Contents

1 one

1.1 chapter summary

"'A Melon, an Emerald, a Fox in the Snow': Quantifying Gender in Virginia Woolf's *Orlando: A Biography*"

The burden of this chapter is to describe how to do text analysis, or "distant reading," in a way that attends to queer concepts of gender performativity, and how that method might leverage computational processes to analyze literary material that expresses complex gender ontologies. This chapter proposes a method for text analysis to study the phenomenon of gender in Virginia Woolf's novel, *Orlando: A Biography* (1931).

I begin the chapter by critiquing what I call "the fantasy of falsifiable criticism," which contrasts traditional distant reading practices with experimental methods that deconstruct social categories. My own method of text analysis draws from these deconstructive approaches to create a new way for understanding gender ontolgies in quantitative form. For the middle portion of the chapter, I draw connections between computer programming and gender theory. This chapter looks to moments where computational processes collapse or evacuate the nuances of gender, sex, and sexuality in literary text. Such moments engage what I call the "reproducible" methodologies in current distant reading

scholarship, which largely analyzes gender in text as a binary system and reproduces this assumption in their results. To resist such methods, I explore how distant reading practices might borrow from the "iterativity" of gender theory in Queer Studies. Here, I take Judith Butler's concept of gender performativity as a basis for modeling a new approach that uses text analysis to deepen the study of gender in *Orlando*.First, I delve into python programming, focusing on the structure of the `for loop` and processes for cleaning and regularizing text, with the goal of bringing out the recursive quality of running python code. Then, I dip into Judith Butler's concept of gender performativity, which lends an understanding to the ways that critical processes can subvert dominant structures through iteration, or what she calls "performative citation." The chapter ends with my text analysis of gender in *Orlando* to demonstrate how the terms "man" and "woman" in that text are re-signified from their initial binary structure into a plural understanding. Here, I ground a distant reading of Woolf's text in close-reading to examine the ways that gender is represented as a problem of signification. Here, I apply Butler's notion of displacement through repetition in gender performativity back to text analysis to illustrate how the iterative process of analyzing text can surface new textual structures that re-signify certain elements of that text.

1.2 intro

The novel *Orlando: A Biography* (1928), by Virginia Woolf, famously opens with an assertive gender designation followed by an immediate qualification: "He–for there could be no doubt of his sex, though the fashion of the time did something to disguise it—-was in the act of slicing at the head of a Moor which swung from the rafters" (11). For those who are unfamiliar with the text, *Orlando* is a fictional biography that follows the life of the eponymous 16th-century English nobleman as he undergoes a sex change and lives into the 20th century as a woman. When performing quantitative text analysis on this text, the standard tasks of "pre-processing" the text evacuate the ways that gender is unsettled in this sentence. In order to computationally analyze a text, a process known as "distant reading" which involves calculating and visualizing textual patterns, the text must first be transformed into a computable format. This task of pre-processing (also called "cleaning" or "normalizing") strips the original text of capitalized words, punctuation, "stop words" (such as articles and prepositions), and inflections in word endings, all of which are deemed to be semantically minor, in order to make the text amenable to quantitative analysis. After pre-processing Orlando, the following list of computable words, or "tokens," remains in the first sentence:

'could', 'doubt', 'sex', 'though', 'fashion', 'time', 'something', 'disguise', 'act', 'slicing', 'head', 'moor', 'swung', 'rafter'.

Cleaning this text not only strips it of its pronouns, including the gender assertion in the first word, "He." It also cuts the em dash immediately following this "He," which signals the entrance of a narrator that layers this assertion with conspicuous certitude: "—-for there could be no doubt of his sex. . . ."

This chapter examines how quantitative text analysis works with gender, using Woolf's *Orlando* as a test case. Text analysis borrows from natural language processing methods to do analyses like topic modelling and sentiment analysis and machine learning methods like logistic regression and word embeddings (discussed below), which involves counting, classifying, and predicting textual patterns. Though these methods differ in important ways, they share a faith in using the speed of computation, which can process very precise elements like word frequency or syntactic patterns, to analyze large collections of textual data.

1.3 the fantasy of the falsifiable

1.3.1 TODO streamline problem of scale to falsifiability

Because computers can process hundreds of texts at a time, "reading" at much faster rates than humans, they offer critics an opportunity for solving the problem of literary scale. This opportunity attracts literary critics like Franco Moretti, Matthew Jockers, Ted Underwood, among others, who pose ambitious questions about literary history. According to Moretti, the process of quantification involves steps that reduce or abstract textual elements into computable form in ways that open the potential for analysis: "fewer elements, hence a sharper sense of their overall interconnection" (*Graphs* 1). [1] Moretti, who is largely responsible for popularizing the practice of "distant reading" in English Studies contexts, explores how social and economic forces impacts literary form in the development of the modern novel. While

[1] Some recent developments in distant reading combine with close reading methods, as critics will use the results of quantitative analysis to identify key moments from the text that merit closer attention. Andrew Piper's methodology, which he calls "bifocal" reading, demonstrates how distant and close reading are used together, with distant reading providing the context or framework that guides close reading""We are no longer using our own judgments as benchmarks... but explicitly constructing the context through which something is seen as significant (and the means through which significance is assessed).... It interweaves subjectivity with objects" (Piper, Andrew. Enumerations: Data and Literary Study, 2018, 17).

critiques of Moretti abound (not least the #metoo ones), I want to focus here on an early moment in his career which reveals his motivations for moving toward what I call "the fantasy of the falsifiable." His purview extends to thousands of texts from about 1750 to the present day–a database that would otherwise be too large and unweildy to analyze though non-digital means. According to Moretti, the computer not only enables large-scale analysis but creates a formal stucture for understanding texts, whereby "distance ... is not an obstacle, but a condition of knowledge" (*Distant Reading* 48).

Underlying this desire for analyzing literature at scale is another desire to shore up the reliance on critical subjectivity in literary analysis. Moretti's early essay, "The Soul and the Harpy" (1983), lays out the reasoning behind developing a new methodological approach for literary criticism, which will eventually flourish into "distant reading." Moretti's positioning in this essay betrays a deep suspicion about the indulgence of critical subjectivity in analysis, which he casts as a concern about current methods of literary criticism, which he calls "a sort of cultivated accompaniment to reading" (21). By this, he means that the analytical practices of contemporary literary critics, which "revolve around concepts such as 'ambiguity' and the like" will "always be pushed into multiplying, rather than reducing, the obstacles every social science encounters when it tries to give itself a testable foundation" (22). Moretti offers a caricature of the "student" of literature to illustrate his critique:

> The [student]–unless desirous of turning into that legendary figure whose only pleasure lay in contemplating his own reflection–must concentrate on the dissimilarities and ruptures: on what has been lost and become irretrievably unfamiliar, and which we can 're-familiarize' only by doing such violence to it that we distort the objective, material consistency of every work which it is the task of scientific knowledge to reconstruct and 'salvage'. 14

According to Moretti, literary criticism encourages the critic's narcissism. Moretti's alternative methodology borrows from the sciences to maintains the "objective, material consistency" of the work by minimizing the potential for difference and disagreement. Grounding the critical process on a more solid foundation, Moretti calls for a "falsifiable criticism" that can "test" literary interpretations (21). *Falsifiable* here means that it must be contestable–one must be able to imagine an alternative to ensure hypothesis can be meaningfully tested. Falsifiable criticism pursues interpretations as "coherent, univocal, and complete," where the "results" can be compared with "data" which "appear contradictory or inexplicable in the light of the hypothesis

itself" (21). Moretti here envisions a literary criticism that would steadily progress toward irrefutable knowledge: "The day criticism gives up the battle cry 'it is possible to interpret this element in the following way', to replace it with the much more prosaic 'the following interpretation is impossible for such and such a reason', it will have taken a huge step forward on the road of methodological solidity" (22).

