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

## Introduction

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). When performing quantitative text analysis on this text, a fictional biography of a 16th-century English nobleman who undergoes a sex change, the standard tasks of "pre-processing" evacuate the ways that gender is complicated in this sentence. In order to perform quantitative analysis, a process which involves calculating and visualizing textual patterns, the text must first be transformed into a computable format.[[1]](#footnote-1) 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, and would affect the analysis of more substantial features like nouns, verbs, adverbs, and adjectives. Afterward, the following list of words, or “tokens,” remain 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's 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. It explores an experimental approach that deconstructs gender binaries by drawing connections between computer programming and gender theory. This analysis contrasts the principle *iteration*, central to both text analysis and gender theory, with current *reproducible* methods in quantitative analysis. It concludes by proposing a text analysis methodology that iterates through distant and close reading of the terms "woman" and "man" in *Orlando*, and considers the limitations of this method within the larger trajectory of Gender Studies.

## The Fantasy of the Falsifiable

Because computers can process hundreds of texts at a time, "reading" at much faster rates than humans, they attract critics like Franco Moretti, Matthew Jockers, and Ted Underwood who pose ambitious questions about literary history. Moretti, for example, explains that the process of quantification reduces the complexity of text to open the potential for its analysis: "fewer elements, hence a sharper sense of their overall interconnection" (*Graphs* 1). This approach toward computation represents what I call the "fantasy of the falsifiable." Leaving aside critiques of Moretti, particularly in light of the credible allegations of harrasment and assault by graduate students (see Lauren F. Klein, "Distant Reading After Moretti"),[[2]](#footnote-2) I want to focus on Moretti's early essay "The Soul and the Harpy" (1983), which lays out the reasoning behind developing a new methodological approach for literary criticism. Moretti in this essay betrays a deep suspicion about the analytical practices of the contemporary literary criticism that "multipl[ies], rather than reduc[es], the obstacles every social science encounters when it tries to give itself a testable foundation," and the literary critic that he likens to a kind of Narcissus, "whose only pleasure lay in contemplating his own reflection" ("Soul" 22, 14). Moretti proposes an alternative methodology, a "falsifiable criticism," that grounds the critical process in hypotheses that "test" literary interpretations ("Soul" 21). *Falsifiable* here means verifiable, pursuing answers that are "coherent, univocal, and complete" ("Soul" 21). The goal, according to Moretti, is to reach irrefutable conclusions: "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" ("Soul" 22).

As Moretti's falsifiable criticism eventually grows into "distant reading," he streamlines his critical method into a recursive process of posing hypotheses, assembling data, making inferences, and occasionally, reframing the original hypotheses. For example, when studying the effect of market forces on book titles from 1740-1850, Moretti finds an interesting relationship between the size of the market and the length of titles: "As the market expands, titles contract; as they do that, they learn to compress meaning" (204). When explaining his process, however, his word choice diminishes the interpretive moves he makes throughout his critical analysis:

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… (181-2; emphasis mine)

The word choices here repeatedly diminish the interpretive moves Moretti makes at each stage of his analysis. As Stephen Ramsay points out, Moretti tends to present his insights as an objective description of reality, which is portrayed through his graphs and other visualizations. It is as if "data is presented to us… not as something that is also in need of interpretation" (Ramsay 5).

Ironically, the faith in falsifiable criticism gets its strongest expression in a famous detraction by Nan Z. Da, who 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). Da emphasizes her point with an experiment in "Topic Modeling," a machine learning method that generates a number of "topics," or keywords, from large collections of text. Da attempts to verify the results of a topic modelling experiment by replicating the process on her own computer, but she fails to produce the same results and concludes that the method is ineffective. However, as Ben Schmidt explains, the difference in her results is due to the use of different parameters and software that Da uses to run her experiments. Additionally, he points out that, "Far *more* than anyone I’ve seen in any humanities article, she asserts that scientists do something arcane, powerful, and true.[[3]](#footnote-3) Despite their vastly different views on the role of quantitative methods for studying literature, Da and Moretti appear to agree that these methods ought to provide results that are, at the very least, reproducible.

