COMPUTATIONAL CLOSE READING: A CRITIQUE OF DIGITAL LITERARY METHODOLOGY

Damiano Consilvio
University of Rhode Island, dconsilvio@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation
https://digitalcommons.uri.edu/oa_diss/1521

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.
COMPUTATIONAL CLOSE READING: A CRITIQUE OF DIGITAL LITERARY METHODOLOGY

BY

DAMIANO CONSILVIO

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN ENGLISH

UNIVERSITY OF RHODE ISLAND 2023
DOCTOR OF PHILOSOPHY DISSERTATION

OF

DAMIANO CONSILVIO

APPROVED:

Dissertation Committee:

Major Professor:    David Faflk

Ryan Trimm

Joan Peckham

Jennifer Jones

Harrison Dekker

Brenton DeBoef
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2023
ABSTRACT

*Computational Close Reading* intervenes in digital humanities and American literary scholarship to negotiate a significant institutional and disciplinary problem: the devaluation of close reading and interpretation in digital humanities applications to literary studies in favor of more distant modes of reading. To remedy this issue, I pursue a core question meant to bring the literary and the digital into a more perfect confluence, as a demonstration of possibilities meant to run counter to these preconceptions about computational reading: how “close,” if close at all, can digital technologies bring us to the interpretive elements of literary texts? I argue that they can bring us quite close, opening a perceived gap in knowledge about the ‘hows’ that are possible in computational reading. As a matter of necessity, I also speak to the questions that underlie this core issue in the humanities, such as why the computerization of literary studies has allegedly moved away from interpretation and close reading in favor of digital curation and digital tool-building; what other modes of reading computerization has given way to, and how these differ from traditional approaches (close, distant, hyper, machine, etc.); what is so different about traditional literary studies and the resulting digital turn, and why this disparity came to be; and, perhaps most importantly, how the analytical affordances of the digital humanities can be leveraged to advance and preserve the original goals of literary studies as a discipline, instead of enabling their supersession. Put plainly, *Computational Close Reading* challenges the notion that digital humanities necessitates new approaches to literature—rather, I am interested in how digital technologies can enhance our original approaches to literature and reaffirm their universal relevance and importance. This volume intends to argue, at its core, that the age
of computerization does not make obsolete or short-sighted the traditional methods of literary analysis in the face of new approaches to reading, but rather encourages us to pursue those same original disciplinary goals with renewed perspective and capability.
ACKNOWLEDGMENT

In the completion of this dissertation, I would like to acknowledge the guidance and mentorship of the great professors Carol J. Singley and Chris Fitter.
DEDICATION

This work is done in Honor of

The Great Nicola Consilvio

And it is dedicated to the following people who were there every step of the way,

reminding me what mattered most

My parents

Damian Consilvio (Dad)               Daniela Consilvio               Nicole Mastantuno (Mom)

My siblings

Nicholas Consilvio (Big Nick)         Giavanna Consilvio (Ms Pia)

My friends

Walker                               Jimmy Jackson                  Zack Romeo

                                  Pinhead

My girlfriend

Jaslynn Martinez

And my animals, who sat with me while I completed my comprehensive examination

Oogen                               Murry                               Hisenberg
PREFACE

If I had to put it in one sentence, I would say that the purpose of this dissertation is to encourage literary scholars to learn to use computer programming code in order to enhance their capacities for close reading. But, to be more elaborate...

This title is not some ornate metaphor. It’s quite literal, actually, and is meant to capture the scope and ambition of this project; and the hope is that by the end of this reading, the phrase “Computational Close Reading” will make perfect sense to the literary scholar without a humanities computing background. My goal is to make such a concept abundantly clear to the curious as well as the apprehensive. Overall, it rests on a simple premise: what if we could computationally automate close reading? And moreover, how would such an achievement reconfigure the position of the literary scholar within the digital humanities? My argument is that it would be for the better, all the while ostensibly relegating a fundamental—and traditionally human—literary task to the computer, bolstering its usage by executing the task at a greater speed, precision, and breadth than has been formerly possible.

Some elaboration is necessary in order to situate what is to come. The phrase from which I take my title—Computational Close Reading—signifies what I argue to be an appropriate, disciplinarily conscious, marrying of the fields of literary studies and the digital humanities, one that constitutes an interdisciplinary borrowing of knowledge that gives proper respect and priority to the disciplinary proclivities of literary studies by harnessing computerization to pursue a literary goal—the close reading of texts and the formulation of interpretive theories—not a technological one. The invention of new
reading platforms and technologies is the crux of the digital humanities on which I rest the warrants of this project.

My desire to codify close reading through a data analytic software springs from a critical observation I’ve made during my graduate career (or more accurately, my lifetime, as I am only 26 years old at the beginning of this writing) studying the digital humanities. I have found that as computational tools have been brought to bear on matters of textuality and interpretation, focus has shifted away from the traditional literary interest in textual interpretation and toward new-digitally dependent phenomenological and hermeneutic methods of reading and studying texts, prioritizing large-scale, global observations of text, mass-digitizations of archives and authorial or generic corpora; additionally innovative ways of viewing a text through digitized platforms; and critical studies of born-digital artifacts, meant to establish new multimodal perspectives on post-computerization textuality. Moreover, the notion of “distant” reading (Moretti) has been cemented in digital humanities theory as the de-facto computational method of textual study, resulting in a (false) dichotomy that supposes close reading as attuned to the human cognition and distant reading as distinctly computational in nature.

By the end of this volume, by executing the command, Computational Close Reading, this study intends to stand against these preconceptions about computer programming and the way it can sustain readings of literary texts. It also digs into fertile ground, carving out more firmly a place for digital applications to literary studies that preserves the disciplinary proclivities that have comprised the field of literary studies since its inception, while still embracing the technological affordances of the modern age.
Literary scholars concerned with the erosion of close reading and interpretation in the advent of computerization will take comfort in observing my thesis: that we can use these digital tools to close read in the same way we’ve always known and loved.

As a literature scholar interested in digital technologies and their impact on the field of literary studies, and by implication the humanities at large, my research deals with exploring possibilities of collaboration between the two disciplines, the humanities and the sciences, charting pathways by which the humanities can make use of computational methods and frameworks to enhance the breadth, depth, and efficiency of humanities inquiry through the use of computational tools.

Much of this work involves revising disciplinary boundary lines in interdisciplinary humanities work that makes use of digital technologies, involving critique of the field of “digital humanities” as it currently operates. In my dissertation I allege that digital humanities work of the early 2000s and into the present day has moved away from a core humanities method, the close interpretation of text and language, in favor of scientific goals like reducing text to figures of data for large-scale analysis, and the building of tools meant to disseminate texts on digital platforms. I argue that this disparity constitutes a “scientizing” of the humanities that threatens the relevance of its traditional disciplinary orientation toward the close reading of text and the understanding of interpretive phenomena in literature.

The premise of my research, then, is vested toward exploring how computational tools can also be used to close read and interpret text, and not simply view text as figures of data or invent tools for digital curation. As such, this research is meant attract the more traditionalist literary scholar who is apprehensive toward computational application for
the same reasons I locate in my research problem, assuring them that computation can serve literary studies from within its own scholarly framework rather than assuming the likeness of a more data-based, scientific discipline.

Challenging the notion that computer application to the study of text necessitates large-scale, data-based perspectives on language and writing, my dissertation culminates in the presentation of Python programming codes, a platform traditionally used for data analysis, that are capable of retrieving passages from literary texts based on their content and type to be used as the subject of a focused interpretive study. Similar to a keyword search, these codes can extract quotations from a text based on contextual commands in Python based on the type of passages the user is interested in examining. For example, it can be used to retrieve a list of all passages of dialogue in a literary text for a scholar interested in examining their specific details. This can be expanded, through further experimentation in Python, to retrieving details like metaphor and other forms of figurative language, as well as acute literary features like narrative viewpoint and internal character contemplation.

In this way, I harness computerization to expedite an analytical endeavor particular to literary studies, rather than view the material of literary studies through the frameworks of the sciences. A literary scholar can make use of these Python functionalities to gather quotations for their own analytical purpose, whatever that may be, and use it to perform interpretive work on those passages for the purpose of a broader literary inquiry. The invention of these codes legitimizes my oppositional stance toward the digital humanities, that computational approaches to the study of text necessitate large-scale data-based observations of texts because computers are predisposed to
handling large multitudes of information. Rather, I set out to prove that computers can be used to study the minute details of text as well, and in a more efficient way than if we hadn’t used them. It aims to assure the field of literary studies that we can make use of computation in our scholarly endeavors without changing the field as we know it. A more disciplinarily conscious form of technological innovation, my research innovates the original methods of my native field, instead of inventing and introducing new ones.

This dissertation adheres to the University of Rhode Island Graduate School’s standards for thesis and dissertation formatting.
## TABLE OF CONTENTS

Abstract ........................................................................................................... ii
Acknowledgements ......................................................................................... iv
Dedication .......................................................................................................... v
Preface .............................................................................................................. vi
Table of Contents ............................................................................................. xi
List of Figures .................................................................................................. xii

Introduction: Problems with Digital Method .................................................. 1

Chapter One: History of Digital Literary Studies ............................................ 32

Chapter Two: Methods of Computational Close Reading .............................. 81

Chapter Three: Examples of Computational Close Reading .......................... 125

Conclusion: Computational Close Reading ..................................................... 163

Appendix .......................................................................................................... 181

Bibliography ..................................................................................................... 183
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>24</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>75</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>84</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>93</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>105</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>107</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>115</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>118</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>119</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>120</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>121</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>122</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>129</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>130</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>143</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>151</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>173</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>175</td>
</tr>
</tbody>
</table>
Introduction: Problems with Digital Method

“Print books are far too hardy, reliable, long-lived, and versatile to be rendered obsolete by digital media. Rather, digital media have given us an opportunity we have not had for the last several hundred years: the chance to see print with new eyes.”
- N. Katherine Hayles, *Writing Machines*, pp. 33, 2002

“At its core, then, digital humanities is more akin to a common methodological outlook than an investment in any one specific set of texts or even technologies.”
- Matthew Kirschenbaum, ADE Bulletin (150), pp. 56, 2010

“The next generation of literary and aesthetic theorists who will most matter are people who will be at least as involved with making things as with writing text.”

“… this book is aimed at bringing readers into this world, but always with an eye to concerns within literary scholarship.”
- Andrew Piper, *Enumerations: Data and Literary Study*
The Problem of Interpretation: What we Read, and How

Methodology in literary studies has long been stewarded by the hermeneutic circle, teaching us ‘what’ to read in a literary text—its context, the figurative subtleties of language, its connection to or interconnection with other texts which embody historical or philosophical relationships to the fictive representations. Methodology in literary studies dictates how we read these things: distantly (distant reading), closely (close reading), at the surface (surface reading), hyper and machine, and others of note. When computation was introduced into the fray so came preconceptions about how computation as a method can fit into the analytical imperatives underscored by the hermeneutic circle: ‘how’ we can read the ‘what’s’ of literary texts.

Hyper and machine reading are easier to address here, as they typically entail reading from a computer screen, but are still relevant for how they meet the demands of hermeneutics: hyper reading is often composed of hyperlinks which draw attention to connections between and among relevant texts, and machine reading is more of an automaton, endowing the computer with the ability to ‘read,’ or more precisely, to extract and scan meaning from texts. Distant reading has long been viewed as the defacto digital method—the ‘how’ of digitally applied literary studies, that because of the nature of computation and its ability to aggregate and retrieve information en masse that this is the mode by which computer tools ought to be applied to the study of texts. In a number of studies that I survey here, close reading is pushed to the background in favor of distant, large scale studies of textual features.

*Computational Close Reading*, then, intervenes in digital humanities and
American literary scholarship to negotiate a significant institutional and disciplinary problem: the devaluation of close reading and interpretation in digital humanities applications to literary studies in favor of more distant modes of reading. To remedy this issue, I pursue a core question meant to bring the literary and the digital into a more perfect confluence, as a demonstration of possibilities meant to run counter to these preconceptions about computational reading: how “close,” if close at all, can digital technologies bring us to the interpretive elements of literary texts? I argue that they can bring us quite close, opening a perceived gap in knowledge about the ‘hows’ that are possible in computational reading. As a matter of necessity, I also speak to the questions that underlie this core issue in the humanities, such as why the computerization of literary studies has allegedly moved away from interpretation and close reading in favor of digital curation and digital tool-building; what other modes of reading computerization has given way to, and how these differ from traditional approaches (close, distant, hyper, machine, etc.); what is so different about traditional literary studies and the resulting digital turn, and why this disparity came to be; and, perhaps most importantly, how the analytical affordances of the digital humanities can be leveraged to advance and preserve the original goals of literary studies as a discipline, instead of enabling their supersession. Put plainly, *Computational Close Reading* challenges the notion that digital humanities necessitates new approaches to literature—rather, I am interested in how digital technologies can enhance our original approaches to literature and reaffirm their universal relevance and importance. This volume intends to argue, at its core, that the age of computerization does not make obsolete or short-sighted the traditional methods of
literary analysis in the face of new approaches to reading, but rather encourages us to pursue those same original disciplinary goals with renewed perspective and capability.

In doing so, *Computational Close Reading* intends to speak at the intersection of the digital humanities and literary studies as a piece of methodological theory, but also as a demonstration of what I will term throughout this volume as computational reading, a mode capable of operative at both distant as well as close proximities to the text’s more subtle interpretive details that have long undergirded the methodology of literary studies, at once appeasing the traditionalists as well as encouraging the visionaries. As such, it is meant to serve as a road map for negotiating the problem of the digital and the literary, of the difference between the digitally enhanced analytical scholarship I have observed in the digital humanities and the traditional close interpretive work of literary studies, suggesting ways that they can be brought together as they seem, I argue, to be persistently pushing each other apart. While not necessarily purporting to have devised “the way,” this volume intends to suggest “a way,” or, more humbly, to model a semblance of ways in which literary interpretation can be preserved as a scholarly endeavor even while we continue to embrace the technological and institutional evolution that is the digital humanities. The first way to go about doing this, as I outline in this introduction, is by resituating the primacy of literary interpretation in the humanistic study of texts, digital or otherwise.