As Moretti's falsifiable criticism eventually grows into "distant reading," he streamlines his critical method into a recursive process of posing hypotheses, collecting and assembling data, and making inferences. The results are often unexpected, and sometimes bring Moretti to reframe his hypotheses. This move relegates the literary critic to the role of explaining results. For example, in "Style, Inc.: Reflections on 7,000 Titles (British Novels, 1740-1850)," Moretti plots book titles on a series of graphs in order to explore how market forces influence the size and content of the titles. He finds that titles are quite sensitive to the market: "As the market expands, titles contract; as they do that, they learn to compress meaning; and as they do that, they develop special 'signals' to place books in the right market niche" (204). Moretti's incisive interpretation, which is borne by the various graphs of title lengths over time, obscures the fact that it is, fundamentally, an interpretation. A closer look at his language betrays the extent to which he makes interpretative moves at every step:

> in what follows, I focus on three moments of this history: first, I *describe* a major metamorphosis of eighteenth-century titles, and *try to explain* its causes; next, I *suggest* how a new type of title that emerged around 1800 may have changed what readers expected of novels; and finally, I *make a little attempt* at quantitative stylistics, *examining* some strategies by which titles point to specific genres. Three sections, three pieces in the large puzzle of the literary field. (181-2; emphasis mine)

1.3.2 TODO add to Ramasy's critique? whats the point here? recursivity?

Moretti's word choices here diminish the subjective work of critical analysis– he *describe[s]*, *suggest[s]*, *examin[es]*, etc. As Stephen Ramsay points out, this language presents the literary critic's work as an objective description of unproblematic reality, as if his conclusions are supplementary to the graphs and other visualizations. According to Ramsay, Moretti's claims that his insights are "independent of interpretation" suggest that "data is presented to us... not as something that is also in need of interpretation" (*Reading*

Machines 5). The computer's ability to quantify textual data emboldens Moretti toward making seemingly objective claims about literary history, all the while diminishing his role as in analysis.

The faith in the "falsifiable" gets its strongest expression in a famous detraction of quantitative methods in literary studies by Nan Z. Da. Da argues that quantitative methods, which trade "speed for accuracy, and coverage for nuance," reveal a "fundamental mismatch betwen the statistical tools that are used and the objects to which they are applied," (620, 601). In her critique, which cites "technical problems, logical fallacies," Da emphasizes the lack of *reproducible* results–that one researcher's process can be reproduced by another researcher with identical output (601). She demonstrates her point with an experiment in Topic Modeling, which generates a number of "topics" from a textual corpus and is often used to "speed-read" massive datasets to get a sense of their content. Da attempts to verify the results of a Topic Modeling experiment by replicating the process on her own machine, a replication which fails. She concludes that, "if the method were effective, someone with comparable training should be able to use the same parameters to get basically the same results" (628-629). As Ben Schmidt points out, however, Da in fact uses different parameters and software to run her experiments, which explains her results. His critique points out that Da, in taking down the quantification for literary studies, reinforces its role in positivist inquiry:

> Rather than pooh-pooh statistical reasoning, she elevates it by incanting the language of quantification against itself. Far *more* than anyone I've seen in any humanities article, she asserts that scientists do something arcane, powerful, and true. [2]

1.3.3 TODO reinforce/reframe w/r/t/ reproducibility

Despite their vastly different committments and conclusions, Da and Moretti align on the value that they place in the objective capacieis of computational analysis. For both scholars, there is a implicit faith that the computer can do some of the analytical work that has been traditionally ascribed to the human.

Unlike Moretti and Da, Ted Underwood demonstrates a working awareness with computational bias. Underwood's project uses machine learning methods, in which a sample of data is used to make predictions about larger groups

[2] For a more thorough critique of Da's aims and methodology in this article, please see Ben Schmidt's "A computational critique of a computational critique of computational critique," *Ben Schmidt*, Dec 5, 2019. `https://benschmidt.org/post/critical_inquiry/2019-03-18-nan-da-critical-inquiry/`

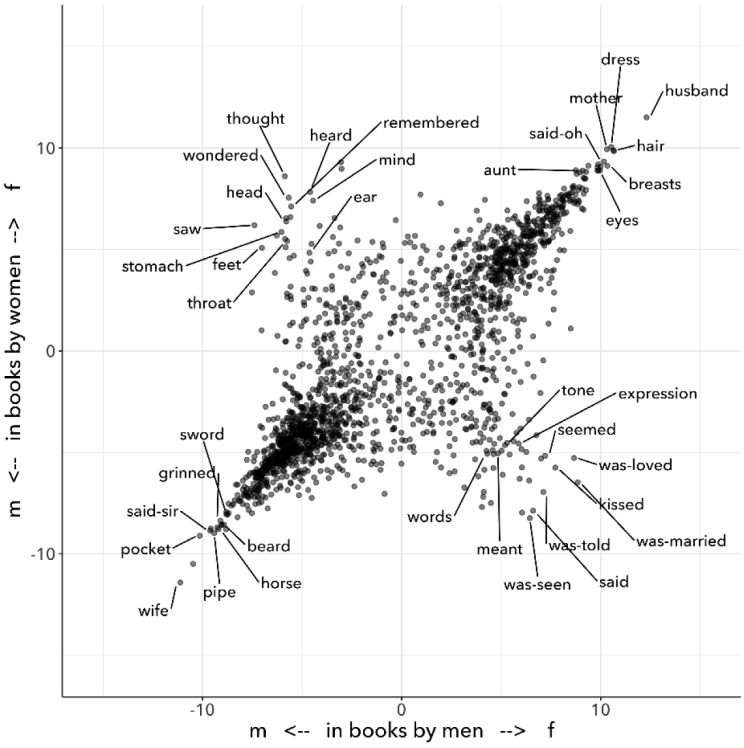
of data, to visualize the "distant horizon" of literary trends across centuries. One of his projects concerns studying how gender markers in novels change over time, using a process he calls "perspectival modelling." Underwood explains that "Machine learning algorithms are actually bad at being objective and rather good at absorbing human perspectives implicit in the evidence used to train them" ("Machine Learning and Human Perspective" 92). According to Dan Sinykin, this method allows Underwood "to leverag[e] the human prejudices built into modeling toward humanistic ends" (par. 4). For example, in his study of gender, Underwood uses a model that predicts the the sex of a fictional character based on the words associated with that character. Underwood explains the process of training the predictive program:

> We represent each character by the adjectives that modify them, verbs they govern and so on–excluding only words that explicitly name a gendered role like *boyhood* or *wife*. Then, we present characters, labeled with grammatical gender, to a learning algorithm. The algorithm will learn what it means to be 'masculine' or 'feminine' purely by observing what men and women actually do in stories. The model produced by the algorithm can make predictions about other characters, previously unseen. *Distant Horizons* 115

This particular model reveals that that, over time, gender roles in novels become more flexible while the actual number of female characters declines (*Distant Horizons* 114). One of the many explanations for this result, Underwood reasons, is that the practice of writing becomes more commonly pursued as a male occupation in the middle of the 20th century than it was previously (*Distant Horizons* 137). This fact, coupled with the tendency of men to write more about men than women, suggests why less women writing would led to a decline in female characters. This explains how Underwood's seemingly paradoxical conclusion, that gender roles become more flexible while the actual prevalence of women dissapates from fiction, might be possible.

Underwood's methodology, however reinscribes the binary that he attempts to deconstruct. This becomes most clear in his study that measures the "gendering of words used in characterization" ("Machine Learning and Human Perspective" 95). Here, Underwood uses uses logistic regression analysis, which is an algorithm specifically designed to study binary values, to see whether words align with masculine or feminine characters. This algorithm, which is an entry-level machine learning method, is designed for making predictions on a scale of probability, from 1 to 0, for example, between yes/no, pass/fail, win/lose, etc. In Underwood's case, the probability is

male/female, so the output necessarily reifies this binary structure of gender. For this project, Underwood admits that he needs a "simple" model in order to bring into relation the dynamics of gender (See Fig. 2). He admits that "gender theorists will be frustrated by the binary structure of the diagram" which "reduce[s] the complex reality of gender identification to two public roles: men and women" ("Machine Learning" 98). In aiming for simplicity, Underwood indicates that his initial assumptions will affect the final result.