Interested in the interpretive possibilities of quantification, Ted Underwood explores what he calls "perspectival models" of literary data. 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). First, Underwood trains an algorithm with sample data, such as descriptions of characters tagged with gender identities. Then, he feeds this data into the algorithm, which "learns" what words are typically associated with women and which words are typically associated with men. At the end of this training process, the resulting model can then predict the gender of characters in new text. Underwood calls this approach "perspectival modelling," because it represents a single, rather than objective or universal, perspective of the data. Dan Sinykin explains that this method enables Underwood "to leverag[e] the human prejudices built into modeling toward humanistic ends" (par. 4). In one project, Underwood uses a logistic regression algorithm to calculate and visualize the terms typically associated with each gender in books by men and books by women (See fig. 1). Here, each axis represents a different perspective on gender. The verticle axis visualizes words by women, and the horizontal by men, with positive numbers signifying overrepresentation of certain words. The terms on the left side of the graph describe men, with the top-left corner and bottom-left corner denoting books by male and female authors, respectively. The terms on the right side of the graph describe women, with the top-right corner and bottom-right corner denoting books by female and male authors, respectively. Underwood explains that, "I needed a simple picture, frankly, in order to explain how a quantitative model can be said to represent a perspective" ("Machine Learning and Human Perspective" 98).

A graph showing different words

Description automatically generated

Fig. 1: Underwoods logistic regression model.

Underwood's methodology, however, reinscribes the same binary that it attempts to historicize. Here, he uses uses logistic regression analysis, an entry-level machine learning algorithm that makes predictions on a scale from 0 to 1. The problem is that this probablistic computation reifies gender as either/or, in other words, as a binary opposition, which Underwood admits himself when he says that "gender theorists will be frustrated by the binary structure of the diagram" ("Machine Learning" 98). Here, the concept of femininity is deliberately consolidated and computed against that of masculinity, which is precisely the purpose of a binary classification algorithm–to collapse all possible answers between a scale of yes/no. Feeding gender terms into such an algorithm effectively imposes a binary structure onto those terms.

In my view, such an approach undershoots the potential of quantitative methods for studying gender, not to mention the goal of attaining some kind of overarching knowledge about literary history, whether that be a "distant horizon" or "the great unread." Distant reading methods might instead harness what Ramsay describes as "the objectivity of the machine," to destabilize the binary (x).[[4]](#footnote-4) Resisting the temptations of falsifiable criticism, work by critics like Susan Brown, Laura Mandell, Richard Jean So, and Edwin Roland apply distant reading methods toward deconstructing social categories like gender and race. Brown and Mandell 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).[[5]](#footnote-5) 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 becomes a tool for exploring how language constructs (and can deconstruct) social categories.

In a critique of distant reading methods, Mandell illustrates how the study of gender often reifies gender stereotypes, "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 "constructed both by the measurer and the measured" (par. 38). Computation offers, in Mandell's words, "parallax, multiple perspectives for viewing a very complex reality” (par. 38). To deconstruct gender, Mandell turns to genre, another category which will allow scholars to see the reductive constitution of categories generally. Here, Mandell uses the popular stylometry measurement, "Burrow's Delta," which visualizes the "distance" between writing styles by creating branches (or "deltas") between different texts. She finds that the stylistic qualities of a female writer, Mary Wollenstonecraft, shares with those of comparable 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). Drawing gender into conversation with genre, Mandell creates 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).

Just as quantification can be harnessed to deonstruct the M/F binary, so it can deconstruct what Edwin Roland and Richard Jean So describe as "the machine's initial binary understanding of race" (68). Roland and So deconstruct racial categories by experimenting with an algorithm that evaluates an author's race 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 result, they infer that white authorship, as a category, only coheres against the variance of black authorship. Whiteness, they explain, *depends* on the characterization of blackness.

This quantitative exercise points Roland and So 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. This misclassification is attributable to a single word, "appalled," which the computer (mistakenly) reads as proof of white authorsip. Going back to the text, Roland and So discover that this term occurs only once, in an early scene where the narrator David 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 Roland, So 71). Noting the connotations of whiteness in "appalled," which has the middle French root, "apalir," meaning "to grow pale," Roland and So posit that this term suggests a relation between gender and 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, as critics have observed in the scholarship on this novel, whereby racial markers are displaced in favor of an implicit whiteness. Taking the computer's mistake as a starting point, Roland and So's analysis thus contributes to the ongoing debate about the complex relationship between gender and race in the novel.

