Reading DT Leaves:
A Digital Analysis of "Lyrical" Fiction

By Timothy Schott

Presented to the Department of English in Partial Fulfillment of the
Requirements for the Degree of Bachelor of Arts with Honors

University of Virginia
Charlottesville, Virginia
April 12, 2019

Abstract

The class of novels referred to as "lyrical" by critics and readers deserves comprehensive inspection and evaluation. Virginia Jackson creates space for me to start this analysis as she recognizes a fraught conception of "lyrical" poetry.
Friedrich Nietzsche and Ralph Freeman inform my view of potential formal features that signal whether or not a novel is "lyrical." My project takes up this task from an untraditional perspective. I implement state of the art machine learning algorithms and computational analysis to understand the formal elements that make a novel "lyrical." Franco Moretti's school of "distant reading" looms large in this type of analysis. This project is enabled through my creation of a database containing every word, sentence and paragraph of fifty novels – half "lyrical," half detective fiction as a control set. I complement my digital work with traditional close readings. Through my paper, I hope to show that digital methods of scholarship are readily compatible with time tested styles of critique. I advocate for further digital scholarship in this space, but of the proper variety: balanced and transparent. Ultimately, I make the case that the most salient feature of "lyrical" novels is their reliance on anaphora.

Table of Contents

List of Figures

"Scientific knowledge is a kind of discourse. And it is fair to say that for the last forty years the 'leading' sciences and technologies have to do with language."

– Lyotard, *The Postmodern Condition*

"To be considered poetry a piece of writing must make significant use of rhythm and metaphor."

– incorrect answer, *Barron's SAT Practice Test Book*

"*Very like a whale."*

– *Hamlet* 3.2.411

Introduction

An amalgamative exercise follows. I apply traditional tools of literary studies in combination with modern digital methods on a group of fiction that may or may not be written in a "lyrical" mode. A balance of statistics and hermeneutics supports my effort. Critics, publishers, and hypnagogic undergraduates alike laud the corpus of books I analyze as "lyrical." *But what does that term mean when applied to a novel?* I'm skeptical of the book jacket that tells us Cormac McCarthy's prose in *The Road* is resoundingly "lyrical," for instance (Book jacket, *Blood Meridian*). I interrogate this set of canonical fiction through a multitude of analytical mediums to shed light on this anodyne accolade.

The most interesting, recent criticism of lyric poetry underscores its precarious nature. Virginia Jackson points out the ways the lyric permeates our interpretation of poetry (Jackson, *Lyric* 826). Daniel Albright flatly dispenses with any attempt to coherently define the lyric. "A lyric is that which resists definition," he insists, characterized by an operation tantamount to "magic spells" (Albright vii, 67). Stephen Burt argues that universities insist upon teaching disparate poetic compositions "as if they were what we call 'lyric' now, whether or not the shoe fits (and whether or not the cobbler wanted it there)" (Burt 436).

The instability of the "lyrical" label in the realm of poetry signals that a similar ambiguity could inhere its application to fiction. Gabrielle Starr contends that the lyric and the novel share a common lineage (Starr 11). Using lyricism with new rhetorical strategies such as free indirect discourse, novelists in the 18th century "create fictions of consciousness to focus readers' movements from the world of reading to the worlds of

novels" (Starr 14). These developments continue beyond the 18th century; conventions from the lyric inform modern fiction writers. Readers, too, cling to the lyric. We invoke lyricism as we praise texts in *Amazon* reviews and fawn over powerful, "lyrical" paragraphs in our marginalia.[1] How should we treat the lyric in an analysis of fictional prose? What are the implications of bounding potentially unrelated works of art with words like "lyrical"?

In this project I press canonical fiction to its limits. I rely on a combination of algorithms and familiar modes of close reading to identify the patterns and characteristics of "lyrical" prose. The potential to close this paper with a new, "data-driven" definition of what makes a novel "lyrical" is certainly alluring. But there is also a chance I antagonize the "lyrical" affixation to the point of oblivion and leave you with nothing but the frayed, unraveled ends of a once mighty knot. I don't fixate on either of those extremes, though. The probable result will be something less dramatic. Somewhere in between those two poles—and significant nonetheless. Of course, there is only one way to find out: let us begin.

The structure of this thesis is as follows. First, I describe how my corpus came to be. Next, I move into Jackson's history of lyric poetry and her intriguing concept of lyricization. I emphasize her insights on poetry that relate to my area of study and complicate it. To follow, I interface with the scholarship of Franco Moretti, Friedrich Nietzsche and Ralph Freedman. I accompany these encounters with a digital point-of-entry that quantifies my corpus. I garner word frequencies from Moretti, sonic depth from

[1] "Woolf's often *lyrical* prose conveys the sights and sounds of life on the island at the same time that it also enlivens the highly philosophical but very personal portrait of family life." –Mary Whipple's five-star review of *To the Lighthouse* (Whipple, emphasis mine).

Nietzsche and imagery from Freedman; simplicity, sound and vision. Innumerable hours spent coding and executing experiments culminate in robust information regarding the syllabic structure, sonic qualities and richness of imagery present in each text, among other metrics.

I take advantage of these measures right away in an effort to remain concretely connected to my corpus. To do this, I close read my texts to press whether or not the values I calculate shed light onto relevant passages from my corpus.[2] Moretti crystalizes this ebb and flow between the digital and the conventional, as do all the other digital scholars this project is indebted to such as our English department's wonk-in-residence, Brad Pasanek.

This survey is largely informed by the Digital Humanities (DH). DH closely aligns time-tested critique with the power of computer programming and statistics. I believe a holistic analysis of a vast amount of data is a promising method for providing insight into the underlying logic of the "lyrical" novel. Humanists, writes Ted Underwood, shine in this type of endeavor, as "we're already familiar with one central application of machine learning—the task of modeling fuzzy, changeable patterns implicit in human behavior" (Underwood, *Why an Age of Machine Learning Needs the Humanities*). Principles imported from DH thus profitably steer this effort.

The writing of Martin Heidegger supports the balanced, iterative structure I adhere to in this project. His claims in *Being and Time* about the conflict of praxis and theory are worth a mention here:

---

[2] Ideally, following the calculations I glean from each scholar with close readings serves as a palette cleanser for those only *tepidly* enthusiastic about the nuances that stem from applying statistics and computer programming to a large body of texts...

> ... action must apply theoretical cognition if it is not to remain blind. Rather, observation is a kind of taking care just as primordially as action has *its own* kind of seeing. Theoretical behavior is just looking, noncircumspectly. Because it is not circumspect, looking is not without rules; its canon takes shape in *method* (Heidegger 69, emphasis author's).

I pair critical voices with digital calculations in hopes of a robust understanding of the concept of a "lyrical" novel. It is important not to get swept into the provocative ideas of Culler, Jackson, Starr and others. Theory is not a substitute for method.

In addition to repeated close readings using my data points, I group the data into a matrix. Ultimately, I associate each text with 30 relevant quantities. Each information point can be called a "feature," a piece of information useful to my analysis.[3] Each row lists the relevant features for a particular text in my corpus. Row 12, for example, is the home of *Pale Fire*. Throughout this paper I refer to this matrix as my "feature matrix," as is standard nomenclature in the realm of statistical inquiry. It is available in its entirety in the Appendix (Figures A.3 and A.4).

With this feature matrix in hand I implement a machine learning algorithm to properly classify my corpus into my camps of interest: "lyrical" and not. In his 2017 overview of digital trends, *Radical Technologies*, Adam Greenfield defines machine learning as, "the process by way of which algorithms are taught to recognize patterns in the world, through the automated analysis of very large data sets" (Greenfield 216). "Algorithm" is not so scary a word, either: it is a set of instructions provided to a computer to help carry out tasks.

---

[3] For example, I create a feature called *i_frequency* that tracks the appearance of the first-person "I" throughout a text.

The machine learning algorithm I use in this project involves "binary classification," the separating of a data set into two distinct camps. I use Decision Trees in tandem with a Support Vector Machine classifier, as detailed in the second half of this paper. Two courses I took this fall, CS 4501 *(Machine Learning)* and DS 4001 *(Practice of Data Science*), provide me with guidance in this area. Indeed, I come to you painfully fluent in the tedious mathematical bulwark of vector norm optimization, information gain, etc. I assure you, then, that it is in good faith and conscience that I condone this algorithm ripping apart our culture's most celebrated novels.

Just as before, I complement this (particularly involved) strain of "data analytics" with a return to the traditional style of close reading we expect as scholars of literature. The classification model I use provides predictions as to what texts in my entire corpus are "lyrical"; to push back, I examine how my computer-generated predictions align with truths I tease out from a reading of my own. This again closes the gap between digital efforts and traditional exercises in critique and, hopefully, fosters transparency.

To cap this thesis, I consider the implications of affixing the "lyrical" label to what seems like every notable work of fiction possible. I sample Ludwig Wittgenstein's discussion of resemblances to elevate these closing considerations.

Corpus and data taxonomy

A selection of 26 English language works published between 1838 and 2006 comprise my corpus. The majority originate from the 20th century. Most of the corpus is novelistic fiction. However, the corpus also contains *Eureka: A Prose Poem* as well as *The Narrative of Arthur Gordon of Nantucket*, authored by Edgar Allan Poe, along with the short stories *In the Heart of the Heart of the Country* and *Billy Budd*, authored by William H. Gass and Herman Melville respectively. A few factors motivate a text's inclusion in my corpus: first, a "lyrical" labeling by critics, professors and even book jackets. For example, Arizona State professor Bert Bender's intriguing essay *Moby Dick, An American Lyrical Novel*, pushed *Moby Dick* into my corpus (Bender 346). Also, I received lists of commonly so-called "lyrical" fiction from UVa faculty, Paul Cantor and Pasanek. I started with a list of around 40 "lyrical" novels. I searched for each text in public repositories such as *Project Gutenberg* and *The Internet Archive*, which serve as the digital libraries for my project. These websites provide free copies of machine-readable texts. Not every novel on my initial list survived this process; I whittled down a group of 40 candidates to 26 based on the availability and quality of their digital texts.

I use late 19th and early 20th century English language detective fiction as a control set. I do not intend this thesis to wade into the heated debate surrounding genre-fiction – its literary merit, the efficacy of its tropes, etc. I simply chose this group of fiction as a control set because this group of literature is frequently used as a ground-truth in experiments by scholars such as Moretti. Genre fiction's rhetorical stability and predictable narrative patterns make it a popular focus in humanities surveys. Its featured in studies as diverse as Umberto Eco's classic *Narrative Structures in Flemming*, a

semiological analysis of the plot devices used in the James Bond novels, to 21st century efforts at the Stanford Literary Lab that contrast the word frequencies between the Bildungsroman and the Gothic (Eco; Heuser & Long). This control group allows my classifier to make accurate predictions. A complete list of the "lyrical" and detective novels I use is attached in the Appendix (Figures A.1 and A.2).

I consulted Matthew Jockers' seminal companion to digital text analysis, *Text Analysis with R for Students of Literature* throughout this project. He supplies an excellent blueprint for parsing digital texts. Jockers even provides a pithy pep talk to start the work: "Computational text analysis has a way of bringing into our field of view certain details and qualities of texts that we would miss with just the naked eye" (Jockers viii).

For my project I coded in R as well as the Python programming language (R Core Team; Python Software Foundation). Both are suitable for manipulating linguistic data. Jockers guides me through the transformation of digital blocks of texts into individual paragraphs, sentences and words; the syntactical and semantic units that scholars of literature are interested in. This process is known as "cleaning." Cleaning can be viewed as a labor of love – as well as an exercise in futility. The professor (eagerly awaiting accolades for an innovative digital project) and the research assistant (stuck with the grunt work of transforming byte after byte of raw data into machine readable texts) tend to fall on opposite sides of this spectrum. This project allowed me to spend time in both camps.

Each text presents promise and problems. Junk characters, page numbers and chapter headings are just a few of the surprises that might lurk inside a text, even when

sourced from a reputable repository like *Project Gutenberg*. A report from IBM notes, "80 percent of a data scientist's valuable time is spent simply finding, cleansing, and organizing data, leaving only 20 percent to actually perform analysis" (Gabernet and Limburn). This is a statement I fully endorse. While I'm hardly a data scientist, I spent over a month cleaning my corpus!

The work of manipulating a digital text does not end when you finish cleaning it, unfortunately. Next comes that dreadful "organizing" Gabernet and Limburn allude to. Indeed, you have to put all the information you extract somewhere. To this end, I channeled the work of two UVa faculty: Rafael Alvarado and Paul Humphreys. Their recent publication, "Big Data, Thick Mediation, and Representational Opacity," expounds the importance of proper data maintenance digital scholarship: "the *database* occupies the central, critical path through which all data must eventually flow… It is difficult to overestimate the significance of this role" (Alvarado & Humphreys 736, italic's mine). A "database" is, at its simplest level, a series of connected spreadsheets. Rows and columns that track entities in a digital system. If you have ever made a packing list for vacation that divided necessary items and quantities into helpful categories such as "Food," "Clothing" and "Electronics," you created a database. For this project, I followed Alvarado and Humphreys's "data lake" model (733). All relevant contents are stored in a single database table; all the data lives in one place. I used SQLite, an open-source database management system, for this project (Muller et al). SQLite integrates nicely with R and boasts impressive query speeds. A snappy first-mate!

I saved every paragraph, sentence and word from my corpus in this database.[4] This provided benefits. When running experiments with my corpus, using a database ensures that once a text is cleaned once, I do not have to do it again. The time I spent cleaning pays off. If I accidentally turned a working copy of *Pale Fire* into digital ash or deleted the entire text, I simply request a fresh copy from my database.[5]

After text cleaning, the database contained 4,689,007 words, 301,852 sentences and 99,188 paragraphs. Saturated by raw data, it came time for me to figure out what to do with it. This process is referred to as "data mining" and entails "finding patterns, correlations, or anomalies in large relational databases" (Abu-Mostafa et. al 15). Data mining felt a lot like going to Costco wholesale. You're engulfed by the familiar, yet everything feels distant when it appears so vast on the shelf. For instance, a cursory search for "tree" across my corpus reveals 655 usages. How could one possibly derive meaning from a repository so massive? Immanuel Kant's speculations surrounding the *mathematical sublime* – "imagination for the presentation of concepts of number," he intones in *The Critique of Judgement* – reverberate as you write code that alters over 5,000,000 rows of data (Kant §26).

This is exactly why I created a definitive plan to carry out that wonderful, Heideggerian "method" before I dove into this trough of data. The first step was interfacing with the scholarship of Virginia Jackson, which provides explanations and critique of the historical developments that embed modern understanding of the lyric. Her work shapes my understanding of on the lyric. Next, I expand my scope to the

---

[4] Indeed, I practice a painfully literal brand of *Deconstructionism*.
[5] All of the relevant code to parse my corpus and contribute to my database can be found in this project's *Github* repository: https://github.com/timschott/dmp.

three more scholars: Nietzsche, Moretti, and Freedman. Their ideas inform the calculation of salient features. I investigate how these numbers operate in my corpus through close readings. I follow this work with advanced machine learning analysis. Now, let us proceed to Jackson's conception of the lyric and interrogation of its legacy.

Jackson's lyricization

Jackson's scholarship on the lyric is highly controversial. However, her entry on the lyric in the *Princeton Encyclopedia of Poetics* brims with straightforward truths about the lyric's development over time. She begins, "In Western poetics, almost all poetry is now characterized as *lyric*, but this has not always been the case" (Jackson, *Lyric* 826, emphasis author's). This quote sets the stage for an exposition of lyric's history. Overall, in the context of this thesis, Jackson's discussion of lyric poetry serves as an appropriate point of departure for the investigation of the characteristics of "lyrical" novels.

In her piece, Jackson provides a historical development of the lyric as a style, genre, category and phenomenon. Jackson ascribes brevity, subjectivity, sensuality, passion, and the expression of personal feeling to the lyric. The canonization and acceptance of lyric as a genre did not take place in Ancient Greece. For instance, Jackson points out that "*lyrikos*" is not invoked by Aristotle in *Poetics*. As such, the poetry of masterclass verse-smiths at this time was never referred to as "lyrics" (826).

Jackson posits modern readers formulate their working sense of the lyric in the sonnet, not the poetics of the ancients. That is, in England, "lyric" as a term gains momentum when the sonnet rises to popularity in the 16th century. The sonnet comports with the lyric label because of formal qualities such as the dramatic shift of a well-crafted volta, its memorable final couplet and emphatic rigidity (827). Poets blend a "lyrical" presence into their sonnets to fashion the complex, intensely personal lyric their modern readers come to crave. Indeed, sonnets work to resolve personal complexities through their formal elements; the volta, for example, structures and resolves formal tension. Solving a word puzzle of sorts aligns with modern "lyrical" notions such as intense,

momentary bursts of self-expressive rhapsody. Importantly for Jackson, critics—not readers or poets—inject the word "lyric" into common discourse.