Caption: Underwood's logistic regression model. The verticle axis visualizes the representation of words by women, and the horizontal by men, with positive numbers signifying overrepresentation of these terms. So terms on the top right are words that are used often by men and women writers, and terms in the upper left and lower right are ones used most often by women and men, respectively.

Collapsing of gender into a single graph might bring to the surface the various "perspectives" on gender markers across time, but it does so within a larger reification of the categories of male and female. Underwood himself

admits the possibility that he finds a structural tension between gender "because [he] explores gender, for the most part, as a binary opposition" (*Distant Horizons* 140). Asking a machine to compute the conscription of gender as male or female for the purpose of seeing how male and female roles in novels change over time only *reproduces* a model of gender that is "simple" enough to be computed.

Without a doubt, reproducing conceptions of gender is useful for historicizing gender identities and ideologies over time. In my view, however, these approaches fail to harness the potential of both computation and gender. Regardless of the methodology, it seems that the goal of establishing some kind of knowledge about literary history, whether that be a "distant horizon," or "the great unread," side-steps some of the more novel and novel insightful processes a computer might undertake. Distant reading methods might, for example, harness what Stephen Ramsay calls "the objectivity of the machine," to destabilize the binary, readings that are inescapably partial and speculative(x). Drawing from the deformative critical methods of Jerome McGann and Lisa Samuels, Ramsay proposes that researchers harness the enabling constraints of computation to "unleash the potentialities" of the text, offering opportunities for new readings (33).

Resisting the temptations of falsifiable criticism, work by critics like Susan Brown and Laura Mandell apply distant reading methods toward deconstructing the historical concepts of gender. In their introduction to *The Journal for Cultural Analytics*'s "Identity Issue," Brown and Mandell situate feminist debates around identity politics as a necessary context for understanding how computational processes engage gender identity. They explain that, "The goal is to acknowledge the subjective effects of belonging to an identity constituted historically through oppression without believing that the identity itself exists independently from historical conditions" (Mandell and Brown 6). In other words, because identity labels are historically constructed, the computer can be used to study this construction as a historical phenomena. Crucially, this position places computational methods within a discursive frame, aligning it with debates from post-structuralist feminist theory that explore and provoke the representative capacities of language. The computer can become a tool, not for verifying/reifying what we know, but for exploring how language constructs (and can deconstruct) categories.

Laura Mandell, for example, uses distant reading to deconstruct what she calls the "M/F binary," which is the reduction of gender into data. In her critique of Matthew Jockers and Jan Rybicki, Mandell demonstrates that the M/F binary reifies stereotypes in their premises, by "presenting conclusions about 'male' and 'female' modes of thinking and writing as if the M/F terms

were simple pointers to an unproblematic reality, transparently referential and not discursively constituted" (par. 5). Mandell's examination marshalls key findings from feminist theory, drawing from Judith Butler, among others, to assert that gender is a socially constituted category which is "constructed both by the measurer and the measured" (par. 38).

To deconstruct gender, Mandell turns to genre, another category which will allow scholars to see the reductive constitution of categories generally. To study genre, Mandell uses the popular stylometry measurement, "Burrow's Delta," which visualizes the "distance" between writing styles by creating branches (or "deltas") between different texts. Her experiment finds that the stylistic qualities of a female writer, Mary Wollenstonecraft, shares with those of male writers: "Wollstonecraft's sentimental anti-Jacobin novels most resemble [William] Godwin's sentimental anti-Jacobin novels... whereas her essays most resemble [Samuel] Johnson's writings" (par. 29). Just as quantification can deconstruct what So and Roland describe as "the machine's initial binary understanding of white and black," so it can deflate the M/F binary (So and Roland 68). Moving beyond deconstruction, however, Mandell encourages rearchers "to experiment with new taxonomies of gender," creating new categories that reflect gender as a multiplicity (par. 37). Mandell emphasizes the potential for complex data models to "break the strength of the signal" by creating categories such as "'men writing as men,' 'women writing as women,' 'women writing as men,' 'men writing as women,' 'unspecified (anonymous) writing as men,'" and so on (par. 35). She points out that the computer allows researchers to "animate numerical processes rather than fixing their results as stereotype" (par. 7). It offers, in Mandell's words, "parallax, multiple perspectives for viewing a very complex reality" (par. 38).

Deconstructing binaries also works with race. Edwin Roland and Richard study explore the constructedness of racial categories by experimenting with an algorithm that evaluates whether an author is white or black based on diction. Analyzing a large corpora of novels by white and black authors, they find that, black authors generally display more varied vocabulary than white authors (66). From this they infer that white authorship, as a category, only coheres against the variance of black authorship. Whiteness, in other words, *depends* on the characterization of blackness.[3]

This quantitative exercise, rather than draw So and Roland toward making

[3] Tie this relationship on the white/black binary to Eve Sedgwick's points about binaries containing an oppostional dynamic in which the subordinated term props up the dominant term.

general conclusions about race and authorship, points them toward a peculiarity in the results: that the algorithm wrongly categorizes James Baldwin's novel *Giovanni's Room* (1956) as being written by a white author. Apparently, the computer reads Baldwin's use of the term "appalled" as proof of white authorsip. Going back to examine this word in the text, So and Roland discover that this term occurs only once, in the early scene where David (the narrator) describes his strained relationship to his father: "I did not want to be his buddy. I wanted to be his son. What passed between us as masculine candor exhausted and *appalled* me" (my emphasis; Rpt. in So and Roland 71). Noting the connotations of whiteness in "appalled," which has the middle French root, "apalir," meaning "to grow pale," So and Roland insightfully conclude that this term indexes an intersection of gender with race: "the moment David develops a troubled relationship to normative masculinity [as] also the moment he becomes 'white'" (71). The computer's misclassification, as they point out, reinforces this text's notorious elision of explicit references to race, whereby racial markers are displaced in favor of an implicit whiteness, as critics have observed in the scholarship on this novel. Taking the computer's mistake as a starting point, So and Roland's analysis thus contributes to the ongoing debate about the complex relationship between race and sexuality in the novel.

Here, in direct opposition to the "falsifiable" position, computational error becomes a starting point for analysis. Because race is a social construct, and machines only impute meaning that is encoded into them, So and Roland reason that machines are be ideal instruments for studying the construction of race (60). In particular, the machine errors surface a yet unexplored fulcrum around which the binary of race turns:

> Our reading's destabilization of the machine's logic of white and black arises directly from the novel's expression of queerness. By queering the machine's color line, Baldwin's novel challenges our initial classifications of the novels as white or black, which had necessarily effaced a more sophisticated, intersectional view of social identity. In their current form, our data and model are not robust enough to handle this kind of intersectionality. 72

In this case, a single computational error opens a site for more daring leaps of speculation about how whiteness gestures toward a troubled understanding of sexuality. So and Roland find that queerness here operates as an articulation (both structurally and semantically) of race. In the next section, I lay the groundwork for computationally analyzing queerness by turning to the inaugural moment in Queer Theory, gender performativity.

1.4 iteration

1.4.1 Gender Performativity

So, Roland, Mandell and Brown demonstrate how computation can be reworked toward deconstructing social categories. Mandell's work, in particular, opens up the consideration of how gender theory, Judith Butler's theory of gender performativity, might influence computational analysis.

She points out that both gender and genre "are... highly imitable" (par. 30), asserting that "Anyone can adopt gendered modes of behavior, just as anyone can write in genres stereotypically labeled M/F" (par.30). While this reading of Butler echoes a common assumption about performativity, which Butler has been careful to clarify in her writings since *Gender Trouble*, it also perceives a generative alignment between gender and computation. In what follows, I explore this alignment between gender and computation, which evoke similarities that are productive for text analysis. As Mandell points out, "Computation enables complexity" (par. 36), and gender, like computation, contains rules and protocols that build toward higher levels of complexity.