Because race is a social construct, and machines only impute meaning that is encoded into them, Roland and So reason that machines are ideal instruments for studying the construction of race (60). In direct opposition to the "falsifiable" position, computational error here becomes a starting point for analysis. Furthermore, the machine error surfaces a yet unexplored aspect related to race, that of sexuality:

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

Like Mandell, So and Roland use computational methods to destabilize, rather than reify, binary categories like male/female, or white/black. In this case, a single computational error opens a site for speculation about how whiteness suggests a troubled understanding of sexuality, where queerness might articulate with race.

## Iteration

Mandell asserts that both gender and genre "are… highly imitatable," so that "anyone can adopt gendered modes of behavior, just as anyone can write in genres stereotypically labeled M/F" (par. 30). While this interpretation echoes a common misunderstanding of Butler's theory (explained further below), gender performativity remains a useful heuristic for quantitative text analysis. First, the common misreading of Butler's theory is that gender performativity denotes an act or series of acts that can be imitated at will.[[6]](#footnote-6) Rather, as Butler emphasizes in her follow up book, *Bodies that Matter: on the Discursive Limits of Sex* (1996), performativity is a compulsory process that precedes and constitutes subjectivity–it is a mechanism through which the subject can emerge: "a process of reiteration by which both 'subjects' and 'acts' come to appear at all" (*Bodies* xviii). Butler here makes the argument for gender as purely discursive, where what is experienced as the physical body, from sex to sexuality, only materializes through the repetition of gender norms in which each act signals a prior, authorizing norm.[[7]](#footnote-7) This thread of discursivity, and its implications within a larger trajectory of Queer Studies, is picked up again in this chapter's conclusion.

To better understand performativity as a discursive phenomenon, it is helpful to situate Butler's work within the context of second-wave feminism and its deconstruction of gender binaries. Here, Butler draws from the work of feminist theorist Luce Irigaray, who asserts that influential Western thinkers like Plato, Aristotle, and Freud have defined feminity "on the basis of masculine parameters" (Irigaray, *The Sex Which Is Not One* 23).[[8]](#footnote-8) Irigaray argues that the association of "woman" with "matter" (associated to binaries such as "rationality/emotion" and "mind/body"), and its subordination to to male "form" erases the possibility of representing woman at all. Rather, this binary "produces the feminine as that which must be excluded for that economy to operate" (Butler, *Bodies* 10). This "domesticated" feminine term contrasts to the excessive feminine, the "necessary outside" of the domesticated feminine (which is also its enabling condition), that creates a "field of disruptive possibilities" (Butler, *Bodies* 13). However, this "unspeakable" element cannot be invoked directly without subscribing itself to the ruling structure (Butler, *Bodies* 12).[[9]](#footnote-9) Butler asks, "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?" (*Bodies* 11). This question–how to express what is not there, what is refused by the system of the visible–leads Butler to her theory of gender subversion.

For Butler, theorizing subversion begins by positing the origin of linguistic signification. She wonders, "Can language simply refer to materiality, or is language also the very condition under which materiality may be said to appear?" (*Bodies* 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: "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" (6). This reasoning leads Bulter to a major realization: "the mimetic or representational status of language…. is not mimetic at all. On the contrary, it is productive, constitutive, one might even argue performative" (*Bodies* 6). If language produces the reality that it seems to merely reference, it means that subjects are always interpellated, and in fact brought into subjectivity, by a discourse prior to their their participation in it.

Within this regulatory structure, this significatory circle, lies the possibility of resistance, the possibility of *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 a "repudiated" meaning. Butler offers the famous example in the resignification of the term "queer," which has been transformed from a term of abjection to one of empowerment. "Queer" is resignified when it harnesses its own repudiation, which is an implied but "disavowed abjection [that] will threaten to expose the self-grounding presumptions of the sexed subject" (*Bodies* 3). Each time that "queer" is used, it draws from this domain of abjection which is repudiated by heterosexuality. Butler proposes that one "consider this threat and disruption… as a critical resource in the struggle to articulate the very terms of symbolic legitimacy and intelligibility" (*Bodies* 3). By citing the repudiated meaning, the term "queer" "resignifyi[es] the abjection of homosexuality into defiance and legitimacy" (*Bodies* xxviii).

