“Queer Tools for Reading, Editing, and Archiving” Filipa Calado

*My dissertation explores how digital methods and tools for studying text engage with queer literature. I critique digital methods and tools by posing computation, where textual data is cleaned and structured for electronic processing, against the complexity of queer subjecthood and affects expressed in textual style, form, and voice. While tools like quantitiative text analysis, for example, transform (and necessarily reduce) qualitative elements of gender and sexuality into numerical data such as word frequencies or concordances, I argue that this reduction opens up possibilities for interpreting the formal qualities of queer literature which expresses queerness in narrative and figurative structures. Just as digital fomats transform and manipulate text into data, so do literary forms figure queer identity and experience. Approaching queerness and computation as formal phenomena challenges scholars to defamiliarize their knowledge about technology and to rethink the “digital” as something at once tangible, slippery, and distinctly queer.*

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, the standard tasks of “pre-processing” the text evacuate the ways that gender is unsettled in this sentence. In order to computationally analyze a text, a process known as “distant reading” or “text mining” which involves calculating, generating, and visualizing textual patterns, the text must first be transformed into a computable format. This task of

pre-processing (also called “cleaning” or “normalizing”) strips the original text of capitalized words, punctuation, “stop words” (such as articles and prepositions), and inflections in word endings, all of which are deemed to be semantically minor, in order to make the text amenable to quantitative analysis. After pre-processing *Orlando*, the following list of computable words, or “tokens,” remains in the first sentence:

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

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

My research looks to such moments where computational processes collapse or evacuate the nuances of gender, sex, and sexuality in literary texts. Then, I explore how digital tools might borrow concepts from Queer Studies to forge new methods for studying queer subjectivity and affect in these texts. To study the novel *Orlando*, whose protagonist undergoes a sex change, I use a powerful text analysis too, called the Natural Language Toolkit (NLTK) to count, classify, and visualize gender terms and associations throughout the novel. Text analysis with NLTK works by creating programmatic “loops” that iterate through textual data to carry out specific tasks, such as removing capitalized letters or punctuation, or counting words that occur in proximity to one another. My research looks to ways that programmatic concepts like iterativity in text analysis evokes ideas from Queer Studies, in this case, from Judith Butler’s concept of gender performativity. I take this concept of gender performativity, in which Butler defines gender as a series of repeated (or *iterated*) acts in which the subject “cites” heteronormative regulatory schemas, to model a new approach for using text analysis to study gender in *Orlando*. Using NLTK, I write a computer program that iterates through the terms “woman” and “man” to generate words that function similarly throughout the text.1 Then, I feed the output (the words which are similar to “woman” and “man”) from this computation as new input into the program, and run it again. Iterating a number of times through the initial binarization of “woman” and “man” yields a list of terms which expands this binarization into a plurality of gender

1 NLTK computes “similarity” by generating words that appear in similar “contexts” (that is, surrounded by the same words) as the target word.

significations from the novel. For example, after iterating three times through word similarities for “woman” and “man,” the top words associated with “woman” are “summer,” “rust,” “burst,” and “shape,” and for “man,” the top words are “ship,” “turn,” “stood,” and “pushing.” My analysis considers how this notion of gender “iterativity” in text analysis expands the study of gender beyond the tendencies of what I call “reproducibility” in current scholarship on distant reading, which largely analyzes gender in text as a binary system and therefore reproduces this assumption in their results.

Like this example with text analysis, my dissertation, “Queer Tools for Reading, Editing, and Archiving,” explores how a variety of digital tools engage with the formal complexity of sex, gender, and sexuality in literary texts. I demonstrate how the limitations of working with text in digital formats allows critics to probe literary representations of queer gender, sex, and sexuality. This project applies computational processes, strict data formats, and abstract programming logics to study the multiplicity, fluidity, and dynamicity of queerness. Here, I unpack how the constraints of technology, which disambiguates and fixes data might be

re-worked to reveal the complex and ambiguous forms of queer identity and experience in text. Like the above example with *Orlando*, I find that the constraints of computational tools reinforce queerness as something formally challenging and complex. I emphasize that what these tools reveal are not solutions for understanding or “fixing” gender or sexuality, but opportunities for exploring their ever shifting permutations.