In her groundbreaking book, *Gender Trouble: Feminism and the Subversion of Identity* (1990), Judith Butler famously disrupts essentialist views of sex and gender in contemporary feminist thought: first, that sex is biological while gender is constructed; and second, the gender, as a construction, is a self-expression of the subject. Because sex and gender are both constructions that exist prior to identity. In fact, according to Butler, there is no such thing as a subject that exists prior to gender expression, as a subject only comes into being by participating in a gender norm. Here emerges the common misreading of performativity, that performativity denotes an act or series of acts that can be imitated at will, to be put on and off like clothing. As Butler emphasizes in her later work, performativity is compulsory and habitual, a process that *precedes* and *constitutes* the subject. Gender is a mechanism that allows the subject to emerge: "construction is neither a subject nor its act, but a process of reiteration by which both 'subjects' and 'acts' come to appear at all" (*Bodies* xviii). This process of *reiteration* is fully delineated in her follow up book, *Bodies That Matter* (1995), where she gives it the term "performative citation." Here, Butler argues that what is experienced as the physical body, its boundaries and its sexuality, only materialize through the repetition, the "citation," of gender norms, whereby each act signals an authorizing norm.

As I will demonstrate, Butler's theory of gender performativity has a lot

to lend to the study of computational text analysis. Before moving to the details of this theory, however, it is important to understand what such a theory can and cannot do. For example, common critiques of Butler point out the limits of this theory for working with discursive notions of gender and sexuality.[4] Jay Prosser, coming from the field of Trans Studies, problematizes Butler's "deliteralization of sex," a critique that he applies to Queer Studies more generally. Prosser explains that because Butler's analysis attends to performativity as a discursive phenomenon, it elides the real-world concerns of the body's materiality. Prosser offers the example of Butler's reading of *Paris Is Burning*'s Venus Xtravaganza who, Butler argues, occupies a space of transgression due to her inability to attain her sex change. According to Butler, a sex change that would "make [her]self complete" would also fulfill the desire for a masculine body would reinscribe heterosexual hegemony (45). Prosser points out that this reading fails to reckon with the material body and its real and precarious existence, as Venus's death illustrates (55). Here, Butler's "metaphorization of the transgender body" demonstrates one crucial way that Queer Theory has subsumed, without fully accounting for, transgressive desires in cross-gendered identifications. This thread of critique is picked up in the conclusion, where it instigates the next move within a larger trajectory of Queer Studies presented in this dissertation.

To understand the constraints of performativity as a discursive phenomenon, it is helpful to situate Butler's work within the context of second-wave feminism and its post-structural approach toward gender binaries. Here, Butler draws from the work of feminist theorist Luce Irigaray, whose critique of gender undermines what Jacques Derrida's defines as "phallogocentrism," the idea that man, symbolized by the phallus, is the center and focus of knowledge. Irigaray asserts that influential Western thinkers, like Plato, Aristotle, and Freud, for example, have defined women and feminity "on the basis of masculine parameters" (Irigaray, *The Sex Which Is Not One* 23). The resulting binaries that associate "woman" with "matter" (such as "rationality/emotion" and "mind/body"), and set it subordinate to male "form," effectively erase the possibility of representing woman at all. Rather,

[4]Another popular critique comes from Political Philosophy, and concerns a logical inconsitency in the way that Butler theorizes subjectivity. If the resistance to signification comes from outside the cycle of signification, from where does that external resistance emerge? Does it not imply a pre-discursive identity or at least desire for resistance? Geoff Boucher writes that Butler locates the potential for subversion "in a disembodied intentionality that appears to stand outside of the culturally-scripted subject positions that the individual occupies" (115). He aptly questions: "Who (or what) decides 'how to repeat'? On what basis is the decision to subvert power made?" (119).

the binary actually "*produces* the feminine as that which must be excluded for that [gender] economy to operate" (10; my emphasis). The produced "domesticated" feminine term contrasts the excessive feminine which cannot be expressed within the terms of the binary (13). This "necessary outside" of the excluded feminine, which is in fact is the enabling condition of the binary in the first place, creates a "field of disruptive possibilities" (13). However, this "unspeakable" element cannot be invoked directly, "through the figures that philosophy provides," without subscribing itself to the ruling structure (12). Butler illustrates this quandry with a hypothetical: "how can one read a text for what does *not* appear within its own terms, but which nevertheless constitutes the illegible conditions of its own legibility?" (11). For Butler, this question–how to express what is not there, what is refused by the system of the visible–will guide her theorization of gender subversion, what she calls resignification, through performativity.

The process of resignification begins by positing a body that exists prior to signification, that is, a body that has not yet been imbued with meaning through language. Butler wonders, "Can language simply refer to materiality, or is language also the very condition under which materiality may be said to appear?" (6). Butler finds that, in order to refer to a body, language must first assume a body. Therefore, she reasons, the signification of the body actually creates the body which it appears to reference: "This signification produces as an *effect* of its own procedure the very body that it nevertheless and simultaneously claims to discover as that which *precedes* its own action" (emphasis original; 6). Butler explains that "the mimetic or representational status of language. . . . is not mimetic at all. On the contrary, it is productive, constitutive, one might even argue performative" (6). This point, that language produces the reality that it claims to merely reference, has two crucial ramifications: first, that subjects are always interpellated, and in fact brought into subjectivity, by a discourse prior to their their participation in it; and second, that this productive power of language nonetheless offers a way out of the significatory circle.

For, amid this regulatory structure lies the possibility of what Butler describes as *resignifying* meaning. Because language transcends a merely representative function, because it works to *produce* meaning, language can be resignified toward subversive usages by "citing" what Bulter calls the "repudiated" meaning implied by signification. Butler offers a rather famous example in the resignification of the term "queer," which has been transformed from a term of abjection to one of empowerment. "Queer" achieves this resignification by harnessing its own repudiation, which is an implied but "disavowed abjection [that] will threaten to expose the self-grounding presumptions

of the sexed subject" (3). In other words, each time the term "queer" is used, it draws from that abjection which is repudiated in every identification with heterosexuality. Butler proposes that one "cite" this repudiation as a resource for resignification: "to consider this threat and disruption... as a critical resource in the struggle to articulate the very terms of symbolic legitimacy and intelligibility" (3). Here, the concept "citation" indicates an act of signification that draws from the authorizing power. By citing the repudiated meaning, the term "queer" "resignifyi[es] the abjection of homosexuality into defiance and legitimacy" (xxviii). The resignification works because this "performative citation" takes on the repudiation as its signification.

Because language is productive, it also offers a possibility of resistance from within the signification system. Butler illustrates how Irigaray achieves this resistance by performing the phallogocentric language of the thinkers that she criticizes: "she mimes philosophy... and, in the mime, takes on a language that effectively cannot belong to her" (12). Butler reads Irigaray's use performative citation as a strategy of undermining his authority through repetition: "She cites Plato again and again, but the citations expose precisely what is excluded from them, and seek to show and to reintroduce the excluded into the system itself" (18). Through repetition, Irigaray displaces the logic of phallogocentrism, introducing something external to the system while remaining within its terminology. Narrating what Butler imagines to be Irigaray's thought process in an invigorating monologue, she lays out the process of resistance:

> I will not be a poor copy in your system, but I will resemble you nevertheless by miming the textual passages through which you construct your system and showing that what cannot enter it is already inside it (as its necessary outside), and I will mime and repeat the gestures of your operation until this emergence of the outside within the system calls into question its systematic closure and its pretension to be self-grounding" (18).

In this description of resistance within the cycle of signification, where deception emerges from resemblance and insubordiation through subservience, the key is repetition. Resistance looks like repetition, a continual activity, the miming of the authorizing norm, which displaces it by introducing what is outside the logic of phallogocentrism.

1.4.2 Python, NLTK, and Word Embeddings

Now that we have a sense of gender peformativity, we turn to Python to get a closer look at how its syntax might evoke the process of iteration.

To do common text analysis tasks, where text passed through an automated seive to find patterns, many distant reading projects use the Python programming language, which offers a number of custom "libraries," or collections of code for specific tasks, such as analyzing textual data. The most popular text analysis library in python is the Natural Language ToolKit (NLTK), which contains useful computational "methods" and "functions" that count, categorize, and visualize textual patterns.