Here, repetition is key, enabling the introduction of what is external to the binary into the system. Irigaray achieves this resistance by "mim[ing] philosophy… and, in the mime, tak[ing] on a language that effectively cannot belong to her" (Butler, *Bodies* 12). Irigaray undermines authority through repetition, by "cit[ing] 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" (Butler, *Bodies* 18). Irigaray introduces something external to the system, displacing the logic of phallogocentrism, while remaining within its terminology. Butler imagines Irigaray's thought process here:

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" (*Bodies* 18).

Through repetition, deception emerges from resemblance, and insubordination through subservience. The key is iteration, a continual miming of the authorizing norm.

In what follows, I examine how iteration also emerges in the Python programming language as it is used for text analysis tasks. Here, I will deconstruct the cleaning and analysis processes to surface the ways that Python's syntax evokes this quality of iteration.

The Python programming language offers a number of custom librarries, suchas the Natural Language ToolKit (NLTK), for common text analysis tasks that clean, count, and visualize textual patterns. Python handles text data in the form of words, or strings, contained within groupings called lists. Then, Python goes through each item in the list to perform a task. One way of going through lists is a consruct known as the for loop, which repeats a single action to each item, each string, within the list. At a very basic level, much of text analysis consists of looping, or iterating, over bits of text and performing actions to standardize the text for analysis. Such actions include tokenizing, cleaning, and regularizing, which help to eliminate pieces of text that will skew the 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, pronouns, and auxiliary verbs, which are deemed to be semantically minor compared to nouns, verbs, and adjectives, and adverbs. The task of preprosessing text forces words into existing boxes, so to speak, stripping text of the significance portrayed by capitalized words, rhythms of language in stop words, inflections in word endings, and so on. This reduction of text is a necessary trade-off in order to reduce idiosyncracies and to make meaningful computations on the text, making them amenable to analysis.

The first step in cleaning usually involves removing things like punctuation and capital letters. The following code block loops through a text, saved here as full-text, to filter out punctuation and transform any capital letters into lowercase forms:

normalized = []

for word in full-text:

if word.isalpha():

normalized.append(word.lower())

Here, the loop begins by creating an empty list, normalized, where words will be dropped after passing the filter. 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 a Boolean (a True or False) condition in which only alphabetic characters will pass through (that is, characters which do not contain numbers or punctuation). If the word fulfills this 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 into lowercase format. The final list, therefore, will only contain alphabetic and lowercased letters.

The next step is removing stop words. Here, we use another loop, compressed into one line of code, in a syntax that is called 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 in stops, a list of stop words. If the word is *not* a stop word, then it will be added to a new list, no\_stops. Running the first sentence of *Orlando* through the loop will return the following list of words:

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

The next (and final) step of cleaning involves stripping word inflections to get the root. Here, there are two possibile choices, which differ in how much computational processing each requires. The first one, called "stemming", simply cuts the endings from the word. For example, "rafters" will be stripped to "rafter." What this method gains in speed, however, it loses in precision, and can sometimes cut letters that are intrinsic to the word. The other possbility, called "lemmatizing," involves looking up each word, one by one, in a dictionary to find its appropriate root. Below is the code for lemmatizing the text:

clean = [WordNetLemmatizer.lemmatize(word, word) for word in no-stops]

At this point, the text is ready for analysis. A good first step is to begin with a target word, in this case, "woman." One useful function, called concordance() returns the context, that is, the immediate words surrounding, the target word. Below, the word "woman" is run through a concordance of the words in *Orlando*:

alities which the old woman loved the more the mo

scarlet . For the old woman loved him . And the Q

les . The old bumboat woman , who was carrying he

h , whether boy 's or woman 's , for the loose tu

boy it must be – no woman could skate with such

eadth off . She was a woman . Orlando stared ; tr

, until now ? An old woman , he answered , all s

and some old country woman hacking at the ice in

and pity the poor old woman who had no such natur

man 's beard and that woman 's skin ; of a rat th

the sight of the old woman hobbling over the ice

ght coming or the old woman or whatever it was ,

tainly not those of a woman bred in a cattle-shed

e world for a Cossack woman and a waste of snow -

erating . There was a woman in white laid upon a

Based off the contexts surrounding the target word, Python can make further analyses. Another method, called similar() returns a list of words used in similar contexts to the target word. To compute the results of similar(), NLTK first takes the context of the target word from concordance(), then it searches the text for other terms which contain the same surrounding words. 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 in similar contexts to the chosen word, this method might reveal words that have semantic resemblance to the target word. It is important to point out, however, that the computer does not impute meaning to the words. Rather, it only counts each word as a string, that is, as a piece of data composed of alphanumeric sequences. It takes "woman," then notes all of the words in proximity to "woman," and then searches the rest of the text for other words that have similar proximities.

Basic NLP tasks offered by libraries like NLTK contrast with "deep learning" methods that work in more sophisticated ways to count and analyze language. Many of these methods use "word embeddings" to ascribe machine-interpretable meaning to strings. Like similar() and concordance(), word embeddings build off patterns of word similarity based on context. Unlike the NLTK methods, however, word embeddings encode a value (actually, a list of values) to a given word based on its context. The value of any given word is a numerical representation known officially as a "word vector." A vector for a single word, "woman," for example, will contain a list of numbers that represent a similarity score between "woman" and another word. As numerical representations, these values enable further quantitative exercises that can analyze the relationship between "woman" and other words. The classic example for introducing the potential of word vector math is the formula, "King - Man + Woman = Queen" (Mikolev et al. 2). Here, gender (between "Man" and "Woman") is isolated as a computable component which enables one to derive the difference between "King" and "Queen".[[10]](#footnote-10)

For example, the vector which represents "woman" contains a list of numbers that score the similarity "woman" to other words. Here, the word "woman" is most closely associated to the word "child," with a similarity score, or "weight," of .93, or 93%, then with "mother," with .92, then "father," with .90.[[11]](#footnote-11) Below is a word vector of words calculated to be most similar to "woman":

[('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)]

Commonly, word embeddings are organized into a matrix, or tabular, format. Using this matrix format, further mathematical operations are possible using statistics, linear algebra, and calculus, which are the building blocks of deep learning methods. In deep learning, the labels accompanying the numerical representations do not matter, only the list of numbers themeslves, which together represent the word vector. The word "woman," therefore, would be represented with the following vector: .937. .921, .917, .915, .914, .909, and so on. This representation demonstrates that, even when removing labels, words are still assigned meaning by their relation, their proximity, to other words. In the following section, I use these word vectors as a starting point to explore terms related to each gender in *Orlando*, starting with the terms "woman" and "man."

## Queer Distant Reading

I now turn to Virginia Woolf's novel, *Orlando: A Biography*. This novel is ideal for a computational study of gender for two reasons. First, published in 1928, it is perhaps the earliest salient example of transgender narrative. Second, as many critics have noted, its characterisitic modernist experimentation with limits of language works toward destabilizing normative concepts of identity and gender. Jane de Gay, Jill Channing, and Christy L. Burns, for example, assert that Woolf deploys imaginative elements, magical realism, and parody, respectively, to resist realist narrative expectations. De Gay describes Woolf's writing as "feminist historiography" that "reject[s] Victorian patriarchal metanarratives" and instead "use[s] 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 coherence between gender, identity, and even race and nationality. For example, 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). Pamela Caughie analysis agrees, arguing that *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).

In what follows, I pursue a text analysis method that I call "iterative" for the way it moves between close and distant reading, a process similar to what Andrew Piper calls "bifocal" reading. This process, in Piper's words, "no longer us[es] our own judgments as benchmarks… but explicitly construct[s] the context through which something is seen as significant (and the means through which significance is assessed)" (17). My method feeds the output of computations into close reading analysis, using the computer to identify words that I then examine in context and with detail.