This project engages the emerging field of *Queer DH*, which troubles the intersection of two seemingly incommensurable fields—the Digital Humanities (DH) and Queer Studies. Queer DH resists the increasing investments in using digital tools and methods to answer questions, resolve ambiguities, or verify hypotheses in Digital Humanities scholarship. In “Toward a Queer

Digital Humanities,” authors Bo Ruberg, Jason Boyd and James Howe explain that, “At a time when the digital humanities promises to make sense of the world through supposedly objective computational tools, queerness refuses to allow us to stop reflecting, stop challenging, and stop questioning” (123). Much of DH scholarship is enticed by the “normativizing” pull of digital methodologies, which borrow from the social sciences and are used by humanists to verify, correct, or establish knowledge about cultural materials. Resisting these practices, *Queer DH* emphasizes instead how digital tools and methods might productively complicate the legibility and stability of knowledge, identity, and experience.

Oriented around the question of “queerness” in technology, I position my work in between two approaches in Queer DH: the first approach aims to disrupt formal systems in computation by imagining alternative ones, while the second approach maintains that queerness is built into computing and is inherent in computational logic. Often, the first approach consists of speculative or critical making projects that problematize the constructed nature of technical objects. For example, Zach Blas and micha cárdenas’s speculative codebase, *transCoder*, describes hypothetical functions inspired by Queer Theory, such as the “destabilizationLoop” and “nonteleo().” And in the influential essay, ‘Queer OS: A User’s Manual’, the authors imagine various components of an operating system, which “[do] not yet exist and may never come to exist [… do] not yet function and may never function.” This theoretical operating system includes an interface that “disappears but is not naturalized,” and “promiscuous” applications that “move and interact across platforms, devices, users, and geographical regions

unrestricted” (Barnett et al). The other side of the debate explores how technological systems and tools contain elements that already encourage queer modes of analysis. For example, work by Jacob Gaboury explores how the “NULL value” in database computing offers a “refusal to

cohere, to become legible” as a built-in option in computational systems (“Becoming NULL”). Gaboury asserts that “there exists a structuring logic to computational systems that, while nearly totalizing, does not account for all forms of knowledge, which excludes certain acts, behaviors, and modes of being” (“Queer History”). Similarly, Textual scholar Julia Flanders finds that queerability is built into certain digital tools, like the TEI (discussed below), because it offers the potential for customization and dissent. According to scholars like Gaboury and Flanders, it is from within the structuring logics of computer systems that queerness finds the space to operate.

Cutting between these debates, my work searches for the structural constraints of digital tools then experiments with how these constraints might be re-imagined to engage queerness.

From the various tools and methodologies in DH scholarship, I focus on text analysis, electronic editing, and media archeology. My methodology is to closely examine each digital tool for the way it transforms and abstracts textual data, and then to explore how this abstraction surfaces formal structures that problematize conceptions of gender, sex, sexuality, and desire, as they have been theorized in Queer Studies. In my analysis of each tool, I first identify how the process of digitization reduces the complexity and nuance of queer identity and experience in text. Here, I probe the displacement between what we see on the screen and what is inscribed in the computer’s hardware (what Matthew Kirschenbaum calls the levels of “formal” and “forensic materiality”) to deconstruct how digital tools operate across various levels of abstraction. Then, I look to how this reduction enables one to proceed with analyzing the coded forms of queerness in text. I redirect the computational processes, which abstract the complexity of the real world into structured data, toward a speculative experimentation of “queer form.” My concept of queer form draws from theorists like Kadji Amin, Amber Jamilla Musser, and Roy Pérez, who read queer literature for the way it deploys formal and narrative strategies toward an “aesthetics that

moves persistantly around the visual,” with the purpose of “make[ing] difference a little less knowable, visible, digtestible” (235). In my work, queer form emerges in aesthetic strategies that problematize the compulsory visibility of queer subjects, and seek out ways that these subjects resist heteronormative structures that demand sexual and gender transparency.

