

CHAPTER ONE:

“Text Analysis”

Doubt

The novel *Orlando: A Biography* (1928), by Virginia Woolf, famously opens with an assertive gender designation, which is 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).

However, in order to perform quantitative *text analysis* on this text, the standard tasks of “pre-processing” would evacuate this sentence of its gender qualification. Text analysis, a process which involves calculating and visualizing textual patterns, requires transforming the source text into a computable format.¹ Pre-processing the text (also called “cleaning” or “normalizing” the text) would strip 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. None of these features, however, are minor in a text about a fictional biography of a 16th-century English nobleman who undergoes a sex change.

Cleaning the first sentence not only would strip it of its pronouns, including the gender assertion in the first word, “He.” It would also cut the em dash immediately following this “He,”

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.

signaling the entrance of the narrator and his conspicuous certitude: “—for there could be no doubt of his sex....” Only the following list of words would remain:

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

This chapter examines how quantitative text analysis works to analyze gender by using Woolf’s *Orlando* as a test case. It explores an experimental approach to text analysis that deconstructs gender binaries by drawing connections between computer programming and gender theory. This analysis emphasizes the principle of *iteration*, central to both text analysis and gender theory, with current “reproducible” methods in quantitative analysis. It concludes by proposing a text analysis procedure that iterates through “distant” and close reading gender terms in *Orlando*, and considers the limitations of this method within the larger trajectory of Gender Studies.

Falsifiable

This experiment pushes against an approach toward text analysis that I call the *fantasy of the falsifiable*. Because computers can process hundreds of texts at a time, “reading” at much faster rates than humans, they tend to attract critics who pose ambitious questions about literary history. Franco Moretti, for example, famously explains that quantification reduces textual complexity to open up its potential for its analysis: “fewer elements, hence a sharper sense of their overall interconnection” (*Graphs* 1). Leaving aside critiques of Moretti, particularly in light

of credible allegations of harassment and assault by his graduate students,² I turn to one of his earliest essays, “The Soul and the Harpy” (1983), which lays a new methodological approach for literary criticism. Moretti in this essay betrays a deep suspicion about the analytical practices in contemporary literary criticism that “multiply, rather than reduce, the obstacles every social science encounters when it tries to give itself a testable foundation,” and the literary critic, “whose only pleasure lay[s] in contemplating his own reflection,” like Narcissus (“Soul” 22, 14). He proposes an alternative methodology, a “falsifiable criticism,” that grounds the critical process in hypotheses to “test” literary interpretations, pursuing answers that are “coherent, univocal, and complete” (“Soul” 21). The goal, according to Moretti, is for literary criticism 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). Moretti’s falsifiable criticism eventually develops into “distant reading,” a critical method that involves posing hypotheses, assembling and analyzing data, making inferences, and occasionally, reframing the original hypotheses.

When describing his process in a project that studies title lengths and the book market:

[F]irst, 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)

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, and Lauren F. Klein, “Distant Reading After Moretti.”

He repeatedly understates his interpretive moves at each stage of his analysis: “try to explain,” “suggest,” “make a little attempt.” As Stephen Ramsay points out, Moretti tends to present his insights as a description of reality, as if “data is presented to us... not as something that is also in need of interpretation” (Ramsay 5).

Ironically, falsifiable criticism gets its strongest expression in a famous detraction, by Nan Z. Da, that argues that quantitative methods reveal a “fundamental mismatch between 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 failing to produce the same results, she concludes that the method is ineffective. However, as Ben Schmidt explains, “Far *more* than anyone I’ve seen in any humanities article, she asserts that scientists do something arcane, powerful, and true.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.

Unlike Moretti and Da, however, some text analysis practitioners push against the apparent objectivity of big data methods. Ted Underwood, for example, creates what he calls “perspectival models” of literary data, for example, on different “perspectives” of gender in literary data. Underwood explains that “Machine learning algorithms are actually bad at being

³ Schmidt rightly points out that Da uses different parameters and software to run her Topic Models, which explains the discrepancy in results. 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/

objective and rather good at absorbing human perspectives implicit in the evidence used to train them” (“Machine Learning and Human Perspective” 92). In one study, he uses a logistic regression algorithm to calculate and visualize the terms typically associated with gender in books by men and books by women (See Figure 3). In the chart below, each axis represents a different perspective on gender: the top half of the graph represents words written by women, and the bottom half by men, with positive numbers signifying overrepresentation of words to describe women, and negative numbers signifying overrepresentation of words to describe men. 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).