First, I begin with a list of terms computed similar to woman and man in the text. Unlike the word embedding of "woman," from my previous section, the embeddings here are trained on Woolf's novel, and therefore reflect an understanding of gender markers based on this specific text. Additionally, to get distinctive results for each gender, I modified the code to remove any words with strong associations to the opposite gender. For example, I compute words that are most positively associated with "woman" and most negatively associated with "man." Though this analysis, like Underwood's, begins with a binary formulation of gender, this binary quickly destabilizes as I move deeper into the close reading analysis, more like Richard Jean So and Edwin Roland's study of race. By *iterating* through distant and close reading, the terms swell with significations that pluralize the binary and work toward resignifying the initial results for "woman" and "man." Below are the word embeddings for the term "woman" and "man," respectively.

distinct\_w = model.wv.most-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)]

distinct\_m = model.wv.most-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)]

At first glance, the terms for each list appear to align with existing conceptions of femininity and masculinity, such as "soft" and "frilled" for "woman," and "fact" and "defend" for "man." But as I examine these words in context, the gender binary becomes less clear cut. From the "woman" list, I examine the term "delicious," which only appears after Orlando has transitioned into a woman in the story. As I explain below, this term reveals a relationship to queerness characterized by what I describe as a distinctly feminine mode of resistance. Then, from the "man" list, I examine a term that only appears once in the novel, at a significant moment which initiates what I call the novel's "crisis of signification" with language. Following this thread, I examine further passages that develop this theme as it spreads into Orlando's interior narration. I conclude with a close reading of a rather dramatic passage that contains words from both the "woman" and "man" lists, and constellates this crisis with themes of gender ambiguity and homosexual desire.

I begin with the word "delicious," which occurs only after Orlando transitions into a woman. Three of this term's five occurances appear in a single passage, when Orlando is sailing from Turkey back to her native England. The ship captain offers Orlando a bit of beef, which sends her into a rapturous speculation about the joys of womanhood:

'A little of the fat, Ma'm?' he asked. 'Let me cut you just the tiniest little slice the size of your fingernail.' At those words a *delicious* tremor ran through her frame. Birds sang; the torrents rushed. It recalled the feeling of indescribable pleasure with which she had first seen Sasha, hundreds of years ago. Then she had pursued, now she fled. Which is the greater ecstasy? The man's or the woman's? And are they not perhaps the same? No, she thought, this is the most *delicious* (thanking the Captain but refusing), to refuse, and see him frown. Well, she would, if he wished it, have the very thinnest, smallest shiver in the world. This was the most *delicious* of all, to yield and see him smile. 'For nothing,' she thought, regaining her couch on deck, and continuing the argument, 'is more heavenly than to resist and to yield; to yield and to resist. 114

Here, "delicious" describes a refusal, then a yielding–the vacillations of what appears to be a passive form of pleasure, ostensibly opposed to the active pleasure of pursuit which Orlando enjoyed as a man. Below, I will return to this active kind of pleasure, and the scene of its emergence when Orlando meets Sasha. Here, although the word "delicious" describes a distinctly feminine experience of pleasure, about withholding and, eventually, submitting to the active force, it is a pleasure rooted in what is not quite passivity and not quite power.

To get a deeper understanding of this term, I run another similarity search with "delicious" as the target word. The top result, the word most related to "delicious" in the text, is "culpable." I then turn back to the text to examine when this word appears, which happens twice, both times in the same scene on the ship, during Orlando's ruminations about the pleasures between the sexes. Here is one of the few times in the novel that Orlando considers her homosexuality explicitly:

And as all Orlando's loves had been women, now, through the culpable laggardry of the human frame to adapt itself to convention, though she herself was a woman, it was still a woman she loved; and if the consciousness of being of the same sex had any effect at all, it was to quicken and deepen those feelings which she had had as a man. 119

"Culpable," from the Latin "culpa," meaning fault, denotes a body deserving of blame for its refusal to conform. Here, the word modifies "laggardry," which describes the obstinacy or persitancy of Orlando's love for women, despite that she is now a woman herself. It seems that, for Orlando, desire is defined by a sense of guilt about refusing to conform. The next usage of this term occurs soon after, when Orlando reprises her earlier thoughts around the time that "delicious" appeared:

'To refuse and to yield,' she murmured, 'how delightful; to pursue and conquer, how august; to perceive and to reason, how sublime.' Not one of these words so coupled together seemed to her wrong; nevertheless, as the chalky cliffs loomed nearer, she felt culpable; dishonoured; unchaste, which, for one who had never given the matter a thought, was strange. 120