To accomplish this theoretical goal in regard to close reading and the digital humanities, I use my computer application to pursue a secondary, practical literary thesis: the close reading of immersive narrative structure in late nineteenth-century women’s realism. In addition to my being a postbellum Americanist by trade and comfortable
using these texts for this demonstration, this question is also warranted by the literature at hand, a further demonstration of how these new methods can serve a traditional analytical endeavor. While American realism is concerned on one level with depicting the internal psychological machinations of characters as they experienced varying social, economic, and political tribulations in their lives, literary regionalism, in large part written by women, was equally concerned with the interpretive proximity by which readers receive the aesthetic of place as a thematic element, endeavoring to convey its fictive effects by immersing the reader in the cognitive experiential viewpoint of the narrative. The close reading, then, of interpretive proximity on various sensory levels provides a unique opportunity to also study the closeness and depth of reading attainable through the use of computation; each implicate the other in a dual thesis that contributes to literary as well as technological knowledge, as a truly interdisciplinary project should and could. By demonstrating that narrative interpretive proximity can be studied through the use of computational application, I also demonstrate that digital humanities frameworks can be utilized to serve close reading.

A useful caveat worth explaining however would be that my digital humanities thesis about close reading through digital application is foregrounded in this volume, and the applied literary inquiry serves as a possible example of the ‘types’ of close reading possible through frameworks similar to mine, as I view them to be innumerable. I argue that digital humanities method can be reconfigured to prioritize traditional humanistic inquiry like close reading without adapting or changing its methods and interests to resemble scientific or statistical scholarship, or to foreground technological innovation as its primary goal. I do this by contradicting original preconceptions about digital
humanities application as being intrinsically suited to distant readings of large-scale
textual features, showing its use in the close reading of a precise choice of literary texts. I
take the study of interpretive proximity as a *translatable* concern to the closeness and
depth possible when reading in digital modes. If digital analytical application can be used
to study the textual manipulation of interpretive proximity in women’s regional writing,
then it has also consequently aided in the process of close reading.

The purpose of this volume, then, is in one part to identify a historical issue in the
Humanities, the ways digital application have enabled a movement away from close
interpretation of language and into the macro analysis of text as figures of data, noting its
permutations and subsequent development; but it is also in a second part to model an
interdisciplinary analytical process that is attentive to this disparity, that demonstrates a
balanced collaboration of disciplinary knowledge without eclipsing one or the other from
their original disciplinary proclivities. Abstractly, this volume intends to boast that digital
technologies can change the way we study texts methodologically without changing the
distinct and important conclusions that we draw from them.

The problem that *Computational Close Reading* speaks to is a clear one: in
Digital Humanities work involving applications to Literary Studies, there is not a clear
balance of digital method and literary interpretation, between engineering computer
methods for analysis versus interpreting and theorizing a close reading of literary texts, a
problem implying that Digital Humanities projects have the potential to privilege
 technological innovation and toolmaking and not hermeneutic literary interpretation
(Dobson; Berry). James Dobson alludes to this problem, saying that “The extraction of
‘information’ about culture from texts, in short, does not fully capture the methodological
variety of humanistic practices” (ix); and David Berry similarly suggests that “the use of
digital technologies can also problematize where disciplinary boundaries have been
drawn in the past, especially considering the tendency of the digital to dissolve traditional
institutional structures” (4).

This problem comes in part through the conception that literary criticism is
grounded in close reading while computer-assisted analytical methods are considered
characteristically distant modes of reading. Scholars attest to the difference between these
two modes of reading as hinged to the difference between the human and the computer:
close reading is defined generally as “the thorough interpretation of a text passage by the
determination of central themes and the analysis of their development,” while distant
reading “aims to generate an abstract view by shifting from observing textual content to
visualizing global features of a single or multiple texts” (Janicke, Franzini, Cheema,
Scheueumann), most commonly executed through data sets and data visualizations.
While thought of as a “macro” approach particular to computer processing compared to
close reading as a “micro” approach perceivable because of the innate human sensitivity
for language (Janicke et. al.), my study challenges this dichotomy to suggest that the data
analysis frameworks particular to distant reading can still bring us intimately close to the
interpretive elements of literary texts; or, more plainly, that digital tools are not only a
distant purview through which to analyze language, but that they can also be used for
close reading.

The digital humanities has long been methodological in nature, concerned with
technological innovation and its effect on the study of texts, and its products are tools
which present new methods for the reading and, at times, the curating and disseminating
of texts (I evidence this observation from the works of McGann; Hayles; Piper; Kokensparger; Gardiner and Musto; Igual and Segui; Driscoll and Pierazzo; Burnard, O’Brien O’Keefe, and Unsworth, who are addressed further in chapter 1.). Looking at the methodological bend of the Digital Humanities, it is worth considering the meaning of interdisciplinarity when it is applied to literary studies as a complicit issue. It is my observation, evidenced in the literature consulted for this research, that pure methodology and toolmaking at times in digital humanities projects takes the place of hermeneutic close reading, aligning the projects more with the disciplinary side of computer science than literary studies for the way they conclude with the presentation of a tool, a digital artifact, or a dataset, in place of an elaborated critical interpretation that explains a close reading. It seems that digital humanities method amounts to, at times, aggregating the material for interpretation and putting it in view in a dynamic, digitized way, while not necessarily interpreting the text, but curating it so that it can be interpreted.

*Computational Close Reading* remedies this problem by demonstrating the use of digital tools and applied computer methods to pursue interpretation, foregrounding the literary *rather than* the digital.

Matthew Kirschenbaum affirms the first, limited approach, suggesting that by way of its interdisciplinary nature “digital humanities is more akin to a common methodological outlook than an investment in any one specific set of texts” (56). In *A New Republic of Letters*, a book foundational to this project, Jerome McGann defines the digital humanities in a similarly methodological way when considering it in the context of the age of computerization, suggesting that the digital turn created a “great obligation facing literary studies…: the reediting, for online environments, the entirety of our
cultural inheritance” (157). In Radiant Textuality, an earlier work, McGann similarly defines the digital humanities as methodologically curatorial in focus, suggesting that “[t]wo lines of work dominate the period [of the digital humanities]: first, the creation of databases of humanities materials—almost exclusively textual materials—for various types of automated retrieval, search, and analysis; second, the design and construction of statistical models for studying language formalities of many kinds” (3). N. Katherine Hayles, another scholar from which my research takes note, views the development of digital humanities studies in phases, quantitative and qualitative, revolving around the early development of databases infused with search and retrieval functionalities and then followed by a move into statistical data-based modelling which, she argues, is attentive to “the Humanities core methodological strengths: attention to complexity, medium specificity, historical context, analytical depth, critique and interpretation” (26). But these scholars speak of digital tools that are sensitive to these interpretive elements of texts, as if the development of the tool serves as the culmination of the project. I argue that it does not: the substance of interpretation must be paramount in order for the project to be literary in nature.

Viewed in this light, it would seem that digital humanities application has developed in a way to leverage computational toolkits in the service, or assistance, of literary studies and the humanities, as projects that operate and end as a survey of tools for further study (as in the volume Introduction to Digital Humanities, by Gardiner and Musto, which charts a survey of applied data programming methods for textual observation as new approaches). This technological bend is a result, I would suggest, of the impulse for rapid technological innovation influencing the need for new approaches
to reading rather than focusing on the innovation of longstanding methods. Databases can simultaneously house large breadths of material while making them retrievable at a speed and precision ostensibly impossible by hand and in print; statistical modelling and visualization through similar computational platforms create new ways and new contexts through which to view and interpret texts; and hypertext and other multimodal editions of literary texts present expansive and dynamic ways to feature and disseminate texts online, making them more widely available through open access web platforms and more comprehensive in their use of editorial apparatus that elucidate their textual and literary features. Be that as it may, there is a persistent, underlying problem on which my research focuses: a database, a statistical model, and a digital edition, while analytically based, are different from the scholarly monograph that explicates interpretations, the goal of literary criticism, but are merely the raw material for interpretation, tools an interpreter would consult before conceiving an interpretation; thus, the project should not end at the creation of a tool, but at the arrival of an interpretation deduced through the use of those tools, as will mine. Digital application has yet to influence this area of scholarly production.

Digital humanities scholarship throughout the 1990s and 2000s has been interwoven with mediations of this disciplinary tension, tension between toolmaking and interpretation that amounts to the preservation of the disciplinary proclivities of literary studies at the risk of their supersession in favor of technological innovation and toolmaking, and vice versa. As such my research follows suit with the current theoretical trends in modern digital humanities scholarship: “It is vital,” David Berry writes in the 2019 edition of *Debates in the Digital Humanities*, “that the creation of tools drawn from
outside the humanities do not simply supersede the theoretical principles that … should inform DH practice. Yet one complaint about the digital humanities is that, too often, tool making is seen as a substitution for hermeneutics” (65). The crux of this issue lies in the place of interpretation, or lack thereof, in digital humanities projects that purport to be pieces of literary studies; as James Dobson similarly notes, “humanists should not replace interpretive work with tool building” (15). Needed, then, is a study that investigates the key methodological and theoretical liminalities between digital humanities projects as they currently operate and the traditional interpretive procedure. It is the intention of this volume to identify what aspects of literary interpretation that tool-based digital humanities studies essentially neglect in the context of traditional literary theory, and suggest ways new projects can foreground interpretation over their presentation of tools, making them primarily literary in nature rather than computer scientific. My study essentially responds to this important methodological need.

While deriving from the impulse for rapid technological innovation, this project also takes its warrants in part through responding to the way scholars have thought about the act of reading in physical versus digital modes (or through digital curation). Close and distant reading (Moretti) and close, hyper, and machine reading (Hayles) inform an aspect of digital humanities thinking to which my research responds: that digital tools cannot close read as well as humans can, simply because computers have characteristically been used for the mass-aggregation of data. Challenging this assumption, I suggest that digital tools can help the human close read more sensitively or expeditiously than they would otherwise. My digitally informed approach to the close reading of texts seeks to complicate the hypothesis that the digital modelling of textual
data is a characteristically distant mode of reading, and that close reading cannot be rendered as efficiently in these digitally engineered ways. As this distinction is a presumption I have located in digital humanities scholarship, it is unsurprising how digital humanities projects have designed themselves as tools and methods peripheral to interpretation itself: as databases, digital editions, statistical models, and as volumes that survey tools for their usability. The history of the digital humanities that I have observed has been one in which projects have sought to lay bare the material for interpretation, or suggest tools useful in doing so, without progressing further to actually positing an interpretive theory vested in a sensitivity toward the finer interpretive elements of language and text. I argue that, in essence, digital humanists who do not leverage interpretation as the centerpiece of their projects are actually acting as computer scientists or software developers.

In order to investigate and then remedy these two core issues of the field—the reconciliation of tool-making and interpretation in digitally-assisted literary studies, and the extent of close reading possible in digital modes—the principal literature of this research, in addition to the primary texts serving as the subjects of my close readings, are current and prominent digital humanities projects as well as current works of digital humanities theory, providing a perspective on what is being done and how, and, more importantly to my project, to what extent. Integral also are works theorizing method in the digital humanities and literary studies as well as works detailing our current understanding of close and distant reading, all of which are detailed in the succession of my chapters and their respective bibliographies. Prominent projects serve as case studies to be critiqued alongside corresponding works of methodological theory. The extent of
the toolmaking versus interpretation-based focus of digital humanities projects is criticized specifically for where these projects stop: a database, a digital edition, or a statistical model (as what I see in McGann and others like him) is a digital product but not necessarily one of literary criticism unless it simultaneously performs and disseminates the content of an interpretation. Looking for more than just a reediting of literature for online environments, my research will question how and to what extent digital tools can be used as a vehicle for literary interpretation, constituting a digital humanities model that grounds itself not in a passive and peripheral technological assistance to literary studies but in the original disciplinary goals of literary criticism: the interpretation of texts.

As this volume negotiates the disciplinary tension revolving around the digital humanities and literary studies, it attempts to reshape the ways we think about reading in digital modes by providing a model for the active close reading and interpretation of literary texts through digitally applied methods. This critique of interdisciplinarity figures largely into the continued sustainability of literary studies and the humanities. The digital humanities as composed purely of digital methods and digital artifacts will, I argue, quickly dissolve into computer science as it purportedly creates technological products that the sciences have been producing already. Or worse, it will cause an encroachment of the computer sciences onto literary studies that results in the devaluation of critical and hermeneutic interpretation in favor of toolmaking and curation on digital platforms. Needed is a model, or a demonstration, that establishes how digital tools could perform humanities work if they want to be regarded as pieces of literary criticism.
In addition to its disciplinary implications, exploring the differences between toolmaking and interpretation, and close and distant reading, has large- and small-scale theoretical implications for the fields of literary studies and the digital humanities. On a theoretical level, it brings to bear a nuanced perspective on close reading in digital modes, suggesting a model which balances the collaboration of disciplinary knowledge rather than favoring either-or, marking a difference between computationally assisted interpretation (my model), and digitally curated repositories for interpretation (the former, which I identify as the edition, the database, or the statistical model). To remedy the toolmaking dilemma in digital humanities projects and re-centralize literary critique within them, this study establishes literary criticism as its original, primary goal, and merely approaches this goal through the use of computer tools. In this way, by situating literary interpretation as its fundamental goal, digital humanities approaches to literary studies can embrace the affordances of technological innovation without diverting or separating the original discipline of literary studies from its distinct methodological goals: sensibility towards language and the critique of culture through textual analysis. This study will essentially seek to synthesize and increase the digital humanities toolbox, arguing that it can do more than create tools and platforms for reading, that digital tools can be used to close read literary texts.