The 17th century finds the lyric's curriculum vitae growing apace. One important development in this century is the definitive conflation of poet and poem. Indeed, a crucial aspect of lyric poetry is the collapse between the speaker's subject and the self. Here we begin to see modern attitudes towards the lyric take shape, especially the gradual identification between the author and poem. Next, Jackson pokes holes in the prevailing critical sentiment that the lyric's popularity and practice wanes in the 18th century. The lyric sheds its anachronistic qualities—its (specious) connection to the ancients, its thirst for unheard music. It becomes difficult to conceive of the lyric at all. It continues to garner more robust (and variegated) aesthetic traits; century by century, the lyric loses any remaining sense of a singular, recognizable harmony.

The 19th century marks the beginning of the end for anyone seriously pursuing a well-bounded definition of the lyric. Johan Wolfgang van Goethe, for example, includes the lyric in his tripartite classification of the core elements of poetic genres: lyric, epic, dramatic. The grouping dissolves the lyric into mythos of poetry, through which "the entirety of literary possibility" mediates…somehow (832). This diverts the lyric from "an idea ... into an aesthetic ideal" that William Wordsworth, Samuel Taylor Coleridge and the other Romantics appropriate with unrivalled ability (833). This confusion ultimately leads to a complete collapse of the distinction between the lyric and poetry itself. Short poems now take on the name "lyric" regardless of their formal content or imagistic resonance. Jackson would have us believe the process is chaotic, and aesthetically

baseless. She is correct, to a certain extent. Of course, it is unsurprising that the merger of two concepts as nebulous as "poetry" and "lyric" did not resolve in absolute alignment.

Our modern age presents no great demystification for the lyric, either, as the term becomes co-opted by publishing houses and academics alike. The critical academy firmly orients itself towards the coupling of the lyric and poetry at large and renders the lyric "frozen by literary criticism" (835). Jackson's timeline documents the inflection points along that journey.[6]

Jackson's critical agenda cannot be disregarded when reading this study. Indeed, for Jackson, this timeline exhibits deliberate, observable instances of "lyricization." Jackson coins this term in her survey of Emily Dickinson's poetry and manuscripts, *Dickinson's Misery: A Theory of Lyric Reading*. This work makes the case that, "from the mid-nineteenth through the beginning of the twenty-first century, to be lyric is to be read as lyric—and to be read as lyric is to be printed and framed as a lyric" (Jackson, *Dickinson's Misery,* 6). Jackson takes issue with this orientation; through the force of lyricization, we consider disparate forms of poetry *as* lyric and implicate most all verse as apparently contextless, highly personal and expressive regardless of the nature of its form or content. Dickinson, the study's subject, exemplifies this phenomenon.

Clear signs of lyricization prevail, for example, in Dickinson's publication history. Astute *ENGL 3820* students recall Dickinson composes the bulk of her poetic output in luminous bursts. A napkin, a newspaper scrap and the back of an envelope are all likely places to uncover her sparkling verse. Problems arise when critics collect these

[6] This is not exactly unique to lyric poetry, as timelines of the novel often contain similar lamentations and speculations. Further reading in this space: Lukacs *The Theory of the Novel* and Watt's *The Rise of the Novel*.

scribblings or her "fascicles" (collections of Dickinson's ephemera and manuscripts), removing them from their original context and installing them collected editions of her work (Oberhaus). Jackson laments editors who have "actively cultivated a disregard for the circumstances of Dickinson's manuscripts circulation" (Jackson, *Dickinson's Misery*, 21). It is fictive, for instance, to wrangle 15 irregularly styled lines from their original context—the inside of a self-addressed envelope— and arrange them in an anthology. This dims Dickinson's lustrous glow. The process arbitrarily standardizes salient elements of Dickinson's oeuvre such as the size of individual words. It erases her penmanship and suffocates her poetic voice.

To clarify, Jackson does not use *Becoming Lyric* to react against the process of printing material that originates in manuscripts or other handwritten forms in general. This, of course, would be an untenable position to advocate, as the process of turning a manuscript into printed matter remains central to distributing literature to a widespread audience. Jackson, rather, detests the particular process of lyricization—of transforming organic, unfiltered verse into bonafide lyrics—because readers do not receive the poet's voice at face value. Jackson demonstrates that printings of Dickinson such as those put forward by nineteenth century editor-magnate Thomas Higginson fail to mention the intricacies and unique elements of the fascicles they are reproducing. He reduces her poetry to lyric because it is the simplest path forward. Critics like Higginson step between the reader and the poet's work. This produces an interpretation of verse *as* lyric regardless of its creative circumstances or formal features. This is lyricization: Jackson's call to arms. Her investigations lead me to believe a similar lyricization – this time, of the *novel* – infiltrates our understanding of fiction.

For example, consider one of the members of my corpus, *The Great Gatsby*. *Gatsby*, for all its insight into the hollow promises inherent in the American Dream, suffers from shoehorned similes and cheap repetition. Carraway's narration, too, leaves much to be desired. However, walk into any Barnes & Noble, pick up the ($18.00!) 2004 Scribner's printing, and bear witness to the synopsis on the back of the dust jacket: "For his sharp social insight and breathtaking *lyricism*, Fitzgerald is heralded as one of the most important American writers of the twentieth century" (Paulino, italics mine).[7] How reductive!

Gatsby is not a bad book, by any stretch of the imagination. But labeling large swaths of unrelated novels with an amorphous term poses problems. Indeed, "lyrical" is just one term at our disposal to categorize poetry (such as epigram, ode, elegy, etc.) yet it unduly distorts our attitudes about poetry (if you believe Jackson) and novels (if you believe me). Moreover, falling back on this "lyrical" label excludes works that do not enjoy the same amount of critical attention as canonical works. The same works that populate "Best Of" lists routinely find themselves distinguished as "lyrical"; *Gatsby* fittingly sits at #2 on the oft-cited "100 Best" 20th century novels list from *The Modern Library* (100 Best Novels). This exacerbates the divide between the remarkable and the overlooked in our culture's conception of authoritative literature. Overall, I believe that viewing the majority of memorable literature our culture consumes under this umbrella category of "lyrical" cheapens the value of the novels we seek to praise.

Jonathan Culler's *Theory of the Lyric* claims to stabilize the "generic status" of lyric poetry and contains a worthwhile response to Jackson. Culler provides insights into

[7] My second book jacket inspection. This decidedly "low-tech" analysis provides promising results and, personally, newfound angst towards the publishing industry.

the genre's history. He points out lapses in our modern interpretations and attitudes towards the lyric such as the overwhelming imposition to engage in hermeneutics. We fail to appreciate the *a priori* elucidation lyrics provide. For instance, Culler contrasts this suspicion with our attitude towards another popular artform, music. "We listen to songs without assuming that we should develop interpretations" (Culler 5).

Sweeping in scope, Culler dissects canonical lyrics, attitudes towards genres at large, the implications of our modern readings of lyrics: the account is exhaustive. Of particular interest to this thesis is Culler's response to lyricization. Culler contends Jackson conflates two distinct historical processes in her claim that reader of poems of all varieties transform them into lyrics. The first is "the process in the nineteenth century where the expressive lyric — lyric as the expression of the poet — becomes the norm" (86). Wordsworth encapsulates this phenomenon in his preface to *Lyrical Ballads*. "While [a Poet] describes and imitates passions, his situation is altogether slavish and mechanical, compared with the freedom and power of real and substantial action and suffering" (Wordsworth 104). The second process, according to Culler, entails the critical apparatus that developed during the mid 20th century. This group "takes the poem away from the historical author and treats it as the *speech* of a persona" (Culler 84). Culler pushes back against Jackson resting her claims at the nexus of these processes. Culler, here:

> To lump them together as "lyricization" seems historically irresponsible. It is the first that produces the historical reduction of the importance of various lyric subgenres in the nineteenth and early twentieth centuries and the second that establishes in the mid-twentieth century a distinctive mode of reading poems as the utterance of a fictional persona (85).

Regardless of which of these authorities on the lyric you subscribe to, the tension between two well-regarded scholars on the subject demonstrates a need for continued study in this space. For me, Jackson's theory is more persuasive than Culler's pushback. Jackson's careful research and specific examples of the process that morphs Dickinson's ephemera into an oeuvre is provocative, clever and thoughtful. She also includes plenty of engagement with the advent of the modern, expressive poet as well as the rise New Criticism. I don't read this as conflation, rather, an honest attempt to encapsulate what she views as a problematic trend negatively impacting our historical aesthetic recall. Above all, the work of these two scholars reminds us that pointing out the instability of a system—in this case, the lyric—as a response to a problematic, ill-contrived "consensus" can be done in more ways than one.

Moretti's *Distant Reading*

We must zoom out from Jackson's granular level of analysis to gather more information regarding the lyric. Franco Moretti, while not a scholar of this group of poetry, provides creative methodologies for studying my corpus of "lyrical" novels. Moretti coins the concept of "distant reading" in *Conjectures on World Literature* (from his anthology, *Distant Reading*) as a response to faithfully studying a massive body of world literature (Moretti, *Distant Reading,* 44). Moretti stakes out an unorthodox position when it comes to analyzing literature *en masse*. He claims that popular criticism, especially in the United States, spends an inordinate amount of time making judgment calls about an "extremely small canon" (48). Moretti balks at the idea that traditional, close reading alone can result in a comprehensive understanding of why certain works become canonized. The sheer size of these sorts of problem forces literary scholars to reckon with their research methods. Moretti, and a growing number of computer scientists, poets and statisticians alike, turn to computers in order to tackle this development

Moretti's medicine for these challenges is distance. He uses space, both physical and technological, to analyze literature. You would not be wrong to balk at importing "distance" into an analysis of the lyric, a mode associated with intimate personal expression. However, in a wonderful paradox, this closeness is not lost when we view the lyric from an elevated platform. The hunt for patterns across numerous lyrics at once points us towards astonishing revelations within individual lyrics. In the context of my work, this means that digitally analyzing a slew of "lyrical" novels prompts me to investigate specific paragraphs and sentences. Moreover, detectable linguistic structures

such as the variety of words used in a corpus can be organized into a "feature space" and serve as fodder for slick machine learning algorithms. And this is precisely what I practice in this paper. I calculate measures related to my corpus, and hold them up against the passages they allegedly describe. This allows me determine the utility of these statistics and gain a deep understanding of the repeated patterns within novels.

Importantly for Moretti, these discoveries could go unnoticed without the aid of distant reading. He invites us to "make a little pact with the devil" and deny our instincts for close reading (47). Moretti continues: "Distant reading: where distance, let me repeat it, *is a condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text." (48, emphasis author's).

I want to unpack this quote. Moretti's quote underscores the diverse outcomes possible through distant reading. The "smaller" and the "larger" components of literature relative to a text itself are devices on one side of the scale and meta-movements such as genres and cultural practices on the other. This analysis hones in on the "smaller" components such as repetition, alliteration and the use of imagery in a formalist fashion. Additionally, with this message, Moretti is not telling us to forget everything we learned in our undergraduate literature seminars: you can keep your highlighters and sticky notes, fellow bookworms, because distant reading and our traditional close reading instincts cooperate just fine. The practice of distant reading supplants—but cannot replace—literary acumen when evaluating a single text for hidden meanings and significance.

Moretti's contentions are not without their pitfalls. Critics of distant reading often point to the narrowness of the conclusions Moretti yields from this process. For instance, in *Planet Hollywood*, some of Moretti's findings don't square with the massive technical

undertaking needed produce them. In this essay, Moretti investigates the streamlining of popular movies across global markets. He assesses box office data for 46 countries (93). Moretti's process here does not involve "data analytics" in any sophisticated sense. He calculates simple measures such as the percentage of Hollywood-produced movies that land as box-office hits in other countries; approximately 85% in more than half of the countries that he looked at. The true significance of this finding, though, should not be overstated. Moretti merely sheds a light upon an obvious claim: Hollywood movies, due to their inflated budgets, world-renowned acting and production ensembles, enjoy most of the spoils of the global cinema landscape. [8] Not exactly the type of robust conclusion one would hope for after poring over movie data for markets as tiny as Slovakia.

Moretti bolsters the scope of his analysis with ideas from a range of disciplines. He leans on them to supplement his digital sleuthing. As Moretti puts it, "Evolution, geography, and formalism—the three approaches that would define my work for over a decade" (2).[9] Now, that formalist element inside Moretti's toolbox is a common point of criticism. Viktor Shklovsky registers a heavy influence on Moretti's Formalism. Both scholars believe that "the literariness or artfulness of a work of literature, that which makes it an aesthetic object, resides entirely in its devices, which should also form the sole object of literary studies" (*Johns Hopkins*).

Moretti holds no qualms obliterating the settled mediums that readers absorb literature through, such as the background and upbringing of an author and the social context surrounding a work. This can present frightening implications for readers and

---

[8] The Pareto Principle at work, it would seem (Moore).

[9] I would add computational linguistics to this list. For instance, Moretti has spent the past decade overseeing the Stanford Literary Lab, an incubator for natural language processing heavyweights such as friend-of-the-thesis Matthew Jockers.

critics who habitually connect plot events and textual content with exterior factors such as an author's biography. Indeed, this is another sticking point in the popular reception of *Distant Reading.* Moreover, in an interview with Moretti for the *Los Angeles Review of Books*, Melissa Dinsman probes Moretti on a range of criticisms of his work and DH as a whole. Dinsman forces him to reckon with the savior complex that critics of the field contend pollutes DH. She notes, "People often speak of digital work (and more frequently the digital humanities) as a means of making the humanities relevant for the 21st-century university" (Dinsman). I believe this phenomenon is unfortunately true. Faculty feel pressure to incorporate the newest technologies into their curricula but lack comprehensive training to properly integrate these new practices into their classrooms.[10] Plus, those versed in DH aren't always adept at sharing that knowledge with their "old-school" colleagues.

A closely related criticism I would tack onto this exchange is the high barrier to entry for DH work. It can be difficult to get started on the types of projects Moretti would view as revolutionary if you do not possess the proper funding or time. Unfortunately, there is seldom time or money to carry out ambitious digital projects. NEH, for example, funded barely 10% of applications to professors seeking "Mellon Fellowships for Digital Publication" (neh.gov).

In step with Moretti, this thesis interfaces with the unadulterated text of canonical novels. For some critics, lyricism lies within the author's upbringing and circumstances—not their prose. However, attempts to distill the subjective biographical experience of an author to a point on a graph are troubling trouble me. I don't pack the

[10] If UVa had funded *Fabrikit* perhaps this would not be the case.

tragedy of Virginia Woolf's death into a data frame. I leave the quantification of unspeakable horror to our feudal big-tech gatekeepers that fund the curriculum and dictate the professional applications of my Computer Science classes[11].

Moreover, although DH detractors claim that packing the syntactic traits of an author such as their prosody or lexicographic habits down into a matrix or vector is problematic, I disagree. I find parsing through a large corpus of works through digital techniques intuitive and effective.

Moretti's distant reading has its flaws, as previously mentioned, but the concept and Moretti's publication history serve as worthwhile frames of reference. Space spurs my inquiries, while looping back to close reading bolsters my claims. Let's pause here and return to literature. Moretti's ideas led me to create a host of data points for my corpus, such as average paragraph length, total number of exclamation points and number of commas per sentence. I'll describe a case where these numbers conflict with a novel's embedded structure.

Comparing *Pale Fire* to *Lolita* demonstrates how rudimentary calculations do not always properly explain the nuances of a text. *Lolita* contains roughly 24 more words per paragraph.[12] At first glance, that is. Kinbote's commentary in *Pale Fire* requires further inspection. His descriptions of Professor Shade and Zembla are far from laconic. For example, his addition to Line 172 (*books and people)* begins with an 81-word sentence.

> In a black pocketbook that I fortunately have with me I find, jotted down, here and there, among various extracts to please me (a footnote from Boswell's *Life of*

[11] I recall with horror this department-wide email promoting a Cybersecurity event: "In addition, you'll be able to network with recruiters and industry professionals from companies and organizations such as Raytheon, Capital One...and the NSA" (Smith). Could anyone pick a more ghoulish trio?

[12] Given the unique structure of Pale Fire, I did not analyze the titular poem or the index. I also omitted the introductory headers for commentary passages ("Line xxx + what follows) in my calculations. I counted each stanza of poetry as a single paragraph (such as the draft lines invoked on page 167).

> *Dr. Johnson*, the inscriptions on the trees in Wordsmith's famous avenue, a quotation from St. Augustine, and so on), a few samples of John Shade's conversation which I had collected in order to refer to them in the presence of people whom my friendship with the poet might interest or annoy (Nabokov, *Pale Fire*, 154-5).

The paragraph continues with a pair of large sentences. This trend holds throughout the novel. Pithy, one-line notes provide relief from Kinbote's aggrandizing overtures. The notes to Lines 213-214 (*a syllogism)* spectrally remark: "This may please a boy. Later in life we learn that we *are* these 'others'" (164, emphasis author's). The presence of both enormous paragraphs and compact counterparts injects noise in the words per paragraph measure.[13] It would be inconsistent, then, to claim *Pale Fire* presents a choppier, more streamlined reading experience than *Lolita* because on average its paragraphs are shorter. Likewise, celebrating a presence of lyricality in *Lolita* because Nabokov adheres to a more consistent syntactical fabric would be incorrect. *Pale Fire*, after all, exists only within the universe of its titular poem, and moreover is a novel packed with poetry inside of itself. The lemniscate in Shade's poem ("a unicursal bicircular quartic," according to Kinbote's dictionary) fits nicely into this discussion (136). Reflexivity—and some kind of infinity—fan this novel's flames. Fittingly, then, in this text packed with paradoxes and deception, the most poetic parts *aren't* the near-forgotten drafts of Shade's poem.