Rehearsing the conventional roles of the sexes, roles which Orlando fails to fit into, she feels (in addition to "culpable) "dishonour[ed]" and "unchaste"–words associated with an ideal of virtuous femininity. This feminizing language deepens the relationship between desire, guilt, and deliciousness, the last of which was previously characterized as a distinctly feminine pleaure in passivity that is not quite powerless. Here, Orlando's refusal to conform, for which she feels culpable, constitutes another kind of passivity, a form of refusal. The connection between "delicious" and "culpable," seem to define queer desire as a distinctly feminine mode of resistance.

Now, I move to Orlando's experience as a man, returning to the original list of similar terms. I begin with "chequered," which appears only once, at the very beginning of the story, when Orlando makes his entrance, stepping into "the yellow pools chequered by the floor" (12). This moment is the first of many in which the narrator calls into question his credibility as a biographer, a self-described "scribe," who distinguishes his role from that of the poet that embellishes and exagerrates through figurative language. However, the narrator's committment to straightforward description soon gallops into full-fledged figuration when he tries to describe Orlando's beauty:

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 so compelled to produce a few lines above this one. This slip into figurative language eventually grows into a crisis of signification that recurrs persistently throught the novel.

The crisis even spreads to Orlando's internal thoughts, where it first emerges during a period of depression following his love affair with Sasha, a Russian princess. Here, I take the term, "despaired" which, like "chequered," occurs only once in the novel:

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

Orlando, deep in depression, struggles to understand the role of figuration in language. Much like the narrator in the previous passage, Orlando also questions the truthfulness of figurative elements. First, he attempts plain language, "the sky is blue", "the grass is green," but these prove insufficient for describing a sky that moves "like the veils which a thousand Madonnas have let fall from their hair" and grass that "fleets and darkens like a flight of girls fleeing the embraces of hairy satyrs from enchanted woods." Orlando, who has just been abandoned by a woman, sees in nature flight and modesty, qualities which he finds "false."

In a final example, I examine the co-occurance of words from both lists within a single passage, the scene where Orlando meets Sasha. The words, "curiosity," which is associated with "man," and "seductiveness," which is associated with "woman," appear in a moment of tumultuous desire about Sasha's gender incomprehensibility, where gender becomes intimately coordinated to language's ability to signify. 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. (For though we must pause not a moment in the narrative we may here hastily note that all his images at this time were simple in the extreme to match his senses and were mostly taken from things he had liked the taste of as a boy. But if his senses were simple they were at the same time extremely strong. To pause therefore and seek the reasons of things is out of the question)… 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. 27-28

The passage constellates the crisis of signification within the larger issue gender ambiguity. Orlando uses seemnigly arbirary metaphors, "A melon, an emerald, a fox in the snow," indicating that at the same time which he cannot place Sasha's gender, he also cannot find the right words to describe her. As Sasha's gender oscillates between male and female throughout the passage, so the narrative voice alternates between Orlando's interiority and the narrator's commentary. Taking this crisis of signifcation to the level of deitic narrative, the narrator's "pause" indexes the constructed nature of the scene. This constructed quality is reinforced by the narrator's attempt to explain Orlando's choice of words to describe Sasha, which have no "reason," and "were mostly taken from things he had liked the taste of as a boy." Language, like gender, thus becomes a tool for representation and performance. As Pamela Caughie, asserts, the gender transgression is intimately connection to the narrator's experiments in figuration and form: "Woolf brings out the arbitrariness of [sexual] 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 (42). This passage, with its "switching" and "shifting" discourse, which asserts that word choices are arbitrary and flowing, implies that gender is also a fluid phenomenon.

## Conclusion

The argument that *Orlando*'s subversiveness is a discursive one, opens the text to numerous critiques[[12]](#footnote-12), particularly from Trans Studies. According Jay Prosser, Woolf's experimentation with language and narrative form belies the physical the embodied reality of transsexuality. He argues that "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 which is bound by the physical and social world. *Orlando's*'s transgressiveness results from a play of language and literary form that elides the specificity and the lived reality of the "sexed body." 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).