In its clearest sense, this project endeavors to revise our perspective on toolmaking and interpretation in the digital humanities, establish the primacy of close reading in interdisciplinary literary projects, and demonstrate the use of interdisciplinary method to arrive at a singular disciplinary goal: that is, the use of computer methods to perform close reading. In my textual exegesis I borrow technique from interdisciplinary
fields—such as data analytic frameworks relative to data science such as topic modelling and textual collation—and demonstrate their use as vehicles for interpretation, not merely survey and suggest usable tools. *print(Close Reading)* is a reflective project that posits, supports, and advances an interpretive theory, but that then reflects on its use of computational methods to pursue its textual analysis. The project at the same time grounds itself in the disciplinary proclivities of literary studies, and by its interdisciplinary connections foregrounds innovative methods that are useful or expeditious in the pursuit of the original disciplinary goal. In its culmination, I present a new feasible model for interdisciplinarity that marks the difference between digital humanities tool-making and traditional literary interpretation, makes a case for digitally enhanced close reading, and demonstrates how digital tools and computational methods can be adopted in the service of literary studies without losing sight of the core interpretive aspect of the field: by using digital tools to Close Read.

**Feasible Interdisciplinarity: Implications for Literary Criticism**

This project would not be true to itself if it were not for my second, literary thesis that is necessitated by my practical and theoretical intentions in regard to the digital humanities. Here I will address how and why my methodological theory is symbiotic with my interpretive one: because examining interpretive closeness through digital tools (i.e., close reading) provides a forum in which to study interpretive proximity in literature, specifically late-nineteenth-century American women’s realism. This interpretive ground corresponds with my theoretical intentions because of the statements I wish to make about the digital humanities in a general sense, and projects which wish to
follow suit may or may not formulate their theories in such a symbiotic way. Mine, however, relies on this relationship between its dual theses to drive its central point that digital and literary concerns can coexist without competing with one another. The body of literature I’ve chosen for this study calls for an investigation into its interpretive proximities, and the digital analytic methods I’ve assembled provide a useful tool in doing so. American realism of the late nineteenth century served as a forum for discussions of conflict between personal passion and social law. In doing so it strived to emulate aspects of the individual experience at various levels of the social structure to expose moral inequity in conventional social practice. Realist narratives sought to be experientially immersive to evoke empathy and influence social change or understanding (Bell; Kaplan). Amy Kaplan teaches us that “Realists do more than passively record the world outside; they actively create and criticize the meanings, representations, and ideologies of their own changing culture” (7). Kelli V. Randall, in a similar light, says that “realist writers used realism as a strategy for defining their own social position as authors” (7). Realist narrative forms, then, mediate the reader’s interpretive perception of the elements of the realist experience—their cognitive “distance” from or proximity to the perspective of the individual that is the interpretive ground of realism. As realism connects the reader to the individual perspective through narratively mediated cognitive space, the forms and functions of the realist narrative correspond to the interpretive distance that I would like to explore in regard to close and distant reading.

With my literary interest vested in degrees of interpretive proximity, I choose regional writing, the majority by women, as my subject matter for its fixation on the depiction of place and time and the experiential details therein. Looking at other forms of
interpretive proximity, I thought it logical to stay with women writers as my study also examines narrative depiction of internal psychology. The two pairs chosen for my study—the regional narrative proximities of place and the realist rendering of narrativized internal psychology—provide a comprehensive purview for my literary thesis: that American women’s writing of the late nineteenth century was especially grounded in these immersive narrative techniques as a way to levy their thematic effect and convey their fictive intentions. Taken together, my endeavor to test the interpretive closeness attainable through digital augmentation is necessitated by the immersive functions of the realist narrative. Looking at one implicates the other.

To further express my project’s literary implications, my analytical interest in the interpretive space by which readers narratively receive characters in their fictive environments comes in response to the scholarly study of space and place in realist fiction which has, by my observation, focused fixedly on characters literal navigation of spaces within the story itself, and not the reader’s occupation of space within or outside of the text. Realist scholars like Jill Bergman, William C. Snyder, and Gary Totten have looked at perceived spaces as the vehicles of representation that advance the social critique inherent in the Realist genre (i.e., conceptions of social spaces imbibed with thematic significance). Foundational to this line of spatial criticism are the theoretical works of Robert T. Tally, Yi-Fu Tuan, Phil Hubbard, Edward Soja, and Henri Lefebvre, among others, who author some of the seminal work on space and literary studies.

Their work, to which this research responds with its nuanced perspective on the interpretive proximities of the reader in relation to the narrative, rather than the literal spatial navigation of the characters, draws specific focus on defined social spaces as they
reflect underlying conflicts of human passion. In the fiction of Charlotte Perkins Gilman, particularly “The Yellow Wallpaper,” Jill Bergman observes the home as emotionally connotative by pointing to the nature of narrative perception and aesthetic description, but does not place the reader within the fabric of these perceptions. William Snyder similarly interprets the narrator’s belated descriptions of the bedroom as a type of semiotic play whose fragmented and incoherent nature underscores the mental degradation of the female character as she is confined to the bedroom, but again without recourse to the interpretive place of the reader and the phenomenology of these acts of narrative perception. Gary Totten, as well, similarly discusses natural space as it exists in conflict with consumer culture, and points to the ways consumer culture influences characters’ utilization of spaces to achieve their consumer needs. Similar to the conception of spaces embraced by realist scholars, he again points to the social and emotional connotations of character-navigated space to make inferences about the social conflicts depicted in the text.

My analytical stance on narrativized interpretive space in realism is different in that it steps outside of the text, looking away from the characters and instead back at the reader: the space, distance, and proximities by which the narrative is received by the reader as a mode of aesthetic immersion and narrative viewpoint. It aligns more closely with the work of Susan Sniader Lanser, who is interested particularly in narrative form and narrative voice as a critical locus of authority in women’s writing. “[N]arrative structures and women’s writing are determined not by essential properties or isolated aesthetic imperatives,” Lanser argues, “but by complex and changing conventions that are themselves produced in and by the relations of power that implicate writer, reader,
and text” (5). If women writers constructed narrative forms in careful consideration of the standards of authority in the literary marketplace, then examination of narrative form in women’s writing provides a means by which to study textual methods of conveying their realist conviction to their readers. It is my argument that one of the ways women writers of the late nineteenth century garnered such authority (i.e., the way they established authenticity and accuracy of depiction as a staple of the genre, the Howellsian “truth” that marks American Realism) is by endeavoring the reader to identify directly with the characters, seeing the fictional world through their eyes rather than as an empathetic yet passive observer of a “picture” of life, by narratively immersing them in their emotional, cognitive, and psychological spaces as if they had experienced it themselves. So although characters’ literal navigations of spaces within the text are emblematic of the work’s thematic underpinnings, the reader’s interpretive proximity to the narrative and its characters’ cognitions marks the foundation upon, or the lens through which, the characters and their fictive tribulations are to be observed, understood, interpreted.

Taken together, this project advances a series of theses: firstly, is that digital humanities scholarship has turned away from close reading and interpretation in favor of toolmaking and textual data science, and predominantly distant readings, in part due to the impulse for rapid technological innovation but also for the way digital analysis methods have been thought of as distant modes of reading. This project affirms the first and challenges the latter: that the digital humanities has in fact become more tool-based in focus than traditionally interpretive, and I suggest pathways to negotiate this disparity by demonstrating a digital tool utilized exclusively for close reading, arguing through the enactment that digital tools can in fact be used for close textual analysis, and that the
digital turn in the humanities need not constitute a movement away from traditional methods of analysis, but rather a rejuvenation of the former. Secondly is my literary thesis that I demonstrate as being attainable through my use of computation, running counter to most data-base approaches: that American women’s Realism (namely Regionalism) is narratively immersive in nature, grounded in a narrative form by which authors sought to convey their narrative effect by striving to immerse the reader in the cognitive, emotional, and material aesthetic of their fictive environments. This literary thesis is driven, guided, by my digital one, testament of the symbiotic relationship of the literary and the digital that marks the interdisciplinarity that I argue for in this project. As such, my close readings examine passages that attempt to bring the reader intimately close to the experiential details of the story, simultaneously pursuing this literary thesis as well as testing the interpretive closeness attainable through digital application tools, using them to examine the interpretive proximity of the reader in relation to the characters and the aesthetic of the text. My project supports the notion that digital analysis methods can be used for close reading by using them to explicate techniques of interpretive proximation in American women’s writing. By using digital analysis tools to close read interpretive proximity, I support (and reiterate) the notions bemoaned by the digital humanities scholars mentioned above: that digital humanities should not (or rather, need not) move away from interpretation in favor of tool-building.

Chapter Outlines: Progression and Development

The chapters of this study progress in a way that effectively supports these practical and theoretical notions in regard to the digital humanities and literary studies.
Collectively, the volume serves as a demonstration of feasible interdisciplinary literary study, but it also brings to light aspects of the history of the digital humanities that I view as important to note for the way they have essentially capitulated to this disparity between interpretation and tool-building. Responding to the historical observation of disciplinary method, I provide through the use of computer tools the exact type of close reading and deep textual analysis that has formerly been envisioned as outside the bounds of digital analytical methods, and I reflect on those tools and methods in an instructive way, showing the literary scholar precisely how to utilize these tools in the pursuit of a literary goal and not a technological or data-based one. The study then culminates in a theoretical statement about the nature of reading in digital modes, critiquing the extent of interpretive distance inherent in their use and arguing for the closeness attainable with them in light of what is demonstrated in the preliminary chapters.

As such, chapter one of this volume is termed “Digital Humanities from the 2000s to the Present Day,” and is spent essentially proving my initial allegation: that digital humanities projects have indeed moved away from interpretation in favor of tool building and data science. I support this allegation by first reviewing the history of digital humanities theory embodied in significant scholarship and theoretical works. I highlight statements by Jerome McGann and N. Katherine Hayles, among others, who argue for a new realm of digital curation and new experimental modes of reading and writing in their view of the development and potential of the digital humanities. I point to these theories of textual innovation as enablers of the movement away from close reading as they implicitly focus on the ways the constitution of textuality has changed, and presume that
the expansion of digital methodology means prioritizing large-scale data-driven analyses of literature and writing.

In addition to the explication of theory, in chapter one I also survey prominent digital humanities projects from major digital humanities institutions and scholars, revealing the production side that corresponds with the theoretical. It is my intention in this chapter to survey digital humanities work from the late 1990s to the early 2000s and catalogue them in a series of types that illustrates my concern about digital methodology. It will be my argument that digital humanities projects fall into the following categories that set them apart markedly from the traditional interpretive work particular to literary scholarship: as databases containing the corpora of an author or authors; as editions with experimental capability at featuring supplementary editorial apparatus; datasets of large scale textual features; and as data visualizations, i.e., geo-spatial mapping and digital illustrations of literary settings and features. Digital Humanities projects that do not fall into this format are more often datasets or surveys of data-driven approaches to linguistic features of literary texts, unique in their own right but different from my interest in traditional close reading, or they appear as volumes (at-times edited volumes) that survey the use of tools for their affordances and capabilities. Chapter 1 will survey and catalogue digital humanities projects in these various types to point out what they do but more importantly what they don’t do, comparing their digitally aided outputs with the traditional interpretive procedure.

The second chapter, “Computational Close Reading,” explicates the digital method behind the readings that I will present in chapter 3, as being examples of the type of interpretively-driven digital literary studies that I am trying to elucidate and advocate
for. Here I give a precise overview of Python scripts used to aggregate the raw material for the interpretation I deliver in the third chapter. The purpose of this chapter is to show the use of interdisciplinary method while prioritizing an original disciplinary goal, so description of Python coding functionalities are particularly tailored to be understood within the context of literary studies specifically. By doing the interpretive work first, and then reflecting on the digital tools used to do so, this chapter effectively demonstrates that digital tools can perform humanities work without losing sight of interpretation in favor of toolmaking, or obfuscating literary inquiry with data. It begins situating a model affirming that digital humanities projects can create tools, but that they must also show what they do in addition to explaining how they work.

Here I provide a preview of this coding exhibition as a way to contextualize what will be done at a larger scale in the chapters themselves. Python programming is accessed through the desktop client known as the Anaconda Navigator, which houses the Jupyter Notebook wherein codes can be written to analyze and manipulate data. For the purposes of this project, my textual datasets are the Gutenberg editions of the texts I’ve selected, as these texts are out of copyright and Gutenberg makes them easily available for download in various file formats that will be readable in Python, namely the portable document format (PDF).

A useful, preliminary example of programming code for textual analysis used to advance a literary inquiry can be seen in a textual collation process executed through a Python script. In this instance, we have a classic task in textual criticism: the adjudication of a copytext decision based on evidence of authorial or non-authorial revision. One detects evidence of revision by parsing successive editions of a text for differences,
making inference about their significance and type. In the case of Edith Wharton’s *Ethan Frome*, a text dealt with in this volume, two important editions encompass its editorial lifespan: the 1911 and 1922 Scribner’s editions, each published during Wharton’s lifetime, and therefore the subject of possible authorial revision.