As illustrated in the opening example in this chapter, the process of preparing a text for text analyis always requires a reduction of data in which some semantic value has escaped. In this example, "cleaning" the first sentence of Woolf's novel, *Orlando*, strips it of its pronouns and punctuation which has the effect of surpressing the gender qualification: "He–for there could be no doubt of his sex, though the fashion of the time did something to disguise it—-was in the act of slicing at the head of a Moor which swung from the rafters" (11). After processing, the following words remain:

'could', 'doubt', 'sex', 'though', 'fashion', 'time', 'something', 'disguise', 'act', 'slicing', 'head', 'moor', 'swung', 'rafter'.

For analyzing text, Python works with data in the form of words, or `strings`, contained within groupings called `lists`. Then, Python *iterates* through the list, that is, it performs a similar task to each item in the list. For this purpose, an expression called the `for loop` consists of six words over two lines which instruct Python to do something to each item in the list, in other words, to "loop" through data, carrying out some specified action to each peice. The first line of the loop (`for word in sentence:`) specifies each word in the list, and the second line (`print(word)`) instructs the computer to display each word in the sentence. Essentially, this loop will go through each item in the data, in this case, each word saved in the list `sentence`, and it will `print` or display that data.[5] The the output will appear thus:

sentence = ['He', '–', 'for', 'there', 'could', 'be', 'no', 'doubt', 'of', 'his', 'sex', ',', 'though', 'the', 'fashion', 'of', 'the', 'time','did', 'something', 'to', 'disguise', 'it', '–', 'was', 'in', 'the', 'act', 'of', 'slicing', 'at', 'the', 'head', 'of', 'a', 'Moor','which', 'swung', 'from', 'the', 'rafters']

for word in $\text{first}_{\text{sentence}}$: print(word)

[5] In JavaScript, for example, the `for loop` is more convoluted:
for (i = 0; i < word.length; i++) { text += word[i] + "
"; }

['He', '–', 'for', 'there', 'could', 'be', 'no', 'doubt', 'of', 'his', 'sex', ',', 'though', 'the', 'fashion', 'of', 'the', 'time', 'did', 'something', 'to', 'disguise', 'it', '–', 'was', 'in', 'the', 'act', 'of', 'slicing', 'at', 'the', 'head', 'of', 'a', 'Moor', 'which', 'swung', 'from', 'the', 'rafters']

These kinds of iterative computations, which are central to programming tasks, are a core component of working with text. At a very basic level, much of text analysis consists of iterating over bits of text and doing something to each bit. In preprocessing, the main tasks include tokenizing, cleaning, and regularizing the text, which helps to eliminate pieces of text that will skew or slow results of analysis due to their high frequency and low semantic value. Tokenizing the text means separating the text into workable units, or `tokens`, that are easier to clean and regularize. Once the text is tokenized, it can be stripped of capital letters, punctuation, and what are called "stop words," which consist of prepositions, articles, and related terms, such as "he," "for," "there," "be," "of," "the," and "did" in the above example. The following code block loops through the text to remove punctuation and capital letters:

normalized = [] for word in $\text{full}_{\text{text}}$: if word.isalpha(): normalized.append(word.lower())

Before moving forward, there are two aspects about the cleaning and regularizing process that merit some attention: the first is recursion. The cleaning and regularizing process is highly recursive, doing the same action to each item to the list of words that make up the text. The logic of the code reinforces this recursiveness, especially in the loop which iterates through items in a list, doing the same thing to each item, one by one. Additionally, the code's nested expressions reinforce recursion, as each line specifies another action to be performed on each word. For example, in the following code block, the first line isolates a word from the list, the second line checks if that word contains only alphabetic characters, and the third transforms that word to lowercase. Each of the three lines performs an additional task on the same word.

The second notable aspect about the cleaning and regularizing process is reduction. These tasks of preprosessing text force words into existing boxes, so to speak, in order to make them amenable to analysis. The effect of this preprocessing therefore strips text of some of its semantic meaning, which can be contained in capitalized words, rhythms of language in stop words, inflections in word endings, and so on. This is not to say that preprocessing ought to be avoided, but that the researcher should be aware of how certain textual reductions have the potential to affect meaning.

Here, it begins by creating an empty list, `normalized`, where words will

be dropped after filtering through them. The next line begins the `for loop`, which iterates through each word in the `full_text` list of words. The third line, an `if statement` creates the condition specifying alphabetic characters (containing no numbers or punctuation), and if the word fulfills that condition, then it passes to the fourth line, which will add that word to the `normalized` list. At the moment that this word is added to the list, its letters will be transformed to lowercase format. The final list, therefore, will contain words that are all lowercase and contain no punctuation.

The next step involves removing stop words, then stemming/lemmatizing. For this process, the `for loop` can be compressed into a `list comprehension`:

no_{stops} = [word for word in normalized if word not in stops]

This expression takes each word in a list, in this case, `normalized`, and checks to see if that word is also contained within the list of stop words in `stops`. If the word is *not* a stop word, then it will be added to a new list, `no_stops`. Once this filtering is done, the final list contains all lowercase words without punctuation or stop words. For example:

['could', 'doubt', 'sex', 'though', 'fashion', 'time', 'something', 'disguise', 'act', 'slicing','head', 'moor', 'swung', 'rafters']

After cleaning the text in this way, the next step involves stripping the grammatical structure to get the word root. One of these processes, called "stemming", involves cutting off the endings from the word. For example, "rafters" will be stripped to "rafter." In another process, called "lemmatizing," the computer will look up each word, one by one, find its appropriate root, and then revert to that root.

clean = [WordNetLemmatizer.lemmatize(word, word) for word in no_{stops}]

At this point, the text is ready for analysis. The NLTK library comes packaged with a series of "exploratory" methods that offer quick analyses of textual patterns. At the base of many of these analyses are word frequencies based on the context surrounding a given word. For example, `concordance()` method returns the context, that is, the immediate words surrounding the word "woman" from the text of *Orlando*:

charm – all qualities which the old woman loved the more the more they failed yed her cheeks scarlet . For the old woman loved him . And the Queen , who knew rladen with apples . The old bumboat woman , who was carrying her fruit to mark a figure , which , whether boy 's or woman 's , for the loose tunic and trouser , for alas , a boy it must be – no woman could skate with such speed and vigo s not a handsbreadth off . She was a woman . Orlando stared ; trembled ; turned mult of emotion , until now ? An old woman , he answered , all skin and bone . e for sea birds and some old country woman

hacking at the ice in a vain attempt h their heat , and pity the poor old woman who had no such natural means of tha agan ; of this man 's beard and that woman 's skin ; of a rat that fed from her of melancholy ; the sight of the old woman hobbling over the ice might be the c en waters or night coming or the old woman or whatever it was , and would try t anners were certainly not those of a woman bred in a cattle-shed . What , then st career in the world for a Cossack woman and a waste of snow – it weighed no arms and vociferating . There was a woman in white laid upon a bed . Rough tho y , and when the Moor suffocated the woman in her bed it was Sasha he killed wi the cobbles , or at the rustle of a woman 's dress . But the traveller was onl hant , making home belated ; or some woman of the quarter whose errand was noth in water he hurled at the faithless woman all the insults that have ever been obinson by way of making a Christian woman of her , understood what they were a ght or the left ? The hand of man or woman , of age or youth ? Had it urged the with sobs , all for the desire of a woman in Russian trousers , with slanting

- like a dog chasing a cat or an old woman blowing her nose into a red cotton h

to talk about – a dog , a horse , a woman , a game of cards – seem brutish in out somehow to allude to this humble woman and her milk-pails , when the poet f

Building from the same concept as the `concordance()` method, another method, called `similar()` calculates words which are used in similar contexts as the target word. To compute the results of `similar()`, NLTK first takes the context of the term from `concordance()`, then it searches the text for other terms that contain similar contexts. The result for running `similar` on the word "woman" is the following:

man moment night boy word world child pen ship door one room window light little lady table book queen king

By searching the text for words that appear *similarly* to the chosen word, this method reveals words that function in semantically similar ways across the text. It is important to point out, however, that the text itself does not impute meaning to the words. Rather, it can only count words as "strings," that is, bits of data composed of the same characters. It takes the string "woman," takes notes of all of the strings in proximity to "woman," and then searches the rest of the text for *other* strings that have similar proximities. This method is based on counting frequencies of characters that occur near each other.

