests LC’s tendency to discuss personhood in more abstract, historically extended terms (topic 5), whereas AM tends to discuss people in terms of contemporary critical regard (topic 1) and GR tends to discuss personal preferences in an immediate context (topic 4). In topic 1, for example, “author” and “woman” appear alongside rhetoric suggestive of *The New York Times* and other book reviews and literary prizes. In topic 4, the first-person pronoun “I” appears alongside the discourse of affective assessment (“good,” “don’t,” “feel,” “felt”). These discursive distinctions also appear in the time-oriented topics, with topic 2’s discourse of domestic temporality particularly determining the distribution of AM and PR documents, topic 8’s extended historical arc particularly determining that of LC documents, and topic 9’s attention to social worldbuilding particularly determining that of GR documents.

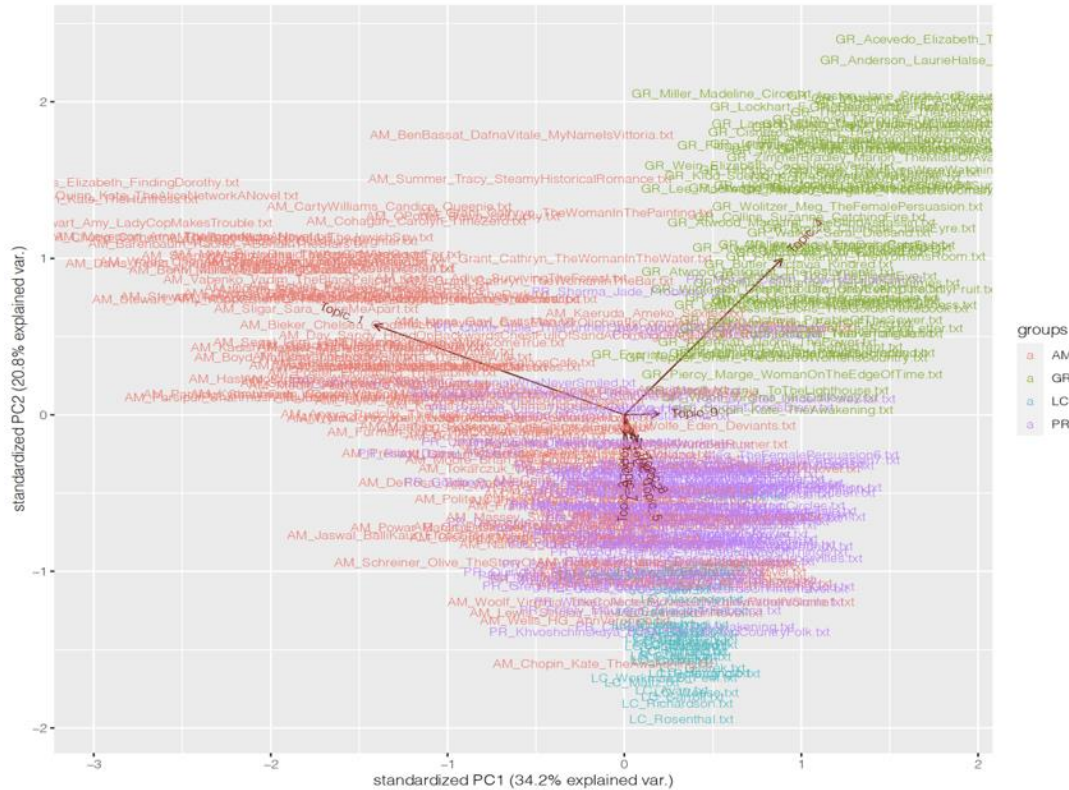


Figure 7: A PCA biplot of the 10-topic model of the full corpus.

More specifically, topics tend to coalesce around corporeal experience; social structures of gender, race, class, sexuality, and nationality; and concerns of past, present, and future. However, different interpretive communities use different language to indicate these same topical fields. Figures 8 and 9 suggest that social readers privilege immediate individual experience in their assessments of feminist fiction, while literary critics emphasize more abstracted social collectives over longer stretches of time. Figure 8 juxtaposes two 5-topic topic models: one of the 1-star and another of the 5-star reviews of the GR subcorpus.⁴⁸ The figure features directive negative (1-star) and laudatory positive (5-star) rhetoric alongside attention to plot and character—combinations that suggest readers recommend novels based on their embodied literary encounters. In the 1-star model, topic 1 warns others against reading a novel with words like “don’t” and “bad” alongside “character” and “plot” as key determinants in these assessments. Topic 3 highlights that issues of racial representation (“white,” “black”) often undergird readers’ negative reviews—an observation borne out by our close reading of these reviews. Topic 4 addresses high or “classic” “literature,” with “boring” and “didn’t” constellating responses to novels frequently assigned in school. By contrast, the 5-star review topic model

manifests affective and emotional positivity (in topic 0: “yeah,” “girls,” “women,” “amazing,” “important,” “love,” “loved”) and captures the value of personal identification (in topic 1, the pronoun “I” and gerund “reading” are tightly knit with “real” “character” and “characters”).

1-star reviews

Topic	Top terms
0	it's katniss people world make she's love didn't point characters series peeta games bad plot supposed collins person kids there's
1	book read don't it's books rev feel i'm felt girl make thing character people thought bad main fact it's plot
2	women men feminist feminism atwood society power tale world woman gender male man work author rev narrator female women's religious
3	book story rev characters writing author white time reading good black people lot character back i'm life it's give interesting
4	read rev time reading life woman jane love boring years find literature part classic didn't pages man husband found i've

5-star reviews

Topic	Top terms
0	book yeah girls women review time world amazing people love feminism woman important things loved white rape books feminist girl
1	book read rev it's story love reading i'm characters don't good books things loved people makes character real lot make
2	story life time read years black woman mother back family atwood past people experience rev stories children lives white grace
3	women world men society book power gender female read time male left future story hand human work fiction people man
4	life love jane woolf time mrs woman marriage thought man circe lighthouse made orlando character virginia day janie moment beautiful

Figure 8. A Goodreads 1-star 5-topic model vs. a Goodreads 5-star 5-topic model.

The 10-topic model of the LC subcorpus in Figure 9 suggests that literary critics discuss historically contextualized collectives in their analyses of feminist embodiment.⁴⁹ For example, topic 7 addresses issues of racial representation by highlighting the histories of slavery and its ethics (“slavery,” “white,” “black,” “slave,” “enslaved”). Topic 1 registers a range of minority experiences, particularly within geopolitical conflicts (the “south,” “africa,” “black”), while topic 2 discusses sex and sexuality in terms of liberal politics. Topics 5, 8, and 9 highlight the temporal dimensions of collective human experience. Topic 5 highlights the domestic novel’s class contours as well as one of its key historical periods, the Victorian era. Topic 8 addresses broad questions of feminist history, while topic 9 raises questions of utopian futurity. Suggesting LC’s tendencies to abstraction, topics 0, 4, and 9 treat concerns of literary personhood more theoretically (“narrative” and “discourse”) and make them more social than individual (“patriarchal,” “world,” “bodies”).⁵⁰ The GR and LC topic models thus suggest the feminist priorities of each interpretive community: LC emphasizes broad historical phenomena and validates embodied experience by rendering it an intellectual concern; GR highlights the visceral impacts of these phenomena on individuals and renders feminism a project of urgent feeling.