In answering the question of which is the more authoritative edition, editors look for substantive variants, revisions across the texts that affect the substance of its content and representation, as these are attributable to an author and not a compositor because substantive revisions are thought to be exclusively within the author’s domain (Greg, Bowers). The following Python code allows us to parse and detect these differences instantaneously, which is more efficient than doing so by way of sight checking, using online tools with specific file format preferences, or using machinery of the Hinman Collator, which is accessible only at the Folger Shakespeare Library and at select research university libraries:

```python
import textract
text1 = textract.process('ethanfrome11.pdf', method='tesseract')
book1 = (text)
from nltk.tokenize import sent_tokenize, word_tokenize
text2 = textract.process('ethanfrome22.pdf', method='tesseract')
book2 = (text)
from nltk.tokenize import sent_tokenize, word_tokenize
import difflib
import nltk
import re
text1 = textract.process('ethanfrome11.pdf', method='tesseract')
book1 = text.decode()
text2 = textract.process('ethanfrome22.pdf', method='tesseract')
book2 = text.decode()
from nltk.tokenize import sent_tokenize, word_tokenize
from IPython.core.display import HTML
words1 = nltk.word_tokenize(book1)
words2 = nltk.word_tokenize(book2)
htmldiff = difflib.HtmlDiff()tbl = htmldiff.make_table(words1, words2, context=True)
HTML(tbl)
```

(Figure 1.1)

Innovative as compared with digital collators currently available, this script is able to read any file regardless of its format. In essence, this script processes a photocopy
image of a page of text as textual data, allowing editors to use it regardless of the file format that happens to be available digitally. In this instance, the two respective editions are processed as portable document format facsimiles of the original editions, named here as “ethanfrome11.pdf” and “ethanfrome22.pdf.” Downloading and saving them in the appropriate directory makes them accessible to Python. The textract.process in the beginning of the code instructs Python to view the .pdf image as a body of textual data—a process known as tokenization, marked in the code process nltk.tokenize. Once Python recognizes the .pdf images as textual data, .difflib instructs the program to juxtapose the two datasets and detect differences. The final process, tbl = htmldiff.make_table, instructs the program to output these differences in a table graph aligning them side by side for easy comparison. The parenthetical code (words1, words2, context = true) contains the two datasets in what are called variables, or abbreviated ways of naming them without writing the entire path, and “context = true” means for the program that it will output data meeting the criteria of “true difference” via difflib.

The result is a table useful to consult for a copytext decision based on the detection of substantive or accidental variants. The collation code provides a scrollable table, a list essentially, of all instances of variation surrounded by five words on each side. An abbreviated image of the table below illustrates this as a digital method that expedites the literary analytical process. Advancing the literary goal, these instances of variation can be read to interpret the development of the text’s particular thematic moments:
The snapshot here shows the program’s detection of a significant substantive variation across the two texts that implies revision on part of the author. With the 1922 edition on the left and the 1911 on the right, we see a markedly different rendering of Ethan Frome’s representation as a character that evidently changed across the two editions. Here the program allows us to dispense with a close reading to answer a literary question about the text’s composition. The absence of the word “strong” in the ’22 after its presence in the ’11 implies that Wharton thought carefully about presenting Frome as a “bleak and unapproachable” “ruin of a man” (3), and that she removed any such phrasing that would imply the opposite about her main character. Observing other instances of variation that follow this trend—i.e., redactions or additions that solidify thematic effect and characterization—allow me to infer that the 1922 edition is more authoritative than the ’11 because it was more substantively revised.

In Chapter two of this project, apart from this example, I utilize Python programming for similar purposes but more so in the realm of literary interpretation.
rather than textual or editorial criticism, although such editorial notions are no less interpretive and dependent on close reading. Using the Gutenberg editions as my materials since copytext is not a primary concern, I merely aggregate textual details through different context commands to be assembled as the reference points of my close reading, a form of applied topic modelling. In this case I use codes for topic modelling to extract passages that serve as the subject of my textual exegesis, or for examining passages for their representation of internal psychology and narrative viewpoint. Similar to the previous code which extracted passages based on a keyword feature and surrounding figures, commands are written to extract passages that involve certain textual characteristics particular to the author’s rendering of character psychology or narrative viewpoint, commands and specificities which need to be translated into the coding processes that I write. Like the process above, it strives to expedite the analytical process by organizing the textual data based on the warrants of the analysis at hand. In each instance, the running argument is that the digital tool is used to assist in a primarily literary task, and paramount to the study is the literary conclusion drawn from the use of the digital tools in question. Taken together, using Python code to collect passages based on their subtle linguistic features will support the notion of digital tools and data analytics applicable as a vehicle for close reading and textual analysis.

After supporting and illustrating my initial allegation about the digital humanities, and introducing my digital method as a more feasible alternative, Chapter 3 moves towards suggesting pathways for the negotiation of this conflict by elaborating an interpretive theory deduced through the use of computer tools, namely the Python programming language for data analysis. The third chapter “Narrativized Interpretive
Proximities” is very traditional in nature, involving close readings of American women’s Realism from the late nineteenth century, namely The Awakening, by Kate Chopin; “The Yellow Wallpaper,” by Charlotte Perkins Gilman; The Country of the Pointed Firs, by Sarah Orne Jewett; and Ethan Frome, by Edith Wharton. To support my thesis about the literary and the digital and to model the interdisciplinary process, I elaborate upon the interpretation drawn from the topic modelling and coding processes overviewed in the previous chapter.

I choose these texts because they provide a comprehensive spectrum of the ways in which realist narratives render subjectivity in narrative form and attempt to immerse the reader in the experience—cognitively, psychologically, and emotionally—of an individual character. Taken as pairs, The Awakening and “The Yellow Wallpaper” each strive to emulate the internal psychology of a woman character experiencing emotional and otherwise sexual oppression at the hands of dominant male figures, yet they render these depictions differently albeit for the same purpose: The Awakening is narrated by a third person, selectively omniscient narrator while “The Yellow Wallpaper” is a first-person monologue. I examine each narrative for how they variably attempt to bring the reader intimately close to the experiential details of their characters by way of their differing narrative forms.

Ethan Frome and The Country of the Pointed Firs will be approached differently but for the same concerns for interpretive proximity. In this case, taken again as pairs, the two texts constitute narratives concerned with place and its effect on an individual’s internal psychology—Ethan Frome is oppressed by the dismal landscape of Starkfield while the unnamed narrator of The Country of the Pointed Firs is rejuvenated by the
scenery of Dunnet Landing. As such, their third person omniscient narratives fixedly function to immerse the reader in the character’s aesthetic surroundings, emulating their emotional affect. Another form of narrativized interpretive proximity, they seek to bring the reader into a physical place in time so that it can be perceived on various sensory levels through the perspective of the character.

For the purpose of variety, I will also model in this chapter a reading of a single text, *Ethan Frome*, studying its aesthetic immersion but through a different interpretive purview; that is, studying its revisions across multiple iterations. This method presents a unique opportunity to view the text in its composition through text comparison capabilities particular to computer code. I introduce an additional opportunity for computationally assisted close reading by using the program to compare two different editions of *Ethan Frome*, the 1911 and 1922, both by Scribner’s, and observing the changes that took place over the course of its editorial lifespan. This analysis, like the former, relies on the computational platform to detect differences and contextualize them in their form and type, as I will illustrate in their exhibition. While the initial readings will observe the operation of these immersive techniques, this singular, concluding reading will show them being developed and coming into fruition in process.

It is necessary to foreground these readings because they will later support the applicability of computer tools to the close reading of texts, a central argument in this study. Following an analytical framework similar to traditional literary critique, I demonstrate how the same interpretive aspects considered in traditional close readings can also be located as objects of data analysis (in my case, textual data for the analysis of narrative form) through methods formerly thought of as distant in nature and not close.
By using methods thought of as “distant” in nature to read these texts “closely,” I attempt to reconcile literary interpretation and digital method by turning these assumptions on their head, so to speak, revealing them as something of a fallacy rather than a reality.

The fourth and final chapter, or conclusion, is largely theoretical in nature. Drawing on digital humanities scholarship, data science, and literary theory, I discuss our current understanding of close reading and the digital humanities and retrospectively challenge the notion that the study of texts through digital application necessitates a characteristically distant mode of reading. An additional reflection on chapters two and three, this chapter discusses works that theorize the difference between close and distant reading and their applicability to digitally enhanced textual study, namely that of Franco Moretti, who coins the term. I align the interpretive work done in chapters two and three with these theories and argue that the demonstrations from the prior chapters constitute a nuanced sense of “closeness” attained through the use of computer tools, that I had used them to perform a close reading in the context of these current theories. Additional support for the “closeness” attained by my digital analytics will be in discussions of data science as a vehicle for interdisciplinary connection between the literary and the digital. Readjusting against previous applications, I similarly argue that viewing texts as data can be similar to the way texts are viewed in the context of traditional literary theory as long as we prioritize the type of knowledge we wish to draw from these materials, and that the objects of literary interpretation and close reading can be reduced to data if harnessed properly.

In its culmination, *Computational Close Reading* presents an additional conception of the ways we can approach literature through digital application. By no
means do I intend to undervalue or disregard the digital humanities work that has characterized its inception and movement into the twenty-first century. It has done so for a reason, no doubt, and the nature of the digital humanities as it has come to be known is testament to its inherent value as an academic institution in the humanities. I only suggest that it has, at times, looked too far ahead at too fast a pace and has run the risk of forsaking the values and principles that led to its inception in the first place. Rather than new forms of textuality creating new ways of reading, my attention is turned to new methodology that can enhance the ways we have been reading all along, giving space for innovating the traditional act of reading rather than fundamentally changing it in its entirety. It is a more historically sensitive form of technological innovation, one that does not do away with the present in favor of the new, but that uses the new to enhance the present.
Chapter 1: History of Digital Literary Studies

“When we use books to study books, or hard copy texts to analyze other hard copy texts, the scale of the tools seriously limits the possible results. ... In a similar way, electronic tools in literary studies don’t simply provide a new point of view on the materials, they lift one’s general level of attention to a higher order.”

- Jerome McGann, Radiant Textuality, pp. 55, 2001

“This book [Enumerations] offers a new perspective on the significance of quantity for the study of literature. Inspired by the emerging fields of natural language processing, machine learning, and text and data mining, as well as a host of colleagues beginning to work in this area, Enumerations explores the quantitative dimensions within texts, the ways in which the repetitions of language lend meaning to our experience as readers.”

- Andrew Piper, Enumerations, pp. 3, 2018

Quantitative formalism, distant reading, algorithmic criticism, macroanalysis, computer-aided text mining: whatever we call and however we conceptualize these methods, they add, in one way or another, forms of computation to our understanding of the interpretive act. ... This book takes as its founding assumption that the exploration and investigation of humanities texts and data with sophisticated computational tools can serve the interpretive goals of humanists.

- James E. Dobson, Critical Digital Humanities, pp. viii, 2019
Introduction: The Phases of the Digital Humanities

This chapter argues directly that the digital humanities as a field has historically neglected the traditional interpretive work of the traditional humanities in favor of curation and tool building that, while innovative, has moved away from the core methodologies for which I advocate in this work, that it has embraced datafied and quantitative approaches of the study of text rather than enforcing the qualitative, evaluative nature of hermeneutic study that marks literary scholarship. I support this by explicating the digital humanities’ historical development, pointing to the disparities as they arise and noting how they have developed the field into what it is today.

Of primary concern is the field’s attentiveness to and use of the hermeneutic circle as a theoretical backdrop to its scholarship. I argue firstly that the act of digitization that characterizes the first phase of digital humanities does not go so far as to contribute to a direct understanding of a text’s hermeneutic composition, but merely raises the text up to a platform for observation, curation, and that the ideas of “mass digitization” that sprung from this initial stage are what led to the “breadth over depth” that marks latter forms of DH work in data-esque approaches to literature. This latter form is of special concern for theoretical as well as disciplinary reasons. Amid making literary study appear more as a statistical science rather than a phenomenological art, which I view as equally problematic, it has moreover created a different type of hermeneutic questioning that replaces a text understood in its precision and immediacy with a breadth of analysis that seeks to grasp texts and textual features in a broad scope, which as we know, runs opposite with what we associate with close reading, the close analysis of subtle and finer interpretive elements by line, rather than by entire corpus.
As a matter of precise focus, I will trace the development of a specific line of thought that I argue characterized the late computerization of the 2000s and persists into now, whereby the curatorial and editorial nature of early digital humanities application originally led to the mass aggregation of textual materials for the purposes of digitized editions and archives, and then led to the resulting interest in large-scale analysis as a follow-up to large-scale digitization. I see the period in question as integral to creating the digital humanities that I allege us dealing with in my introduction chapter, because the origins of computerization as a tool for digitization led to its use in archival database construction which led to its further use in, primarily, macro data analyses of texts. It was a moment in which the field began thinking bigger and bigger and bigger of itself and its capabilities—first with larger and more comprehensive editions of texts, then the same with databases of historical and authorial corpora, and finally in pushing the *breadth* of critical observations possible (at the expense, I argue, of depth), and the desire for closeness was put on hold in favor of this newfound interest in increasing our analytical *scale*. In this view, the digital humanities was a field interested in questioning how much more analytical material we can aggregate now than has been previously possible in print. Jerome McGann is especially telling in this respect, as the central axioms to much of his work that I cover here stress that “Computerization made much more information (and much more varied information) available—vast amounts of data in forms, relational as well as facsimile, that were previously unimaginable” (*Radiant Textuality* 16). This initial disposition that came about during the first and second phases of digital humanities went on to dictate and influence the nature of digital humanities that we have today.
The third “phase” of digital humanities work that I will observe, those interested in macro-level analyses of texts—or distant readings—will be of special concern for the theoretical presuppositions they have of themselves. I also acknowledge, part out of courtesy and respect but also a genuine belief, that the digital humanities was not consciously aware of the disparity it was creating between close and distant reading, but that late digital humanities practitioners, particularly Andrew Piper and others advocating for a “data-centered” or “text-as-data” approach to literature, believed that these distant reading techniques acted as complements to the traditional modes of literary analysis, priding themselves on the new forms of knowledge that they brought to the table. I can accept this notion because it is very well true that, for example, Andrew Piper is able to observe the entire historical genre distribution of novels from the late eighteenth century to the early twentieth through plot and topic models, providing insights on literary reception and canon, or in his words “the way this type of novel has mattered to readers” (Enumerations 78) of this period. These are important insights not as efficiently attainable through print and hardcopy means, and without interdisciplinary borrowing of established data science frameworks for the manipulation and investigation of such large bodies of knowledge. While comprised of exciting, new approaches to new literary questions, my response to them is a simple one: in the composition of these projects, where is the close reading? Where is the contribution of new computational machinery to traditional, longstanding methods? It is the argument in this chapter that amid the various incarnations of digital humanities that I will observe here, a close reading through computational means has not been performed in actuality, or in the way I advocate in future chapters. Instead, digital application enlarged the scope on the various scales of
literary scholarship: through larger editions, larger archives, and larger analytical purviews.