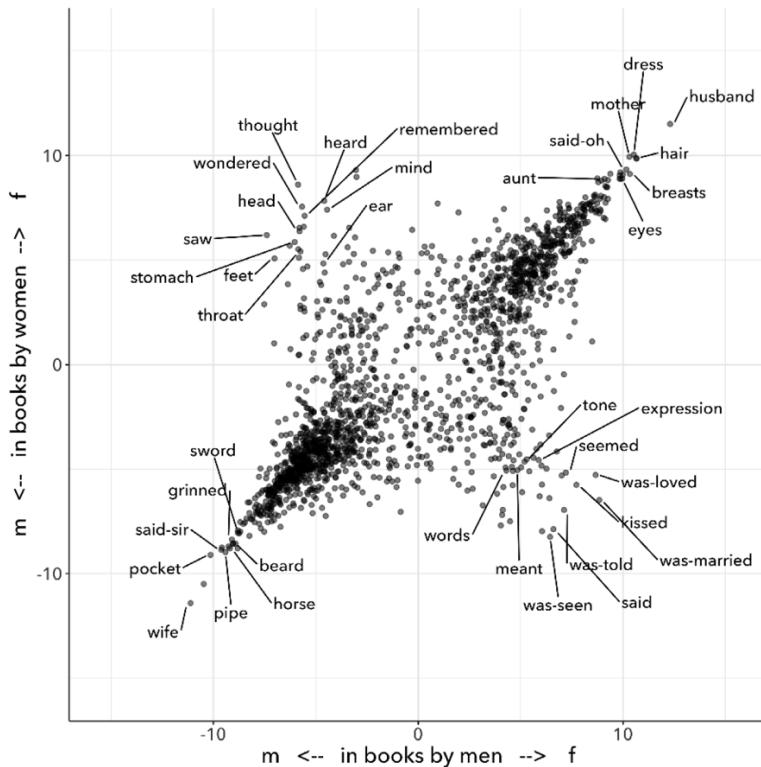


Figure 3: Underwoods logistic regression model. The axes indicate books written by men (x-axis) and by women (y-axis). On each axis, positive numbers indicate that a word is overrepresented in descriptions of women and negative numbers indicate that it is overrepresented in descriptions of men.

This methodology, however, reinscribes the gender binary as stereotypically oppositional. Here, Underwood uses logistic regression analysis, a binary classification algorithm that makes predictions on a scale from 0 to 1—in other words, into one of two options that are structurally in opposition. Underwood himself admits this when he says that “gender theorists will be frustrated by the binary structure of the diagram” (“Machine Learning” 98). But he doesn’t address how this “binary structure” deliberately consolidates and computes femininity against masculinity. The diagonal line moving from the bottom-left to the top-right quadrant represents the common cluster of gender terms used by both male and female authors to describe male and female characters. But the inverse line, a more sparse cluster moving from top-left to bottom-right, represents terms used by women to describe women (top-left), using terms like “thought” and “wondered,” and those by men to describe women (bottom-right), using terms like “kissed” and “was-married.” The binary nature of the chart, therefore, not only imposes an opposition, but an opposition between how male and female authors describe women. The effect is to reinforce a stereotypical active/passive binary (among other binaries) between men and women.

While there may be valid reasons to study gender as a binary system, this approach overlooks the potential of computational methods to bring something new to the surface. Toward this end, work by critics like Susan Brown, Laura Mandell, Richard Jean So, and Edwin Roland use text analysis to deconstruct social categories. In their introduction to *The Journal for Cultural Analytics*’s “Identity Issue,” 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” (6). 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 for the purpose of deconstructing social categories. For example, in one project, Mandell uses genre to study gender. 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 Wollstonecraft, 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 through 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).