Topic	Top terms
0	world story social female make sense experience patriarchal fact time narrative relationship place form back past suggests point reality real
1	south african oil war gender africa violence political community social women country male slovo society black children popular end struggle
2	sexual women women’s feminist radical women’s sex feminism york fear sexuality liberation feminists movement flying political abortion freedom ibid oppression
3	room kate university body mira dream feminist housewife queen art york life waste ibid summer memoirs death novels dreams french
4	female gender material body bodies desire male irish knowledge girls discourse beauty feminist power matter agency process terms cultural mind

5	love press victorian domestic sexual middle-class wife image young isabel london reading free rosalind female dix george figure hall british
6	hetty translation reta joanna frankie shields mother female carl womb translator mrs male clones view technologies mitchell dorval power reta's
7	family slavery white fenwick jimmy black slave fenwick's jimmy's marriage blacks enslaved human people crake mother theme space ethical children
8	women woman life fiction men man work novels writing feminist historical history lives husband writers society early love years family
9	unnatural narrative feminist utopian texts end utopia future time narratives big ways page film narration relationships eden voice edge text

Figure 9: A 10-topic model of the LC subcorpus.

How We Talk *Distinctively* When We Talk About Feminist Novels

Most distinctive word (MDW) analysis allows us to further investigate the interpretive communities' interests. As Sarah Allison et al. describe, MDW analysis calculates the “average, or ‘expected,’ frequency” of all the words of a large corpus, then calculates “the actual—or ‘observed’—occurrence” of each word in different component parts of the corpus, in order to highlight “those instances that reveal[] a significant observed-over-expected ratio.”⁵¹ MDW tests are independent tests of each word versus the rest of the vocabulary. We tested the significance of each odds-ratio with a Fisher’s exact test, producing a P-value; we only returned results with P-values under 0.05. This method indicates which words appear more often than statistically expected in one subcorpus versus another and so, in the terms of our project, captures which words are most distinctive of a given interpretive community. We filtered our MDW results by subcorpus and in order of statistical significance. We then looked for patterns in the resulting MDWs.

Figure 10 showcases the people who distinctively characterize each subcorpus, including characters, novelists, political figures, and other critics. Our reading of these patterns suggests that LC and PR distinctively highlight the historically living

beings who produce and professionally discuss feminist literature, while GR and AM lean into the fictive worlds these historical figures create. These distinctions may well be consistent with the broader differences between these four interpretive communities and their respective priorities, but our point is not about how the communities uniquely discuss feminist fiction as compared to other genres of literature. Rather, MDW analysis allows us further purchase on how these communities highlight different kinds of persons, which in the case of our dataset are people who variously interface with feminist concerns, past and present. LC distinctively mentions political figures and other critics (e.g., “lukacs,” “irigaray,” “millet,” “theorists,” “critics”), while PR references authors and then characters (e.g., “heti,” “kraus,” “didion”; “zoe,” “prue,” “lena”). GR reverses PR’s priorities, distinctively discussing characters more prevalently than authors—but with a markedly different cast of individuals (e.g., “rochester,” “katniss”; “atwood,” “austen”). AM, like GR, prioritizes characters and then authors, though again with a different cast of people (e.g., “oliphant,” “beowulf”; “mackintosh,” “quinn”). These differences between professional and social readers echo recent feminist literary criticism that recenters character as a topic of scholarly discourse and does so in part to counter narratives of social readers’ supposedly naïve tendencies to reader-character identification.⁵²

Subcorpus	MDW	Interpretive label	Observations	P-value	Odds-ratio
LC	lukacs	political figures and other critics	19	0.007315694	3.166666667
LC	irigaray	political figures and other critics	20	0.009577385	2.857142857
LC	millet	political figures and other critics	22	0.008061164	2.75

LC	theorists	political figures and other critics	30	0.001110368	3
LC	critics	political figures and other critics	77	0.000266736	1.974358974
PR	heti	author	121	3.06E-20	6.722222222
PR	kraus	author	89	2.79E-14	5.933333333
PR	didion	author	20	0.00077161	5
PR	zoe	character	37	2.21E-07	7.4
PR	prue	character	35	6.90E-07	7
PR	lena	character	49	1.35E-08	6.125
GR	rochester	character	32	0.001828375	2.666666667
GR	katniss	character	98	7.49E-08	2.648648649
GR	atwood	author	380	5.27E-17	2.065217391
GR	austen	author	51	0.004384043	1.888888889
AM	oliphant	character	14	0.035154439	7
AM	beowulf	character	17	0.035154439	5.666666667
AM	mackintosh	author	10	0.035154439	5
AM	quinn	author	17	0.035154439	2.833333333

Figure 10. MDWs by subcorpus illustrating how these communities highlight different kinds of feminist persons. While all MDWs can be examined in our results, MDWs are here organized by subcorpus according to our close reading of trends. We list entries for each subcorpus—and within each subcorpus, each subjectively determined “interpretive label”—in descending order by odds-ratio.

The MDWs in Figure 11 indicate a second preoccupation of these interpretive communities: the process of reading and review, or what one might call community practice. Each subcorpus's discussions of community practice prioritize the intersubjective network within which these literary texts are evaluated, though their modes of engaging with these collectives range from impersonal and self-distancing to intimate and disclosing. LC and PR express the former. LC distinctively focuses on literary critical assessment, including periodizations of "nineteenth-" and "twentieth-century" along with other academic terminology that "argue[s]" about "hegemonic" values, "performativity," "narratology," and "poetics." PR foregrounds authorial microhistory, discussing the "novelist" at hand and drawing attention to the hallmarks of publishing success, such as the numbers of "copies" sold, "publications" to date, and "television" appearances. These communities' prose trends minimize individual reviewers' selfhood by emphasizing external contexts of literary history or authorial biography.

By contrast, GR and AM are more self-referential in their distinctive vocabularies. GR emphasizes the personal experience of reading and review, including abbreviations of popular terms such as "dnf" (did not finish) and "pov," self-conscious references to "rating," "stars," and "hype," and a distinctive mix of negative and positive assessments, ranging from "annoying," "depressing," "boring," and "disappointed" to "enjoyed" and "appreciate." AM emphasizes other, largely digital external platforms of review (e.g., "buzzfeed," "lithub," "esquire," "booklist," "bustle," "tribune," "bookpage," "kirkus"). Interestingly, GR is the only subcorpus with a MDW that comes from community practice—"pov" followed by "dnf"—whereas the other three prioritize people.