A decade later, Omise'eke Natasha Tinsley writes about the problem of gender fluidity as a metaphor. In her essay, "Black Atlantic, Queer Atlantic: Queer Imaginings of the Middle Passage," Tinsley argues for the imbrication of sexuality and race through the lense of the Black Atlantic. By sexuality, Tinsley does not necessarily mean "same-sex" desire, but relationships from the Middle Passage, that "mak[e] disruption to the violence of the normative order… connecting in ways that commodified flesh was never supposed to" (199). Reading for relation rather than desire, her critique re-works the trope of fluidity which, drawing from the ocean, "is not an easy metaphor or queer and racially hybrid identities but for concrete, painful, *and* liberatory experience" (192-193). For Tinsley, fluidity is an opportunity for "a return to the materiality of water to make its metaphors mean more complexly, shaking off settling into frozen figures" (212). Reading from Dionne Brand's book, *Map to the Door of No Return* (2001), on the Middle Passage, Tinsley theorizes fluidity as a "social liquidation," being stripped by the water, particulars of identity washed away in the current. She explains that "brown bodies are gender fluid not because they choose parodic proliferations but because they have been 'washed of all this lading, bag and baggage'" (209).

Tinsley's critique surfaces the ways that gender fluidity, as a trope for queerness, obscures the very physical connotation of corrosion. While this chapter, following Butler, has proposed iterativity, which is a form of fluidity, as a kind of enabling constraint that creates a possibility for resistance, I wonder how Tinsley's evocation of materiality might deepen this analysis, opening the potential of Queer Studies-inflected frameworks for text analysis. One might think more deeply about the concept of iteration and its connection to the productive power of language, the ways that language physically executes action, for example, in computer code. Python differs from other languages (like markup languages HTML and XML, which I explore in the next chapter) in that it is an executable language. The code not only defines instructions, but also enacts them. How might this active quality of Python programming influence the way that we study gender as an active, embodied phenomenon?

1. Text analysis borrows from Natural Language Processing and Machine Learning methods to do analyses like Topic Modeling, Sentiment Analysis, and Logistic Regression (discussed below). Though these methods differ in important ways, they share in basic tasks of counting and classifying words and other textual elements with the goal of predicting and visualizing patterns in text. [↑](#footnote-ref-1)
2. In the wake of the #metoo movement in 2017, three graduate students accused Moretti of harassment and assault. Stanford university claimed to be reviewing the case with no formal proceedings or other action being taken. See Liu, Fangzhou and Hannah Knowles. [↑](#footnote-ref-2)
3. 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," [↑](#footnote-ref-3)
4. In his book *Reading Machines*, Ramsay draws from the deformative critical methods of Jerome McGann and Lisa Samuels to harness the enabling constraints of computation that "unleash the potentialities" of the text, offering opportunities for new readings (33). [↑](#footnote-ref-4)
5. 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. [↑](#footnote-ref-5)
6. In *Gender Trouble: Feminism and the Subversion of Identity* (1990), Judith Butler famously disrupts two 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. 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. [↑](#footnote-ref-6)
7. Common critiques of Butler point out the limits of this theory for posing gender and sexuality as discursive. From the field of Trans Studies, Jay Prosser 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 precarious existence, as Venus's death illustrates (55). 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. [↑](#footnote-ref-7)
8. Irigaray here critiques Jacques Derrida famously defines as "phallogocentrism," the idea that man, symbolized by the phallus, is the center and focus of knowledge. [↑](#footnote-ref-8)
9. Irigaray's concept of the "necessary outside" seems to anticipate another popular critique of Butler's theory, from the field of Political Philsophy, which claims a logical inconsistency in Butler's theorization of subjectivity. If the resistance to signification comes from outside the cycle of signification, does this 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). [↑](#footnote-ref-9)
10. It almost goes without saying–the formula reproduces gender as a binary structure, where "Queen" is computed through its relation to "King" and "Man." [↑](#footnote-ref-10)
11. The language model for this computation comes from Word2Vec's "glove-twitter-25" dataset, which is based on Twitter data. [↑](#footnote-ref-11)
12. 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). [↑](#footnote-ref-12)