Applying text analysis to racial categories, Edwin Roland and Richard Jean So experiment with an algorithm that evaluates an author’s race based on diction. Analyzing a large corpus of novels by white and black authors, they find that black authors generally display more varied vocabulary than white authors. From this result, they infer that white authorship, as a category, only coheres against the variance of black authorship (66). However, 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 due to a single word, “appalled,” which the computer mistakenly reads as indicative of white authorship. 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” (Rpt. in Roland, So 71; emphasis mine). Given the connotation of whiteness in “appalled,” which has the middle French root, “apalir” (to grow

pale), Roland and So posit that this term suggests a relation between gender and race. The computer's misclassification enables these researchers to pinpoint a specific moment that race and gender become imbricated in the novel: "the moment David develops a troubled relationship to normative masculinity [as] also the moment he becomes 'white'" (71). Additionally, the this misclassification of the novel as written by a white author engages with a common observation by critics that it lacks explicit references to race.

In direct opposition to the "falsifiable" position advanced by Moretti and others, then, text analysis can also surface new and unexpected readings of even critically well-trodden texts. The necessary constraints of certain algorithms impose structures that pull out particular social constructs, but not for the purpose of their reification as if, in Mandell's words, these constructs "were simple pointers to an unproblematic reality, transparently referential and not discursively constituted" (par. 5) But rather, they can surface new contexts and operations for such constructs. Indeed, as I argue below, because machines must comply by rigid and reductive rules, they are ideal instruments for interrogating and revising similarly restrictive social constructs.

Iterative

Mandell asserts that both gender and genre "are... highly imitable," 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 gender performativity (explained further below), this theory remains a useful heuristic for quantitative text analysis. First, the common misreading of gender performativity from Judith Butler's book, *Gender Trouble: Feminism and the Subversion of Identity* (1990), is that it denotes an act or series of acts that can be imitated at will. 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, 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. Common critiques of Butler point out the limits of this theory for posing gender and sexuality as discursive.⁴ From the field of Trans Studies, Jay Prosser problematizes Butler’s “deliteralization of sex,” which he applies to Queer Studies more generally, for the way it elides the real-world concerns of the body’s materiality and its precarious existence in cross-gendered identifications. This thread of discursivity, and its implications within a larger trajectory of Queer Studies, is picked up again in this chapter’s conclusion.

As a discursive phenomenon, performativity is situated within second-wave feminism and its deconstruction of the gender binary. Butler draws from the work of feminist theorist Luce Irigaray, who asserts that influential Western thinkers like Plato, Aristotle, and Freud have defined femininity “on the basis of masculine parameters” (Irigaray, *The Sex Which Is Not One* 23). Irigaray argues that the association of “woman” with “matter” (associated to binaries such as “rationality/emotion” and “mind/body”), and its subordination to male “form,” “produces the feminine as that which must be excluded for that economy to operate” (Butler, *Bodies* 10). This excessive feminine, the “necessary outside” contrasts with the “domesticated” feminine within

⁴ Irigaray’s concept of the “necessary outside” seems to anticipate another popular critique of Butler’s theory, from the field of Political Philosophy, 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).

the system. As Butler points out, this outside or “unspeakable” element cannot be invoked directly without subscribing itself to the ruling structure (*Bodies* 12). So she 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 refused by the system—leads Butler to her theory of gender subversion. For Butler, subversion begins by questioning 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 Butler 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 participation in it. Within this regulatory structure, this significatory circle, lies the possibility of resistance by *resignifying* meaning. Language can be resignified toward subversive usages by “citing” what Butler calls a “repudiated” meaning. Butler offers the famous example in the resignification of the term “queer,” which is resignified when it harnesses its own repudiation, an implied but “disavowed abjection [that] will threaten to expose the self-grounding presumptions of the sexed subject” (*Bodies* 3). By citing the repudiated meaning, the term “queer” “resignify[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. Turning back to Irigaray's writing, Butler explains that 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). In an ironic simulation of Irigaray's own thought process, Butler posits that,

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, Irigaray displaces the logic of phallogocentrism, introducing something external to the system while simultaneously remaining within its terminology. Deception emerges from resemblance, and insubordination through subservience. The key is iteration, a continual miming of the authorizing norm.