Keeping a critical awareness of the tool’s limitations, I redeploy the tool to mark and visualize this queer form, particularly the elusive identities, repressed desires, and other coded elements of queerness in text. My primary texts center on auto/biographical fiction or life-writing (such as Gloria Anzaldúa’s *Borderlands/La Frontera: The New Mestiza*, Carmen Maria Machado’s *In The Dreamhouse*, and Virginia Woolf’s *Orlando: A Biography*), hand-written manuscripts (Oscar Wilde’s *The Picture of Dorian Gray* and James Baldwin’s *Giovanni’s Room*), and autofiction (Caitlin Fisher’s *These Waves of Girls* and Anne Garretá’s *Not One Day*). In these texts I examine how formal experimentation occludes the legibility of sex, gender, sexuality, and the writing self. My analysis draws from canonical readings in the field of Queer Studies, including Judith Butler, José Esteban Muñoz, Eve Sedgwick, and Heather Love. These thinkers pose frameworks for examining how formal and narrative strategies can obfuscate easy understandings of gender, sexuality, and other marginalized elements of identity, in literature. For example, Muñoz and Sedgwick offer an approach for analyzing queerness as associated with potentiality to read affect as an aesthetically productive force that creates new structures and forms for delineating (or attempting to delineate) queerness. My understanding of affect is deepened by Heather Love’s work, which gives full rein to the “backward” affects such as despair and self-hatred, without trying to resolve or “recuperate” them into more palatable formulations. Love’s work crucially opens a space for queerness to resist the neoliberal specters progress and affirmation, as she encourages critics to confront (without attempting to redeem)

painful queer histories. The work of these theorists guide my usage of digital tools for studying the “coded” forms of gender and sexuality in figurative and narrative strategies. For example, I find that the strict data structure of the TEI (Text Encoding Initiative, an “encoding” method for transcribing and editing electronic text), allows scholars to think productively about social experiences of exclusion or impossible love that may seem beyond redemption. This careful, minute work of “encoding,” which is as much manual as intellectual labor, confronts one simultaneously with the compulsory containerization of textual data and the limitations of such containment. In my research, I connect this quality about the strict structures of textual containment to the ways that queer critics negotiate uresolvable and unyielding representations of queer suffering in their critical projects. This close and careful work with TEI allows me to make this connection between affective and data structures, which offers an opportunity for rethinking my usage of the tool.

Like computation, queer form relies on abstraction; but unlike computation, queer form does not normalize or reduce the complexity of data. My work aims to show precisely where and how the reductions of the computer might engage queer form as complex and slippery. For queerness, as Muñoz says, is “not yet here,” but perpetually on the horizon (1). In this project, queerness remains as a target beyond reach, as a fount for future subversions, exemplified with the term “queer” itself, which Butler says “never fully owned, but always and only redeployed, twisted, queered from a prior usage and in the direction of urgent and expanding political purposes” (173). Therefore, this work offers an approach for studying queer material that does not fully circumscribe to a general methodology. My goal is not to build reproducible schemas and models for analyzing queerness. Rather, I look to harness opacity and unintelligibility as resources for resisting inclusion in what Muñoz describes as “the ossifying effects of neoliberal

ideology” (22). My project therefore posits queer form as a kind of technology of resistence, which digital tools can help to surface. I hope this experimental work will encourage the further developments in reading our queer literary heritage, which, as Butler says, “begin, without ending, without mastering, to own—and yet never fully to own—the exclusions by which we proceed” (25).