I must say I identify with Kinbote at this juncture of the project. We both pry apart the writing of literary greats. Hopefully, though, my remarks boast a marked lucidity relative to his. I'll now investigate Nietzsche's poetic scholarship in *The Birth of*

---

[13] "Of course there's noise, it's an average!" – Yes! Right you are. However, averages possess advantages for this type of calculation because they are easily interpretable and easier to calculate – no bloated temporary data structures per book. Medians are readily employed in this project, just not here.

Tragedy to gain further clarity on the lyric and cultivate more fodder for digital experiments.

Nietzsche's lyric

Nietzsche goes unmentioned in Jackson's survey. However, Culler highlights Nietzsche's fixation on rhythm, which is precisely the angle of Nietzsche's analysis I would like to explicate in this section. Indeed, Nietzsche's treatment of the lyric in *The Birth of Tragedy* strikes me because of the sonic angle of his analysis. He contends that lyric poetry is the 'imitative effulgence of music in images and concepts" (Nietzsche 34). The words of the lyric create the illusion of melody, part of the set of supposed features of the lyric. Moreover, the lyric poet constructs images in a reader's head through the use of rhythmic language just as the pure sound of a Bach composition or Steve Reich experiment evokes a wellspring of images. Nietzsche also recognizes that lyric poets operate in an individualistic fashion. "The lyric poet," according to Nietzsche, "always says 'I'" (34). This trait of the lyric is echoed across the critical landscape; its inclusion in *The Birth of Tragedy* is thus encouraging.

Furthermore, Nietzsche believes the lyric poet crafts images that stir passions and unsettle desires. Of course, because the images that lyricists manufacture "have no distinctive value," according to Nietzsche, their unique "lyrical" footprint can at first appear difficult to contain in a coherent framework (34). Given the wealth of critics in this space, it can be difficult to decide which scholarly voices to go to war with. In my view, the work of DH scholar Holst Katsma at Stanford's Literary Lab provides ample fodder for a digital examination of Nietzsche's hypothesis.

In Katsma's thesis-turned-book-chapter, *Loudness in the Novel,* the author performs quantative computational analysis on a corpus of 19th century texts to determine whether we can quantify the sound of novels. He claims, "the main revelation is the

discovery that loudness is perceivable and measurable within the novel...Written language codifies loudness; the word becomes its own type of gramophone-record" (Katsma 145). Katsma's scholarship breathes life into Nietzsche's (unsurprisingly obtuse) definition of the lyric. Tracking the loudness of a text is an important step to fleshing out the text's musical qualities; its inner melody; the very rhythm Nietzsche fixates upon. Moreover, Katsma's "gramophone-record" metaphor corroborates Nietzsche's view that the language of the lyric possesses a musical edge that strikingly separates the lyric from other modes of verse.

Katsma inspects the speaking verbs used in "tagged dialogue" across his corpus to fuel his work. Tagged dialogue, to clarify, is dialogue that explicitly includes a speaking verb (117). He breaks tagged dialogue into three categories: quiet, neutral and loud. Simple summations across a text—such as the proportion of loud to quiet dialogue—allow Katsma to construct a profile of the text's sound. These calculations lead to powerful conclusions such as, "Loudness appears to be acutely organized in the third volume of Pride and Prejudice" (133).

This effort echoes Moretti's *Distant Reading*; for the most part, Katsma's work is no more complex than combing addition and division with objective, detectable elements of works of literature. Katsma makes an important conclusion—that we can detect loudness in a novel—from an intuitive application of distant reading.

Katsma's analysis operates as a literary decibel meter. Unfortunately, implementing this type of scheme across my entire corpus would require immense computational lifting. Although unseen in the study, a significant amount of back-end computation and manpower is required to generate the three different levels of dialogue

for each text in Katsma's corpus. Plus, I don't enjoy a retinue of eager research assistants to crunch through the tedium of this type of task. Lastly, I wish to identify defining features in "lyrical" novels; not to create the next *Shazam* for novels. It would be inappropriate to spend significant amount of time recreating Katsma's work.

Instead, I adapt Katsma's work by calculating objective traces such as syllabic distribution, repetition and range of vocabulary. For instance, I use the *qdap* package in R to calculate the incidence of monosyllabic and polysyllabic words across my corpus (Rinker). I also calculate the frequency of anaphora across my texts. Lastly, I implement a schema to track the amount of dialogue utilized across my corpus. The homegrown nature of my dialogue tracking includes numerous special cases, if-statements, etc. on account of the manifold representations of quotations across my corpus. Some digital books use curly quotation marks, others vertical and some nothing at all. The difference between dialogue set off with " " versus " " does not alter the digital reading experience for humans, but this minor inconsistency is consequential for computers. I achieve fairly consistent and accurate measures of dialogue across my corpus – Joyce's em-dashes be damned[14].

Let's view some of the numbers I calculated because of Nietzsche. *The Sound and the Fury* is ranked in the bottom 5 across my corpus in the following three measures: comma usage per sentence (0.565), average number of syllables per word (1.230) and average polysyllables per word (0.0288). That last measure is more easily interpreted when viewing its inverse, the average number of monosyllabic words separating two

[14] The precise dialogue measurements for works by McCarthy and Gass are not readily calculatable. Their texts do not mark off dialogue with punctuation. To remedy this, I randomly sampled 10% of the pages in the offending works, counted the frequency of dialogue on each page by and imputed the rest.

polysyllabic words. This inverse is a staggering 35! To continue, most of works commas appear in exchanges of dialogue. Faulkner tags the majority of dialogue with "said"; its 1683 appearances comprise a sizeable 1.75% of the text. However, in practice, that comma usage per sentence metric is often lower than one comma per two sentences. This book offers more than a patchwork of conversations padded with prose segues, because characters routinely soliloquize with no regard for grammar. I want to inspect a passage that demonstrates how these measures operate during one such sequence.

The work's second chapter, "June 10, 1910, " details Quentin's caustic response to Caddy's sexual impurity. He obsesses over time in this chapter, clawing for a chance to travel backwards in time and prevent his sister's violation. Eventually, these thoughts drive Quentin to suicide. The prose detailing the ruminations preceding the end of his life lacks commas. Faulkner opts for receive enormous blocks of unstructured text that lack punctuation or capitalization. Faulkner dismisses the rules of grammar and toys with our expectations for closure and coherence. The multisyllabic words in this sequence link function as the necessary punctuation to accentuate Quentin's turmoil. Faulkner's syllables are sparing and unforgiving; he releases them together in groups and then pulls them back in. The following graph visualizes this strategy at work. It displays the syllabic profile for the opening portion of Quentin's closing sequence, a jumbled flashback to the conversation with his father concerning Caddy's sins.[15] The syllables for the first one hundred words comprise the top row, the second hundred words are in the second row and so on. On a strange linguistic note, it almost looks like Morse Code.

[15] Starting on page 195 with "and he we must just stay" and ending with "even time until it was" on page 197

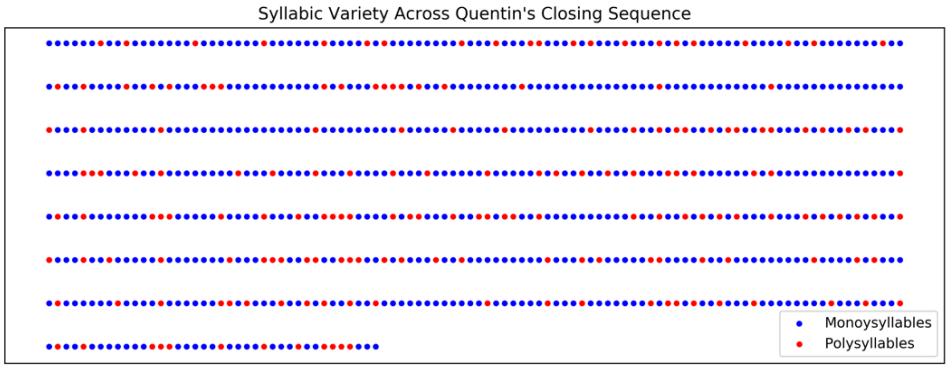

Figure 1: Syllabic Variety Across Quentin's Closing Sequence

Concerning Faulkner's prose, polysyllables travel together. On average, just four words separate polysyllabic words. Moreover, this 739-word passage has a total of 30 *consecutive* polysyllables.[16] Faulkner establishes a rhythm whose energy draws us inside Quentin's deteriorating mind. The speech patterns of Quentin and his father contrast. Quentin's words are exasperated. He repeats himself ("i wasnt lying i wasnt lying" and later "i was afraid to i was afraid") in an effort to reclaim his mind from trauma (195). Quentin attempts to take the blame for Caddy's wrongdoing by falsely claiming the two participated in an incestuous relationship that resulted in her pregnancy. His father, though, does not buy into the lie and instead attempts to impress sense into his son. Overall, Quentin's words are monosyllabic. The longest stretch of consecutive one-syllable words in this passage occurs as Quentin attempts to rationalize his dishonesty.

Quentin does not possess the elevated vocabulary of his father. Indeed, his father speaks in earnest platitudes. He quips, "you are still blind to what is in yourself" and later, "every man is the arbiter of his own virtues" (195, 197). Helpfully, this unique

[16] I calculated these numbers in Python to save you from counting dots.

mode of analyzing *The Sound and the Fury* wonderfully encapsulates Nietzsche's musically-inflected lyric.

For further explication of lyricality at work in the novel I will now probe Ralph Freedman's study, *The Lyrical Novel*. This will be the last time I bundle a critical survey with a digitally inflected close reading in this thesis. To follow, I take up advanced machine learning.

Ralph Freedman's "lyrical" novel

Freedman's explicit focus on lyricality in the novel sets his work apart from the wealth of writing on lyric poetry. To start *The Lyrical Novel*, he posits that the "lyrical" novel consists of "the paradoxical submersion of narrative in imagery and portraiture" (Freedman vii). A "lyrical" novelist blends modes of art such as poetry and painting into a final product that sublimates the experience of its characters into the world around them. The images "reach a specific intensity" and culminate in the realization of the novelist's vision (7). Self-reflexivity is another quality of the "lyrical" novel. Characters in "lyrical" novels possess a strain of passivity that generates "images" from the objects that surround them (9). "Lyrical" novels are thus generative fictions that render images and sensations with ease. Freedman is also aware of the conflict between language and narrative in "lyrical" novels. He posits that "lyrical" novelists take advantage of the expectation for narrative through a repurposing "as the object of [the novelist's] deformations" (10). Freedman is hard to grasp here.[17] However, his core contention concerning plot is "lyrical" novels promote verisimilitude through creative usages of narratives. These works deviate from expectations through the usage of poetic devices and tilted points of view. Logically, then, Freedman wraps techniques such as stream of consciousness into this contention; for instance, he states proper stream of consciousness generates "a design of images and motifs emerges from associations of the mind" (11).

[17] Not only is Freedman hard to understand, but the vagueness conflicts with my project's goal of remaining explicit. John Unsworth, UVa Dean of Libraries, channels a publication on knowledge representation to focalize the "ontological commitments" of DH: "The commitments are in effect a strong pair of glasses that determine what we can see, bringing some part of the world into sharp focus, at the expense of blurring other parts" (Unsworth).

In light of Freedman's framework, I believe an exemplary moment of lyricality in fiction is Lily Briscoe's triumph at the end of *To the Lighthouse*:

> With a sudden intensity, as if she saw it clear for a second, she drew a line there, in the centre. It was done; it was finished. Yes, she thought, laying down her brush in extreme fatigue, I have had my vision (Woolf 154).

Woolf cleverly manipulates syntax in the work's final sentence. Woolf's prose runs over itself as the sequence ends. Her writing even mirrors Lily's fatigue through the use of the past perfect tense. Woolf deftly employs simplistic language to describe the brush stroke, rendering it before the mind's eye. Woolf's unceremonious conjuring of imagery echoes Freedman's principal of "paradoxical submersion" (Freedman vii). Freedman in fact praises Woolf on account of "the ruminations of her individual characters" whose voices culminate in a "monologue spoken in the language and couched in the imagery of the omniscient author" (13). Nothing, then, escapes the magnetic energy of lyricism that pulses its way through the work's characters and scenes.

Helpfully, Freedman explicates what a "lyrical" novel is not, too. He claims that "lyrical" novels do not set themselves apart through a particular "poetic style" and "purple prose" (1). Freedman in fact believes that "lyrical" novels preclude mechanical decadence — that there is no perfect formula for composing a "lyrical" novel. Lyrical novels come with a few other caveats. These works, recall, spurn narrative in favor of brooding, meditative journeys through the minds of a mosaic of complex characters. When considering the specific points of view of the characters, Freedman coins the "lyrical point of view," in which "such a world is conceived, not as a universe in which men display their actions, but as a poet's vision fashioned as a design" (7). Freedman imports this concept from lyric poetry. Freedman aligns with scholarly

tradition in his view that the poet's "I" (or, the "lyrical" self) dominates lyric poetry. This creates an environment where narrator refines the author's voice and views.

There is no best practice for applying his contentions. As such, Freedman's proclamations appear baseless at times. Joyce is put onto Freedman's list of "lyrical" experimenters gone awry. Freedman pins *Ulysses* as a "lyrical fiction rather than being a lyrical novel itself" (12). He approves of the collapse of the work's trio of narrators—Bloom, Stephen and Mollie—into a divided, "triple lyrical self" with a divided point of view and enjoy a shared moment of recognition at the novels end (12, 13). However, a "lyrical" novel, according to Freedman, cannot be reliant on underlying patterns of logic in the way that *Ulysses* borrows the Homeric scheme. Plus, Joyce's unsettling, unsteady style of prose (e.g. his ribald mockeries of the canon) disrupt the somewhat placid qualities Freedman ascribes to "lyrical" novels. For this reason, Freedman claims *Ulysses* is pulled in too many directions to exist as coherently "lyrical" (13).

From a formal level, Freedman values imagery over the sonically rich and elliptical prose I discussed alongside my reading of *The Birth of Tragedy.* He is clearly suspicious of attempts to delimit "lyrical" novels through the recognition of florid and excessively ornate prose. Freedman's "lyrical" attitude would thus appear unlikely to translate into a set of formal features. This poses problems for the digital humanist; I'm approaching these "lyrical" texts with the intent of funneling their words and measurable qualities into the rows of my feature matrix. So, Freedman's contentions complicate Moretti's advocacy for the transformation of the novels into observable pieces of linguistic data. On the other hand, Freedman's refusal to integrate formalism into his study would appease Nietzsche, whose relationship to historicism rarely skews positive.

Additionally, Freedman's claim that "purple prose" does not ultimately prop up "lyrical" novels is incongruous with my work. Instead, I conjecture that distinguishable vocabulary and attending markers (such as syntactical structures and forms) exist within these imagistic narratives. As such, Freedman's fixation on imagery sets me off in pursuit of computationally-rigorous methods of unearthing imagery. Recent scholarship supports me in this quest.

In their paper, *A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry,* Stanford faculty Justine Kao and Dan Jurafsky utilize rigorous computational methods to distinguish between amateur poetry and the verse of award-winning wordsmiths. Their treatment of imagery aligns with Freedman's profile of "lyrical" novels. In order to analyze the imagery of their corpus the authors used dictionary-based analysis. Dictionary-based analysis simply means counting the occurrences of a special group of words in your text. Kao and Jurafsky sourced a dictionary of words that signal imagery from the Harvard General Inquirer (HGI). The HGI was developed in the 1960s in *General Inquirer: A Computer Approach to Content Analysis* and digitized in the 1990s (Stone et al). The HGI standardizes thematic analyses across large bodies of text:

> The General Inquirer is basically a mapping tool. It maps each text file with counts on dictionary-supplied categories. … Each category is a list of words and word senses. A category such as "self references" may contain only a dozen entries, mostly pronouns (General Inquirer Usage).

The authors predict that adept poets rely less upon abstract generalizations and more on concrete invocations of easily imaginable objects.[18] Their study culminates in fascinating conclusions. Their final model relies on a total of 8 unique features including,

[18] Virginia Jackson is vindicated at this juncture as this assumption brims with lyricization.

significantly, the three imagistic categories sourced from the HGI. In other words, their model (a simple logistic regression) gained enough information from the frequency of these types of words to include it in its decision-making process.

I incorporate the work of Kao and Jurafsky in my project. There are five relevant categories from the HGI I test on my corpus: *Perceptual* words, *Objects*, *Abstract* vocabulary, words indicative of *Time* consciousness, and signifiers of *Relationships*. These five categories provide me with a set of markers to evaluate Kao and Jurafsky's conclusion, "poems written by professional poets contain significantly more words that reference objects and significantly less words about abstract concepts and generalizations." (Kao and Jurafsky 8). The possibility of applying their work to fiction is fascinating.