All four interpretive communities thus prioritize selfhood in talking about feminist fiction, but center different selves in these discussions. The professional communities of LC and PR once again hang together in these discussions. These subcorpora prioritize similar types of people—those who read and discuss literature—even as their community practice largely obfuscates reviewers' individual selves. By contrast, GR and AM focus on how the people inside the pages speak to readers' individual subjectivities—personally and retrospectively for GR ("I felt X about the protagonist") and generically and prospectively for AM ("buy this book and you, too, will experience these feelings").

Subcorpus	MDW	Interpretive label	Observations	P-value	Odds-ratio
LC	nineteenth-century	academic terminology	27	0.006925075	2.454545455
LC	twentieth-century	academic terminology	21	0.02138453	2.333333333
LC	hegemonic	academic terminology	13	0.024519375	3.25
LC	performativity	academic terminology	25	0.002275179	3.125
LC	narratology	academic terminology	59	0.0000032	3.105263158
LC	poetics	academic terminology	15	0.020693148	3
LC	marxist	academic terminology	14	0.031782086	2.8
LC	argue	academic terminology	65	0.000218758	2.214285714
PR	copies	publishing milestones	15	0.009603484	3.75
PR	publications	publishing milestones	10	0.046138945	3.333333333
PR	translated	publishing milestones	15	0.02069127	3
PR	interview	publishing milestones	23	0.010027417	2.555555556

PR	novelist	publishing milestones	32	0.00567258	2.285714286
PR	television	publishing milestones	18	0.037752807	2.25
PR	famous	publishing milestones	27	0.019231991	2.076923077
GR	dnf	personal experience	14	0.031782326	2.8
GR	pov	personal experience	19	0.014478209	2.714285714
GR	audiobook	personal experience	29	0.003212565	2.636363636
GR	spoiler	personal experience	71	0.00000505	2.62962963
GR	rating	personal experience	62	0.0000254	2.583333333
GR	re-read	personal experience	15	0.039174776	2.5
GR	recommend	personal experience	75	0.00000333	2.419354839
GR	stars	personal experience	199	3.14E-12	2.369047619
GR	hype	personal experience	16	0.046567216	2.285714286
GR	annoying	personal experience	36	0.001300647	2.571428571

GR	depressing	personal experience	33	0.002266299	2.538461538
GR	boring	personal experience	94	5.92E-07	2.473684211
GR	enjoyed	personal experience	114	3.16E-07	2.28
GR	disappointed	personal experience	57	0.000267168	2.28
GR	appreciate	personal experience	64	0.00018274	2.206896552
AM	buzzfeed	external review platforms	7	0.035154439	7
AM	lithub	external review platforms	7	0.035154439	7
AM	esquire	external review platforms	7	0.035154439	7
AM	booklist	external review platforms	39	0.035154439	6.5
AM	bustle	external review platforms	13	0.035154439	6.5
AM	tribune	external review platforms	19	0.00042755	6.333333333

AM	bookpage	external review platforms	12	0.035154439	6
AM	kirkus	external review platforms	41	0.035154439	5.857142857

Figure 11. MDWs by subcorpus suggesting each subcorpus's discussions of community practice.

Why Read Feminist Novels?

MDW results also allow us to posit that the rationales different interpretive communities offer for reading feminist novels are driven by the forms of subjectivity they value. Our interpretation of the MDW results indicates that literary critics value feminist subjectivity as a long-standing sociopolitical identity committed to equity, inclusion, and diversity. LC's MDW results indicate that both "feminist" and the corollary "feminism" are observed in LC around 1.88 times more likely than expected.⁵³ Close reading of individual LC documents elaborates the critical values that this keyword magnetizes. For example, Melanie Micir's article "The Impossible Miss Woolf: Kate Atkinson and the Feminist Modernist Historical Novel" (2017) situates Atkinson's *Life After Life* (2013) as "belonging to a genealogy of feminist historiography," arguing "that Atkinson's methodology descends from both modernist and feminist arguments about how to write the history of marginalized subjects."⁵⁴ Micir's "feminist" thus links Atkinson's novel to a lineage of text-based recovery projects that are understood to enrich models of historical experience as well as enhance the accuracy of claims about cultural production. Micir's use of the term further associates her own scholarship with this esteemed ongoing project.

MDW analysis combined with such close reading also indicates that LC and GR tend to evaluate exclusionary models of feminist sensibility, though they employ different language to do so. "Nonfeminist" only appears fourteen times in the entire corpus and primarily (12 times) in LC.⁵⁵ Close reading of LC documents suggests that literary critics employ this language to compare narratological strategies and reader responses according to a feminist/nonfeminist binary, with both narratives

and readers identified according to this logic. One such example, Ellen Peel's "Unnatural Feminist Narratology" (2016), differentiates "feminist and nonfeminist readers" of Joanna Russ's science fiction novel *The Female Man* (1975), only to assert that the novel itself challenges a monolithic heteronormative model of feminism: "the text's values are clear: it sympathizes with feminisms."⁵⁶ GR registers a similar pluralizing logic with more popular politicized language. For example, a 2-star review of *Moxie* asserts, "Seek out other feminist voices after you read this: queer feminists, POC feminists, trans feminists, disabled feminists. Feminism is intersectional and includes all these (and other) marginalized people, or it's bullshit."⁵⁷ As these distinct ways of talking about feminism's collective project suggest, social readers mobilize the concrete, individual-embodied and temporally specific language of race and skin color to advocate for intersectionality, while literary critics employ more abstract lexicons to address the histories and ongoing presence of oppression—even as their own rhetoric is at times less inclusive than the vision they elsewhere propound.

While LC and GR assume their own evaluative capacities in assessing and promoting a novel for its feminism, AM contends one should read feminist novels because other trusted persons say so. AM's MDW results invoke prize-awarders, consumers, and literary critics, suggesting one should read these novels because they are "award-winning" and "bestselling." These superlatives presumably reflect AM's broader marketing aims and are not necessarily specifically distinctive of feminist novels on the platform. However, the language importantly distinguishes AM's discourse about feminist fiction from that of the other three subcorpora. Leigh Stein's runaway hit *Self Care* (2020) is a case in point. The featured blurbs include *Elle*'s description as "a novel of manners for our 280-character era," at once evoking historical precedent (classic and timeless) and pitch-perfect contemporaneity (mordant and timely).⁵⁸ Figure 12 lists other AM MDW results, all of which apparently seek to convince potential customers to purchase novels because they provoke positive affective and intellectual responses to which other people, including Oprah Winfrey, can attest, having already bought and read them.⁵⁹ Oprah's distinctiveness here merges the discursive practices of LC and AM: as both a representative of nonwhite feminism and the consummate embodiment of corporate feminism, Oprah embodies a superior evaluative subjectivity and a subjectivity whose evaluation makes or breaks a book.