My first chapter, “Reading: Quantifying Gender in Virginia Woolf’s *Orlando: A Biography*,” examines how computational text analysis grapples with gender ontology in Woolf’s novel, *Orlando*, which features a transgender protagonist. This chapter reworks the

iterability of text analysis and gender codes to surface new forms of gender as multiply-sexed in Woolf’s novel. It begins with a critique of current practices of what I call “reproducible criticism” in distant reading, in which the results of the critical process often reproduce the critic’s assumptions from the outset. Then, I examine the Natural Language Toolkit (NLTK), a powerful code library for carrying out text analysis with Python. Using NLTK consists of repetitive tasks of “cleaning” and counting textual elements, an iterability that evokes Judith Butler’s concept of gender performativity, in which gender is theorized as a repetitive habit.

Here, I draw from Butler’s concept of performativity to model a new method for text analysis that subverts gender reproducibility through iteration. The chapter ends with my reading of Woolf’s *Orlando* alongside a visualization of gender which demonstrates how the terms “woman” and “man” can be re-signified from binary to multiple.

My second chapter, “Encoding Queer Erasure in Oscar Wilde’s *The Picture of Dorian Gray*,” uses the TEI editing format, which subscribes textual data to rigidly contained hierarchical “tagging” structures, to encode the homoerotic elements that Wilde edited during his revision of *Dorian Gray* (1890). My analysis finds that the TEI works best with data which is

discrete, rather than smooth data, like the homoeroticism obscured by Wilde's pen. Like my critique of text analysis, the inner logics of the TEI reveal a connection to queerness: As a labelling tool, the TEI surfaces moments where queerness, which is plural and permeable in this text, threatens to spill over the bounds of its data structure. I discover the TEI’s constraint when attempting to mark up and analyze this text's homoeroticism, which I group into the general themes of ‘intimacy’, ‘beauty’, ‘passion’, and ‘fatality’, as well as the pen strokes that Wilde uses to strike these elements from the text. The functionality of the TEI as a tool that separates smooth data of these revisions into discrete elements led me to explore (without resolving) the indeterminate boundaries of these queer themes in the text. This practice in “queer encoding” tags the homoerotic elements in such a way that allows them to retain some of their elusiveness.

Whereas the first two chapters are about “Reading” and “Editing” tools, or text analysis and text encoding, respectively, my third chapter is about “Archiving,” and engages media archaeological methods with discussions on New Materialisms from Queer of Color Critique and Black Feminist studies. This chapter (which is currently being researched) draws from Matthew Kirschenbaum’s “media archeological” method, which deconstructs the material specificities of what are commonly conceptualized as “immaterial” digital processes. I deploy his approach to study the material qualities and of Electronic Literature, such as hypertext and interactive fiction. I engage Kirschenbaum’s methods with accounts from New Materialisms, predominantly from within Black Feminist studies, which offer frameworks for thinking through the physical, vital, and sensual aspects of media, and of the perceptual engagement between user and computer.

Black feminist authors like Jayna Brown, Zakiyyah Iman Jackson, and Amber Jamilla Musser offer crucial counter-lineages for theorizing materiality and sensuality, which I apply to digital media.

My dissertation offers a blueprint for educators and students using digital tools to work with queer literature. It employs an experimental and accessible format to serve as a guide for non-experts in technology. Each chapter is supplemented by a small digital project that demonstrates in practice how to use each digital tool for “Reading,” “Editing,” and “Archiving” in literary studies. I use open-access tools like NLTK and TEI, and share all of my code in online repositories, so that readers can download, use, and build from my work. Over the past four years, my work as a Digital Fellow, Program Coordinator, and Student Advisor at the Graduate Center Digital Initiatives (GCDI) and the Digital Humanities Research Institute (DHRI) allowed me to cultivate a method of hands-on learning in my teaching which guides my approach in writing about technology. In my teaching, I encourage my students, who are often total beginners in using digital tools, to learn through practice, and to consider the limitations of digital tools as opportunities for analysis. In particular, I emphasize the unfamiliar and unexpected ways that these tools might affect their critical and creative processes. Bringing this exploratory attitude to my dissertation, my overall aim is to empower readers to experiment with these tools, so that they might gain purchase over the structures that determine meaning-making in the humanities.

Thank you for your consideration.