To summarize, the HGI dictionaries provide me with a computationally sound method of tracking the usage of imagery. To carry this out, I keep track of the frequencies of each HGI word in my five categories across my corpus.

As I blaze a path towards advanced machine learning the complexity of my calculations increases. My experiments brim with promise and complement my analysis of the "lyrical" scholarship of Moretti, Nietzsche and others. My programming work culminates in a detailed evaluation of my corpus through machine learning as well as requisite close readings at each step in the process. Here's my last experiment in feature engineering before I move onto machine learning.

To pushback against Freedman's dismissal of "purple prose" I implemented a robust measure of lexical variety, Moving-Average Type-Token Ratio (MATTR). Type-Token Ratio (TTR) measures the diversity of diction. TTR is the number of unique words

used in a text divided by the total number of words[19]. Unfortunately, there are issues with applying this ratio to a large body of texts. As a piece of writing unfolds, the author naturally introduces new words into the mix. Covington and McFall write, "TTR is not a good measure of lexical diversity because it is always lower with longer documents" (Covington and McFall). Conventional TTR would point to "richer" vocabulary usage in "lyrical" texts, which contain approximately 107,000 words per book versus the approximate 80,000 words per book found in my detective set. To fix this problem, I calculated a *MATTR* measure for my corpus. A moving average serves as a "widely used indicator in technical analysis that filter[s] out the "noise" from random ... fluctuations" (Hayes). For my project, this allowed me to measure the lexical variety of a text in small bursts at a time.[20]

It is perhaps no surprise that *Gravity's Rainbow* trumps the rest of the field with a median MATTR of 52.44. Fascinatingly, *The Sound and the Fury* has a much lower median MATTR of 44.56. The inspection of this single metric in turn materializes a crucial difference between *The Sound and the Fury* and Pynchon's doorstopper. Faulkner exposes the shallow limits of the human capacity for empathy. Faulkner's repetition lulls us into a gradual understanding of the mixed consciousness' of his characters; think back to Quentin's frantic, repetitious stream of consciousness. Pynchon, meanwhile, utilizes his capacious vocabulary to push the limits of fiction, one equation, diatribe or fever dream at a time.

---

[19] For example, consider the sentence, "The dog ran very, very fast." The TTR comes out to 0.8 – 4 unique words out of 5 total.

[20] Specifically, I calculated the median MATTR across a text using a window size of 64 words (the average length of a paragraph in my corpus).

I made graphs to visualize this ratio across my corpus. Here is *Mrs. Dalloway*, below, through the lens of its unique word usage. You will note how Woolf gradually modulates the amount of unique words as the novel completes. As the character's slowly lose their grip on reality, Woolf increases her repetition. With this, she motions towards the failure of expression and utter silence that results from extreme mental trauma. A cascade of unique words in the work's closing sequence could muffle the muted agony. Instead, she presses us with the familiar to demonstrate the depersonalized and stunted aspects of the work's characters.

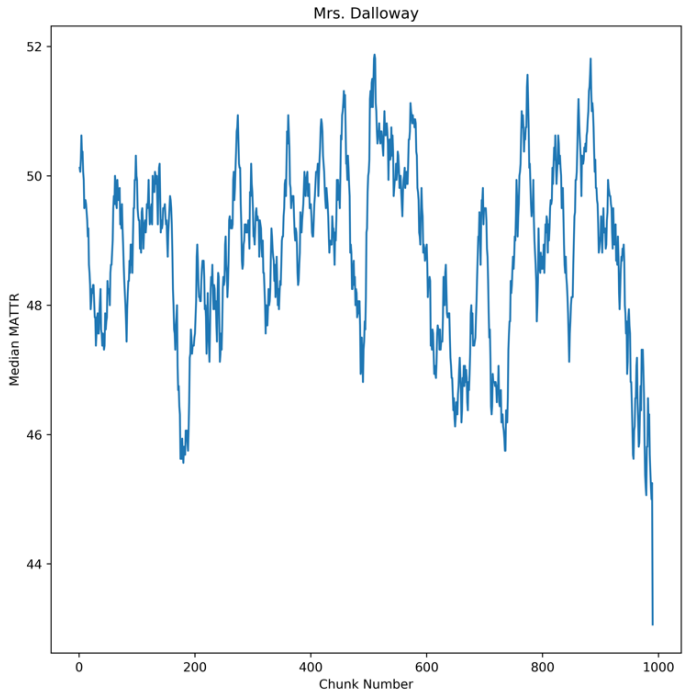

Figure 2: Median MATTR, Mrs. Dalloway

Machine Learning: Methods and Results

With my circuit of iterative, computationally informed analyses and close readings completed, my next and final tactic for attacking the "lyrical" novel is the implementation of machine learning processes. As I explained in the introduction to this paper, machine learning can remedy the opacity of spreadsheets through the exposure of trends and patterns that might be otherwise overlooked. It is the most sophisticated technique I use to explore "lyrical" novels. Thus, I saved it for last. Now, you'll also recall that I took our Computer Science department's machine learning course. I'm acutely aware of the technical complexity that travels alongside this field. Importantly, I also know what it is like to feel lost in these discussions. I aim for clarity, here, for my readership's sake.

Concepts from machine learning address one of the core issues in the way humans learn: as a species, we bore easily.[21] MacArthur et. al, in a robust qualitative analysis of acoustical poetry, discuss the problem of recognizing patterns across a diverse sample set as follows: "Too little unpredictability bores us. Too much confuses us and isn't rewarding— humans, instead, attend to the reasonably unpredictable" (MacArthur et. al).

This gravitation towards what is "reasonably unpredictable" can feel strained when mired inside of a database containing millions of words. Machine learning, while not a panacea, steadies our steps as we journey from data to knowledge. One of the field's founders, Thomas Dietterich, describes the goal of a machine learning system as the creation of "computer systems that can adapt and learn from their experience" (Qi 1.32).

---

[21] --this thesis notwithstanding!

Adaptability is key. These systems ought to be flexible in order to understand new, unseen data points down the road. The amount that these systems can truly "learn" is highly contested and outside the scope of this thesis. I do not claim that the systems I use independently outstrip the knowledge of humanists familiar with my corpus. Rather, the most striking advantage of approaching humanities learning with machine learning in hand is the sheer *capacity* of these systems to parse, evaluate and judge. "What digital humanities in general — and computationally assisted literary studies in particular — offer is a new set of methods for dealing with such abundance," writes Matthew Wilkens in an overview of recent trends in DH (Wilkens 11). A potent antidote to Kant's reckoning with the infinite, it would seem.

The field of machine learning is undoubtedly headed towards a promising future. One of the most interesting areas of further study that could be applied to my corpus is the application of neural networks. "Long-Short Term Memory" neural networks mimic the forces of human cognition and repeatedly simulate the act of reading thousands of books.[22] They form the backbone of services such as *Google Translate* (Korbus). I tried implementing this type of neural network in hopes of generating a list of words most closely aligned with "lyrical" novels. After 400 lines of code and numerous (unanswered) forum posts, I decided I was in over my head.[23]

Alternatively, I carried out a much more intuitive process known as "binary classification." This process trains a model that separates each book into its proper "lyrical" or detective camp. But before I performed this type of classification, I needed to

---

[22] For further reading on these models, I suggest Ilya Sutskever's Doctoral Thesis, *Training Recurrent Neural Networks* (Sutskever).

[23] An example of one of my cries for help: https://github.com/keras-team/keras/issues/4962#issuecomment-475036588

distill my feature matrix down to its most relevant columns. Now, there is something beguiling about an enormous feature matrix; self-contained, unassailable—why not use the whole thing? The phenomenon known as "the curse of dimensionality" instructs us otherwise. This hex arises when the number of columns in a feature matrix is large relative to the number of rows (Mitchell 170). In my case, I have 50 rows and 30 columns. I would need *thousands* of books in my corpus if I were to tackle the curse of dimensionality head on and use my entire feature set. Using every column in a small feature matrix saturates any type of algorithm and creates skewed results. Moreover, the highly correlated nature of my feature matrix exacerbates these negative outcomes.

There exist numerous, well-documented approaches to reducing dimensionality. I elected to use Variable Importance via Decision Tree (DT) learning.[24] DT's are staples of statistical learning. DT's classify data based on a series of one-variable decisions (Qi 18.9). They feature in fields outside of statistics because they mimic the decision-making process of humans. For example, you might go through a series of internal checks before ordering food at a restaurant. First you decide whether you're in the mood for Italian or Thai, next you consider how much money you're willing to spend. You could represent this disjunction of conjunctions with a tree structure; first creating a slot to decide what type of cuisine you want and below it how much money you budget. Every possible outcome resides on a "leaf node."

All DT's are predicated upon the principle of "Information Gain." In simplest terms, this means that the first decision represented in the tree needs to be the most consequential (Qi 18.20). In terms of my data, it would be foolish, for example, to

[24] Other suitable strategies include: LASSO, PCA, Ridge Regression and Stepwise Feature Selection.

place “Median MATTR” at the top of my tree since its median is 49 and standard deviation is just 1.56; as such, this statistic does not vary all that much across my corpus.

Numerous measures exist that record the importance of the columns in a feature matrix. I utilized the Gini coefficient because it is compatible with DT’s. Another advantage of using DTs is the ability to stack numerous DT’s together to form a “Random Forest.” This method polls a committee of randomly generated, decorrelated trees and uses the majority vote to classify an inputted data point (Qi 18.53). This creates a robust system that cuts through the noise in data and recognizes its most salient features. Let’s take a look at how the Random Forest model ranked my features.

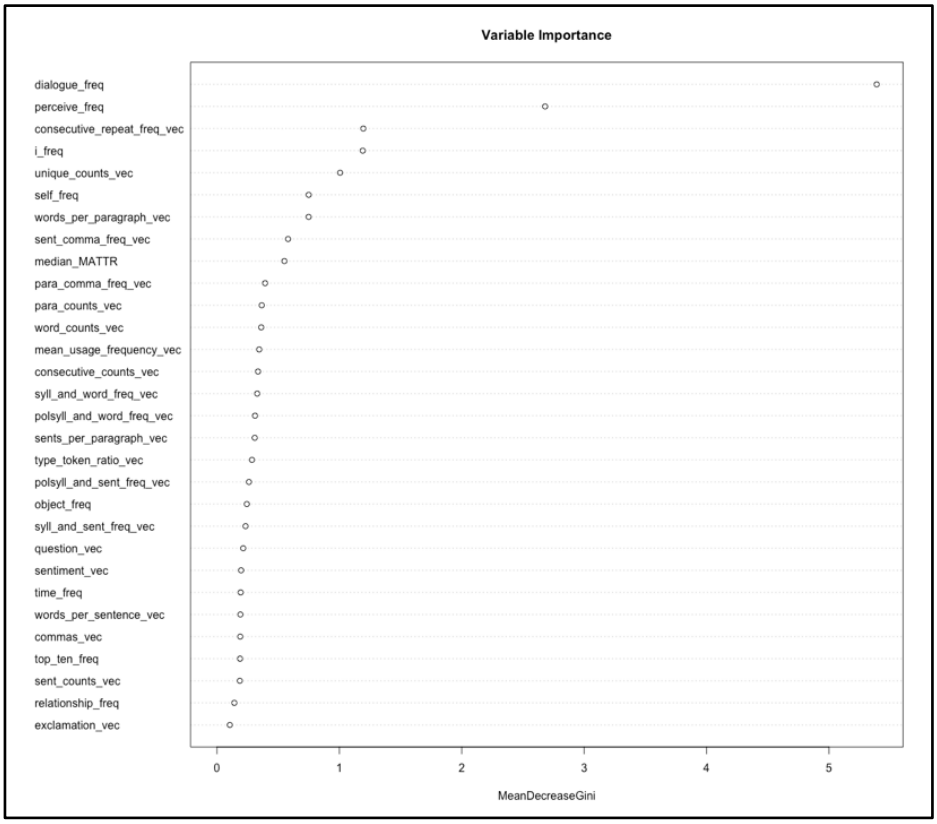

Figure 3: Variable Importance (per Feature)

As you can see, the ensemble model selected dialogue frequency as the most elucidating feature in my dataset. This is fascinating when held up against the scholarship surrounding sound and voices in the novel.[25] The next three most-important features selected by the model were: frequency of perception words, frequency of anaphora and frequency of "I." These results vindicated a winter spent calculating features as I hoped to unearth underlying truths about "lyrical" novels. For example, the frequency of "I" is of paramount interest in the context of my research. Critics from myriad backgrounds – Nietzsche, panglossian philologist, Culler and Freedman, renegade academics – find common cause as they invoke the "lyrical-self." That my ensemble model highlighted this feature as consequential corroborates the views of critical authorities in this space.

Four features across a corpus of fifty books still presented a packed crowd. I needed to eliminate more. The hefty weight attached to dialogue, over twice that of the second-place feature, felt suspicious. To observe interactions between these four features I carried out a few classification experiments, again using Random Forests. I created six models, each using a different pair of these four features.[26] These models each learned the difference between "lyrical" and detective novels from the levels of two features. The three models that used dialogue greatly outstripped the three that did not. In the most extreme case, the classifier that used perception and dialogue achieved a classification accuracy of 100%.[27] This set off alarm bells. "100% accuracy" is problematic in every discipline.

---

[25] For instance, Bakhtin's "voices" in his readings of Dostoevsky (Bakhtin 3).
[26] Professor Ma, if you're reading this, that's 4 choose 2!
[27] The specific accuracy measure invoked here is "Out-of-Bag" error.

This result means the model correctly placed every single book in my corpus into its true class. This signals a known problem in the machine learning domain known as overfitting, "an analysis which corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably" (OED). I am not entirely certain why dialogue leads to a faulty model. Errors in the calculation are the most likely suspect. As previously stated, my code to generate the "dialogue frequency" statistic is messy. This led me to drop dialogue and hand off the remaining three features to my binary classifier. Interrogating outcomes should not to be overlooked during ostensibly rigorous studies such as this thesis. To summarize, dialogue was not interacting well with my other features so I eliminated it from this advanced analysis.

Nixing dialogue left me with three features to inspect using Support Vector Machines (SVM). SVM was first introduced in 1992 and rose to popularity through its success in handwriting recognition. (Qi 11.1-2). SVM extends the paradigmatic linear model to find the "maximum margin" that separate two classes in high dimensional vector space. It is analogous to performing a logistic regression that outputs a class identity rather than probabilities.

For my project, I implemented an SVM using the *e1071* package in *R* (Technische Universität Wien). I built a final SVM using the two most extreme variables from an initial trial run. The levels *anaphora-frequency* and *perception-frequency* across my corpus were placed into this model. I used "Leave-One-Out-Cross-Validation" (LOOCV) and a cost parameter of 2.5 to the tune of 84% accuracy. Here is a brief technical explanation of my validation technique, for the sake of reproducibility. In data analysis, it

is common to separate your dataset into a training and testing set. Your model learns information about your training set that allows it to, ideally, make the correct decision about the testing data you set aside. Cross-Validation enhances this process.

With Cross-Validation, you repeatedly train your model with a subset of your data (Qi 5.59). You split your data into *k* "folds." For example, with *k* = 10-fold cross validation, divide your data as follows: 9 chunks for training, and set aside a 10th chunk for testing. Repeat this process 10 times, each iteration setting a different chunk as your test data. Total error is cumulative error for all trials divided by *k*. This process is beneficial because it allows you to build models that learn from more of your data. LOOCV is special case of Cross-Validation. With LOOCV, you set *k* equal to the total number of observations in your dataset.[28] This maximizes the amount of data you train your model with. You end up validating with a single data point at a time: the observation that is "left out."

In 42 out of 50 trials, the test novel was placed in its true class. This is where that 84% accuracy comes from. According to my classifier, novels with high levels of anaphora and low levels of perception belong in the "lyrical" class. The opposite is true for detective fiction. This is fascinating. Aligning "lyrical" novels with this type of parallelism buttresses critical arguments that contend this scheme of lyric poetry seeps into "lyrical" novels. Plus, my control set's allegiance with high levels of perception words readily demonstrates detective novelist's reliance on words connoting ambiguity and logical problem solving.[29] These words, such as "appearance," "glimpse" and "hidden" align nicely with the themes and motifs of detective fiction.

---

[28] I set *k* = 50, then, to perform LOOCV.

[29] The complete HGI "perception" list is available at: http://www.wjh.harvard.edu/~inquirer/Perceiv.html

Not every work fits this mold though. The following table contains works my model incorrectly classified. It displays the two data points the models considered when classifying them. I included the "z-score" for these data points, which measures how many standard deviations away from the mean an observation is. Any work with z-score with an absolute value near or above 1.0 can be viewed as particularly deviant from its peers. This measure is commonly used in statistics to capture the amount of distance between a sample observation and its respective mean.