These phases of scholarly pursuit that I locate on the practical side of the digital humanities were also foundationalized by the field’s core theoretical underpinning, gleanable most clearly from N. Katherine Hayles, who I view as a seminal figure, who stewards the endeavor to understand the ways in which the constitution of textuality has changed with the advent of computerization. An additional act, as I argue, of literal and metaphorical expansion, Hayles encourages us to implicate computation’s multimodal representational nature into hermeneutic understandings of textual expression, reception, and interpretation, which dove tails with the “reediting, for online environments, the entirety of our cultural inheritance” (A New Republic of Letters 57) that Jerome McGann describes on the practical side of things in regard to archives and editions, suggesting that the digital environment’s representational abilities allow it to harness uniquely large breadths of information. This theoretical backdrop, I argue, encouraged the field’s persistent pursuit of “newness” on the various scales of scholarly work: with the way works were prepared for dissemination (i.e. representationally dynamic digital edition and archive construction), how and where works were stored (digital archives and databases), and finally in how they were studied (distant reading and data-centric techniques). My observing the production of digital humanities scholarship in this chapter must also account for the corresponding theory that coincided and influenced its production. As such, while surveying the projects that were produced and advocated for, I will also give recourse to their theoretical orientations, questioning the nature of the contribution that the respective authors claim they make to the field of literary studies. I
will critique these claims for the recourse or lack thereof that is provided to close reading as a foundational disciplinary method, whether they function to enable its innovation or supersession.

Phase One: Transmedial Modes of Representation, the Late 90s

At the outset, as these works presuppose of themselves, scholars like Jerome McGann seem to have a thesis similar to mine, appearing vested in the humanistic interest of harnessing computerization to deepen our interpretive capacities. The thesis in McGann’s preface to *Radiant Textuality* from 2001 is telling in this respect in the way he speaks about humanistic engagement with digital technology and the importance of the humanities’ interpretive procedures:

> Digital technology used by humanities scholars has focused almost exclusively on methods of sorting, accessing, and disseminating large bodies of materials, and on certain specialized problems in computational stylistics and linguistics. In this respect the work rarely engages those questions about interpretation and self-aware critical reflection that are the central concerns for most humanities scholars and educators. Digital technology has remained instrumental in serving the technical and precritical occupations of librarians and archivists and editors. But the general field of humanities education and scholarship will not make use of digital technology seriously until it demonstrates how its tools improve the ways we explore and explain aesthetic works—until, that is, they expand our interpretational procedures (xiii)
In this preface McGann does well to locate the problem with which I am also concerned, only viewing it as a reluctance of humanities scholars to use digital tools that do not adequately engage interpretive questions, rather than connecting it to the field’s natural developmental tendencies, as I would like to. By “precritical occupations” McGann refers to the editorial and curatorial methods that underlie and structure existing conceptions of the potential of the digital humanities, noting its universal acceptance among librarians, archivists, and editors, essentially presupposing that scholars or producers of scholarship are either separate from or not an integral part of this mix of people. Stating that “the general field of humanities education and scholarship” will not embrace digital technology until “it” demonstrates its contribution to interpretive capacities is telling to me, because it rhetorically implies that the technology and tool will prove this aspect of itself autonomously of the scholar that uses it. I leverage that obligation to the scholar, not the machine, and the direction of the scholarly ambitions that the scholar puts forth in their use of computation; so in this respect while I loosely agree with how McGann locates the problem, I find special issue in his suggested remedy—that we use technology to “expand” interpretive procedures (add more avenues and variety of inquiry)—as if only the new approaches are warranted and required by the use of computation.

In respect to using technology to innovate longstanding methods versus using them to conceive of new methods amounting to what we know as distant reading, McGann follows this path by way of an interest in developing theoretical models for multimodal textuality in order to comprehend the digital environment in a semiotic context. With almost a revolutionary air he warrants that “Ideas about textuality that were once taken as speculative or even imaginary now appear to be the only ones that have any
practical relation to the digital environments we occupy every day. So that now all of aesthetic, literary, and humane studies appear brinked for major changes in the ways they will be studied, analyzed, and interpreted (2). Or, more dramatically, that “These changes will bring to the center of scholarly procedures theoretical models that have been perceived until now as odd, idiosyncratic, nonnormal” (1). McGann is keenly interested in using the computer as a multimodal reading environment to give an expanded view on the nature of textuality in the advent of digital multimodal representational capabilities. This is certainly a necessary improvement to our understandings of aesthetic, but in its essence, I argue, it is already gesturing away from the core focus on the interpretation of written language that I advocate prioritizing in my digital analytic. But I see this even more so from his own edict on the difference between librarians, archivists, and editors, and the general field of humanities education and scholarship that he divides from one another, when he chooses as his basis of demonstration the Rosetti Archive, a multimodal digital database of the writings of Dante Gabriel Rosetti, an editorial project, nonetheless.

McGann claims he developed the Rosetti Archive as an experiment in digital textuality that would best illuminate “the apparitions of text—its paratexts, bibliographical codes, and all visual features” and “the social intercourse of texts—the context of their relations,” interpretive and hermeneutic factors conceived in his view as “part of the text itself” essential to “an adequate critical grasp of the textual situation” (12). As McGann describes the digital archive, he seems to conflate depth of analysis with breadth of information aggregated, repeating the persistent axiom of digital humanities, that “hypermedia networks and digitization have the means to study visual materials and the visibilities of language in ways that have not been possible before. This
archive was built to harness those capabilities” (13). As such, he goes on to praise the archive for the great wealth of primary and secondary supplementary material that it makes available to users, and the multimodal dynamism that it demonstrates in its ability to show, in color, the visual features of Rosetti’s manuscripts and other creative work, exalting what a “great advance computerization makes in this case” (14) as compared to existing archives that were limited to black and white manuscript facsimiles.

McGann is nonetheless apt to understand the limitations of even this editorial and archival achievement, envisioning that the project and its digitally applied methods could still to an extent model contributions to interpretive and hermeneutic knowledge. He says of his archive reflectively that

Computerization has made much more information (and much more varied information) available—vast amounts of data in forms, relational as well as facsimile, that were previously unimaginable. … As a tool for rethinking these materials, however, whether through structured or randomized searches of data, the computer continually disappointed the high hopes it had raised. … More than anything else, the making of The Rosetti Archive has exposed the gulf that stands between digital tools and media, on the one hand, and the regular practices of traditional philosophy, ‘theory,’ hermeneutics, and arts/literary/cultural criticism, on the other (17)

His suggested remedy for this disparity is what he calls “a procedure for critical thinking that calls for digital implementation” (19), envisaged through a differentiation between the current functioning ground of computational tools and our “other permanent scholarly function” (18), which he calls “the remembrance of things past,” located in the digital
archive and editorial side, versus critical thought toward “imagin[ing] what we don’t know in a disciplined and deliberated fashion” (18). In an effort to involve digital implementation in critical inquiry, as “prosthetic extensions of that demand for critical reflection” (18), McGann theorizes the digital markup of hardcopy documents for search and analysis as a critical and interpretive act imbedded within all digital curatorial projects, something he terms the rationale of hypertext, a theoretical underpinning to the digitization of humanities materials akin, in his view, to interpretation itself.

McGann refers to his rationale of hypertext as a framework to “exploit digital tools to augment critical reflection” (213), but earlier in his argument he shows the same recursivity to multimodal digital forms rather than hardcopy text. “To function in a ‘hyper’ mode,” McGann argues, “an editing project must use computerization as a means to secure freedom from the analytic limits of hardcopy text,” stressing the same expansionist edict that I highlighted before: that “Hypertexts allow one to navigate through large masses of documents and to connect these documents, or parts of the documents, in complex ways” (57), implying that his interest and target of digital application is still within curation and textual editing—for McGann merely speaks to interpretation as it is implicated in the archival and editorial process through the composition of editorial and supplementary apparatus. But I argue that this is merely an act of providing a digitized platform to facilitate interpretation by an audience. It is not a digitally applied project that advances an interpretation in its constitution. Markedly different from the close reading that I advocate for in digital application, McGann’s rationale of hypertext seems still to be meant for the purposes of multimodal editions and archives, even though he chooses to view the act of editing and curating as interpretive.
McGann’s arguments in *Radiant Textuality* culminate in his discussion of another example of digitally-applied critical augmentation in what he calls “The Ivanhoe Game,” a digital platform for critical reflection along topical lines of inquiry, facilitated by the platform’s host of scholarly materials. “The procedure,” as McGann calls it, “is performance based and collaborative and operates in the discourse field of specific historical or literary works or events” (218). Modelling the process of critical reflection around a single textual focal point, the game process facilitates the illumination of “the inter-and-extratextual contexts and relations that these works (as it were) half create and half perceive” (213). So a digital platform for bibliographic research, essentially, “[t]he game ‘moves’ involve the production (the writing) of texts that integrate with and simulate the materials in the discourse field of the game. Players produce texts in response to the opportunities and problems raised by the texts produced by the other players” (218).

While McGann endeavors and does well to view the radiance of textuality in the age of computerization, I argue that his work at this time does not go farther than treating archival, editorial, and bibliographic concerns, and he harnesses computerization primarily to produce “machine-generated interpretive forms” (214) i.e., multimodal editions, digitally curated repositories of information, and digitally-efficient search and retrieval platforms, without giving thought to the interpretation or close reading of text itself through the use of digital machinery. As he says: “The point [of The Ivanhoe Game] is not as such to arrive at a reading or interpretation of *Ivanhoe* but to refashion and reshape its discourse field in ways that bring to the fore ‘possible worlds’ latent in the work and in the materials that transmit the work to us” (221). So if the radiant textuality
of post-computerization means new platforms for editing and archiving corpora of texts, and newly efficient digitally facilitated platforms for search, retrieval, and bibliographic connection, what does that do for our understanding of the radiance of language? I argue very little, for by now we have merely curated text, made connections to other texts, but close reading through the use of these tools and methods seems eclipsed from the discussion.

In a later work, N. Katherine Hayles proclaims *New Horizons for the Literary* in the subtitle to her book *Electronic Literature*, which essentially works to inaugurate a similar interpretive digital landscape. Whereas McGann is demonstrational of methods and case studies in his theories, Hayles is more historically grounded in her approach to digitality, asking “what large-scale social and cultural changes are bound up with the spread of digital culture, and what do they portend for the future of writing?” (2). Hayles mounts her answer to this question in working to situate electronic literature in our overall historical understanding of “signifying strategies” (2) in writing, which she acknowledges stem from print. Hayles draws important correlations meant to expand our knowledge of meaning making in digital forms that, as she argues, occupy a space in our knowledge of literary tradition. Her central purpose in *Electronic Literature* is “a wide-ranging exploration of what electronic literature is, how it overlaps and diverges from print, what signifying strategies characterize it, and how these strategies are interpreted by users as they go in search of meaning” (2), suggesting that “electronic literature can be understood as both partaking of literary tradition and introducing crucial transformations that redefine what literature is” (3). As before, such understandings are crucial to account for the historical development of literary traditions, and our knowledge of their forms and
permutations, but I still feel that a volume dedicated to exploring how “electronic literature challenges us to rethink what literature, and the literary, can do and be” (42) still constitutes an unconscious impulse to move away from traditional methods of analyzing print in favor of a new interpretive landscape that presupposes print reading ancillary or arcane from the place it once occupied, and that digitality “has become the textual condition of twenty-first-century literature” (186), or “that print should be considered a particular form of output for digital files rather than a medium separate from digital instantiation” (159). The alterity of digitality creates a separation from print culture whereby innovation is not applied to print based methods.

Leading the field, Hayles moves toward acknowledging the disparity that I point to, the textual analysis of print writing apart from multimodality, when she considers *How we Think* in 2012, exploring *Digital Media and Contemporary Technogenesis*. In it, Hayles notes that “needed are approaches that can locate digital work within print traditions, and print traditions within digital media, without obscuring or failing to account for the differences between them” (7), which sounds promising, but I am compelled to take issue with her conception of synergism between print and digital media as apparatus for interpretation. Hayles harkens to the problems mentioned thus far when she says: “As scale grows exponentially larger, visualization tools become increasingly necessary. Machine queries frequently yield masses of information that are incomprehensible when presented as tables or databases of results. Visualization helps sort the information and make patterns visible. Once patterns can be discerned, the work of interpretation can begin” (33). As McGann sought to accommodate the interests of the “Traditional Humanities” (Hayles 33) by nuancing theories of textuality so that they
could be applied to thinking about composition in the digital mode, Hayles attempts to establish a framework for leveraging computer tools to aid in print-based textual interpretation, but through the use of databases and visualizations that, I argue, are on one level tools—amounting to toolmaking, as I had discerned before—and on another level, datasets, each of which I have argued are ancillary to the literal act of positing interpretation because they merely lay bare the material for interpretation, and do not render one solely on their own unless consulted—categorized excerpts of full texts.

Hayles reaches this point self-consciously through her own theoretical justifications. In doing so she also creates a gulf between digital augmentation and the interpretive argumentation that I advocate, which I view as problematic, because she presupposes that the two cannot co-operate as I will demonstrate them to do. She refers to interpretive argumentation under the umbrella term of “narrative,” and aggregation of material for interpretation she refers to as “database,” saying that “Because databases can construct relational juxtapositions but are helpless to interpret and explain them [which I view as untrue], it needs narrative to make its results meaningful [equally untrue, because the computer’s content retrieval functions can be directed by meaning]. … If narrative often dissolves into database,” she continues, “database catalyzes and indeed demands narrative’s reappearance as soon as meaning and interpretation are required” (176). While it is true that narrative is necessary to interpret a database of knowledge material, I feel it untrue to think that database aggregation is the only pathway to digitally applied analysis. Moreover it seems to presuppose my original thesis on the digital humanities and distant reading, that digital tools cannot interpret in the way humans can. Digital tools can
however perform correlatable interpretive functions similar to the act of interpretation that humans perform.