Subcorpus	MDW	Observations	P-value	Odds-ratio
AM	suspenseful	18	0.035154439	6
AM	bestselling	88	4.79E-14	5.866666667
AM	engrossing	11	0.035154439	5.5
AM	must-read	27	0.0000565	5.4
AM	page-turning	16	0.002211937	5.333333333
AM	irresistible	16	0.035154439	5.333333333
AM	page-turner	21	0.035154439	5.25
AM	unforgettable	26	0.000096	5.2
AM	award-winning	19	0.001299274	4.75
AM	captivating	22	0.000756509	4.4
AM	oprah	14	0.031779944	2.8
AM	compelling	39	0.001332458	2.4375

Figure 12: MDW results for AM subcorpus, again listed in descending order by odds-ratio.

In sum, LC is more likely to label novels or people “feminist” in order to value them as part of an historically informed sociopolitical cultural project, while GR discusses these commitments using more recently emergent politicized rhetoric. By contrast, AM relies on vaunted collectives’ preferences as it promotes novels based on their prizewinning or bestselling success.

In contrast to AM’s assertive extremes, LC and GR register emotional and intellectual uncertainty and complexity as valued modes of feminist subjectivity—or at least as thoughts and feelings that come along with valuing feminist subjectivity (Figure 13). The words “subjectivities” and “subjectivity” are distinctive of LC. A close reading of LC’s MDW results suggest the ambivalence of this psychic

formation, as the list is rife with terms that connote conflict, multiplicity, and indecision (e.g., “alternatives,” “alienation,” “anxieties,” “limitations,” “boundaries,” “failure,” “crisis,” “both,” “between,” “impossible”). These MDW results add nuance to the positive attributions of feminist across the subcorpus. Even as LC engages feminist as an affirmative descriptor, the term describes conflicted, often highly contested, contexts and registers novels’, authors’, and characters’ contradictory responses to these challenging circumstances. These LC MDW results also remain associated with a valued white masculine subjectivity that dates back to the rise of New Criticism.⁶⁰ This language can therefore be seen to confer intellectual prestige, though the vexed embodied experiences the rhetoric seeks to describe are hardly to be envied.⁶¹ For its part, GR registers such conflicts in terms of readers’ mixed feelings, featuring borderline subjective responses like “guess,” “confusing,” and “seem” alongside more definitive evaluative terms like “totally” and “absolutely.” As noted above and detailed below, social readers also express anxiety about using the adjective “feminist” to describe a novel, in part because they imagine the genre to operate in complex, less straightforward relations to such political investments. Across these subcorpora, ambivalence emerges as a persistent experience of feminism, alternately vaunted, valued, rued, and regretted.⁶²

Subcorpus	MDW	Observations	P-value	Odds-ratio
LC	subjectivities	13	0.024519375	3.25
LC	subjectivity	64	0.00000164	3.04761905
LC	alternatives	12	0.038404521	3
LC	alienation	22	0.01472272	2.444444444
LC	anxieties	16	0.046566855	2.285714286
LC	limitations	22	0.025048345	2.2
LC	boundaries	43	0.002575101	2.15
LC	failure	39	0.009165843	1.95
LC	crisis	36	0.031362421	1.714285714

LC	both	480	3.46E-11	1.610738255
LC	between	541	3.75E-12	1.600591716
LC	impossible	67	0.013918653	1.558139535
GR	guess	95	0.00000123	2.375
GR	confusing	21	0.021384827	2.333333333
GR	seem	127	0.018601893	1.336842105
GR	totally	63	0.000630235	2.032258065
GR	absolutely	77	0.00017531	2.026315789

Figure 13. Select MDW results for LC and GR, again listed in descending order by odds-ratio.

Who We Talk About When We Talk About Feminist Novels

NER yields a more granular view of the selves who are mentioned in each interpretive community's discussions of the feminist novel. To generate this data, we used David Bamman's BookNLP tool for NER, producing semantically parsed spreadsheets for each document that we then filtered for "person" to analyze the named entities who appear therein.⁶³ This gave us a measure of the number and identities of people mentioned within each document of each subcorpus. We then calculated the frequency of named entities in each subcorpus relative to its total word count. Finally, we were especially curious about the nature of the people each interpretive community mentioned, especially in discussions of the same novel, so we selected several texts that appeared across multiple interpretive communities, reviewed their NER results, and analyzed them qualitatively via close reading. Our results thus extend the anthropocentric findings of our MDW analysis by capturing the individuals associated with feminist novels as they appear within each subcorpus.

LC had the highest proportion of named entities at 3.92%, closely followed by PR at 3.89% and AM at 3.48%. These numbers are roughly double that of GR, where named entities constitute only 1.78% of total word count. The relatively high proportion of references to people in LC, PR, and AM suggests that in these three

interpretive communities, novels accrue cultural value by connecting subjectivities within a densely populated intergenerational critical and creative network. By contrast, as we have seen, *Goodreads* readers tend to emphasize first-person experience and therefore, we hypothesize, place less value on critical and creative forebears in favor of engaging other social readers as interlocutors. A close reading of GR's NER results, however, indicates that certain novels also accrue value in this fashion. For example, Kraus's *I Love Dick* appears in AM, GR, and PR. Reviewers across these interpretive communities highlight a wide range of literary luminaries and cultural interlocutors as they assess the generic hybridity of Kraus's text, which interweaves highbrow philosophizing with more traditional fictional characterization. The combined 5-star *Goodreads* reviews of the novel name drop Anaïs Nin, Hannah Wilke, Joan Didion, Jack Kerouac, W. G. Sebald, Kate Zambreno, Eileen Myles, Tom McCarthy, Marcel Proust, and Jacques Lacan, among other figures. Meanwhile, Elaine Blair's *New Yorker* review of the novel alone references Pierre Marivaux, Gilles Deleuze, Michel Foucault, Paul Virilio, Sheila Heti, Lena Dunham, Lorde, Jill Soloway, Don DeLillo, Toni Morrison, Philip Roth, Amy Hempel, Deborah Eisenberg, Mark Leyner, Kathy Acker, Mary Gaitskill, Karl Ove Knausgård, Ben Lerner, Geoff Dyer, Philip Valdez, Mary McCarthy, Amy Schumer, Nan Goldin, Simone Weil, and more.⁶⁴ These findings both suggest the intersections of these interpretive communities—Kraus's hybrid text encourages this kind of reading across all communities—and highlight their differences: even when a text is densely populated in this fashion across its 5-star *Goodreads* reviews, it's still a far cry less populous than a single laudatory review in PR.