	Title	Author	Classified Label	Correct Label	Frequency of Anaphora	z-score	Frequency of Perception Words	z-score
1	Billy Budd	Herman Meville	detective	lyrical	0.03518	-1.69	0.013597	0.69
2	Eureka: A Prose Poem	Edgar Allan Poe	detective	lyrical	0.041651	-1.44	0.011179	-0.86
3	Heart of Darkness	Joseph Conrad	detective	lyrical	0.076887	-0.05	0.012584	0.04
4	The Sound and the Fury	William Faulkner	detective	lyrical	0.093796	0.61	0.01236	-0.1
5	Wide Sargasso Sea	Jean Rhys	detective	lyrical	0.069218	-0.35	0.013172	0.42
6	The Big Sleep	Raymond Chandler	lyrical	detective	0.103745	2.68	0.013809	-0.88
7	The Circular Staircase	Mary Roberts Rinehart	lyrical	detective	0.064207	0.44	0.013025	-1.33
8	The Mystery of Room 75	Fred Merrick White	lyrical	detective	0.073456	0.96	0.013804	-0.88

Figure 4: Incorrectly Classified novels

A few elements stand out from this table. To start, every "lyrical" work except for *The Sound and the Fury* is quite short; they are all less than 50,000 words. Importantly, Culler and Jackson both contend that brevity is associated with lyricality. This squares with a central quality of lyric poems: they're short. Most of Wordsworth's lyrics are miniscule. *My Heart Leaps Up*, for example, addresses the complexities of childhood in a tidy nine lines. Culler's invocation of music in *Theories of Lyric* brings David Bowie's song "Eight Line Poem" to mind. Bowie emphasizes the compact nature of his lyric to signal its poetic slant. It is remarkable, then, that my classifier incorrectly assessed 4 of the 5 shortest "lyrical" works in my corpus.

At first, I suspected that the length of the works was unfairly influencing my classifier – there had to be an underlying reason for these results. Upon further investigation, my suspicions are quelled, but not totally answered. For example,

Melville's *Billy Budd* and *Moby Dick*, which was classified correctly, vastly differ in length and scope. However, they share strikingly similar levels of anaphora; 3.51% of the sentences in Billy Budd feature anaphora, which is staggeringly close to *Moby Dick*'s measure of 3.49%. This demonstrates how authors consistently employ this device across different works in spite of the number of total words they output. It is difficult to write your way out of your own style.

Moving on, let's examine the z-scores for these observations, as these provide a quick measure of how much the books deviate from their group mean. *Billy Budd*, with its large negative z-score, moves into detective territory because of Melville's hesitance to employ anaphora. Raymond Chandler's consistent anaphora places him in the 97th percentile of control novels analyzed in this study. This pushes *The Big Sleep* into the "lyrical" space. Chandler's extraordinary reliance on anaphora cannot be overlooked on even a cursory inspection of this novel. Here he is setting the stage for Marlowe to enter Geiger's house:

> There were low bookshelves, there was a thick pinkish Chinese rug in which a gopher could have spent a week without showing his nose above the nap. There were floor cushions, bits of odd silk tossed around, as if whoever lived there had to have a piece he could reach out and thumb. There was a broad low divan of old rose tapestry. It had a wad of clothes on it, including lilac-colored silk underwear. There was a big carved lamp on a pedestal, two other standing lamps with jade-green shades and long tassels. There was a black desk with carved gargoyles at the corners and behind it a yellow satin cushion on a polished black chair with carved arms and back (Chandler 17).

This prose is remarkable. Almost every single sentence in this sequence begins with "there." Moreover, Chandler adheres to this structure on a clause by clause basis, inserting "there" at every juncture he can. The sentences share almost identical structures. In place of a variable syntax, Chandler utilizes descriptive vocabulary, focusing on colors

and objects, to present the scene to readers. This phenomenon holds throughout the entire novel. It's no wonder, then, that my classifier grouped *The Big Sleep* with "lyrical" novels, as this rhetorical technique is closely allied with my set of interest. That being said, I do not descry anything particularly "lyrical" in this style of prose. Chandler never allows Marlowe to fully supplant the author's voice and take on the type of "lyrical self" lauded by Nietzsche and company. The anaphora used here, while effective in the context of a hardboiled crime novel, cannot reasonably justify an elevation of *The Big Sleep* to the rarified "lyrical" realm.

Not every "misclassification" is as straightforward. My model incorrectly labeled *Heart of Darkness* as outside of the "lyrical" class. However, its levels of anaphora and frequency of perception are consistent with the rest of the "lyrical" works I analyzed. This is an example of the SVM model making a plain mistake. The results of this type of algorithm are never perfect. Overall, though, through the use of feature selection and SVM classification I successfully separated the bulk my data with a high amount of accuracy. My classification algorithms wonderfully evinced trends within my corpus. I am thrilled that the frequency of anaphora, a frequency I engineered out of sheer curiosity, substantially separated "lyrical" novels from their control set.

A Final Close Reading

Up to this point, I have close read passages to determine whether or not my statistics and models are consistent with the textual reality of my corpus. Some of the readings endorse the results of my computational work. The closing of *To the Lighthouse* and Quentin's breathless stream of consciousness sequence from *The Sound and the Fury* fall on this side of the divide. Conversely, my inspections of *Pale Fire* and Chandler's ostensible "lyricality" in *The Big Sleep* both demonstrate examples of programming and statistics alone not adequately describing novels in my corpus. However, not every close read in a digital project should come at the behest of a computer system; distant reading can sometimes miss the mark. Plus, debugging code and scrutinizing spreadsheets gets old.[30] With this in mind I want to pay it forward to Melville, an author I have always considered "lyrical," and inspect a chapter from *Moby Dick*, "Stubb Kills a Whale." Again, no models or spreadsheets this time. I want to revel in Melville's prose without mediating his words through a machine.[31]

For example, notice the lack of punctuation as Ishmael describes Stubb's attack on the whale:

> And lo! close under our lee, not forty fathoms off, a gigantic Sperm Whale lay rolling in the water like the capsized hull of a frigate, his broad, glossy back, of an Ethiopian hue, glistening in the sun's rays like a mirror. But lazily undulating in the trough of the sea, and ever and anon tranquilly spouting his vapory jet, the whale looked like a portly burgher smoking his pipe of a warm afternoon (Melville 220).

[30] You cannot *really* bond with a novel until you open its .txt file an inch away from your eyes and squint to tell whether its dialogue is marked off with " or ". These relationships take a toll.
[31] Not that there's anything wrong with mediating words through machines!

Melville stretches this passage with multisyllabic words and scant punctuation. This allots time and space for the massive whale to captivate us. Moreover, in this passage, his syntactical choices synchronize with his diction. Lengthy, multisyllabic bigrams like “lazily undulating” decelerate the whale and the very act of reading about its movements, too. As the chapter unfolds, Melville overloads sentences with commas to percussively interrupt the capturing of the whale:

> A continual cascade played at the bows; a ceaseless whirling eddy in her wake; and, at the slightest motion from within, even but of a little finger, the vibrating, cracking craft canted over her spasmodic gunwale into the sea (222).

Commas heighten the suspense of the passage. We are left gasping after successive, short clauses, still unsure of the ultimate resolution of the situation. Melville’s punctuation squares nicely with the torrent of water that whips the men aboard into a frenzy. Plus, the alliterative bigrams “continual cascade” and “cracking craft” reinforce Melville’s depiction of delirium. Clauses collide as if we too are subjected to the whim of the roaring sea. Within the space of a few paragraphs, Melville transports us from languor to furor. This remarkable malleability makes *Moby Dick* an effective vehicle for discussing the complex aspects of human relationships; aboard The Pequod, Melville addresses race, sexuality, colonialism, among other challenging topics. This dimension of the novel comes to light in this chapter through the competing treatments of Queequeg versus Starbuck and Stubb.

The rare sighting of a squid precedes this chapter. In nautical circles, squid are viewed as ominous and fateful. Few ships safely return to harbor after a squid spotting. As such, the squid rattles the mates ... most of them, at least. Starbuck views “the apparition of the Squid [as] a thing of portents” (219). Melville makes a point to contrast

this connection of the squid with a remark from Queequeg: "When you see him 'quid," said the savage, honing his harpoon in the bow of his hoisted boat, "then you quick see him 'parm whale" (219). The nuance and lore that dominate nautical voyages are never explained in terms familiar to Queequeg. As a "savage" he exists solely to carry out physical labor and risk his life for the fulfilment of Ahab's twisted desires. Melville floods this line, and the novel as a whole, with the "qu-" sound. Queequeq's alliterative name thus has an affinity with the name of the ship he's chained to, and we hear that sound surface in this line as well. *Moby Dick's* mellifluous prose tracks Nietzsche and Katsma's emphasis on the lyric's sonic dimension.

Like the harpoon that never leaves his side, Queequeg is a tool. He does not enjoy the cultural capital necessary to understand the significance of a squid.[32] The white members of the crew shut him out from discussions of strategy. As the chapter continues, Queequeg's reactions read as caricatures. During the assault on the sperm whale, Queequeg's ravings are so vociferous it is "as if [Queequeq was] smacking his lips over a mouthful of Grenadier's steak" (221). Melville wades into stereotypes here. He focalizes the primal energy of Queequeg and needlessly exaggerates the movement of his lips.

This also clashes with the depiction Stubb enjoys in this passage. His characterization under duress is statesmanlike and incisive. In the face of life-threatening pandemonium, "Stubb ... still encourages his men to the onset, all the while puffing the smoke from his mouth" (221). He directs the crew with poise. Compared to Queequeg, Stubb possesses superior emotional control. He reaps the benefits of subjugating other

[32] "But the most powerful principle of the symbolic efficacy of cultural capital no doubt lies in the logic of its transmission." –Pierre Bourdieu, *The Forms of Capital*

humans; they are “his men” after all. In a wry flourish, Stubb directs the response to this chaotic event with his pipe in his mouth. A pipe for the bourgeois second-mate, then, and a well-worn harpoon for the cannibal: these accessories stand in for the divisions separating the crew of the Pequod.

Melville’s toothsome prose is a delectable *digestif* to round off the diverse platter of computer experiments, visualizations and close readings served in this paper. Let’s move to Wittgenstein’s writings on resemblance to cap off my holistic analysis of “lyrical” prose.

Towards a Promontory

DH projects require spending time beneath the surface of literature. You practice a "sub-formalism" and look for trends in disparate pieces of literature digitally transformed into data. At this juncture, I want to move on from this type of work and ponder the act of labeling sets of literature as "lyrical." We've heard from computer scientists, literary scholars and scores of others about their views on the lyric. I've offered my own analysis that synthesizes digital learning and close reading. However, I want the scope of this project to close on a high note — by scaling a promontory of sorts. Wittgenstein's ambitious conjectures on language in *Philosophical Investigations* provide the necessary elevation to take us there.

In talking about a general concept, Wittgenstein writes: "Instead of producing something common to all ... I am saying that these phenomena have no one thing in common which makes us use the same word for all — but that they are *related* to one another in many different ways" (Wittgenstein § 65, emphasis author's).

A network of relationships bonds together members of the group. The linkages vary greatly; some threads are tenuous, others, nearly concrete. Wittgenstein describes these connections as "family resemblances" (§ 67). The strength of the network is indebted to the numerous, overlapping similarities of its members. How can we describe a concept that adheres to this framework? The best method, according to Wittgenstein, involves a clear description of *one* of its members (§ 67). This draws a boundary around the concept even though it is by nature uncircumscribed. For instance, if someone asks you to describe what a game is, you should answer with examples of games.[33] Next,

[33] Flailing APMA students such as myself ought to agree that examples trump theorems every time.

you'd probably teach your interlocutor how to play one. An obtuse wavering about the underlying nature of any thing that could, under particular circumstances and conditions, be considered a game would not be very particularly instructive.

With my corpus, then, the best way to answer the question of "what is a 'lyrical' novel?" is to enumerate examples of "lyrical" fiction. Discovering what makes a novel "lyrical" is impossible without repeatedly identifying associations in different types of books that fall under this label. Crucially, this is not an "*indirect"* approach – but an optimal method for answering difficult research questions (§ 71, emphasis author's).

It is amusing to speculate about the happenings of the Lyrical household.[34] But the best studies of the lyric blaze past arbitrary proclamations. Jackson's level of detail, for example, in *Dickinson's Misery* is what makes it so effective. Jackson exhausts Dickinson's oeuvre to support her theory of lyric. Likewise, this project preferred focused examples to generalized conjectures. The wealth of data I collected from my census of this "lyrical" family offers an unalloyed level detail not possible in most analyses.

At the same time, Wittgenstein justifies our inclination to affix labels like "lyrical" to works of fiction we read. This grouping offers a natural avenue of explanation for how these texts operate and relate to one another. Wittgenstein's piece presents a compelling rebuttal to my skeptical attitude about indiscriminately utilizing the word "lyrical".[35] Clutching ambiguous labels such as "lyrical" might not be as insidious as I initially presumed.

---

[34] Who is the most annoying at Thanksgiving Dinner? My money is on Joyce.
[35] For further reading on the subject of considerate criticism, Felski's *Uses of Literature* is indispensable (Felski).

Conclusion

Mardi Gras day, Bourbon Street. The sun, hiding behind a murky haze, beats down on garbage-strewn streets. A man holds his head in his hands. Two children argue over whose beads are best. Car alarms so frequent they form a concerto.

Hardly an environment for scholarship.

I peel away from my group of friends in search of an unlikely companion: Ignatius J. Riley. Ignatius is the misguided protagonist of Jeffrey Toole's gobsmacking *The Confederacy of Dunces*, a work I absolutely would have included in my corpus if a reliable digital text were to exist.[36]

Siri tells me a statue of Ignatius looms in front of an upscale department store on the adjacent street. I wander his way. I find him across the sidewalk from a hot dog vendor. The statue is hilarious, though unfortunately it sits behind a chain-link fence on Mardi Gras day.[37] I squint through the links and devour the words of an adjacent plaque. I savor the quote on the plaque, Ignatius' contempt-filled descriptions of his city's residents. Simultaneously, the recollection of the tragic circumstances of Toole's suicide ravage me. A multiplicity of emotions roam through my consciousness. To call back to *The Birth of Tragedy*, in the midst of a Dionysian wasteland, the Apollonian beauty of this display graces me with evanescent clarity.

What else but literature could draw a college kid away from Bourbon Street?[38] What other form of art incites these moments of rapturous delight and melancholic agony in the midst of bacchanalian debauchery? We must exercise excruciating diligence

---

[36] Digital projects expand the scope of inquiry. However, you cannot feasibly analyze every book that you would like. It's crushing, really.
[37] A shame given the holiday's Dickensian sobriquet: "Fat Tuesday."
[38] Don't answer that.

towards the words we toss around to describe literature in order to protect and cherish experience such as these. If we do not, the critical sensibilities *du jour* supplant and muddle these marvelous, resonant experiences.

My work creates new territory for describing the wonderful experiences books offer. Wittgenstein's wisdom notwithstanding, if we cannot explain the joys of our field without wading into ambiguous tropes, we stand the risk of squandering potential connections to literature. This produces disengaged students, half-hearted discussions, fluffy papers, endowing our literary studies to the fortunate few who can cut through the noise.

When my friends asked me why I cared about Ignatius' statue so much, I did not reply that Toole's work was "lyrical" or "tragic" or "epic." That would have been ridiculous. I explained to them how the genius of Toole's pen won more laughter from me than any Hollywood comedy, *Vine* compilation, or *Reddit* post ever could. Laughter, love, loss, poverty, adulation: the myriad, universal emotions literature generates deserve a coherent vocabulary. My experience in New Orleans stuck with me as I rounded the final lap of my DMP. It pushed me to evaluate all that my failures and success over the course of this multifaceted project.

Now, it would not be wrong to point out that facing the most gargantuan of tasks, I failed. I did not formulate an unassailable blueprint for creating "lyrical" novels. I did not synthesize a pithy definition of "lyrical" novels, either. Indeed, *StackOverflow* and *JSTOR* could only take me so far. Virginia Jackson can keep her day job; she is not going to pick up this thesis and call it quits because an undergraduate with too much time on his hands ran experiments on his computers. Keep the champagne bottles corked and

trophies unengraved. The lyric remains elusive, opaque, capacious, unrattled, omnipresent and authoritative.

In spite of this, I'm content with the outcomes of my project. The work carried out in my project injects empiricism and rigor into the claims we make about books. I cleaned a reusable corpus of canonical texts and packaged their rhetorical units in a relational database for future study. I incorporated the ideas of critical theorists. I calculated remarkably minute yet consequential markers of syntax and style in the novel. I focused attention on the invocation of anaphora and demonstrated that this measure is indicative of lyricality in novels.

Importantly, I carried out this work the *right* way. My exacting attention to detail (hopefully) ensures that this piece will not appear in a future iteration of Nan Z. Da's recent bombshell *The Case Against Computational Literary Studies*.[39] In keeping with this, I verified the claims of my experiments against numerous passages from my corpus. Additionally, my sleuthing boasts a low barrier to entry. My diverse readership, from Dickens scholars to computational linguists, derives new truths about the lyric's complex history, computational methodologies, and properly integrating mathematics and critique from my project. In turn, I engender broad discussions about canonization, genre and form—and questions that which lurks deep beneath the glossy surface of the celebrated novels we cling to.