Scattered throughout her larger argument are sentiments that I find similarly discerning for their “futurist” and “revolutionary” outlook on digitization’s effect on humanities scholarship, ones that, like in McGann, imply that the field, its modes of analysis, and its analytical interests, are changing and moving away from traditional imperatives. Throughout, Hayles speaks of the synergy of mass-scale databasing and narrativistic interpretation of data as “digital productions [that] produce and catalyze new kinds of knowledge” (8); she talks of “visible signs of a shift” which imply “the changing nature of research… in nearly every humanities discipline” (20); and a conviction that this shift changes both the analytical goal of the humanities as well as its rhetorical and material modes of knowledge dissemination, that “The emphasis on databases in Digital Humanities projects shifts the emphasis from argumentation—a rhetorical form that historically has foregrounded context, crafted prose, logical relationships, and audience response—to data elements embedded in forms in which the structure and parameters embody significant implications” (39). So not only does “Comparative Media Studies” (9) appear a movement away from close textual analysis and toward large-scale aggregations of global textual features, but it also appears a movement away from the narrative monograph as the primary mode of scholarly production. In essence, it presupposes that a database compiled with an interpretive or analytical intention, rife with multimodal visualizations of knowledge relative to its subject, is a “new kind of knowledge” (8) tantamount to the traditional interpretive work of the humanities. My rebuttal would be a simple one: if it isn’t broken why change it? Or more precisely,
introducing new methods and modes of analysis only in the effort to equate them with longstanding scholarly tradition does not properly harness computerization to enhance the traditions we already have. It further distances and diverts our attention from them. I am still questioning these scholars to provide me with a way to use computerization to enhance close reading, not give me new ways and new reasons for reading texts, especially many at a time, as a database would encourage me to.

Anne Burdick in 2012 continues moving in this direction when arguing for transmedial modes of argumentation in her aptly named volume Digital_Humanities, similarly playing on the title of her work but with a different methodology than mine. Like McGann, Burdick takes the same stance toward what she views to be the persistently changing culture of the humanities: “In contrast with most traditional modes of scholarship, digital approaches are conspicuously collaborative and generative, even as they remain grounded [so she feels] in the traditions of humanistic inquiry. This changes the culture of humanities work as well as the questions that can be asked of the materials and objects that comprise the humanistic corpus” (1). Burdick endeavors to “[take] stock of this new world and [anticipate] future developments in the Digital Humanities,” which she feels, as I have alleged of others, are vested solely in the “curation, analysis, editing, and modeling as central to contemporary humanistic discourse” (5). It will be my argument in regard to Burdick’s work that she prioritizes multimodal curation of text and data as a transmedia mode of argumentation that will become the primary mode of scholarly production in the humanities. On this issue I would offer skepticism at best. Of greater import is her conception of analysis via digital application, which I feel she does
not give adequate attention, in so much as claiming that it has given way to distant reading as a new structure for argumentation.

Burdick prioritizes two things in *Digital Humanities*: multimodal databases for knowledge dissemination, and new transmedial modes of argumentation, and not necessarily new methods for traditional modes of argumentation (i.e. close reading). The impulse for innovative modes of representing argumentation come from, she notes, the potential of digitality to reach wider audiences and students, envisioning the use of “the digital as a way of extending the toolkits of traditional scholarship and opening up archives and databases to wider audiences of users” (8). She arrives at this notion implying that the digital shift constitutes a “reinterpretation of the humanities as a generative enterprise” rather than an analytical-argumentative one, “one in which students and faculty alike are making things as they study and perform research, generating not just texts … but also images, interactions, cross-media corpora, software, and platforms” (10). Like others mentioned above, Burdick feels that these modalities of scholarship are complementary to traditional modes whereas I view them as digressive. She goes so far, even, to explicitly separate digital humanities from conventional scholarship: “It [digital humanities] is simply a different, distinct medium for thinking, communicating, and working, with its own rigors and histories, its own skill-sets and language, and its own freedoms and constraints” (11). I argue that digital humanities is a field rather than a medium, as calling it a medium implicitly refers to it as a tool or a platform rather than a discourse community of scholarly inquiry. It is more than that.

Saying that the medium of digitality is the digital humanities is like saying the monograph is the humanities, while we know the monograph to be a mode of scholarly
production shared across various disciplines; but each field’s monograph is different because it follows its own disciplinary analytical intentions. But perhaps we are reaching the crux of the problem here. Perhaps the digital humanities has always seen itself, above all else, as a new medium of representation, without ever tethering itself to a specific, distinct set of disciplinary traditions. Perhaps such is the reason for its nebulous outlook on analytic intention and interdisciplinarity: it does not firmly enough tether itself to the humanities core interests and values—the close textual analysis of language—and instead used its new multimodal dynamism to more or less bring the humanities up to par with fields like data science, software analytics, and computer science for its innovative quality, imitating their likeness in the type of knowledge it pursues and the ways it presents that knowledge. Most things new in the digital humanities, to that end, seem an unconscious attempt to act like the sciences while holding humanities materials instead of vials and graphs.

Anne Burdick seems even to affirm this scientific move in what digital humanities is meant to do in form and function, suggesting that “the generative aspects of digital humanities thus go a long way to addressing the much-lamented atomization and irrelevance of scholarship” (26). She argues that “Digital humanities scholarship,” on the other hand, “promises to expand the constituency of serious scholarship and engage in a dialogue with the world at large” (26). I am not so brazened as to deny these apprehensions about the future of humanities scholarship all together, I am just attempting to argue that a complete overhaul of humanities methodology is not the way to address it. What is needed instead is a reinforcement of humanities core interests with the advent of digitality, rather than a restructuring of it. If the gaze of intention in this field
can be turned backward just slightly, and the focus can be realigned from creating new parts to upgrading old ones, then I think the results would be fruitful, because the digital humanities would ground itself in a core disciplinary mentality instead of almost formlessly oscillating around various disciplines in the name of a loosely defined sense of interdisciplinarity.

Phase Two: Digital Mediation of Analytical Method

We see this shifting, reinterpreting, reconstituting theme around the digital humanities solidified in a different way in the same year, in David Berry’s *Understanding Digital Humanities* (also 2012), an anthology of voices throughout the field but with mention here being given to Bernhard Rieder and Theo Rhole, along with Berry’s introduction, which charts the general outlook of the rest of the volume quite well.

David Berry starts off by noting that “research is increasingly being mediated through digital technology,” claiming like others that this mediation is “beginning to change what it means to undertake research, affecting both the epistemologies that underlie a research programme (sometimes conceptualized as ‘close’ versus ‘distant’ reading)” (1). By “mediated,” I take Berry to mean “contained within” as connoting the extent of his outlook on digital humanities, that it were a container for “representation … what might be called the digital ‘folding’ of memory and archives” (2). He attributes this to what he also calls the “plasticity of digital forms” (2). Similar to Burdick, Berry seems to view digital humanities primarily for its medium specificity as its primary innovative quality: representing knowledge in new mediums of digitality. Not yet is he willing to consider a digital humanities application with its own learned capacity for thought and
analysis, as I would posit about my applied coding methods. As I argue, a digital tool instructed to search and retrieve textual data based on contextual data commands is very much so and at that point thinking for itself as it parses for information. It is a taught entity. More than just a container for representation it becomes the method and producer of an analysis. This should be the digital humanities’ final phase, as it would follow suit with the trend of automation that underlies all impulse for technological innovation: the machine performing tasks for us. I argue that a “representational” digital humanities is not doing anything other than the book, albeit in multimodal ways. A taught analytic tool on the other hand, more than just archiving texts and expanding into a digital canon, or establishing networked pathways throughout and across texts, is actually capable of producing interpretive knowledge.

I bring up the issue of digital humanities “phases” because they are a recurring theme in DH scholarship, and the way they’ve been conceptualized lends an understanding of the field’s digression away from past methods in favor of new, uncharted pathways of knowledge acquisition and preservation. Berry views the digital humanities phases as beginning with building “digital repositories” (4) for archive purposes, and that the second wave merely “expands the notional limits of the archive to include digital works” (4) which implicated humanities methods in understanding “born digital materials” (4); while a third wave, for which he argues, “points to the way in which digital technology highlights the anomalies generated in a humanities research project and which lends to the questioning of the assumptions implicit in such research” (5). So, in other words, a phase to read digital mediation critically for the ways it reshapes and restructures longstanding assumptions about what the humanities is and can
take as its subject of study, “a focus on the computer code that is entangled with all aspects of culture and memory” (5). Berry argues that a critical understanding of code and its influence on humanities research “serves as a condition of possibility for the many new computational forms that mediate our experience of contemporary culture and society” (17). This understanding of code however has no recursivity to the past. It seems by nature bent on understanding code as a complete and fundamental change to the humanities, which would restructure the field’s aims and interests all together, rather than serve as a complement to its core. I argue that, if a third wave of this nature were necessary—I believe it is, for a full understanding of technological mediation will provide a needed roadmap for feasible digital application—then perhaps a fourth phase will make the reflexive move that I am here to advocate for: a digital application that builds on humanities core methods from within rather than adding new things to the outside. This move would result in a four-phased digital humanities to date: beginning with first-phase multimodal archives and curatorial work in the early 90s, which led to the enthusiasm for born digital artifacts to be aligned with our knowledge of writing in the second-phase in the early 2000s, which then developed into a need to understand technological mediation as a reflection on phases one and two. Phase four would be an additional act of reflection but also theoretical development: questioning that if computerization can house, curate, and disseminate humanities materials for us, and is now a persistent arbiter of our knowledge, then how can we use it to take a step further from containing to creating knowledge? This move would make the field remain attentive to the humanities core disciplinary purposes while also moving into automated
interpretation as a developmental step after the digitizing, representative, archival-esque phase of the early 2000s.

With the theoretical ground that David Berry lays out in his edited volume it is easy to see Bernhard Rieder and Theo Rohle, as exemplars, follow through their discussion with similar preconceptions about the digital humanities. Their chapter, “Digital Methods: Five Challenges” is telling in that it charts a series of important fallibilities related to digital humanities work and its reception as legitimate scholarship. At the outset however, they attribute to digitization the proliferation of a deluge of born-digital and multimodal cultural artifacts that “contribute to a growing mountain of data [their emphasis] to be analyzed” (68), and that because of this “explosion of material available” (68) scholars are increasingly turning towards analytical tools capable of exploring them in great breadth. Like the others, they view digital humanities as necessitating new approaches to the study of empirical material: “Rich graphical interfaces, advanced visualization techniques, and ‘fuzzy’ processing have led some of those who have held numbers, calculations, and computers at a safe distance for a long time to warm up to new computational possibilities” (68), pointing to the digital turn and its resulting data-centric methods as constituting “fundamental transformations that challenge established epistemological paradigms” (68). In the spirit of my study, I argue that digital humanities methods may expand upon established epistemological paradigms in the humanities by implicating textual artifacts in digital form, but suggesting that they “challenge” or change those paradigms leads us down the slippery slope I have been bemoaning in this chapter, an inattentiveness toward disciplinary tradition.
Advocating for this near disciplinary overhaul in their essay, I find it interesting that they should devote their essay to apprehensions about digital work all while bolstering it as the new standard of humanities scholarship. In essence, they argue for a new epistemological digital humanities while pointing to its fallibilities. They call these “five challenges” the 1.) lure of objectivity; 2.) the power of visual evidence; 3.) black boxing; 4.) institutional perturbations; and 5.) the quest for universalism (71). But they speak about these challenges as they pertain mainly to the basis of new digital scholarship, as they more closely apply to those analytical scenarios because of their composition through digital platforms and tools. Reflecting upon them, I argue that these same challenges are not as omnipresent when the digital tool is applied to close reading. Rather, I argue, these apprehensions are present because they choose implicate extra-disciplinary concepts like cultural analytics and graphical interfaces. These complications are alleviated when applied to a more interpretively-driven textual analysis via digital tools, i.e. its application as an expedient to an original disciplinary aim.

By the “lure of objectivity” Rieder and Rohle refer to a mentality behind digital application, that it creates a bridge between the humanities’ interest in interpretive and verifiable knowledge. In their idea of “trans-disciplinarity” (72), they suggest that by embracing “the quantitative orientation of the natural sciences” (72) digital humanities can produce knowledge “that can compete with the natural science son their own terms, by being as ‘objective,’ as ‘rigorous,’ with the help of machines” (72). They go on to note that this ideal about “mechanical objectivity” (72) in regard to digital has been put into question and is not taken as a given in the field, but I would also feel it important to note that the impulse or desire for a sense of mechanical objectivity, and the resulting need to
be cognizant of it when using tools tailored across disciplines, is only relevant to the humanities perspective because it is being used for a cross-disciplinary purpose, i.e., a quantitative analysis not as easily adaptable to a humanities context, and therefore subject these logical difficulties. Inversely, a tool used more consciously and from within the context of a humanities methodology would not subject to similar misconceptions because the myriad sustainability of textual interpretations and the subjectivity inherent in the interpretive process is already understood. In this context the digital tool does not overreach to longstanding disciplinary logic when it is merely used to advance and already established disciplinary goal like close reading.

The issues they raise about the power of visual evidence deals mainly with the use of visual outputs like topographies, data networks, and cartographies as argumentative tools, which I similarly feel is of concern in so much as an image or data network is the desired outcome of the research endeavor. If close reading is the desired outcome by way of computer application, then the substance of argumentation will be the same. Black boxing also worth notice in a similar context. They call this the need for “Transparency,” in interdisciplinary methodological application, “our ability to understand the method, to see how it works, which assumptions it is built on, to reproduce it, and to criticize it” (75). This most commonly applies to a hypothetical research scenario in which an aim at proving a falsified hypothesis involves the presentation of obscure results without concise elaboration on the methods involved. But in a humanities context a lack of understanding of the epistemological preconceptions of a tool’s functionality amounts “a closed-source approach” that is “highly problematic” (76) in humanities research. Use of obscure tools
without adequate epistemological transparency could falsify or at the very least further subjectivize humanities output.