This case study of Kraus's novel also underscores that these feminist networks are importantly transhistorical. Blair traces *I Love Dick*'s reception history from its largely unnoticed publication in 1997 to its 2013 recuperation by Heti, subsequent recommendation by Dunham, Instagram feature by singer-songwriter Lorde, and recent television adaptation. This trajectory suggests how a feminist novel's reception reflects individuals' responses to historical change (Heti to Dunham to Lorde to TV producers) even in contexts that tend to obscure the reviewer's own subjectivity, such as prestige reviews. Yet for all GR's emphasis on immediate subjective experience, this interpretive community also engages in a kind of transhistorical self-erasure by locating novels in historically agnostic or ambiguously historical genres like feminist sci-fi and feminist dystopian fiction and

identifying novels with blurry historical formulations like second-wave feminism and feminist radicalism. Once again, relays between individual, idiosyncratic being and collective life and community belonging are what we talk about when we talk about feminist fiction.

Imperfect Feminisms

The synthesis of quantitative and qualitative analysis we have undertaken in this study allows us to explore at scale how reading novels might contribute to feminist knowledge making, both for individuals and communities. We have seen how each interpretive community's discussions of feminist novels grapple with constructions of personhood in time—through their vocabularies, their modes of self-address, the rationales they offer for reading feminist novels, and the forms of feminist subjectivity they promote. In wrestling with these concerns, we argue, readers offer productive models of feminist reading as a dialogue between personal experience and social belonging, embodied theory and critical practice, and history as recorded and as a dynamically unfolding present. These readers establish contemporary discourse about feminist fiction as both a collective space for imagining prosocial change and an individual space for recalibrating one's own feminist value system as it is lived forward in time.

Just as there are many actively evolving, collectively and individually organized feminisms, our study's interpretive communities recognize these novels for their varied embodiments and articulations of feminism. Read through this lens, even novels that come up politically short can be politically generative. One social reader speaks to this dialectical potential when she calls *Dietland* “disappointing” for “not promoting a feminism that I’m fond of. It shows a feminism that is not intersectional, that is often teetering on racist, that condones slut shaming and the villainizing of sex workers.” Here, “a feminism” implies there is more than one feminism, which dovetails with literary critics’ use of the term “feminisms” discussed above. Social readers who speak to this inclusive multiplicity—who cast what they read as both “a feminist novel” and “not feminist enough”—epitomize a feminist project of active renewal, wrought by individual and collective grappling with ever-imperfect worlds.

This dynamic quality of incompleteness is a testament to feminist literature's diverse interlocutors and their heterogeneous ambitions. Accordingly, there are many directions that further DH research on feminist literature might pursue. As noted above, one limitation of our study is that we cannot determine how the dynamics we observe in these interpretive communities' discussions of feminist novels might obtain in other contexts. Future research might therefore explore how these discussions of feminist novels align with discussions of other kinds of novels in these same interpretive communities. Alternatively, further research might extend this analysis to other branches of these interpretive communities (e.g., Amazon customer reviews, literary monographs) and to other interpretive communities (e.g., para-academic venues such as *Medium*, *Public Books*, popular periodicals such as *O*, *The Oprah Magazine*, *Elle*). Future scholarship might also examine the interplay between feminist novels and other modes of feminist literature, such as memoirs, manifestos, and other non-fiction, in these and other interpretive communities. If future scholarship is to live up to feminisms' inclusive ambitions, it will also need to find creative ways to harness more ephemeral discussions of feminist literature such as those proliferating on social media sites, individual blogs, and podcasts, among other fora.

As the generative dissensus across our interpretive communities suggests, reading feminist novels and discussions thereof does not mean arriving at a final conclusion—here is feminism, or even here is *my* feminism—for how could it? Feminism is an aspirational project of social transformation committed, as philosopher Noëlle McAfee recently put it, to striving to “know what came before, be attuned to what is wrong, and act on [the] desire to make things better ... for all people, past, present, and future.”⁶⁵ This sustained and sustaining work promises to be always in the making. As we have shown, this work takes place wherever we read and wherever we talk to others about what we are reading.

References

¹ Sara Ahmed, *Living a Feminist Life* (Durham: Duke University Press, 2017), 1.

² The authors contributed equally to this piece. We thank Mark Algee-Hewitt, J.D. Porter, Alice Staveley, and Laura McGrath for generously sharing their expertise and crucially facilitating our project's development. The Stanford Literary Lab provided crucial training and institutional support. We are also grateful to this article's two peer reviewers, Andrew Piper, and the attendees of a summer 2020 Zoom research presentation for their thoughtful engagement with our work; their insightful feedback and generative suggestions valuably strengthened this project and manuscript.

³ In *Feminism: A Quick Immersion* (2021), philosopher Noëlle McAfee asserts that "there isn't just one kind of feminism but a variety of feminist approaches and points of view; but all are trying to make sense of the problem that, for most of the past three millennia, if not longer, the world over women have been ... deprived of the ability to chart their own life" (New York: Tibidabo Publishing, 2021), 10. Ahmed articulates a similar position both explicitly and by example, as she offers a series of overlapping definitional statements (e.g., "Feminism: the dynamism of making connections"; "Feminism: how we survive the consequences of what we come up against by offering new ways of understanding what we come up against"; "Feminism is DIY: a form of self-assembly"; "Feminism: on the way to becoming something else" (*Living a Feminist Life*, 3, 22, 27, 186)). See also Lisa Mendelman, "Who Are We? Feminist Ambivalence in Contemporary Literary Criticism," *American Literary History* 32, no. 1 (February 1, 2020): 190-200.

⁴ Barbara Hill Rigney's *Madness and Sexual Politics in the Feminist Novel: Studies in Brontë, Woolf, Lessing, and Atwood* (1978), seemingly the first monograph to employ the key term, avoids defining its eponymous phrase, instead describing various examples that circle around the broad common ground of heteropatriarchal critique (Madison: University of Wisconsin Press). Judi M. Roller's *The Politics of the Feminist Novel* (1986) takes the opposite approach, asserting a tight literary definition—"the central character or characters in a feminist novel must be female and must represent women generally as well as a woman specifically"—and then offering a broad description of the novel's ideological commitments: that "human problems are collective as well as individual" and that "the status of women with relation to [economic, political, and social] structures is that of a minority group" (Westport, Connecticut: Greenwood Press), 4-5. Walker's *Feminist Alternatives: Irony and Fantasy in the Contemporary Novel by Women* (1990) refers to "the feminist novel" as "the women's novel," defining the latter as "the essentially realistic novel, as opposed to the popular romance" (Jackson: University Press of Mississippi), 4, 5. While, for Walker, the women's novel is not "overtly political," its "central message is change: the need for social change and the recognition that social change begins with individual change—the personal is political, but the reverse is also true" (36). Felski's *Literature After Feminism* (2003) offers a more explicit entry into this tradition of indirection, observing that "many feminist critics are wary of any general claims about women, femininity, or the female condition" and that "there is not much agreement about the best way to explain the links between gender and writing" (Chicago: University of Chicago Press), 2.