Turning from gender to computation, iteration also emerges in the Python programming language as it is used for computational text analysis. At a very basic level, much of text processing with the Python programming language consists of what programmers call "iterating" over words in a text. The Python language handles individual words, or strings, within groupings called lists. The `for` loop, a common construct for iterating through lists,

repeats a single action to each string within the list.

Below is an example of a `for` loop that “iterates through,” or cycles through, the words in text with the goal of filtering out punctuation and transforming any capital letters into lowercase forms:

```
clean = []

for word in text:

    if word.isalpha():

        clean.append(word.lower())
```

Here, the loop begins by creating an empty list, `clean`, where words will be dropped after passing the filters below. The next line begins the `for` loop, which iterates through each `word` in the `text`, using the syntax, `for word in text`. The third line is a filter that checks if the word is comprised of only alphabetic characters, `isalpha`. If the word is purely alphabetic, that is, contains no numbers or punctuation, then it passes the filter, and it appended to the empty list, called `clean`. When the word is added to the list, its letters are transformed into lowercase format with the `lower()` function. The final list, therefore, will only contain alphabetic and lowercased letters.

Most of text analysis consists of iterating or “looping” through bits of text and performing actions to standardize the text for analysis. These actions include tokenizing, cleaning, and regularizing. The first step, tokenizing, means separating the strings in the text into workable units, into individual words and punctuation marks, that are easier to clean and regularize. Once the text is tokenized, it can be stripped of capital letters, punctuation, like in the example above.

The next step is removing what are called “stop words,” which consist of prepositions, articles, pronouns, and auxiliary verbs. Due to their high frequency and low semantic value (in comparison to nouns, verbs, and adjectives, and adverbs) stop words will skew the results of analysis, so they must be removed. To do so, I use another loop:

```
no-stops = []  
for word in cleaned:  
    if word not in stops:  
        no-stops.append(word)
```

This expression takes each word in our `cleaned` list, and checks to see if that word is also contained in a list of stop words, that is, words like articles, prepositions, and auxiliary verbs. If the word is *not* a stop word, then it will be added to a new list, `no-stops`. After this processing, the following words remain from the first sentence of *Orlando*:

```
'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 possible 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 alter part of the word root, such as “berries,” which would become “berri.” The other possibility, called “lemmatizing,” involves looking up each word, one by one, in a dictionary to find its appropriate root, or “lemma.” Though it takes more time, this

process is useful to handle words like “berries,” and also more complex plural forms, like “children” and “teeth.” Below is the code for lemmatizing the text, which uses the Python module, called “WordNetLemmatizer,” to lookup the individual lemmas for each word:

```
lemmatized = []

for word in no_stops:

    lemma = WordNetLemmatizer().lemmatize(word)

    lemmatized.append(lemma)
```

In this example, the code iterates through the list of words, `no_stops`, which have already been stripped of punctuation, numbers, capitalization, and stop words, and it lemmatizes each individual `word` using the `lemmatize` function. Then, it re-saves the word to a new variable, called `lemma`, which it appends to the list called `lemmatized`. After iterating through every `word` in the `no_stops` list, and lemmatizing each one by one, the text is ready for analysis.

Text analysis is, at a most basic level, about counting words. A good first step, then, is to begin by counting the words in context, or the immediate words, that surround a given word. The function `concordance()` displays the context window surrounding a given word, in this case the word “woman,” from *Orlando*:

Displaying 15 of 187 matches:

```
ty manly charm quality old woman loved failed growing old w
en without knowing perhaps woman heart intricate ignorance
indow pulled among cushion woman laid worn old made bury fa
red dyed cheek scarlet old woman loved queen knew man saw o
```

finger swore vilely master woman scarcely le bold speech le
ment panic twelve poor old woman parish today drink tea ton
perverse cruel disposition woman broke engagement night eve
verladen apple old bumboat woman carrying fruit market surr
embassy figure whether boy woman loose tunic trouser russia
ed stared boy ala boy must woman could skate speed vigour s
me standstill handsbreadth woman orlando stared trembled tu
d asked tumult emotion old woman answered skin bone trulls
and longer full prying old woman said stared one face bump
save sea bird old country woman hacking ice vain attempt d
ce melt heat pity poor old woman natural mean thawing must

This context window around the usage of “woman” in the text, which only shows its first fifteen occurrences, reveals a text that has been stripped of punctuation, capital letters, prepositions, articles, pronouns, auxiliary words. What remains is largely nouns, verbs, and adverbs. The task is to then count these remains. For that purpose, another function, called `similar()` can then infer words that tend to be used in similar contexts. To compute the results of `similar()`, the program first takes the context of the target word from `concordance()`, then it searches the text for other terms which contain the same surrounding words. Behind the scenes, then, the program is keeping track of many times every word appears next to every other word in the text. So when it takes the string “woman,” it can search for other words that tend to have similar strings in proximity. The resulting “similar” words, then, do not appear in the same context as “woman” in the text; rather, they appear in contexts that are similar to those that surround “woman”:

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

Although this method might reveal words that have an obvious relation to the target word, such as “lady” or “queen” for woman, it is important to point out 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.

Basic text analysis tasks like `similar()` and `concordance()` contrast with “deep learning” methods (discussed in more detail in Chapter Four) that use predictive mechanisms to analyze language. Many of these methods are based on the concept of word embeddings to ascribe machine-interpretable meaning to strings. Like `similar()` and `concordance()`, word embeddings work from patterns of word similarity based on context. Unlike the previous methods, however, word embeddings encode numeric values to a given word based on its context. The value of any given word is known as its “vector,” which represents the location of the word in graphical space. A vector for a single word, “woman,” like the one included below, will contain a list of numbers that represent a similarity score between “woman” and another word. As numerical representations, word vectors enable further quantitative exercises that can analyze the relationship between “woman” and other words. Based off a popular dataset, the vector which represents “woman” contains a list of numbers that score the similarity “woman” to other words in the dataset.⁵ 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

⁵ The language model for this computation comes from Word2Vec’s “glove-twitter-25” dataset, which is based on Twitter data.

“father,” with .90, and so on:

```
'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 vectors are organized into a matrix, or tabular, format. Using this matrix format, further mathematical operations are possible using statistics, linear algebra, and calculus—the building blocks of so-called “deep learning” methods. Significantly, in deep learning, the individual words associated to each probability do not matter, only the probabilities themselves, 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. 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.”

Fluid discourse

Woolf's novel, *Orlando: A Biography*, is ideal for a computational study of gender for two reasons. First, published in 1928, it is an early example of transgender narrative. Second, as many critics have noted, its characteristic modernist experimentation with the limits of language destabilizes 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 to disrupt expectations of plot and narrative to challenge the stability of gender and identity. Doubling down on the role of language, some critics emphasize that Woolf's 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 Caughey 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" (Caughey 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 takes the output of

computations as material for close reading analysis, using the computer to identify words that I then examine in context and with critical subjectivity.

I begin with a list of terms that are computed as similar to “woman” and “man” in the text.⁶ 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. 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 words most closely associated with “woman” and “man,” respectively, each with its own score, or probability. The following are the words most associated with “woman”:

```
[('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),
```

⁶ Unlike the vector for “woman” in the previous section, the vectors here are trained on Woolf’s novel, and therefore reflect an understanding of gender markers based on this specific text.

(‘seductiveness’, 0.30290719866752625)]