The examples they provide are telling: “different graph layout algorithms will push different features of a network into the forefront and thereby suggesting different interpretations of the same data” and “Web crawlers will use different strategies to create a corpus and different criteria to decide where to stop crawling—identical starting points may lead to very different networks” (77). As with the other problematizations raised by these scholars and others I argue that they arise because the digital application has encroached too far upon the research context and method of the humanities. To provide a detailed example and gesture toward what is to come in chapter two, I will allude to the idea that a computer-assisted close reading can be carried out in such a way that avoids these pitfalls because it situates itself firmly within a humanities, interpretation-based mode of inquiry.

By using a data collection tool like Python, for example, one is capable of then processing a document of writing as textual data and making it available to full manipulation by the aggregation platform. But there is no underlying algorithm that acts on its own for any great instance. Instead, it is directed by command functionalities which can be imbued with a far range of contextual particularities. Keyword searches, for example, can be performed in a non-data-centered way—i.e., not to be tabulated, but to be retrieved with the surrounding ten or twenty words, constituting the retrieval of passages and quotes based on a desired context. To expedite a focused study of dialogue, commands could be executed that retrieve all words of text contained between the figures [“] and [”], which when aggregated together, can be used as reference for a critical study.
In instances such as these, the onus for falsification or blackboxing is not vested in a computer research object as the machine generator of already-interpreted data. That lies with the scholar or interpreter that consults the empirical material retrieved by the tool. The tool does not overstep its disciplinary bound. It is used as a vehicle for an interpretation, not a surrogate for an interpreter.

Rieder and Rohle’s statements on institutional perturbations and the quest for universalism are relative to the same disjointed disciplinary outlook that I note here. Both deal with the question of whether the advent of computerization means a wholesale reconfiguration of the humanities and the sciences, with the humanities seeing a reshaping of its discipline through the application of the latter’s methods. They note them as challenges to digital humanities integration because it raises questions such as “will the scholarly memoire hold up to the interactive full-screen visualizations? Could computerized research make more traditional scholarship (look) obsolete?” (78); and the quest for universalism basically refers to the need I address here, the need to preserve longstanding disciplinary methodological traditions as a way to more feasibly situate new developments such as these, only they express it in the sense that disciplinary orientation is and should be ever changing rather than universal. I would only suggest that it can be both, it can preserve its universal truths and use them as a roadmap for properly handling the epistemological complications of future developments.

Phase 2.5: Institutional Reconfiguration

In the same year Patrick Svensson attempts to provide remedy to these issues by reconceptualizing the digital humanities and its place in the disciplinary oscillations of
which it is part, by choosing not to view it as a field at all, but “a trading zone and meeting place rather than a strained ‘big tent’” (42). While I feel it problematic and minimizing to deprive the digital humanities of this disciplinary status, I do feel Svensson’s justifications are valid because they vie closely with my own theoretical priorities: centralizing humanistic values within digital humanities integration. While he feels the best way to do this is to reshape how we think about the digital humanities alongside other disciplines, I have argued that the way to do this is to rethink the place of other disciplines within the digital humanities, moving them to the forefront rather than the periphery. Our agreements and contentions will be elucidated here.

Firstly, I do agree with Svensson’s warrants to this rethinking as they are the same as the warrants of my project, that “the current and actual presumed canalization of resources and interest in the digital humanities can be seen as a threat taking away resource and visibility from other parts of the humanities” (43). This is true, but only in so much as, I argue, digital humanities methods are being used to advance non-humanistic purposes and inquiries, leading to a reduction in their value because they are not made use of—not used to pursue humanities questions but rather scientific ones (like big data).

Svensson also acutely suggests that “we need to acknowledge that the digital humanities are not one, coherent entity, and that there are tensions and epistemic traditions at play” (43)—equally true, but also result of what I have outlined here as a conflation of method with purpose: an assumption in digital humanities work that the use of a scientific, multidisciplinary, or otherwise non-humanistic method also means the same for the nature of its analysis and results. As I have been arguing about close reading
via computer application, this need not be the case. Svensson implicitly states what I
would say in response to this but I do not feel he fully acknowledges and understands the
fact: “Digital humanities is an natural outgrowth and expansion of the traditional scope of
the humanities, not a replacement or rejection of humanistic inquiry” (45). If this is the
case, then the digital humanities is indeed a ‘field’—it is a division of the humanities in
so much as literary studies, history, philosophy, and the like. As any humanities field is
concerned with accessing and producing knowledge of the human condition, they each
differ in their methods and the particularities of their results. Digital humanities is no
different from this set when it is grounded in the pursuit of humanistic knowledge. Again,
our understanding of digital humanities runs astray when we begin to confuse it with the
sciences.

There is a significant crux to Svensson’s rethinking to which I would levy
opposition, and to which I believe it is attributable much of the disciplinary
misconception of the digital humanities. As Svensson continues to argue: “If the digital
humanities are about engaging with technology as tool, object of inquiry, medium of
expression, activist venue and more, and if we see these modes as intrinsically connected,
we presumably need to see the field as a place where these perspectives and epistemic
traditions come together” (54). To speak colloquially in response to this statement, the
digital humanities would not be so complicated if we were to remember its surname,
humanities, perhaps even after this volume striving to call it “the humanities, digitally.”
At first glance, I would remind us that the humanities does not engage with technology,
computer science does. The humanities engages with human thought (by way of text),
and digital humanities therefore would engage human thought, conceived in text, by way
of digital application or otherwise computational expedition. Any more or any less than this disciplinary tether does indeed dissolve the field status of the digital humanities, but only because it distances it from the humanities as its core rationale. It moreover exaggerates digital humanities’ resemblance to the sciences with an underlying misconception that anything technological must also be implicitly scientific. It is not. Instead of a loosely defined technological meeting hub for all disciplines (which would at that point resemble an IT helpdesk more so than a university department) humanistic scholarship should take ownership of digital humanities as a source for innovations in humanistic inquiry.

I think at this point the development of the digital humanities that we have seen has been one where the field had become caught up in the impulse for rapid innovation of its technological capacities. This was evident in the editorial and curatorial phase of the late 90s and early 2000s. This platform-based moment however gave way to what I would identify as a residual “scientization” (to borrow the term from Barbara Hernstein Smith 353), that reached the rest of the humanistic methodological sphere. The “scientizing” that occurred on the side of practical application and tool-use methodologies went on to influence the research outcomes and nature of scholarly inquiry made through the use of those tools and platforms because of their technological implications for managing textual data. That line between tools-use and method versus research outcome and nature of inquiry is where I would draw the theoretical line marking the encroachment point of feasible interdisciplinarity—where methodology is borrowed from another discipline the research outcome and nature of inquiry should remain within the original field. This way disciplines could still engage in
interdisciplinary meeting places but always to the purpose of innovating the *pursuit* of their distinctive scholarly purposes.

**Phase Three: Quantitative Macro Textual Analysis**

It will be prudent now to move into discussion of two editions of *Debates in the Digital Humanities* from 2016 and 2019, as representing the field’s turn from the originary stance I’ve outlined here and into the present day, where I believe the field stands currently. The volume collectively establishes a turn toward textual analysis and encourages its development, and two pieces will be important here for the way they capture the movement. Steven Jones in “The Emergence of the Digital Humanities” talks about how earlier established DH practices began with database building and moved into distant reading techniques of large datasets of humanities materials as corresponding points of development, and that an “eversion” from the database foundation occurred in a “spatial turn” (5) in digital humanities involving “data-layered ‘thick mapping projects’” (5), resulting in a new analytical pathway for digital humanities application. “The intention was to avoid the reductive definition of DH as mere digitization,” Jones notes, when talking about both distant reading and digital mapping. Jones emphasizes that we put the digital “into reciprocal conversation” (5) with cultural artifacts as a method of study.

David Hoover, in “Argument, Evidence, and the Limits of Digital Literary Studies,” offers a vivid model of this DH mentality. Studying narrative voice in a novel consisting of six monologues, he uses a word frequency distribution program to test the similarities and differences in diction to surmise whether they should be read as six
separate voices or one. Using this methodology borrowed “from the field of information retrieval” (233) Hoover hypothesizes that “words with the highest tf-idf scores for each of the six characters [the word’s frequency across the six monologues] suggests a possible answer to whether we should read the novel as six voices or one” (235).

Drawing an important crux in this method of data-gathering to study differences in narrative voice, he notes that “they [the data] still need to be interpreted” (235).

The need for data to still be interpreted marks an important logical point in understanding reach and potential in DH application to humanities pursuits. He calls it “no panacea” (248)—meaning that digital humanities’ “tools and methods can never eliminate the importance of literary intuition and close reading” (248). I would like to interpret this as meaning that digital humanities-based data output of humanities materials can never be fully constituted analytical knowledge without consultation by the human scholar for the real, interpretive purposes underpinning the project. As I have argued previously about some digital humanities projects, data-centric methods of retrieving and organizing aspects of textual materials for analysis can only amount to datasets until they are interpreted by a person using them as the basis for an analytical argument. It only aggregated material to be interpreted in the pursuit of this question in an expeditious way. The computational act of retrieval is not synonymous with the interpretive act that comes afterward. Acknowledging this pragmatic boundary line in methodology, Hoover argues that digital methods “should be integrated into the set of accepted approaches to literary texts” (248), and in doing so essentially foreshadows the nature of digital humanities work that will progress into the present day, those that make
use of computational data analytics in the pursuit of literary questions, and explore their interconnection.

While this sounds like a promising tipping point, I argue for the latter part of this chapter that this conception of text-as-data that has structured digital humanities text-analysis has had a part in the reduction of close reading and traditional approaches because it has not used data methods to pursue literary questions, but has rather merely chosen to view literary texts through the lens of data in order to fit them within the pragmatics of data analysis; not data methods to serve literary questions, but data and statistical questions about literature.

Following this enigmatic path, the 2019 edition of Debates deals with this question thoroughly—“The Digital Humanities and ‘Critical Theory’” and “The Elusive Digital/Critical Synthesis,” By John Hunter and James Baker, respectively, being some of the artful titles considering this question of the confluence of method and theory, and modelling its application. John Hunter calls it a “historical fact,” quite sure of the point, “that the digital humanities developed outside and apart from the concerns of critical theory;” as I have argued previously, “rightly or wrongly, and with no conscious programmatic intent, the digital humanities was institutionally positioned as an entity apart from critical theory, rather than an inheritor of it” (190). This ‘positioning’ as Hunter interprets the development of the field, has significant, and from this standpoint ominous, implications about the place of the humanities in the production, dissemination, and consumption of knowledge in the face of media bodies that are dominating the cultural-intellectual landscape of today—“as Google, Apple, Amazon, et al. redraw the
conditions of knowledge for the world,” Hunter emphasizes, “humanities scholars will not have a guaranteed seat at the table” (190).

Hunter’s proposed remedy for this is a re-positioning of the place of “the critical task,” as he calls it, within the “making/programming/hacking that the digital humanities embodies” (192), suggesting that the two be more closely integrated. He asserts that “digital humanities needs to justify its critical worth, as well as its earnings potential,” the latter referring to the making/curatorial side of early digital humanities.

James Baker and Seth Long reiterate these concerns, and my critique from here on out will be in the way these principles of synthesis between the digital and the critical are carried out in real time application. I note at this point calls for digital application similar to mine, and my argument will be that they are not carried out to the extent that they would like to think because, consequently, the data-centric critical analyses left to mention here are in essence, as I’ve said before, merely reductions of literature to textual data, as a form of cultural knowledge—empty posturing toward interdisciplinarity and not the true cultural criticism/textual analysis particular to close reading and the humanities.

As examples of praxis now, rather than theory, Michael Gavin et. al., in “Spaces of Meaning: Conceptual History, Vector Semantics, and Close Reading,” provide a clear example of the missing-of-the-mark that I would like to allege in this final portion of my discussion. They endeavor for “a method of computationally assisted close reading that draws from two distinct intellectual traditions: conceptual history and vector semantics” (243). Conceptual history being their humanities orientation, they view vector semantics though data analytics as a “point of contact at a theoretical level between computational and humanistic scholarship” (261). “Much like ‘topics’ produced by topic models,” they
specify, “concepts are best understood as structural patterns that recur among words” (243), providing a pathway for understanding the quantitative lexical components of theme in texts, exposing the “semantic fields that underlie the very conditions for thought” (260). But this apparently ‘close’ reading as they refer to it is characteristically a macro-approach to studying thematic construction, or as they note “to study how they are deployed in individual documents, by individual authors, and among historical communities” (243).

In John Hunter’s he decries, or tries to, “the model of scientific rationality underpinning much that falls under the banner of the digital humanities” stressing that “it cannot simply disavow the problems of earlier humanist debates” (188). With Gavin et. al.’s mentions of vector semantics as a “quantitative study of word meaning” (243) and the “quantitative results of text analysis” (243) or their need to synthesize the digital and critical with “areas of overlap that cross the disciplines” (260) I would like to reiterate this concern. The interdisciplinarity of the digital humanities should not mean the substitution of humanistic questions with scientific ones, and the substitution of scientific materials with humanistic ones. Such a macro, quantitatively based analysis of topic distribution in literary texts does just that. It replaces a scientific material with a literary text and treats it as data, construing concept history with a humanistic question. While a ‘form’ of textual analysis, this is not close reading of literary expression, but the reduction of literary expression to objects of data, absent of any critical or phenomenological interpretation of the figurative properties of language. Here the disciplinary borrowing within the digital application consisted of method but also result—data figures about conceptual semantics and linguistic probability rather than
close reading and interpretation. It had produced a scientific answer about a piece of humanities material, and not a humanistic notion through scientific application.