⁵ For instance, Felski argues "that it [i]s a mistake to try to define a 'feminist aesthetic' or general theory of women and literature," asserting, among other things, that a given narrative structure or plot can be "used to argue for diametrically opposed morals, political theories, or social codes: there is no single ideology frozen in a plotline that dictates its meaning and use" (*Literature After Feminism*, 4, 108).

⁶ Felski, *Literature After Feminism*, 109.

⁷ Melanie Walsh and Maria Antoniak, "The Goodreads 'Classics': A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism," *Journal of Cultural Analytics* 4 (2021): 243, 260.

⁸ Janice Radway, *Reading the Romance: Women, Patriarchy, and Popular Literature* (Chapel Hill: The University of North Carolina Press, 1991), 5.

- ⁹ Stanley Fish, *Is there a text in this class? The authority of interpretive communities* (Cambridge: Harvard University Press, 1980), 14.
- ¹⁰ Fish, *Is there a text in this class?*, 14.
- ¹¹ Andrew Piper, *Enumerations: Data and Literary Study* (Chicago: University of Chicago Press, 2018), 116.
- ¹² Piper, *Enumerations*, 115.
- ¹³ Piper, *Enumerations*, 116.
- ¹⁴ Piper, *Enumerations*, 115.
- ¹⁵ Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge: MIT Press, 2020), 18.
- ¹⁶ Elizabeth Losh and Jacqueline Wernimont, *Bodies of Information: Intersectional Feminism and Digital Humanities* (Minneapolis: University of Minnesota Press, 2018), xv.
- ¹⁷ Losh and Wernimont, *Bodies of Information*, xii.
- ¹⁸ Janice Radway, *Reading the Romance: Women, Patriarchy, and Popular Literature* (Chapel Hill: The University of North Carolina Press, 1991). Joan Shelley Rubin, *The Making of Middlebrow Culture* (Chapel Hill: The University of North Carolina Press, 1992).
- ¹⁹ Radway, *Reading the Romance*, 221.
- ²⁰ Rita Felski, *Hooked: Art and Attachment* (Chicago and London: The University of Chicago Press, 2020), xi, viii-ix.
- ²¹ Timothy Aubry, *Reading as Therapy: What Contemporary Fiction Does for Middle-Class Americans* (Iowa City: The University of Iowa Press, 2011), 11. For especial attention to issues of sex and gender, see Chapter Two, "Therapy and Displacement in *Divine Secrets of the Ya-Ya Sisterhood*" and Chapter Five, "The Politics of Interiority in *The Pilot's Wife*."
- ²² Elizabeth Long, *Book Clubs: Women and the Uses of Reading in Everyday Life* (Chicago: The University of Chicago Press, 2003), 114. Long casts this interpretive practice as "literally productive in that it enables women not merely to reflect on identities they already have but also to ... creat[e] new connections, new meanings, and new relationships—to the characters in books or their authors, to themselves, to the other members of the [book club] group, to the society and culture in which they live" (22).
- ²³ Felski, *Hooked*, 47-48.
- ²⁴ Matthew J. Lavin, "Gender Dynamics and Critical Reception: A Study of Early 20th-century Book Reviews from The New York Times," *Journal of Cultural Analytics* (January 31, 2020): 3.
- ²⁵ Lavin, "Gender Dynamics and Critical Reception," 27.
- ²⁶ Karen Bourrier and Mike Thelwall, "The Social Lives of Books: Reading Victorian Literature on *Goodreads*," *Journal of Cultural Analytics* (February 7, 2020).
- ²⁷ Bourrier and Thelwall, "The Social Lives of Books," 25, 27.
- ²⁸ Walsh and Antoniak, "The Goodreads 'Classics,'" 244-245, 253.

²⁹ Walsh and Antoniak, “The Goodreads ‘Classics,’” 276. This attention to the critical and interpretive value of *Goodreads* reviews contrasts Shadi Shahsavari et al.’s interest in using these reviews to automate plot summaries and dismissal of their critical and interpretive merits. See Shadi Shahsavari et al., “An Automated Pipeline for Character and Relationship Extraction from Readers Literary Book Reviews on Goodreads.com,” *12th ACM Conference on Web Science*, Association for Computing Machinery (2020): 277-286.

³⁰ See Suzanne Bost, *Encarnacion: Illness and Body Politics in Chicana Feminist Literature* (New York: Fordham University Press, 2009); Shelley Fisher Fishkin, *Feminist Engagements: Forays into American Literature and Culture* (New York: Palgrave Macmillan, 2009); Sharon Patricia Holland, *The Erotic Life of Racism* (Durham: Duke University Press, 2012); Tanya Ann Kennedy, *Historicizing Post-Discourses: Postfeminism and Postracialism in United States Culture* (New York: State University of New York Press, 2017); Emily Westkaemper, *Selling Women’s History: Packaging Feminism in Twentieth-Century American Popular Culture* (New Brunswick, NJ, and London: Rutgers University Press, 2017); Richard Grusin, ed., *Anthropocene Feminism* (Minneapolis: University of Minnesota Press, 2017); and Robin Truth Goodman, *The Bloomsbury Handbook of 21st-Century Feminist Theory* (London: Bloomsbury Publishing, 2019).

³¹ See, for instance, Amanda Anderson, Rita Felski, and Toril Moi, *Character: Three Inquiries in Literary Studies* (Chicago: University of Chicago Press, 2019).

³² Felski, *Literature After Feminism*, 20.

³³ Data for all subcorpora was collected in May and June 2020. Across the four interpretive communities, there are 17 instances of novels in English translation, for example: Ameko Kaeruda’s *Sexiled: My Sexist Party Leader Kicked Me Out, So I Teamed Up With a Mythical Sorceress!* translated from the Japanese by Molly Lee in 2020 (from AM); Cho Nam-Joo’s *Kim Jiyoung, Born 1982* translated from the Korean by Jamie Chang in 2020 (from PR); and Charlotte Roche’s *Wetlands* translated from the German by Tim Mohr in 2010 (from PR).

³⁴ We excluded interviews from PR because they rarely identified novels as “feminist” or discussed specific novels vis-à-vis feminism; they more often mentioned feminism in its historical context or individual feminist luminaries.

³⁵ Though AM is more heterogenous than the other interpretive communities, its materials all serve a marketing purpose. Amazon’s famously opaque search algorithms also revealed clear marketing aims; in several cases, the same novel appeared multiple times in our search results, sometimes under the label “Sponsored” and sometimes not.