This method, which is a basic natural language processing task, contrasts with algorithmic and "deep learning" methods that work in more sophisticated ways to count and analyze language. Many of these methods use the concept of "word embeddings" to ascribe machine-interpretable meaning to words. Like `similar()` and `concordance()`, word embeddings build off patterns of word similarity based on context. Unlike the NLTK methods, however, word embeddings impute meaning to the contexts surrounding a given word. The meaning of any given word is a numerical representation, actually a list of numbers, in the form of a matrix. The classic example for introducing the power of word embedding methods is the formula, "King - Man + Woman = Queen" (Mikolev et al. 2). Here, gender is isolated as a computable component which enables the formula to derive the difference between "King" and "Queen".

In more technical terms, each word, such as "woman," is assigned a vector representation in n-dimensional space, where each dimension represents the similarity between woman and another word. For example, according to one language model, the word "woman," is calculated according to its similarity (or "weight") to other words, such as "mother" and "father." Here, the word "woman" is more closely associated with "mother," with its weight being .92, or 92%, then "father," which has a weight of .90, or 90%. In simpler terms, any given word is calculated according to its similarity to other words. And the similarity, in turn, is calculated by context. For example, below is a list of words from a popular language model calculated as similar to "woman".[6]

[('child', 0.9371739625930786), ('mother', 0.9214696884155273), ('whose', 0.9174973368644714), ('called', 0.9146499633789062), ('person', 0.9135538339614868), ('wife', 0.9088311195373535), ('being', 0.9037441611289978), ('father', 0.9028053283691406), ('guy', 0.9026350975036621), ('known', 0.8997253179550171)]

A word embedding for "woman," therefore, would contain a list of numbers representing the similarity of other words to it, organized within a tabular format. The word embedding would resemble the following matrix:

Target Word	child	mother	whose	called	person	wife	...
Woman	.937	.921	.917	.915	.914	.909	...

Given this tabular representation, numerous mathematical operations are possible using principles from statistics, linear algebra, and calculus, which is the realm of "shallow learning" methods. Within the more elusive realm of "deep learning" methods, like neural networks, the labels of the numerical

[6]The language model for this computation comes from Word2Vec's "glove-twitter-25" dataset.

representations do not matter. Rather, the only thing that matters is the list of numbers themeslves, which together, form a vector to represent the word. The word "woman," therefore, would be represented with the following vector: .937. .921, .917, .915, .914, .909, and so on. Deep learning methods demonstrate that, even when removing semantic labels, *words are assigned meaning by their relation to other words.* Even with each of these words represented as a vector with the labels removed, the sexism of the formula remains obvious: the woman is computed according to her relation to a man.

1.5 queer distant reading

1.5.1 Woolf's *Orlando*

I now turn to looking at gender in Virginia Woolf's novel, *Orlando: A Biography.* This novel is ideal for a computational study of gender for two reasons: first, it is perhaps the most salient example of transgender narrative in the modernist era, and second, because this narrative traces a crisis of signification of which gender is only one expression. As various critics have noted, *Orlando* deploys a characterisitic modernist experimentation with limits of language toward destabilizing gender norms.[7] In what follows, I pursue an *iterative* text analysis of this text that interweave distant and close readings based on the word embeddings of the gender markers, "woman," and "man."

First, I begin with a list of terms computed similar to woman and man, respectively, in the text. To get these results, I trained a language model on the entire text of *Orlando*, so that the word contexts would be based on how words are used in this specific text. Unlike the word embeddings

[7] Much of the scholarship on this text explores its resistance against normative concepts of identity and gender. The experimental use of language and narrative form creates a narrative that is recalcitrant against coherent understandings of gender and identity. Jane de Gay, Jill Channing, and Christy L. Burns, for example, assert that Woolf deploys imaginative elements, magical realism, and parody, respectively, to resist realism and narrative expectations in her fictional biography. De Gay aligns Woolf's writing with that of Walter Pater and Vernon Lee as a "feminist historiography" that "rejected Victorian patriarchal metanarratives" and instead "used the strategies of fiction to bring history alive and make it live in the present" (de Gay 71). In a similar vein, Burns and Channing both point out that Woolf uses fantastical elements, in the former in the service of parody, and the latter as part of magical realist writing, that disrupt expectations of plot and narrative to challange the stability of gender and identity. Doubling down on the role of langauge, some critics emphasize that the narration purposefully obfuscates any resolution about concepts like gender, identity, and even race and nationality. Victoria L. Smith asserts that "The fantastic content in the novel is directly linked to the undecidability/impossibility of the form of the novel and of the protagonist" (58).

from my previous section, which were trained on Twitter data, the results here therefore reflect an understanding of gender which is specific to Woolf's novel.

The following are words associated with "woman":

[('would', 0.5118660926818848), ('hand', 0.5049053430557251), ('night', 0.4855204224586487), ('though', 0.4815906882286072), ('way', 0.476143479347229), ('foot', 0.4528403580188751), ('orlando', 0.433744877576828), ('said', 0.43140658736228943), ('like', 0.41121190786361694), ('life', 0.4069981873035431)]

And the following are words associated with "man":

[('would', 0.6174017786979675), ('orlando', 0.6018419861793518), ('night', 0.5755824446678162), ('way', 0.5710440874099731), ('great', 0.5492382645606995), ('long', 0.5454811453819275), ('could', 0.53724604845047), ('table', 0.5338666439056396), ('thus', 0.533319354057312), ('said', 0.5238105058670044)]

At first glance, the lists reflect commonly used words, and appear somewhat similar, sharing terms like "would," "orlando," "night," and "way." To get more specific results, I modified the code to remove any words with strong associations to the opposite gender. The results revealed more distinctive words associated with each gender:

$>$ $\text{distinct}_{\text{w}}$ = $\text{model.wv.most}_{\text{similar}}$(positive="woman", negative="man")

[('soft', 0.3692586421966553), ('named', 0.34212377667427063), ('sciatica', 0.3223450779914856), ('frilled', 0.3187992572784424), ('despaired', 0.31375786662101746), ('friend', 0.31238242983818054), ('delicious', 0.30853813886642456), ('winked', 0.30514153838157654), ('notion', 0.3047487139701843), ('seductiveness', 0.30290719866752625)]

$>$ $\text{distinct}_{\text{m}}$ = $\text{model.wv.most}_{\text{similar}}$(positive="man", negative="woman")

[('chequered', 0.4025157392024994), ('fact', 0.3394489586353302), ('denounced', 0.3346075117588043), ('house', 0.33423593640327454), ('curiosity', 0.33144116401672363), ('defend', 0.3284823000431061), ('dancing', 0.3282632827758789), ('marbling', 0.3184848427772522), ('cynosure', 0.3057470917701721), ('rather', 0.3024100363254547)]

This list of results contains words more uniquely associated with each gender. The top terms for each list might align with existing conceptions of femininity and masculinity, such as "soft" for "woman," and "chequered" for "man." The rest of the terms also appear to uphold a binary understanding of gender, with words like "frilled," "delicious," and "seductiveness," associated with "woman," and "fact," "defend," and "denounced" associated with "man."

Beyond these general patterns, however, the results complicate an easy understanding of gender as binary. Rather, they suggest that gender is one expression for a larger crisis of signification in the novel. In what follows, I use some of these words as starting points for close-reading analysis of the text. I begin with unique words from both lists which, appearing only once in

the text, carry significant semantic weight in their relation to gender. Then, I examine words that co-occur in certain passages of the texts–moments which are provocatively indicative of the relationship between gender and language in the text.