And the following with “man”:

```
[ ('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, the term “delicious” 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 a distinctly passive mode of gender subversion. 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 that coordinates 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 occurrences appear in a single passage, when Orlando is sailing from Turkey back to her native England. The ship captain offers Orlando a slice 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. Although the word “delicious” describes an arguably 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 where 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. The first of those appearances occurs within one of the few sentences in the novel in which 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 persistency 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 about the deliciousness of refusal:

‘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 pleasure in a passivity that is not quite powerless. Here, Orlando’s refusal to conform to social expectations, for which she feels culpable, constitutes another kind of passivity, a form of refusal. The connection between “delicious” and “culpable” seems to define queer desire as a distinctly passive mode of resistance.

Now, I move to Orlando’s experience as a man, returning to the original list of similar terms. I return to the list of vectors associated with the word “man,” and begin with “chequered,” which appears only once, at the start 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 the poet who embellishes and exaggerates reality through figurative language. However, the narrator’s commitment 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 his pretense to objectivity. This slip into figurative language eventually grows into a crisis of signification that recurs persistently through the novel. This crisis will eventually affect not only language's ability to signify, but also that of gender. As Victoria L. Smith affirms, "what happens in the novel... and what it thematizes—language's inability to adequately represent the 'thing itself'—mirrors the undecidability of the text—is it a biography, an autobiography, fantasy, etc.—and the impossibility of the form of 'woman'" (58).

The crisis of signification 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 another term, "despaired," from the list of vectors associated with "man." Like "chequered," this term 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 who is suspicious of figurative language, Orlando also questions the truthfulness of figuration. 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 flight and modesty, qualities which he finds “false,” in nature. The word “despaired” here, like “chequered” above, appears in a moment that elevates the role of figuration in language. It also marks a crossing of the crisis of signification over the diegetic threshold, into Orlando’s internal world. These terms, which occur in scenes where the lines between fiction and reality, and internal and external narration are blurred, elevate the importance of language’s imaginative capacities in the novel.

In a final example, language’s ability to signify becomes closely coordinated to gender. I examine the co-occurrence of words from both the “woman” and “man” lists within a single passage, the scene where Orlando first meets his beloved Sasha. The words “curiosity,” which is associated with “man,” and “seductiveness,” which is associated with “woman,” appear in a tumultuous meditation on Sasha’s gender incomprehensibility. 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 seemingly arbitrary 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 signification beyond the diegetic narrative, the narrator’s “pause” signals the constructed nature of the scene. This constructed quality is emphasized by the narrator’s explanation for Orlando’s choice of words, 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.

The iterative reading practice, which bases close attention to language from the results of distant reading analysis, brings to the surface the relationship between gender and imagination, particularly imaginative forms of language, in the novel. As Pamela Caughey asserts, Orlando’s gender transgression is intimately connected 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). The list of words offer a starting point for isolating passages throughout the text, which then offer opportunities for finding new words to read in context. As the reader moves from distant to close, from a word vector to its context, the meanings of certain words gain new significations. And what happens with language in the novel also happens with gender. The above passage, with its “switching” and “shifting” discourse, which asserts that word choices are arbitrary and flowing, implies that gender is also a fluid phenomenon.

The argument that *Orlando*'s subversiveness is a discursive one, opens the text to numerous critiques,⁷ particularly from Trans Studies. According Jay Prosser, Woolf's experimentation with language and narrative form belies 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/her narrative propels h/her through four centuries of history, Orlando is free to move beyond h/her body—quite queerly, to break through the limits of the flesh" (Prosser 168). By "the sexed body," Prosser means the physical body that exists in a physical world. *Orlando*'s transgressiveness, Prosser argues, results from a play of language and literary form that elides the specificity and the lived reality of the "sexed body." Due to its "ambivalence, a wavering around transition," "a transformation of transition into new identity," its "easy androgyny," this text is transgender rather than transsexual (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 lens 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

⁷ 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 perspective. 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 transgressive subjects, the novel "allows the protagonist to pass as respectable 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).

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 connotations of water, especially the physical one of corrosion. While the topic of physical embodiment and racialization is one that I take up in my third chapter, in the concept of flesh, here I have tried to show how attention to iterativity in computer code, and the fluidity it inspires for close and distant reading analysis, surfaces the material components that undergird what is often presented as an immaterial or abstract processes within quantitative analysis.

To think more deeply about this materiality, I would consider iterativity’s connection to the productive power of language, the ways that language physically executes action, for example, in computer code. The Python programming language differs from other languages (like markup languages HTML and XML, which I explore in my next chapter) in that it is an *executable* language. The code not only defines instructions, but it also enacts them. Perhaps, this active quality of Python programming might deepen our understanding of gender as an active, embodied phenomenon.