³⁶ Jordan Weissmann, “The Simple Reason Why Goodreads Is So Valuable to Amazon,” *Business, The Atlantic*, April 1, 2013, <https://www.theatlantic.com/business/archive/2013/04/the-simple-reason-why-Goodreads-is-so-valuable-to-amazon/274548/>. From a literary critical and sociological perspective, Mark McGurl also takes seriously what the rise of Amazon.com means for contemporary literary history; see McGurl, *Everything and Less: The Novel in the Age of Amazon* (London: Verso, 2021).

³⁷ Jennifer Mathieu, “Tank Tops and Short Shorts and Bears, Oh My!”, *Books, The New York Times*, August 1, 2020, <https://www.nytimes.com/2020/08/01/books/review/carrie-firestone-dress-coded.html>.

³⁸ Katherine Bode, “Why You Can’t Model Away Bias,” *Modern Language Quarterly* 81, no. 1 (March 2020): 96-97.

³⁹ D’Ignazio and Klein, *Data Feminism*, 18.

⁴⁰ We arrived at these numbers by evaluating how the discourse around authors assigns gender identity. We visited personal websites, Wikipedia pages, and other online sources, such as author profiles on publisher websites, to determine gender identification for each author. The male-identified authors are: in AM, Rudolfo A. Anaya, David Blixt, Daniel Taylor, H.G. Wells; in GR, Nathaniel Hawthorne, Khaled Hosseini, Stieg Larsson; in LC, André Brink, Nuruddin Farah, Zulifkar Ghose, Abdulrazak Gurnah, Mohsin Hamid, Ben Okri; and in PR, Jorge Barón Biza, Daniel Riley, Tim O’Brien, Anthony Quinn, Ngugi wa Thiong’o.

⁴¹ Mark Algee-Hewitt and Mark McGurl have observed a dramatic increase in twentieth-century anglophone publishing using data from the Bowker *Books in Print* database, noting that “the number of unique fictional texts published per year in English worldwide has grown from 7,948 to 278,985” between 1969 and 2014 alone (“Between Canon and Corpus: Six Perspectives on 20th-Century Novels,” *Pamphlets of the Stanford Literary Lab* 8 (2015): 2).

⁴² Andrew Kachites McCallum, “MALLET: A Machine Learning for Language Toolkit,” 2002, <http://mallet.cs.umass.edu>.

⁴³ Benjamin M. Schmidt, “Words Alone: Dismantling Topic Models in the Humanities,” *Journal of Digital Humanities* (2012). Mark Algee-Hewitt, “Computing Criticism: Humanities Concepts and Digital Methods,” *Debates in the Digital Humanities* (University of Minnesota Press, forthcoming).

⁴⁴ Schmidt, “Words Alone,” paragraph 5.

⁴⁵ Mark Algee-Hewitt, “Computing Criticism,” 13.

⁴⁶ This framework echoes Walsh and Antoniak’s approach to topic modeling as “an interpretive tool,” not “an objective magic wand” and a tool that involves “significant and subjective decision[s]” such as the number of topics in each model (“The Goodreads ‘Classics,’” 264).

⁴⁷ We ran full corpus topic models by downsampling and reconfiguring the GR subcorpus to produce word counts and document numbers in the same order of magnitude across all four subcorpora. We constructed this GR subcorpus sample by placing reviews of each novel across all five star categories in the same document, yielding 70 documents containing a total of 358,551 words. We then ran a number of topic models using different numbers of topics and found that the 10-topic model balanced relative coherence of topics with descriptive breadth across our large and heterogenous corpus.

⁴⁸ The full GR subcorpus is composed of a selection of 1-, 2-, 3-, 4-, and 5-star reviews of each of the 70 novels in this subcorpus. For example, all of the 1-star reviews of Leni Zumas’s *Red Clocks* (2018) are in one document, and so on for other star ratings. For Figure 4, we ran a 5-topic model of the 5-star reviews of all 70 novels (70 documents; a total of 921,343 words) and compared it to a 5-topic model of the 1-star reviews of these same 70 novels (again, 70 documents; a total of 617,513 words). We used the MALLET topic model toolkit to measure the coherence of each topic in both the 1-star and 5-star review models. Coherence measures whether the top terms of a given topic tend to co-occur together across documents, as Andrew Kachites McCallum explains (“Topic model diagnostics,” 2018, <http://mallet.cs.umass.edu/diagnostics.php>). As Mark Algee-Hewitt describes, coherence “can be used to assess topics for their relative meaningfulness to experts in the field upon which the corpus is based” (“Computing Criticism,” 13). As log probabilities, coherence scores are negative; the larger the negative value, the less often the words co-occur (or the less coherent the topic); the closer to zero the value, the more often the words co-occur (or the more coherent the topic). For the GR 1-star model, the coherence scores are: -221.4674 for topic 0, -41.8934 for topic 1, -34.8486 for topic 2, -97.3759 for topic 3, and -3.1615 for topic 4. For the GR 5-star model, the coherence scores are: -49.8547 for topic 0, -5.5319 for topic 1, -131.3004 for topic 2, -24.0985 for topic 3, and -281.6439 for topic 4. These results echo our qualitative sense that despite the relatively small corpus size (in terms of number of documents and number of words across those documents), the topics cohere. For more on the utility of coherence scores and how the measure is calculated, see David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum, “Optimizing Semantic Coherence in Topic Models,” *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2011).

⁴⁹ Again, we ran a number of topic models on this 40-document subcorpus (composed of one plain text file for each article) and found that the 10-topic model produced the most coherent and legible set of topics, providing the most descriptive overview of this collection of documents based on our close reading of the articles.

⁵⁰ We calculated coherence scores for this topic model according to the method described above. For the LC 10-topic model, the coherence scores are: -38.4437 for topic 0, -146.9340 for topic 1, -141.3961 for topic 2, -189.4120 for topic 3, -112.2622 for topic 4, -209.5487 for topic 5, -489.2012 for topic 6, -306.6074 for topic 7, -40.3141 for

topic 8, and -204.1436 for topic 9. Again, these results support our qualitative sense that despite the relatively small size of the LC subcorpus, the topics cohere.

⁵¹ Sarah Allison, Marissa Gemma, Ryan Heuser, Franco Moretti, Amir Tevel, and Irena Yamboliev, “Style at the Scale of the Sentence,” *Pamphlets of the Stanford Literary Lab* 5 (2013): 10. The MDW method developed at the Literary Lab is one of several methods that attempt to determine which words are significantly more important in one set of documents of a large corpus than another. For example, Christof Schöch et al. have compellingly explored John Burrows’ Zeta, “a measure of distinctiveness or keyness” (Christof Schöch, Daniel Schlör, Albin Zehe, Henning Gebhard, Martin Becker, and Andreas Hotho, “Burrows’ Zeta: Exploring And Evaluating Variants And Parameters,” *DH 2018*, <https://dh2018.adho.org/en/burrows-zeta-exploring-and-evaluating-variants-and-parameters/>). Both Burrows’ Zeta and the Literary Lab’s MDW method are informed by TF-IDF (term frequency-inverse document frequency), a measure used in many systems of modern information retrieval. For more on the latter, see Ho Chung Wu, Robert Wing Pong Luk, Kam Fai Wong, and Kui Lam Kwok, “Interpreting TF-IDF Term Weights as Making Relevance Decisions,” *ACM Transactions on Information Systems* 26, no. 3 (June 2008).