Interestingly, while the top term for the "woman" category, "soft," is used 9 times throughout the text, the top term for the "man" category, "chequered" is only used once, at the very beginning of the story, when the narrator describes Orlando stepping into "the yellow pools chequered by the floor" (Woolf 12). This moment, as Orlando literally steps into the spotlight of the story, is the first of many in which the narrator casts doubt his credibility as a biographer, introducing a crisis of signification that will plague his narration. Soon after Orlando makes his appearance, the narrator distinguishes his role as a biographer from that of the poet, who works to embellish and exagerrate through figurative language. However, the narrator's committment to straightforward description soon unravels when he attempts to describe Orlando's beauty. Here, the language swells to full-fledged figuration:

> Directly we glance at Orlando standing by the window, we must admit that he had eyes like drenched violets, so large that the water seemed to have brimmed in them and widened them; and a brow like the swelling of a marble dome pressed between the two blank medallions which were his temples. Directly we glance at eyes and forehead, thus do we rhapsodize. Directly we glance at eyes and forehead, we have to admit a thousand disagreeables which it is the aim of every good biographer to ignore. 12-13

Here, the narrator's evocative language undermines the pretense to objectivity which he feels compelled to produce. This doubt, which I call the crisis of signification, reoccurs persistently throught the novel. That the usage of "chequered," a uniquely "masculine" term in the story, occurs in this passage, suggests that gender may play a central role in this crisis.

The crisis of signification on the part of the narrator also occurs within Orlando's experience itself. To reinforce this point, I take a from the "woman" list, "despaired" which, like "chequered," occurs only once in the novel. It appears at a point when Orlando, deep in a depression following his desertion by Sasha, the Russian princess, struggles to peice together his beliefs on truth and language:

> 'Another metaphor by Jupiter!' he would exclaim as he said this (which will show the disorderly and circuitous way in which

> his mind worked and explain why the oak tree flowered and faded so often before he came to any conclusion about Love). 'And what's the point of it?' he would ask himself. 'Why not say simply in so many words–' and then he would try to think for half an hour,–or was it two years and a half?–how to say simply in so many words what love is. 'A figure like that is manifestly untruthful,' he argued, 'for no dragon-fly, unless under very exceptional circumstances, could live at the bottom of the sea. And if literature is not the Bride and Bedfellow of Truth, what is she? Confound it all,' he cried, 'why say Bedfellow when one's already said Bride? Why not simply say what one means and leave it?'
>
> So then he tried saying the grass is green and the sky is blue and so to propitiate the austere spirit of poetry whom still, though at a great distance, he could not help reverencing. 'The sky is blue,' he said, 'the grass is green.' Looking up, he saw that, on the contrary, the sky is like the veils which a thousand Madonnas have let fall from their hair; and the grass fleets and darkens like a flight of girls fleeing the embraces of hairy satyrs from enchanted woods. 'Upon my word,' he said (for he had fallen into the bad habit of speaking aloud), 'I don't see that one's more true than another. Both are utterly false.' And he *despaired* of being able to solve the problem of what poetry is and what truth is and fell into a deep dejection. 75; emphasis mine

Here, like the narrator from the previous passage, Orlando interrogates the truthfulness of figurative elements. The failure of the "dragon-fly," the "bedfellow," and "bride" to signify love is only the beginning of such an interrogation: for the crisis of signification extends to seemingly objective realities, like the passage of time. Furthermore, Orlando's rumination here illustrates the way that the outer narration blends into Orlando's interiority, a blending that eventually casts into doubt on the consistency of universal constants –"then he would try to think for half an hour,–or was it two years and a half?" Orlando's doubt about language is crystalized in the sentence which contains the key term, "despaired": "he *despaired* of being able to solve the problem of what poetry is and what truth is and fell into a deep dejection." It seems that, for Orlando, gender has something to do with the authority of language to convey truth in plain terms, of "say[ing] what one means and leav[ing] it." As Victoria L. Smith affirms, various scenes of the novel similarly "thematiz[s] within the text how representation or, rather

more particularly, how literary language finds itself at a loss" (Smith 68).

In a final example, I examine the co-occurance of words from both lists within a single passage. The words, "curiosity," which is associated with "man," and "seductiveness," which is associated with "woman," appear in a passage that portrays desire as driven by gender incomprehensibility. together portray gender a problem of expression that is intimately coordinated with language. The drama begins when Orlando, upon seeing Sasha for the first time, cannot tell whether she is a man or a woman:

> He beheld, coming from the pavilion of the Muscovite Embassy, a figure, which, whether boy's or woman's, for the loose tunic and trousers of the Russian fashion served to disguise the sex, filled him with the highest *curiosity*. The person, whatever the name or sex, was about middle height, very slenderly fashioned, and dressed entirely in oyster-coloured velvet, trimmed with some unfamiliar greenish-coloured fur. But these details were obscured by the extraordinary *seductiveness* which issued from the whole person. Images, metaphors of the most extreme and extravagant twined and twisted in his mind. He called her a melon, a pineapple, an olive tree, an emerald, and a fox in the snow all in the space of three seconds; he did not know whether he had heard her, tasted her, seen her, or all three together.... A melon, an emerald, a fox in the snow–so he raved, so he stared. When the boy, for alas, a boy it must be–no woman could skate with such speed and vigour–swept almost on tiptoe past him, Orlando was ready to tear his hair with vexation that the person was of his own sex, and thus all embraces were out of the question.

For Orlando, the problem of language and gender has to do with signification– he cannot resolve how to express either one–which arises when he first sees Sasha. Within this undefined space, he uses seemingly arbitrary metaphors, "a melon, a pinapple, an olive tree, an emerald, and a fox in the snow" to describe Sasha. At the same time that Orlando cannot place Sasha's gender, he also cannot find the right words to describe her.

As Sasha's probable gender oscillates between male and female throughout passage, Orlando's desire crescendos. The narrative voice and form of the sentences in this scene also shape the building tension: the narration alternates interiority and description a in free indirect discourse that jumps abruptly between narration and interjections, to express a cyclical quality about Orlando's confused mental state. The effect is to mirror with language the tortuous thought process that Orlando undergoes as he guesses then doubts the reality

of Sasha's gender. While the tension thus mounts throughout the passage, the relationship between gender and language comes to a climax:

> But the skater came closer. Legs, hands, carriage, were a boy's, but no boy ever had a mouth like that; no boy had those breasts; no boy had eyes which looked as if they had been fished from the bottom of the sea. Finally, coming to a stop and sweeping a curtsey with the utmost grace to the King, who was shuffling past on the arm of some Lord-in-waiting, the unknown skater came to a standstill. She was not a handsbreadth off. She was a woman. 27-28

Athough the tension finally ebbs as Orlando settles on Sasha's gender, settling on the phrase, "She was a woman," the use of figuration and form in this passage situate gender as something difficult, if not impossible, to grasp. The lessson seems to be that if gender is ambiguous, then language is also imprecise.

In filtering the shared contexts between "woman" and "man," coming closer to a sense of gender *distinctiveness* in this text, it is important to emphasize that gender still descends from a binary system–from the initial analysis of "woman" and "man." However, by *iterating* through distant and close reading, the terms swell with significations that pluralize the binary. Like Butler's account of gender subversion, this kind of computational analysis works toward resignifying the initial understanding of "woman" and "man." Despite the tight constraints of these computational work, there is a freedom in the possibility of working the results into closer-reading analysis. The rule here is iterativity which, as Butler suggests, opens up the opportunity for subversion:

> The compulsion to repeat an injury is not necessarily the compulsion to repeat the injury in the same way or to stay fully within the traumatic orbit of that injury. The force of repetition in language may be the paradoxical condition by which a certain agency—not linked to a fiction of the ego as master of circumstance—is derived from the impossibility of choice. 83

Butler explains that the repetition of language is the condition enables a certain agency to emerge. Through, repetition, dominant or established meaning can be resignified. Taking Butler's concept of "performative citation" as guidance, then, one may repeat the same computation over and over again, with each new result expanding and resignifying the initial understanding of binary gender.