Literary scholars spend entire careers chasing after elusive questions concerning the moral and ethical dimensions of the world we live in and the literature we consume. And what good is their scholarship without loose ends? In a way, then, the nebulousness

[39] "It's basically *Kill Bill*, but the DH people we like citing are the victims" – Brad

that characterizes “lyrical” novels is preferable. However, a comprehensive understanding of “lyrical” novels will likely supplant this uncertainty. Humanists will sharpen their interpretations of these canonical novels, and hackers will accelerate the development of technologies to answer this question from a digital angle. This thesis makes a compelling case for leveraging the compatibility between these two parties – both seek the same outcome and should rely on one another to achieve it. As for me and my salvos, I urge you to be careful: your favorite novel might be the book I dismantle next.

Appendix

	Author	Title
1	Cormac McCarthy	Blood Meridian
2	Cormac McCarthy	The Road
3	D.H. Lawrence	The Rainbow
4	D.H. Lawrence	The Plumed Serpent
5	D.H. Lawrence	Women in Love
6	Edgar Allen Poe	The Narrative of Arthur Gordon Pym of Nantucket
7	Edgar Allen Poe	Eureka: A Prose Poem
8	F. Scott Fitzgerald	The Great Gatsby
9	Herman Melville	Billy Budd
10	Herman Melville	Moby Dick
11	James Joyce	Portrait of the Artist as a Young Man
12	Jean Rhys	Wide Sargasso Sea
13	Joseph Conrad	Heart of Darkness
14	Joseph Heller	Something Happened
15	J. M. Coetzee	Life & Times of Michael K
16	Malcolm Lawry	Under the Volcano
17	Oscar Wilde	The Picture of Dorian Gray
18	Thomas Pynchon	Gravity's Rainbow
19	Virginia Woolf	Mrs. Dalloway
20	Virginia Woolf	Orlando
21	Virginia Woolf	To The Lighthouse
22	Vladimir Nabokov	Pale Fire
23	Vladimir Nabokov	Lolita
24	William Faulkner	Absalom Absalom
25	William Faulkner	The Sound and the Fury
26	William H. Gass	In The Heart of the Heart of the Country

Figure A.1 Lyrical Corpus

	Author	Title
1	Agatha Christie	The Secret Adversary
2	Anna Katherine Green	The Leavenworth Case
3	Arthur Conan Doyle	A Study in Scarlet
4	Arthur Conan Doye	The Sign of Four
5	Arthur J. Rees	The Shrieking Pit
6	Arthur J. Rees	The Moon Rock
7	Arthur J. Rees	The Hand in the Dark
8	Carolyn Wells	The Maxwell Mystery
9	Edgar Wallace	The Angel Of Terror
10	Edgar Wallace	The Daffodil Mystery
11	Emmuska Orczy	The Old Man in the Corner
12	Ethel Lina White	The Spiral Staircase
13	Fred Merrick White	The Lady in Blue
14	Fred Merrick White	The Mystery of Room 75
15	G.K. Chesterton	The Innocence of Father Brown
16	Harrington Strong	The Brand of Silence
17	J.S. Fletcher	The Paradise Mystery
18	J.S. Fletcher	The Rayner Slade Amalgamation
19	J.S. Fletcher	The Scarhaven Keep
20	Mary Roberts Rinehart	The Circular Staircase
21	Mrs. Charles Bryce	The Ashiel Mystery
22	R. Austin Freedman	The Red Thumb Mark
23	Raymond Chandler	The Big Sleep
24	Wilkie Collins	The Moonstone

Figure A.2 Detective Corpus

	Feature Names	Explanation
1	labels_vec	Category
2	word_counts	Number of words
3	sent_counts	Number of sentences
4	para_counts	Number of paragraphs
5	sent_comma_freq	Number of commas per sentence
6	para_comma_freq	Number of commas per paragraph
7	words_per_sentence	Number of words per sentence
8	words_per_paragraph	Number of words per paragraph
9	sents_per_paragraph	Number of sentences per paragraph
10	consecutive_counts	Number of anaphoric sentences
11	consecutive_repeat_freq	Frequency of anaphoric sentences
12	syll_and_word_freq	Number of syllables per word
13	polysyll_and_word_freq	Number of polysyllables per word
14	polysyll_and_sent_freq	Number of polysyllables per sentence
15	unique_counts	Number of unique words
16	type_token_ratio	Type-Token Ratio
17	mean_usage_freq	Mean Usage Frequency (1 over Type-Token Ratio)
18	median_MATTR	Median Moving-Average Type-Token Ratio
19	object_freq	Percent of words in Harvard Inquirer's "object" category
20	relationship_freq	Percent of words in Harvard Inquirer's "relationship" category
21	time_freq	Percent of words in Harvard Inquirer's "time" category
22	self_freq	Percent of words in Harvard Inquirer's "self" category
23	perceive_freq	Percent of words in Harvard Inquirer's "perceive" category
24	i_freq	Frequency of the word 'I'
25	top_ten_freq	Frequency of a book's top ten words
26	dialogue_freq	Percentage of dialogue
27	question_vec	Number of question marks
28	exclamation_vec	Number of exclamation points
29	sentiment_vec	average sentiment across text

Figure A.3 Feature Matrix Explanation

	absalomAbsalom	billyBudd	bloodMeridian	eureka	gravitysRainbow	heartOfDarkness	lifeAndTimesOfMichaelK
labels_vec	lyrical	lyrical	lyrical	lyrical	lyrical	lyrical	lyrical
word_counts_vec	132668	30743	115200	38566	330351	39085	66354
sent_counts_vec	3027	1137	7552	1088	18847	2402	4253
para_counts_vec	569	288	2640	255	5669	197	762
commas_vec	7519	2061	1924	3476	31565	2870	3762
sent_comma_freq_vec	2.4839775	1.8126649	0.2547669	3.1948529	1.6748024	1.1948376	0.8845521
para_comma_freq_vec	13.2144112	7.1562500	0.7287879	13.6313726	5.5680014	14.5685279	4.9370079
words_per_sentence_vec	43.828213	27.038698	15.254237	35.446691	17.528042	16.271857	15.601693
words_per_paragraph_vec	233.15993	106.74653	43.63636	151.23922	58.27324	198.40102	87.07874
sents_per_paragraph_vec	5.319859	3.947917	2.860606	4.266667	3.324572	12.192893	5.581365
consecutive_counts_vec	326	40	811	71	785	227	327
consecutive_repeat_freq_vec	0.10769739	0.03518030	0.10738877	0.06525735	0.04165119	0.09450458	0.07688690
syll_and_word_freq_vec	1.343730	1.525876	1.274479	1.619302	1.411257	1.352156	1.271574
polsyll_and_word_freq_vec	0.06278078	0.13160720	0.04690104	0.17699528	0.08612960	0.07621850	0.04507641
syll_and_sent_freq_vec	58.893294	41.257696	19.441208	57.398897	24.736563	22.002082	19.838702
polsyll_and_sent_freq_vec	2.7515692	3.5584872	0.7154396	6.2738971	1.5096832	1.2402165	0.7032683
unique_counts_vec	8981	5773	10079	4194	26428	5418	6564
type_token_ratio_vec	0.06769530	0.18778259	0.08749132	0.10874864	0.07999976	0.13862095	0.09892395
mean_usage_frequency_vec	14.772074	5.325307	11.429705	9.195517	12.500038	7.213917	10.108775
median_MATTR	47.75	50.69	46.44	47.19	52.44	49.25	48.44
object_freq	0.02477613	0.02462349	0.03440972	0.01641342	0.03223238	0.02776001	0.04066070
relationship_freq	0.02107516	0.02062258	0.01518229	0.01944718	0.01878002	0.01675835	0.01698466
time_freq	0.05749691	0.04231858	0.03967882	0.04208370	0.04932330	0.04190866	0.05187329
self_freq	0.012090331	0.005269492	0.005225694	0.009516154	0.006266062	0.043290265	0.015733791
perceive_freq	0.01297977	0.01359659	0.01137153	0.01407976	0.01117902	0.01501855	0.01258402
i_freq	0.008238611	0.002797385	0.003559028	0.006612042	0.003983642	0.029474223	0.010232993
top_ten_freq	0.2526985	0.2608073	0.3007899	0.2780428	0.2137545	0.2625048	0.2656208
dialogue_freq	0.2478032	0.1996528	0.4671610	0.1431373	0.2667137	0.2994924	0.2257218
question_vec	366	95	518	73	2173	156	401
exclamation_vec	90	37	5	17	784	159	127
sentiment_vec	-0.015000876	0.031769516	-0.043100786	0.091198383	-0.007974449	-0.027648484	-0.024025435

Figure A.4.i Feature Matrix

	lolita	mobyDick	mrsDalloway	orlando	paleFire	portraitOfTheArtist	pym
labels_vec	lyrical	lyrical	lyrical	lyrical	lyrical	lyrical	lyrical
word_counts_vec	112193	214183	64267	79547	68132	84927	100954
sent_counts_vec	5169	7381	3405	3292	2546	4452	3680
para_counts_vec	1189	2434	754	451	971	2208	625
commas_vec	8783	16135	6098	6047	4387	4258	8674
sent_comma_freq_vec	1.6991681	2.1860182	1.7908957	1.8368773	1.7230951	0.9564241	2.3570652
para_comma_freq_vec	7.3868797	6.6290058	8.0875332	13.4079823	4.5180227	1.9284420	13.8784000
words_per_sentence_vec	21.704972	29.018155	18.874302	24.163730	26.760408	19.076146	27.433152
words_per_paragraph_vec	94.35913	87.99630	85.23475	176.37916	70.16684	38.46332	161.52640
sents_per_paragraph_vec	4.347351	3.032457	4.515915	7.299335	2.622039	2.016304	5.888000
consecutive_counts_vec	347	262	265	243	150	421	284
consecutive_repeat_freq_vec	0.06713097	0.03549654	0.07782672	0.07381531	0.05891595	0.09456424	0.07717391
syll_and_word_freq_vec	1.419224	1.374087	1.383183	1.353024	1.449994	1.345332	1.454187
polsyll_and_word_freq_vec	0.09382938	0.08199997	0.07803383	0.07512540	0.10607350	0.06494990	0.10922796
syll_and_sent_freq_vec	30.804217	39.873459	26.106608	32.694107	38.802435	25.663747	39.892935
polsyll_and_sent_freq_vec	2.0365641	2.3794879	1.4728341	1.8153098	2.8385703	1.2389937	2.9964674
unique_counts_vec	14113	16747	7088	9397	11456	9190	9170
type_token_ratio_vec	0.12579216	0.07819015	0.11028988	0.11813142	0.16814419	0.10821058	0.09083345
mean_usage_frequency_vec	7.949621	12.789335	9.067015	8.465148	5.947277	9.241240	11.009160
median_MATTR	51.25	50.12	48.75	49.19	51.00	45.94	50.00
object_freq	0.03149929	0.03419506	0.02724571	0.03012056	0.02972172	0.02664641	0.02808210
relationship_freq	0.01655184	0.01973079	0.01588685	0.01544999	0.01646803	0.01523662	0.01571012
time_freq	0.04527912	0.05097510	0.04204335	0.04242775	0.04262314	0.04193013	0.04627850
self_freq	0.048077866	0.016028350	0.003112017	0.003142796	0.024790113	0.010185218	0.030033481
perceive_freq	0.01029476	0.01213915	0.01177899	0.01087407	0.01108143	0.01254018	0.01107435
i_freq	0.026579198	0.009818706	0.001571569	0.001684539	0.013826102	0.006358402	0.015888424
top_ten_freq	0.2466999	0.2344257	0.2447913	0.2578098	0.2466536	0.2763197	0.2614755
dialogue_freq	0.2380151	0.2002876	0.1942971	0.1873614	0.1266735	0.2504529	0.1200000
question_vec	414	870	361	218	161	556	120
exclamation_vec	262	1493	346	199	103	434	260
sentiment_vec	0.008131482	-0.000804251	0.010923513	0.006561174	0.025613752	-0.010018552	0.000531653

Figure A.4.ii Feature Matrix (cont.)

	somethingHappened	theGreatGatsby	thePedersenKid	thePictureOfDorianGray	theRainbow	theRoad	theSerpent
labels_vec	lyrical	lyrical	lyrical	lyrical	lyrical	lyrical	lyrical
word_counts_vec	190235	48818	23702	79289	187559	58702	172685
sent_counts_vec	14994	3296	2622	6251	12278	6533	13587
para_counts_vec	5124	1589	789	1524	4524	2467	4984
commas_vec	9857	2971	1036	5360	15133	814	15251
sent_comma_freq_vec	0.6573963	0.9013956	0.3951182	0.8574628	1.2325297	0.1245982	1.1224700
para_comma_freq_vec	1.9236924	1.8697294	1.3130545	3.5170604	3.3450486	0.3299554	3.0599920
words_per_sentence_vec	12.687408	14.811286	9.039664	12.684211	15.276022	8.985458	12.709575
words_per_paragraph_vec	37.12627	30.72247	30.04056	52.02690	41.45866	23.79489	34.64787
sents_per_paragraph_vec	2.926230	2.074261	3.323194	4.101706	2.713970	2.648156	2.726124
consecutive_counts_vec	2252	190	269	505	1163	643	1130
consecutive_repeat_freq_vec	0.15019341	0.05764563	0.10259344	0.08078707	0.09472227	0.09842339	0.08316773
syll_and_word_freq_vec	1.333698	1.359027	1.180365	1.333438	1.346056	1.202020	1.337279
polsyll_and_word_freq_vec	0.06804216	0.07380474	0.02096869	0.06389285	0.06914091	0.02580832	0.06715696
syll_and_sent_freq_vec	16.921168	20.128944	10.670099	16.913614	20.562388	10.800704	16.996246
polsyll_and_sent_freq_vec	0.8632786	1.0931432	0.1895500	0.8104303	1.0561981	0.2318996	0.8535365
unique_counts_vec	12694	5959	2442	7113	11237	4776	10491
type_token_ratio_vec	0.06672799	0.12206563	0.10302928	0.08970980	0.05991181	0.08136009	0.06075224
mean_usage_frequency_vec	14.986214	8.192314	9.705979	11.147055	16.691199	12.291039	16.460299
median_MATTR	47.75	50.31	46.12	49.44	46.94	45.75	48.69
object_freq	0.02399138	0.03320497	0.03754957	0.02900781	0.02304875	0.03895949	0.02937719
relationship_freq	0.02102663	0.01775984	0.01987174	0.01788394	0.01608027	0.01557017	0.01876249
time_freq	0.05126817	0.05657749	0.04126234	0.04274237	0.04420476	0.04498995	0.04192605
self_freq	0.074171420	0.038100701	0.045481394	0.028263694	0.006440640	0.012316446	0.014483018
perceive_freq	0.01221647	0.01231103	0.01291030	0.01215805	0.01151104	0.01335559	0.01096795
i_freq	0.045874839	0.024396739	0.030883470	0.017644314	0.003849455	0.008636844	0.008732664
top_ten_freq	0.2251268	0.2423082	0.2523838	0.2342443	0.2871363	0.3011652	0.2530735
dialogue_freq	0.3885636	0.3285085	0.2318841	0.3858268	0.1934129	0.2060125	0.2231140
question_vec	1787	327	221	572	954	633	1098
exclamation_vec	171	124	3	386	261	0	1785
sentiment_vec	-0.013274430	0.001433412	-0.008945749	0.004301973	-0.005002090	-0.010952869	-0.013266229

Figure A.4.iii Feature Matrix (cont.)