Each of these methods of determining distinctiveness has limitations. Weighing these factors, which we gloss below, we use the Literary Lab’s MDW method and take additional steps to mitigate its drawbacks. As the MDW method is sensitive to frequency, it requires imposing some constraints. For example, since very rare words, especially those that show up only in one group of documents, are overweighted by the algorithm, we have imposed a minimum frequency of 5 appearances to avoid capturing a word that appears only a few times as exceptionally highly distinctive. Stopwords usually show up as significant due to their high frequency, so we have removed them from our analysis. The MDW method valuably attaches a significance score to each result (having established a threshold for significance through the Fisher’s exact test), by contrast to TF-IDF and Burrows’ Zeta, which create a score that can rank distinctiveness but do not assign it a significance without further testing. We attend to this significance score, read for trends in groups of related distinctive words to avoid placing undue interpretive weight on our selections, and synthesize our close readings of the results with those of the other methods we employ.

For further discussion of the limitations of other methods of determining distinctiveness, see Juan Ramos, “Using TF-IDF to determine word relevance in document queries,” *Proceedings of the first instructional conference on machine learning* 242, no. 1 (2003); Shahzad Qaiser and Ramsha Ali, “Text mining: use of TF-IDF to examine the relevance of words to documents,” *International Journal of Computer Applications* 181, no. 1 (2018): 25-29; Sean G. Weidman 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): 374-390; and Schöch et al. Many thanks to Mark Algee-Hewitt for help with these references.

⁵² For a recent account of the nuances of reader-character identification, see Rita Felski, “Identifying With Characters” in Amanda Anderson, Felski, and Toril Moi, *Character: Three Inquiries in Literary Studies* (Chicago and London: The University of Chicago Press, 2019), 77-126.

⁵³ “Feminist” (obs = 12, p-value = 1.79E-29, odds-ratio = 1.877155172); “feminism” (obs = 366, p-value = 3.32E-13, odds-ratio = 1.867346939).

⁵⁴ Melanie Micir, “The Impossible Miss Woolf: Kate Atkinson and the Feminist Modernist Historical Novel,” *Modern Language Quarterly* 78, no. 4 (December 2017): 520.

⁵⁵ “Nonfeminist” (obs = 12, p-value = 0.038404521, odds-ratio = 3).

⁵⁶ Ellen Peel, “Unnatural Feminist Narratology,” *Storyworlds: A Journal of Narrative Studies* 8, no. 2 (2016): 103.

⁵⁷ This social reader here evokes the Argentinian cultural critic Flavia Dzodan’s viral assertion: “My feminism will be intersectional or it will be bullshit!” (*Tiger Beatdown*, October 10, 2011, <http://tigerbeatdown.com/2011/10/10/my-feminism-will-be-intersectional-or-it-will-be-bullshit/>). Social readers also engage this terminology to evaluate novels for their feminist relevance over time. One subset of these assessments considers whether a novel previously identified as feminist remains feminist by contemporary standards. For instance, a 3-star reviewer of Charlotte Brontë’s *Jane Eyre* asks, “Was it feminist? Yes. But was it *feminist enough* for a woman of today? In my honest opinion, no.” A 5-star reviewer disagrees, arguing that the character Jane is “a feminist for all ages.” While some social readers say they feel like a bad feminist for rating certain novels poorly,

there is nonetheless a pervasive sense that certain “feminist classics”—from *Herland* (1915) to Joanna Russ’s *The Female Man* (1975)—haven’t aged well. Yet social readers do not confine their disapproval to authors past. A 2-star reviewer calls *The Penelopiad* “a testament to the generational differences between more traditional feminists, like Margaret Atwood, and feminists in my own age group.” While the timeless/retrograde opposition situates novels along a historical axis of reception, these commentaries also establish a more subjective axis of viable models of feminist value systems, which readers recalibrate in a self-reflexive process of articulating the feminism they aspire to live.

⁵⁸ “Self Care: A novel,” Amazon.com, June 30, 2020, https://www.amazon.com/gp/product/0143135198/ref=ox_sc_act_title_1?smid=ATVPDKIKX0DER&psc=1.

⁵⁹ Many social readers critique this tautological capitalist logic (e.g., “you should buy this because others have bought it”), warning “Just don’t believe the hype!” (from a 1-star review of *Their Eyes Were Watching God*).

⁶⁰ See Lisa Mendelman, “Ambivalence and Irony: Gendered Forms in Interwar America,” *Arizona Quarterly* 71, no. 4 (2015): 23-52.

⁶¹ In a parallel vein, Tara K. Menon argues that ambivalence does not necessarily serve a feminist agenda in novels “unless it is undergirded by fundamental feminist principles” (“What Women Want,” *Public Books*, June 24, 2020, <https://www.publicbooks.org/what-women-want/>).

⁶² PR’s MDW results seem to aggregate patterns from other subcorpora, much like PR documents appear in the midst of AM and GR on the PCA biplot. For example, “pleasure” (obs = 36, p-value = 0.009911452, odds-ratio = 2) is distinctive of PR but evokes GR’s commitment to the embodied reading experience in more formal language, while “latest” (obs = 16, p-value = 0.026234952, odds-ratio = 2.666666667) and “current” (obs = 29, p-value = 0.024374837, odds-ratio = 1.933333333) suggest AM’s consumer-oriented impulses. Meanwhile, “complicated” (obs = 19, p-value = 0.026114042, odds-ratio = 2.375), “ambitious” (obs = 19, p-value = 0.026114042, odds-ratio = 2.375), and “seems” (obs = 91, p-value = 0.00916838, odds-ratio = 1.491803279) echo LC’s and GR’s respective registrations of complexity. The word “ambivalence” itself appears in the LC, GR, and PR subcorpora and is not distinctive of any one of them.

⁶³ See David Bamman, “BookNLP,” <https://github.com/dbamman/book-nlp>. For an explanation of how BookNLP works and the pipelines through which it was built, see David Bamman, Ted Underwood, and Noah Smith, “A Bayesian Mixed Effects Model of Literary Character,” *ACL* (2014). Mark Algee-Hewitt helped us write a “batch” code to perform BookNLP on all corpus documents simultaneously.

⁶⁴ Elaine Blair, “A Female Antihero,” *The New Yorker* 92, no. 38 (Nov. 21, 2016): 42.

⁶⁵ McAfee, *Feminism: A Quick Immersion*, 18.