1.5.2 TODO conclusion: on discursivity

The understanding of gender in this text is primarily discursive. Pamela Caughie zeroes in on the indeterminacy language, finding that it purposefully precludes a straightforward understanding of sex and gender; as a result, "sex cannot be separated from text, the grammatical from the gendered" (Caughie 51). According to Caughie:

> "Orlando works as a feminist text not because of what it says about sexual identity but because of what it manages not to say; not because of what it reveals about the relation between the sexes but because of what it does to that relation; not because its protagonist is androgynous but because its discourse is duplicitous" (Caughie 41).

This argument, that *Orlando*'s subversiveness is a discursive one, that it operates through language, has led to further critiques of its political significance,[8] none more incisive than the critique from Trans Studies. From the perspective of critics like Jay Prosser, the discursive understanding of gender is precisely what allows Orlando to transgress the norms of gender and sex in the novel. According Prosser, Woolf's experimentation with language and narrative form belies the physical the embodied reality of transsexuality. He explains: "Orlando is not about the sexed body at all but the cultural vicissitudes of gender. As h/er narrative propels h/er through four centuries of history, Orlando is free to move beyond h/er body–quite queerly, to break through the limits of the flesh" (Prosser 168). By "the sexed body," Prosser means the physical body, what Jack Halberstam describes as the "literal, the real, the intractable flesh" which is bound by the rules and boundaries of the physical and social world (Halberstam 314). That *Orlando*'s transgressiveness results from a play of *language* and *literary form* that elides the specificity and

[8] Jamie Hovey and Jessica Berman both explore how the text challenges the boundaries of national identity through an implicit critique of imperialism, a critique that emerges from the privileged position of the white, British persective. Hovey remarks that *Orlando* is "an ambivalent articulation of English nationalism," a nationalism that intersects with (and depends on) gender and race (Hovey 394). Displacing the oppressive effects of nationalism to racialized and sexually transgresive subjects, the novel "allows the protagonist to pass as respectible and heterosexual" (Hovey 398). Bringing the question of transsexuality to the fore, Berman argue that as a "trans text," *Orlando* utilizes methods of marking and categorizing bodies to interrogate the structures and boundaries of nationality (Berman 218). According to Berman, "The transnational situation as also intrinsically transgender" (Berman 218). Berman's account harps on "the disruptive, critical energy of the prefix 'trans'" to unpack the concept of "nation" and "nationality" (Berman 220).

the lived reality of the "sexed body" precludes the novel from being what Prosser would call a "transsexual" text. Rather, due to its "ambivalence, a wavering around transition", "a transformation of transition into new identity," its "easy androgyny," this text is transgender (Prosser 169). As Caughie asserts, *Orlando*'s transgressiveness comes from its discursive moves: "Far from defeating sexual difference, as many feminist critics claim, Orlando enacts it, enshrines it, exploits it, makes a spectacle of it, but as a playful oscillation not a stable opposition" (Caughie 48). According to Prosser, such strategies frames gender as a discursive (rather than embodied) phenomenon.

Interestingly, Prosser's critique is what aligns this text more closely with Butler's concept of gender performativity, particularly in the way that language is used to produce (and and reproduce) gender identity. In other words, a *discursive* understanding of gender is one that can be destabilized, distorted, and/or reformulated through language. Caughie attributes the emergence of gender transgression in this novel to experiments in figuration and narrative form:

> Sexual identity is assumed in language... Woolf brings out the arbitrariness of that identity, the arbitrariness of language itself, through Orlando's switching from one sex to the other, and from one poetic language to another, as well as through the shifting of her own rhetoric in this novel. Caughie 42

This text, with its "switching" and "shifting" discourse, which at once asserts that language is deficient and that it overshoots the mark, that it conveys plainness and poetry, implies that gender is also a shifting, formal phenomenon. Like Butler, Caughie presents a vew of gender as discursive.

2 Works Cited

Amin, Kadji, Amber Jamilla Musser, and Roy Pérez "Queer Form: Aesthetics, Race, and the Violences of the Social" ASAP/Journal, Volume 2, Number 2, May 2017, p. 235.

Barad, Karen. *Meeting the Universe Halfway.*

Benzel, Kathryn N. "Reading Readers in Virginia Woolf's 'Orlando: A Biography.'" Style, vol. 28, no. 2, 1994, pp. 169–82. JSTOR, `http://www.jstor.org/stable/42946241`.

Berman, Jessica. "Is the Trans in Transnational the Trans in Transgender?" Modernism/modernity, vol. 24 no. 2, 2017, pp. 217-244. Project MUSE, `doi:10.1353/mod.2017.0019`

Bode, Katherine. "Computational modeling: From data representation to performative materiality." *Animating Text Newcastle Univeristy (ATNU) Speaker Series*, no. 3: Thursday, 26th of November 2020. https://research.ncl.ac.uk/atnu/news/atnuiesvirtualspeakerseries202020213.html

Burns, Christy L. "Re-Dressing Feminist Identities: Tensions between Essential and Constructed Selves in Virginia Woolf's Orlando." Twentieth Century Literature, vol. 40, no. 3, 1994, pp. 342–364. JSTOR, www.jstor.org/stable/441560.

Boucher, Geoff, "The Politics of Performativity"

Butler, Judith, *Bodies That Matter*,

Butler, Judith, *Gender Trouble*,

Caughie, Emily Datskou and Rebecca Parker. "Storm Clouds on the Horizon: Feminist Ontologies and the Problem of Gender." Feminist Modernist Studies. 1:3, 230-242. 2018.

Channing, Jill. "Magical realism and gender variability in Orlando." Virginia Woolf Miscellany, no. 67, 2005, p. 11+.

"ContextIndex." NLTK Documentation. Accessed July 4, 2022. https://www.nltk.org/_modules/nltk/text.html#ContextIndex

"ContextIndex.similar$_{\text{words}}$." NLTK Documentation. Accessed July 4, 2022. https://www.nltk.org/_modules/nltk/text.html#ContextIndex.similar_words

Galloway, Alexander. **Protocol**, 2004.

de Gay, Jane. "Virginia Woolf's feminist historiography in Orlando." Critical Survey, vol. 19, no. 1, 2007, p. 62+.

Halberstam, (Jack) Judith. "Second Skins: The Body Narratives of Transsexuality. Jay Prosser Trans Liberation: Beyond Pink or Blue. Leslie Feinberg FTM: Female-to-Male Transsexuals in Society. Holly Devor." Signs: Journal of Women in Culture and Society, vol. 26, no. 1, Oct. 2000, pp. 313–17, https://doi.org/10.1086/495591.

Hovey, Jaime. "'Kissing a Negress in the Dark': Englishness as a Masquerade in Woolf's Orlando." *PMLA*, vol. 112, no. 3, 1997, pp. 393–404. JSTOR, www.jstor.org/stable/462948.

Mikolov, Tomas, et al. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781, arXiv, 6 Sept. 2013. arXiv.org, https://doi.org/10.48550/arXiv.1301.3781.

Mandell, Laura. "Gender and Cultural Analytics: Finding or Making Stereotypes?" Debates in Digital Humanities 2019. Edited by Matthew K. Gold and Lauren Klein. University of Minnesota Press, 2019.

Franco Moretti, "Conjectures on World Literature", *New Left Review* 1 (2000): 54-68, https:// tinyurl.com/moretti2000conjwl.Moretti, Franco. *Distant Reading*

Moretti, Franco. "The Soul and the Harpy." *Signs Taken For Wonders: On the Sociology of Literary Forms*, trad. David Forgacs, New York, Verso, 1983, pp. 1-41.

Prosser, Jay. *Second Skins: The Body Narratives of Transsexuality.* Columbia University Press, 1998.

Schmidt, Ben.

Sinykin, Dan. "Distant Reading and Literary Knowledge." *Post45.* May 6, 2019. https://post45.org/2019/05/distant-reading-and-literary-knowledge/

Smith, Victoria L. "'Ransacking the Language': Finding the Missing Goods in Virginia Woolf's Orlando.". *Journal of Modern Literature*, vol. 29 no. 4, 2006, pp. 57-75. Project MUSE, doi:10.1353/jml.2006.0050

So and Roland.

Underwood, Ted.

Woolf, Virginia. *Orlando: A Biography.*