	theSoundAndTheFury	toTheLighthouse	underTheVolcano	wideSargassoSea	womenInLove	aStudyInScarlet	theAngelOfTerror
labels_vec	lyrical	lyrical	lyrical	lyrical	lyrical	detective	detective
word_counts_vec	96472	69905	138612	47349	182657	43901	63264
sent_counts_vec	9197	3401	7559	4278	12829	2616	4413
para_counts_vec	3208	495	2086	1175	6012	811	2429
commas_vec	5199	6500	12406	2657	15044	2951	4474
sent_comma_freq_vec	0.5652930	1.9112026	1.6412224	0.6210846	1.1726557	1.1280581	1.0138228
para_comma_freq_vec	1.6206359	13.1313131	5.9472675	2.2612766	2.5023287	3.6387176	1.8419103
words_per_sentence_vec	10.489507	20.554249	18.337346	11.068022	14.237821	16.781728	14.335826
words_per_paragraph_vec	30.07232	141.22222	66.44871	40.29702	30.38207	54.13194	26.04529
sents_per_paragraph_vec	2.866895	6.870707	3.623682	3.640851	2.133899	3.225647	1.816797
consecutive_counts_vec	569	319	386	328	888	148	217
consecutive_repeat_freq_vec	0.06186800	0.09379594	0.05106496	0.07667134	0.06921818	0.05657492	0.04917290
syll_and_word_freq_vec	1.232399	1.329676	1.423253	1.240681	1.367092	1.348124	1.346374
polsyll_and_word_freq_vec	0.02889958	0.05753523	0.09435691	0.03970517	0.07704605	0.07580693	0.07342248
syll_and_sent_freq_vec	12.927259	27.330491	26.098690	13.731884	19.464417	22.623853	19.301382
polsyll_and_sent_freq_vec	0.3031423	1.1825934	1.7302553	0.4394577	1.0969678	1.2721713	1.0525719
unique_counts_vec	6123	6773	13781	3876	11265	5665	6102
type_token_ratio_vec	0.06346919	0.09688864	0.09942141	0.08186023	0.06167297	0.12904034	0.09645296
mean_usage_frequency_vec	15.755675	10.321128	10.058196	12.215944	16.214558	7.749515	10.367748
median_MATTR	44.56	48.50	50.75	48.94	48.50	50.00	49.56
object_freq	0.03716104	0.02978328	0.02867717	0.03841686	0.02373301	0.02838204	0.03256196
relationship_freq	0.02030641	0.01700880	0.01903154	0.01757165	0.01721806	0.01961231	0.01632840
time_freq	0.05245045	0.04162792	0.04989467	0.05263047	0.04005869	0.04644541	0.03978566
self_freq	0.048646239	0.002303126	0.010670072	0.066632875	0.013407644	0.033780552	0.026587001
perceive_freq	0.01625342	0.01235963	0.01131937	0.01638894	0.01317223	0.01414546	0.01319866
i_freq	0.034870221	0.001273156	0.006752662	0.044858392	0.009422031	0.021320699	0.017197774
top_ten_freq	0.2415830	0.2520134	0.2292947	0.2483685	0.2441461	0.2492654	0.2480874
dialogue_freq	0.3439838	0.2060606	0.3056088	0.3480851	0.3137059	0.6732429	0.6088925
question_vec	663	391	994	408	1524	208	604
exclamation_vec	100	111	477	47	579	85	146
sentiment_vec	0.008524116	-0.001487740	-0.010723943	0.001937430	0.009939226	0.016638310	-0.001876058

Figure A.4.iv Feature Matrix (cont.)

	theSoundAndTheFury	toTheLighthouse	underTheVolcano	wideSargassoSea	womenInLove	aStudyInScarlet	theAngelOfTerror
labels_vec	lyrical	lyrical	lyrical	lyrical	lyrical	detective	detective
word_counts_vec	96472	69905	138612	47349	182657	43901	63264
sent_counts_vec	9197	3401	7559	4278	12829	2616	4413
para_counts_vec	3208	495	2086	1175	6012	811	2429
commas_vec	5199	6500	12406	2657	15044	2951	4474
sent_comma_freq_vec	0.5652930	1.9112026	1.6412224	0.6210846	1.1726557	1.1280581	1.0138228
para_comma_freq_vec	1.6206359	13.1313131	5.9472675	2.2612766	2.5023287	3.6387176	1.8419103
words_per_sentence_vec	10.489507	20.554249	18.337346	11.068022	14.237821	16.781728	14.335826
words_per_paragraph_vec	30.07232	141.22222	66.44871	40.29702	30.38207	54.13194	26.04529
sents_per_paragraph_vec	2.866895	6.870707	3.623682	3.640851	2.133899	3.225647	1.816797
consecutive_counts_vec	569	319	386	328	888	148	217
consecutive_repeat_freq_vec	0.06186800	0.09379594	0.05106496	0.07667134	0.06921818	0.05657492	0.04917290
syll_and_word_freq_vec	1.232399	1.329676	1.423253	1.240681	1.367092	1.348124	1.346374
polsyll_and_word_freq_vec	0.02889958	0.05753523	0.09435691	0.03970517	0.07704605	0.07580693	0.07342248
syll_and_sent_freq_vec	12.927259	27.330491	26.098690	13.731884	19.464417	22.623853	19.301382
polsyll_and_sent_freq_vec	0.3031423	1.1825934	1.7302553	0.4394577	1.0969678	1.2721713	1.0525719
unique_counts_vec	6123	6773	13781	3876	11265	5665	6102
type_token_ratio_vec	0.06346919	0.09688864	0.09942141	0.08186023	0.06167297	0.12904034	0.09645296
mean_usage_frequency_vec	15.755675	10.321128	10.058196	12.215944	16.214558	7.749515	10.367748
median_MATTR	44.56	48.50	50.75	48.94	48.50	50.00	49.56
object_freq	0.03716104	0.02978328	0.02867717	0.03841686	0.02373301	0.02838204	0.03256196
relationship_freq	0.02030641	0.01700880	0.01903154	0.01757165	0.01721806	0.01961231	0.01632840
time_freq	0.05245045	0.04162792	0.04989467	0.05263047	0.04005869	0.04644541	0.03978566
self_freq	0.048646239	0.002303126	0.010670072	0.066632875	0.013407644	0.033780552	0.026587001
perceive_freq	0.01625342	0.01235963	0.01131937	0.01638894	0.01317223	0.01414546	0.01319866
i_freq	0.034870221	0.001273156	0.006752662	0.044858392	0.009422031	0.021320699	0.017197774
top_ten_freq	0.2415830	0.2520134	0.2292947	0.2483685	0.2441461	0.2492654	0.2480874
dialogue_freq	0.3439838	0.2060606	0.3056088	0.3480851	0.3137059	0.6732429	0.6088925
question_vec	663	391	994	408	1524	208	604
exclamation_vec	100	111	477	47	579	85	146
sentiment_vec	0.008524116	-0.001487740	-0.010723943	0.001937430	0.009939226	0.016638310	-0.001876058

Figure A.4.v Feature Matrix (cont.)

	theLadyInBlue	theLeavenworthCase	theMaxwellMystery	theMoonRock	theMoonstone	theMysteryOfRoom75	theOldManInTheCorner
labels_vec	detective	detective	detective	detective	detective	detective	detective
word_counts_vec	78625	110218	60289	107804	197369	48754	69944
sent_counts_vec	9459	6133	3508	7184	11417	2818	2941
para_counts_vec	1091	3066	1911	2216	3607	758	1308
commas_vec	9556	8610	5120	5546	14420	3508	5396
sent_comma_freq_vec	1.0102548	1.4038806	1.4595211	0.7719933	1.2630288	1.2448545	1.8347501
para_comma_freq_vec	8.7589368	2.8082192	2.6792255	2.5027076	3.9977821	4.6279683	4.1253823
words_per_sentence_vec	8.312189	17.971303	17.186146	15.006125	17.287291	17.300923	23.782387
words_per_paragraph_vec	72.06691	35.94847	31.54840	48.64801	54.71833	64.31926	53.47401
sents_per_paragraph_vec	8.670027	2.000326	1.835688	3.241877	3.165234	3.717678	2.248471
consecutive_counts_vec	546	243	193	462	790	207	158
consecutive_repeat_freq_vec	0.05772280	0.03962172	0.05501710	0.06430958	0.06919506	0.07345635	0.05372322
syll_and_word_freq_vec	1.345564	1.378314	1.359037	1.379912	1.381124	1.334414	1.429787
polsyll_and_word_freq_vec	0.07393323	0.08567566	0.07853837	0.08326222	0.08340216	0.07262994	0.09960826
syll_and_sent_freq_vec	11.184586	24.770096	23.356613	20.707127	23.875887	23.086586	34.003740
polsyll_and_sent_freq_vec	0.6145470	1.5397032	1.3497720	1.2494432	1.4417973	1.2565649	2.3689221
unique_counts_vec	5923	7142	4979	8482	8997	4458	6344
type_token_ratio_vec	0.07533227	0.06479885	0.08258555	0.07867983	0.04558467	0.09143865	0.09070113
mean_usage_frequency_vec	13.274523	15.432372	12.108656	12.709738	21.937201	10.936294	11.025221
median_MATTR	50.31	49.62	50.00	49.25	48.44	48.62	49.62
object_freq	0.02981240	0.03159194	0.02750087	0.02903417	0.02956898	0.03041802	0.02940924
relationship_freq	0.02146900	0.01990600	0.01746587	0.01680828	0.02415273	0.02085983	0.01625586
time_freq	0.04909380	0.04523762	0.04493357	0.04455308	0.05144172	0.05462116	0.05282798
self_freq	0.037774245	0.053403255	0.050556486	0.021965790	0.054785706	0.028120770	0.015855542
perceive_freq	0.01584738	0.01888076	0.01484516	0.01760603	0.01458689	0.01380400	0.01466888
i_freq	0.024228935	0.033070823	0.036723117	0.014173871	0.029584180	0.018521557	0.010837241
top_ten_freq	0.2551860	0.2431726	0.2359966	0.2498238	0.2571984	0.2669935	0.2592074
dialogue_freq	0.7112741	0.7491846	0.6504448	0.6358303	0.5494871	0.5382586	0.5313456
question_vec	936	1129	426	813	1210	201	262
exclamation_vec	271	420	229	169	995	27	109
sentiment_vec	0.035255324	0.003746350	0.016439436	-0.025858065	0.016893728	0.018500188	0.004562986

Figure A.4.vi Feature Matrix (cont.)

	theOldManInTheCorner	theParadiseMystery	theRaynerSladeAmalgamation	theRedThumbMark	theScarhavenKeep	theSecretAdversary	theShriekingPit	theSignOfFour	theSpiralStaircase
labels_vec	detective	detective	detective	detective	detective	detective	detective	detective	detective
word_counts_vec	69944	76869	79730	74609	74609	76046	100006	43445	70956
sent_counts_vec	2941	15307	5081	3474	4831	7809	5194	2847	6163
para_counts_vec	1308	1742	1717	1747	1644	3240	1740	772	3123
commas_vec	5396	18025	6254	5344	5322	4441	6090	3266	5639
sent_comma_freq_vec	1.8347501	1.1775658	1.2308601	1.5382844	1.1016353	0.5687028	1.1725067	1.1471725	0.9149765
para_comma_freq_vec	4.1253823	10.3473020	3.6423995	3.0589582	3.2372263	1.3706790	3.5000000	4.2305699	1.8056356
words_per_sentence_vec	23.782387	5.021820	15.691793	21.476396	15.443800	9.738251	19.254139	15.259923	11.513224
words_per_paragraph_vec	53.47401	44.12687	46.43564	42.70693	45.38260	23.47099	57.47471	56.27591	22.72046
sents_per_paragraph_vec	2.248471	8.787026	2.959231	1.988552	2.938564	2.410185	2.985057	3.687824	1.973423
consecutive_counts_vec	158	530	173	145	175	278	377	179	145
consecutive_repeat_freq_vec	0.05372322	0.03462468	0.03404842	0.04173863	0.03622438	0.03559995	0.07258375	0.06287320	0.02352750
syll_and_word_freq_vec	1.429787	1.369811	1.387232	1.425391	1.363911	1.360295	1.390267	1.317850	1.370497
polsyll_and_word_freq_vec	0.09960826	0.08262108	0.09133325	0.10187779	0.08245654	0.07507298	0.08622483	0.06221660	0.07280568
syll_and_sent_freq_vec	34.003740	6.878944	21.768156	30.612263	21.063962	13.246895	26.768387	20.110292	15.778841
polsyll_and_sent_freq_vec	2.3689221	0.4149082	1.4331824	2.1879678	1.2734424	0.7310795	1.6601848	0.9494204	0.8382281
unique_counts_vec	6344	5314	5714	6854	5813	6757	7234	5643	7335
type_token_ratio_vec	0.09070113	0.06913060	0.07166688	0.09186559	0.07791285	0.08885411	0.07233566	0.12988837	0.10337392
mean_usage_frequency_vec	11.025221	14.465374	13.953448	10.885468	12.834853	11.254403	13.824440	7.698919	9.673620
median_MATTR	49.62	49.69	50.25	49.62	50.81	50.94	48.75	50.62	51.06
object_freq	0.02940924	0.02703300	0.03499310	0.03593400	0.02653835	0.03125740	0.03809771	0.02983082	0.03555725
relationship_freq	0.01625586	0.01735420	0.01660604	0.01952848	0.01704888	0.01760776	0.01834890	0.02209690	0.01533345
time_freq	0.05282798	0.05249190	0.05068356	0.03894972	0.05252718	0.04752387	0.04661720	0.04566693	0.03953154
self_freq	0.015855542	0.023117251	0.023780258	0.039499256	0.022879277	0.029955553	0.025278483	0.039521234	0.015262980
perceive_freq	0.01466888	0.01767943	0.01609181	0.01666019	0.01900575	0.01401783	0.01678899	0.01574404	0.01362816
i_freq	0.010837241	0.014791398	0.014937915	0.025640338	0.014059966	0.020934697	0.016579005	0.025204281	0.006877502
top_ten_freq	0.2592074	0.2353615	0.2297379	0.2591376	0.2267689	0.2213397	0.2677439	0.2382092	0.2486330
dialogue_freq	0.5313456	0.6882893	0.6540478	0.7727533	0.6812652	0.6626543	0.6913793	0.7914508	0.5565162
question_vec	262	2205	704	515	698	967	552	215	580
exclamation_vec	109	2307	701	123	807	647	81	127	5
sentiment_vec	0.004562986	0.044894146	0.048802118	0.034809740	0.035484154	0.029305943	-0.010209293	0.009999532	-0.045097032

Figure A.4.vii Feature Matrix (cont.)

Works Cited

***

Primary (hard copy)

Chandler, Raymond. *The Big Sleep*. Stellar Books, 2013.
Faulkner, William. *The Sound and the Fury*. The Modern Library. Accessed 15 Mar. 2019.
McCarthy, Cormac. *Book Jacket. Blood Meridian*. Vintage, 1992, https://www.amazon.com/Blood-Meridian-Evening-Redness-West/dp/0679728759/ref=pd_lpo_sbs_14_t_1?_encoding=UTF8&psc=1&refRID=FE5NN70FEX8FMBJJZ13H.
Melville, Herman. *Moby Dick*. Macmillan, 2016.
Nabokov, Vladimir. *Pale Fire*. First Vintage International Edition, Vintage, 1989.
Woolf, Virginia. *To the Lighthouse*. Harcourt Brace Jovanovich, 1989.

***

Secondary

100 Best Novels « Modern Library. http://www.modernlibrary.com/top-100/100-best-novels/. Accessed 26 Mar. 2019.
Abu-Mostafa, Yaser, et al. *Learning From Data*. AMLbook, 2012, AMLbook.com.
Alvarado, Rafael, and Paul Humphreys. "Big Data, Thick Mediation, and Representational Opacity." *New Literary History*, vol. 48, no. 4, Autumn 2017, pp. 729–49.
Bakhtin, Mikhail. *Problems of Dostoevsky's Poetics*. University of Minnesota Press, 1984, https://www.amazon.com/Problems-Dostoevskys-Poetics-History-Literature/dp/0816612285.
Bender, Bert. "'Moby-Dick', An American Lyrical Novel." *Studies in the Novel*, vol. 10, no. 3, Fall 1978, pp. 346–56.
Bourdieu, Pierre. *The Forms of Capital*. 1986, https://www.marxists.org/reference/subject/philosophy/works/fr/bourdieu-forms-capital.htm.
Burt, Stephen. "What Is This Thing Called Lyric?" *Modern Philology*, vol. 133, no. 3, Feb. 2016, pp. 422–40, doi:https://doi.org/10.1086/684097.
Covington, Michael, and Joe McFall. *The Moving-Average Type-Token Ratio*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.248.5206&rep=rep1&type=pdf. Linguistic Society of America, Chicago.
Culler, Jonathan. *Theory of the Lyric*. Harvard University Press. Accessed 7 July 2018.
Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry*, vol. 45, no. 3, Spring 2019, pp. 601–39, doi:https://doi.org/10.1086/702594.

Dietterich, Thomas. *Challenges for Machine Learning in Computational Sustainability*. https://datascience.columbia.edu/files/seasdepts/idse/pdf-files/Seminar_Poster__Thomas_Dietterich.pdf. Columbia University.

Eco, Umberto. "Narrative Structures in Fleming." *The Role of the Reader*, Indiana Univ. Press, 1979, p. 273, http://my.fit.edu/~rosiene/eco%20bond.pdf.

Felski, Rita. *Uses of Literature*. Wiley-Blackwell, 2008.

Gabernet, Armand Ruiz, and Jay Limburn. "Breaking the 80/20 Rule: How Data Catalogs Transform Data Scientists' Productivity." *IBM Cloud Blog*, 23 Aug. 2017, https://www.ibm.com/blogs/bluemix/2017/08/ibm-data-catalog-data-scientists-productivity/.

Gee, Philip. *Barron's 6 SAT Practice Tests*. Second, Barrons Educational Series, https://www.amazon.com/Barrons-6-SAT-Practice-Tests/dp/1438009968/ref=asc_df_1438009968/?tag=hyprod-20&linkCode=df0&hvadid=312695551910&hvpos=1o5&hvnetw=g&hvrand=45287473260476950 21&hvpone=&hvptwo=&hvqmt=&hvdev=c&hvdvcmdl=&hvlocint=&hvlocphy=9060485&hvtargid=pla-490432695581&psc=1. Accessed 10 Mar. 2019.

General Inquirer Category Description. http://www.wjh.harvard.edu/~inquirer/homecat.htm. Accessed 3 Dec. 2018.

General Inquirer Usage. http://www.wjh.harvard.edu/~inquirer/3JMoreInfo.html. Accessed 3 Dec. 2018.

Google. *Keras*. 2.2.4., Google, 2018, https://www.tensorflow.org/guide/keras.

Greenfield, Adam. *Radical Technologies: The Design of Everyday Life*. Verso, 2017.

Hayes, Adam. "Moving Average." *Investopedia*, 31 Jan. 2019, https://www.investopedia.com/terms/m/movingaverage.asp.

Heidegger, Martin, and Joan Stambaugh. *Being and Time*. SUNY Press, 2010, https://www.amazon.com/Being-Time-Translation-Contemporary-Continental/dp/1438432763/ref=asc_df_1438432763/?tag=hyprod-20&linkCode=df0&hvadid=312168414377&hvpos=1o2&hvnetw=g&hvrand=17660720388080620249&hvpone=&hvptwo=&hvqmt=&hvdev=c&hvdvcmdl=&hvlocint=&hvlocphy=9008337&hvtargid=pla-436622550668&psc=1.

Heuser, Ryan, and Le-Khac Long. *A Quantative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method*. n+1, 2014, https://litlab.stanford.edu/LiteraryLabPamphlet7.pdf.

Huff Oberhaus, Dorothy. "About Dickinson's 'Fascicles.'" *Modern American Poetry*, http://www.english.illinois.edu/maps/poets/a_f/dickinson/fascicles.htm. Accessed 23 Oct. 2018.

Jackson, Virginia. *Dickinson's Misery*. Princeton University Press, 2005, https://press.princeton.edu/titles/7989.html.

---. "Lyric." *The Princeton Encyclopedia of Poetry and Poetics: Fourth Edition*, edited by Roland Greene et al., Princeton University Press, 2012, pp. 826–34. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/uva/detail.action?docID=913846.

Jockers, Matthew. *Macroanalysis: Digital Methods & Literary History*. University of Illinois Press, 2013, 6/28/2018.

---. *Text Analysis with R for Students of Literature*. Springer, 2014, 3/3/2018.

Kant, Immanuel. *Kant's Critique of Judgement*. Translated by J. H. Bernard, Macmillan, 1914, https://oll.libertyfund.org/titles/kant-the-critique-of-judgement.

Kao, Justine, and Dan Jurafsky. "A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry." *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, June 2012, p. 10.

Katsma, Holst. "Loudness in the Novel." *Canon / Archive*, n+1, 2014, https://litlab.stanford.edu/LiteraryLabPamphlet7.pdf.

Kivilo, Maarit. "Archilochus." *Early Greek Poets' Lives*, Brill, 2010, : https://www.jstor.org/stable/10.1163/j.ctv4cbgkd.8.

Korbut, Daniel. "Machine Learning Translation and the Google Translate Algorithm." *Stats and Bots*, 1 Aug. 2017, https://blog.statsbot.co/machine-learning-translation-96f0ed8f19e4.

Lukacs, Georg. *The Theory of the Novel*. The MIT Press, 1974, https://www.amazon.com/Theory-Novel-Georg-Lukacs/dp/0262620278.

Lyotard, Jean-Francois. *The Postmodern Condition: A Report on Knowledge*. Translated by Geoff Bennington and Massumi Brian, Manchester University Press, 1984, https://monoskop.org/images/e/e0/Lyotard_Jean-Francois_The_Postmodern_Condition_A_Report_on_Knowledge.pdf.

Moore, Sarah. "Pareto's Law for Dummies." *Helix Magazine: Northwestern University*, 5 Nov. 2012, https://helix.northwestern.edu/blog/2012/11/paretos-law-dummies.

Moretti, Franco, editor. *Canon / Archive: Studies in Quantative Formalism From the Stanford Literary Lab*. n+1, 2017.

---. *Distant Reading*. Penguin Random House, 2013.

Muller, Kirill, et al. *RSQLite*. 2.1.1, 2018, https://cran.r-project.org/web/packages/RSQLite/RSQLite.pdf. CRAN.

NEH.gov. "NEH-Mellon Fellowships for Digital Publication." *National Endowment for the Humanities (NEH)*, https://www.neh.gov/grants/research/neh-mellon-fellowships-digital-publication. Accessed 26 Feb. 2019.

Nietzsche, Friedrich. *The Birth of Tragedy*. Translated by Shaun Whiteside, Penguin Random House, 1993.

---. *Untimely Meditations: Use and Abuse of History for Life*. Translated by Ian C. Johnson, Vol. V9R, Malaspina University-College, 1998, http://nietzsche.holtof.com/Nietzsche_untimely_meditations/on_the_use_and_abuse_of_History.htm.

"Overfitting, n." *OED Online*, Oxford University Press. *Oxford English Dictionary*, http://www.oed.com/view/Entry/258314. Accessed 28 Mar. 2019.

Paulino, Nayib. "Prezi: The Great Gatsby." *Prezi.Com*, 31 Aug. 2013, https://prezi.com/yqojsrkla2yj/the-great-gatsby/.

Project Gutenberg Australia. http://gutenberg.net.au/crime-mystery.html. Accessed 25 Feb. 2019.

Python Software Foundation. *Python*. 2.7, http://python.org. Accessed 11 Mar. 2018.

Qi, Dr Yanjun. *Lecture 11: Support Vector Machine (Basics)*. p. 58.

---. *Lecture 18: Decision Tree / Random Forest / Ensemble*. p. 72.

Qi, Yanjun. *Lecture 1: Introduction*. https://qiyanjun.github.io/2018fUVA-CS4501MachineLearning//Lectures/L01-intro.pdf. UVA CS 4501.

---. *Lecture 11: Support Vector Machine (Basics)*. https://qiyanjun.github.io/2018fUVA-CS4501MachineLearning//Lectures/L11-SVM-basic.pdf. UVA CS 4501. Accessed 30 Oct. 2018.
---. *Lecture 18: Decision Tree / Random Forest / Ensemble*. UVA CS 4501.
R Core Team. *R*. R Foundation for Statistical Computing, 2014, http://www.R-project.org/.
Rinker, Tyler. *Textclean*. 0.9.3, 2018, https://cran.r-project.org/web/packages/textclean/textclean.pdf. CRAN.
Rinker, Tyler W. *Qdap (Quantitative Discourse Analysis Package)*. 2.3.2, 2019, https://cran.r-project.org/web/packages/qdap/vignettes/qdap_vignette.html. CRAN.
"Russian Formalism." *The Johns Hopkins Guide to Literary Theory and Criticism*, 2005, https://litguide-press-jhu-edu.proxy01.its.virginia.edu/cgi-bin/view.cgi?eid=227&query=formalism.
Shakespeare, William. *Hamlet*. Folger Shakespeare Library, https://www.folgerdigitaltexts.org/html/Ham.html. Accessed 15 Mar. 2019.
Smith, Jake. *MetaCTF Cybersecurity Competition @UVA on October 20th*. 28 Sept. 2018.
Starr, G. Gabrielle. *Lyric Generations: Poetry and the Novel in the Long Eighteenth Century*. John Hopkins University Press, 2015, https://ebookcentral.proquest.com/lib/uva/detail.action?docID=4398495.
Stone, Philip, et al. *General Inquirer: A Computer Approach to Content Analysis*. MIT Press, 1966, https://www.amazon.com/General-Inquirer-Computer-Approach-Analysis/dp/026269011X.
Sutskever, Ilya. *Training Recurrent Neural Networks*. University of Toronto, 2013, http://www.cs.utoronto.ca/~ilya/pubs/ilya_sutskever_phd_thesis.pdf.
Technische Universität Wien. *E1071*. 1.7-1, 2019, https://cran.r-project.org/web/packages/e1071/e1071.pdf. CRAN.
"The Digital in the Humanities: An Interview with Franco Moretti." *Los Angeles Review of Books*, https://lareviewofbooks.org/article/the-digital-in-the-humanities-an-interview-with-franco-moretti/. Accessed 21 Mar. 2019.
Tyler, Rinker. *SentimentR*. 2.6.1, 2018, https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf. CRAN.
Underwood, Ted. "Why an Age of Machine Learning Needs the Humanities." *Public Books*, 5 Dec. 2018, https://www.publicbooks.org/why-an-age-of-machine-learning-needs-the-humanities/.
Unsworth, John. *What Is Humanities Computing (and What Is Not)?* 8 Nov. 2002, http://www.people.virginia.edu/~jmu2m/texas-hc.html.
Watt, Ian. *The Rise of the Novel: Studies in Defoe, Richardson and Fielding*. University of California Press, 2001, https://www.amazon.com/Rise-Novel-Studies-Richardson-Fielding/dp/0520230698/ref=sr_1_2?keywords=rise+of+the+novel&qid=1553626378&s=books&sr=1-2.
Whipple, Mary. "Review of To the Lighthouse." *Amazon Reviews*, 1 July 2002, https://www.amazon.com/product-reviews/0374272409?ie=UTF8&pd_rd_i=0374272409&pd_rd_r=ebd810aa-0d93-

11e9-9c11-8d51749b796d&pd_rd_w=6psdE&pd_rd_wg=oAFQv&pf_rd_p=b21f843a-654c-40f8-899e-282283dbe728&pf_rd_r=V507S5556X3NRMAMGEJQ.

Wilkens, Matthew. "Digital Humanities and Its Application in the Study of Literature and Culture." *Comparative Literature*, vol. 67, no. 1, 2015, pp. 11–20.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G. E. M. Anscombe, Basil Blackwell, 2009.

Wordsworth, William. "Lyrical Ballads: 1798 and 1802." *Preface to Lyrical Ballads*, Oxford, 2013.

Yang, Zichao, et al. "Hierarchical Attention Networks for Document Classification." *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2016, pp. 1480–89. *Crossref*, doi:10.18653/v1/N16-1174.

***

Primary eTexts (Detective Novels)

Bryce, Mrs. Charles. *The Ashiel Mystery - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/9746. Accessed 10 Oct. 2018.

Chandler, Raymond. *The Big Sleep - EBook*. http://mirror1.booksdl.org/get.php?md5=B3EBA4DD9BBD3CD08A42D55C0882AE26&key=PLFH2FZJC9HSDO55. Books DL. Accessed 10 Oct. 2018.

Chesterton, G. K. *The Innocence of Father Brown - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/204. Accessed 10 Oct. 2018.

Christie, Agatha. *The Secret Adversary - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/1155. Accessed 12 Oct. 2018.

Collins, Wilkie. *The Moonstone - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/155. Accessed 10 Oct. 2018.

Conan Doyle, Arthur. *A Study in Scarlet - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/244. Accessed 10 Oct. 2018.

---. *The Sign of Four - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/2097. Accessed 10 Oct. 2018.

Fletcher, J. S. *Scarhaven Keep - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/9807. Accessed 10 Oct. 2018.

---. *The Paradise Mystery - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/5308. Accessed 10 Oct. 2018.

---. *The Rayner Slade Amalgamation - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/10443. Accessed 10 Oct. 2018.

Freedman, R. Austin. *The Red Thumb Mark - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/11128. Accessed 10 Oct. 2018.

Green, Anna Katherine. *The Leavenworth Case - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/4047. Accessed 10 Oct. 2018.

Orczy, Emmuska. *The Old Man in the Corner - EBook*. http://www.gutenberg.org/ebooks/10556. Accessed 10 Oct. 2018.

Rees, Arthur J. *The Hand in the Dark - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/20546. Accessed 10 Oct. 2018.
---. *The Moon Rock - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/12509. Accessed 10 Oct. 2018.
---. *The Shrieking Pit - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/20494. Accessed 10 Oct. 2018.
Rinehart, Mary Roberts. *The Circular Staircase - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/434. Accessed 10 Oct. 2018.
Strong, Harrington. *The Brand of Silence - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/21891. Accessed 10 Oct. 2018.
Wallace, Edgar. *The Angel of Terror - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/21530. Accessed 10 Oct. 2018.
---. *The Daffodil Mystery - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/20912. Accessed 10 Oct. 2018.
White, Ethel Lina. *The Spiral Staircase - EBook*. Project Gutenberg, http://gutenberg.net.au/ebooks03/0300931.txt. Accessed 10 Oct. 2018.
White, Fred Merrick. *The Lady in Blue*. Project Gutenberg, http://gutenberg.net.au/ebooks12/1201411.txt. Accessed 10 Oct. 2018.
---. *The Mystery of Room 75 - Ebook*. Project Gutenberg, http://gutenberg.net.au/ebooks12/1200221.txt. Accessed 11 Oct. 2018.

***

Primacy eTexts ("Lyrical" Novels)

Coetzee, J. M. *Life and Times of Michael K - EBook*. http://gen.lib.rus.ec/search.php?req=life+and+times+of+michael+k&lg_topic=libgen&open=0&view=simple&res=25&phrase=1&column=def. Libgen. Accessed 13 Sept. 2018.
Conrad, Joseph. *Heart of Darkness - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/526. Accessed 12 Sept. 2018.
Faulkner, William. *Absalom, Absalom! - EBook*. https://archive.org/details/in.ernet.dli.2015.185612. Archive.org. Accessed 14 Sept. 2018.
---. *The Sound and the Fury - EBook*. https://www.fadedpage.com/showbook.php?pid=201410L9. Faded Page. Accessed 14 Sept. 2018.
Fitzgerald, F. Scott. *The Great Gatsby - EBook*. University of Virginia, http://xroads.virginia.edu/~hyper2/Fitzgerald/gatsby.txt. Accessed 12 Sept. 2018.
Heller, Joseph. *Something Happened - EBook*. https://www.amazon.com/SOMETHING-HAPPENED-Novel-Joseph-Heller-ebook/dp/B005IQZ894. Amazon. Accessed 13 Sept. 2018.
Joyce, James. *Portrait of the Artist as a Young Man - EBook*. Project Gutenberg, http://www.gutenberg.org/files/4217/4217-0.txt. Accessed 13 Sept. 2018.
Lawrence, D. H. *The Plumed Serpent - EBook*. Project Gutenberg, http://gutenberg.net.au/ebooks03/0300021h.html. Accessed 12 Sept. 2018.

---. *The Rainbow - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/28948. Accessed 12 Sept. 2018.
---. *Women in Love - EBook*. Project Gutenberg, http://www.gutenberg.org/files/4240/4240-0.txt. Accessed 12 Sept. 2018.
Lowry, Malcolm. *Under the Volcano - EBook*. https://www.fadedpage.com/showbook.php?pid=20170345. Faded Page. Accessed 13 Sept. 2018.
McCarthy, Cormac. *Blood Meridian - EBook*. http://booksdescr.org/item/index.php?md5=53E7BA6FC50597C8F009521A68E1E901. Libgen. Accessed 25 Dec. 2018.
---. *The Road - EBook*. http://libgen.io/search.php?req=the+road+mccarthy&open=0&res=25&view=simple&phrase=1&column=def. Libgen. Accessed 12 Sept. 2018.
Melville, Herman. *Billy Budd - EBook*. Project Gutenberg, http://gutenberg.net.au/ebooks06/0608511.txt. Accessed 12 Sept. 2018.
---. *Moby Dick - EBook*. Project Gutenberg, http://www.gutenberg.org/files/2701/2701-0.txt. Accessed 12 Sept. 2018.
Nabokov, Vladimir. *Lolita - EBook*. http://libgen.io/search.php?req=lolita&open=0&res=25&view=simple&phrase=1&column=def. Libgen. Accessed 13 Sept. 2018.
---. *Pale Fire - EBook*. http://libgen.io/search.php?req=pale+fire&lg_topic=libgen&open=0&view=simple&res=25&phrase=1&column=def. Libgen. Accessed 25 Dec. 2018.
Poe, Edgar Allan. *Eureka: A Prose Poem - EBook*. University of Virginia, http://xroads.virginia.edu/~hyper/poe/eureka.html. Accessed 12 Sept. 2018.
---. *The Narrative of Arthur Gordyn Pym of Nantucket - EBook*. Project Gutenberg, https://www.gutenberg.org/files/2149/2149-h/2149-h.htm. Accessed 12 Sept. 2018.
Pynchon, Thomas. *Gravity's Rainbow*. http://libgen.io/search.php?req=gravity%27s+rainbow&lg_topic=libgen&open=0&view=simple&res=25&phrase=1&column=def. Libgen. Accessed 25 Dec. 2018.
Rhys, Jean. *Wide Sargasso Sea - EBook*. http://libgen.io/search.php?req=wide+SARGASSO+sea&open=0&res=25&view=simple&phrase=1&column=def. Libgen. Accessed 13 Sept. 2018.
Wilde, Oscar. *The Picture of Dorian Gray - EBook*. Project Gutenberg, http://www.gutenberg.org/ebooks/174. Accessed 13 Sept. 2018.
Woolf, Virginia. *Mrs. Dalloway - EBook*. Project Gutenberg, http://gutenberg.net.au/ebooks02/0200991.txt. Accessed 14 Sept. 2018.
---. *Orlando - EBook*. Project Gutenberg, http://gutenberg.net.au/ebooks02/0200331.txt. Accessed 14 Sept. 2018.
---. *To the Lighthouse - EBook*. Project Gutenberg, http://gutenberg.net.au/ebooks01/0100101.txt. Accessed 14 Sept. 2018.