Nevertheless, this tendency toward quantitative analysis of humanities materials persists and is even embraced, which I feel conceptualizes the extent of textual analysis currently going on in present-day digital humanities scholarship: advocating for quantitative textual analysis as a new component of humanistic inquiry. Calling it a “Mixed Methodological Digital Humanities” (title), Moacir De Sa Pereira argues that “the digital humanities can become more sociological without necessarily becoming exclusively quantitative” (409). Mixing methods,” he continues, “need not involve a move from the macro (distant) to the micro (close), but rather can include a mixing performed iteratively over the same objects” (409). This impulse to “bring literary studies closer to the social sciences” (405) as a source of its validation is problematic because it implicitly foregrounds quantitative approaches under the guise of ‘new, uncharted analytical territory,’ further obfuscating the digital humanities from the methods now as well as goals of the traditional humanities. The voices mentioned thus far share in common a sentiment that feasible digital and critical integration constitutes an opening up of the types of conclusions that can be drawn from and the types of arguments that can be made about humanities materials. This essentially takes the humanities materials and brings it all the way to meeting the sciences on its own terms. I argue that we still need to consider bringing the digital to bear on the humanities in terms of the humanities, and not the sciences—computational approaches to distinctly humanistic purposes, and not scientific ones like vector semantics or topic modelling, but close reading fully in and of itself.
Phase 3.5: Data Processing of Literature

What will follow, to begin to conclude, are two important book-length volumes of digital humanities scholarship that attempt to model the disciplinary integration discussed at this moment in the field: Andrew Piper’s *Enumerations: Data and Literary Study*, and James E. Dobson’s *Critical Digital Humanities: The Search for a Methodology*. I mean them to illustrate the theory in practice that I’ve highlighted in these two *Debates* volumes, and serve as a representation of where the digital humanities currently stands. I will review them here exposing the interdisciplinary incongruities that I’ve endeavored to elucidate in this chapter.

In *Enumerations*, Andrew Piper uses what he calls “the emerging fields” of “natural language processing, machine learning, and text and data mining” to look at “the quantitative dimensions within texts,” a form of “distributional semantics” that he feels highlight the ways that “repetitions of language lend meaning to our experience as readers” (2). Through data approaches, Piper argues that observation of lexical quantitative structures and corresponding semantic patterns within texts lends itself to an understanding of the deeper “grooves and channels” (3) of cultural expression. This approach to quantity, Piper suggests, aims to “rethink how the computational study of literature has initially been framed” (3) by positioning his study against what he calls “The emphasis on novelty, but also bigness, empiricism and, as we will see, overly simplified and often binary modes of reading” which he argues have “led us to miss the important ways that computational reading is inevitably tied to the norms and practices of the past” (3). He calls this a form of “misapprehending the disciplinary inheritance” (3).
when digital application, as I have similarly argued, makes peripheral to their concern the
disciplinary norms and practices of the past for innovative experimentation.

Establishing this, he attempts to situate ‘quantity’ as an approach significant to
“questions of cultural distinction and durability” (4), pivoting a theoretical argument
through his quantitative perspective that “literature is not founded on the rare and the
singular, but rather the common and the collective, the fabric of repetition from which it
is made” (4). Viewing quantitative dimensions of texts as well as qualitative dimensions
as “two integral components of a more holistic understanding of human mentality” (5),
Piper is making the case for digital literary studies as a ‘counting discipline’ at the same
time that it is an interpretive one, positioning “the value of translating between letters and
numbers as a newly vital form of humanistic thought” (5). He views the two analytic
standpoints and thus disciplinary perspectives—quantitative versus qualitative inquiry—
complementary rather than oppositional to each other. So like others, he makes use of a
data analytic method to produce quantitative answers about humanities material that he
feels lend themselves to a humanities concern: the distributional semantics of language
repetition in the construction of plot and topic.

Grounding his approach in distributional semantics and what he calls the
“distributional hypothesis” (14), Piper tests four quantitative assumptions: “a) a word’s
meaning is tied to how often is occurs; b) a word’s meaning is tied to how often is occurs
with other words within a given context; c) these relationships are entirely contingent
upon the scale of analysis; d) these relationships can be rendered spatially to capture the
semantic associations between them” (14). By spatially, he means through visual
representations of networks and connections between associations of meaning. Through
this data framework Piper intends to assess, with attention to the construction of plot and topic, the probability of the construction of meaning based on the linguistic semantic associations among repetitive or patterned words, “the probabilistic way we assess meaning through language” and “the importance of special relationships for the construction of meaning” (14). And in this analytical pragmatic I argue in response that Andrew Piper nicely fits humanities materials of literary texts into a data science framework in order to produce quantitative data about literature. He reduces the fabric of meaning making to the quantity of word patterns, taking attention away from the qualities of figurative language. Absent of interpretation, he counts its linguistic signals and signposts and deduces its lexical patterns.

Looking at the construction of plot in narrative form, Piper uses distributional semantics to measure “the relationships that exist between words and how meaning is shaped by such probabilistic distributions.” “As words recur,” he argues, “we begin to develop mental models of what to expect as we read.” Measuring and datafying, then, the probability of word repetition on the level of vocabulary as a structural component of plot on the level of semantics, allows him to “understand what the text is conveying at a more formal or semantic level” (43). He gives one preliminary example-as-hypothesis to contextualize an inference deductible from semantic distributional probability, saying that “If only 5% of Jane Austen’s Pride and Prejudice consists of unique words, then a vast majority of the novel’s significance is contained in the repetitions that run through the fabric of the text” (43). Not necessarily a hermeneutic reading of the texts content—indeed, Piper exerts energy to distance himself from a content-statement based analysis—he assembles data of vocabulary distribution as a correlative to the construction of plot.
Situating his approach, Piper similarly conceptualizes his study as a new method of textual analysis, positioning it as being different from and in addition to traditional approaches. Viewing it as a contribution to the study of narrative (which is duly is, although the digital application does not enhance traditional approaches as I would like it to) he claims to have found it “surprising that until now the theory of narrative has been so little concerned with the problems of narrative enunciating concentrating almost all its attention on the statement and its contents” (43). His approach is markedly different, giving new attention to “how larger formal structures—the organization, not the specific content, of language—convey meaning beyond any specific motif” (43). So a new formalistic approach grounded in semantic distribution, rejecting motif and content in plot construction in favor of an interest in the probabilistic lexical networks underlying the composition of the content of plot.

In one example modelling this statistical measure of lexical probability, Piper uses a line graph to illustrate the distribution of proper names across the 600 pages of Virginia Woolf’s *Mrs. Dalloway*, with the Y axis connoting percentage and the X marking the page range, and a dotted line down the middle marking the novel’s midpoint. The line graph implies that the use of proper names sharply increases after the middle of the novel, suggesting a chance in the narrative perspective. Through the dataset, Piper deduces that “we move from the interior world of momentary thought and marriage (thought, moment, morning, marry)” in the beginning of the novel, “to the more collective space of the party—people, ladies, family and friends” (59). An interpretive deduction, he suggests that “The second half of the novel has considerably more proper names per 1000 word windows, culminating in the legendary cocktail party. As people occupy more space, the
lexical diversity of the novel contracts. There is less room for the kind of cognitive expansiveness that marks Woolf’s fiction more generally” (60). Piper interprets this textual pattern thematically as a “move from deep interiors” to the “shallow surfaces of the cocktail party” implying “a death both symbolic and real” (61) of the protagonist introverted community with one’s self to their extroverted sense of being for others. On this basis of analysis and deduction Piper situates his advocacy for “the ways in which computational models can be a useful tool for thinking spatially about narrative form—for seeing the form of narrative” (61); with ‘form’ in his view, being vested in the probability of lexical patterns rather than content and statement as he himself specifies.

Other examples that Piper provides in further chapters illustrating his data approach to literature involve, to name several, a type of topic modelling that measures “the co-occurrence of words within particular documents” (67); “Network diagrams of topic-to-document relationships in 14,888 passages from German novels” (84); and word clouds illustrating “distinctive features for nineteenth-century fiction” revealing a prominence of “social and biological categories” (109). By aggregating data figures about textual and lexical probability distributions, and illustrating their distributions and probabilities in a variety of visualization techniques, Piper claims that we will be more “able to attend to the repetitions [of language] and their differences that underpin literature” (93).

While Piper’s knowledge contributions are substantial, I still desire to set my approach apart from his. I take issue mostly, as I have and will throughout this volume, with the computational application hinged to a “new” approach to the study of text because it implicitly obfuscates itself from traditional methods which could be otherwise
enhances. Instead of enhancing method, such works enhance the deduction, the nature of knowledge produced. Piper’s approach then has taken bodies of literature and datafied its textual and lexical composition, and viewed the probability of word distributions as a correlative to the construction of plot. I am not saying that these deductions are not true—I believe they are—I am just taking issue with the fact that in doing so it rejects consideration of motif, content, and expression as a component of plot, viewing the statistical correlatives as a defining measure. But correlation does not always mean causation, and there are risks to this approach that would reduce interpretive potential to quantifiable possibilities based on lexical statistics, counterintuitive to literary criticism’s assertion of a text’s ability to sustain interpretive ambiguity. Quite the contrary, this type of approach runs the risk of reducing interpretation to a matter of probability. The practice risks creating a precedent whereby statistical schemata limit the perceived applicability of interpretive considerations to particular texts because of preconceived topical datasets ostensibly suggesting otherwise where that may not always be the case. And ultimately, as I have been arguing, it reduces the figurative elements of text to figures of quantifiable data, taking attention away from the fabric of interpretive language and turning the humanities endeavor from an interpretive into a ‘counting discipline.’

In 2019 James E. Dobson introduces his *Critical Digital Humanities*, another treatise attempting to conceptualize and direct the field among its myriad institutional implications, calling his study “a critique of existing and potential computational methods for the analysis and interpretation of text,” harnessing “Quantitative formalism, distant reading, algorithmic criticism, macroanalysis, computer-aided text mining” for the ways in which “they add, in one way or another, forms of computation to our understanding of
the interpretive act” (vii-viii). Dobson outlines thoroughly and in the foreground of his preface that he aims to use computerization to “serve the interpretive goals of humanists” and he affirms that “these approaches cannot and will not make obsolete other existing interpretive frameworks” (viii) even as he similarly uses computerization to draw new frameworks. “The discourse of the humanities,” Dobson specifies, “is an infinite discourse” (viii). With this I agree, but he goes on to say about method that “Neither the methods nor the objects within such a framing are capable of settling into inertia or abruptly coming to a halt”—this also I agree with because of the infinite sustainability of interpretive ground in text and in figurative language, but where I view the ‘infinite discourse’ in the potentiality for interpretation-as-method, Dobson locates this infinity in the “search for method” (viii) as apart from interpretation of the object.

I would respond at the outset and say that the construed need for an infinite search for new method would imply that existing frameworks are finite at various logical boundary marks, that they are indeed “capable of settling into inertia or abruptly coming to a halt” (viii) because new interpretive frameworks are ostensibly needed to sustain the applicability of computerization to the study of text, as if they are not translatable to traditional methods like close reading. But moreover this axiom of Dobson’s that “It is the infinite search for method that prevents the humanities from staying fixed long enough to say that the objects or the methods have finally given it their all and spoken all there is to say” (viii) misses an important distinction that the humanities endeavor is an infinite search for meaning and not method. Imbedded in the humanities is an infinite potential for method, but within each method is also not a finitude, but an infinite potential for deducing meaning.
If I were to adopt Dobson’s thesis here, I would revise it and say that it is the infinite sustainability of meaning within texts that prevents existing humanities methods from becoming obsoleted in the face of new quantitative approaches, and it is the role of computerization to enhance these approaches so that they can reach new interpretive potentials. As it would in any innovative scenario, computerization ought to enhance the natural disciplinary proclivities of the field in which it is applied rather than replacing it with new ones simply because they are similarly applicable somewhere else. Nevertheless, Dobson leverages computerization in this volume geared towards the latter, similarly embracing the notion of “new insights” and “fascinating new arguments” that “extend the horizon of interpretive possibilities,” quantitatively exposing what he views as “the seen and unseen of the interpretive scene in computational criticism” (x) as a new interpretive framework all together.

Piper premises his applied study on the notion that “the intellectual core of any understanding of the digital humanities” is “the use of computational methods in humanities research and scholarship” (1) and I think it is this conceptualization that underpins the real problem that I am addressing. Like his previous phrase, I would respond to it saying that, no, the core of the digital humanities is computer application to humanities method and scholarship. Method dictates the nature of the research endeavor, and thus the disciplinary agenda. Application is a supplement, enhancing the potential, breadth, and capability of the original method and goal. In Piper’s view, this ‘core’ is “specifically, the use of sophisticated quantitative methods for text and data mining” (1) but that is not the core of the humanities, per se, it is the core of statistics and data science, amounting to another near-sighted attempt to make text fit into the framework of
the sciences, pursuing quantitative inquiries, rather than making frameworks of the sciences fit within humanities interests. Only until we are able to feasibly use quantitative or otherwise data-centric modes of knowledge production to achieve interpretive, humanistic knowledge, will we have truly reached the core of the digital humanities. Otherwise, we are just producing data about text, but not traditional humanities scholarship, but not literary interpretation concerned with the figurative permutations of language.

Turning to a concluding example, one instance of text based analytics in Piper’s volume which he uses to illustrate his theoretical claims is the use of topic modelling through Python to establish “thematic categories” (49) of Martha Griffith Brown’s *Autobiography of a Female Slave*. This study and this use of topic modelling is premised upon “tables of word or phrase frequency or occurrence counts” (49) as a correlative to theme in a written text, in this case an autobiography. Using a CountVectorizer function in a program known as Scikit-learn, a Python module, Piper produces a table firstly of word occurrence count in hierarchical order, resulting in the following table:

- man 226
- negro 196
- people 186
- great 169
- men 162
- country 144
- slave 144
- rev 137
- say 135
- American 134

(Figure 2.1)

In addition to other topic modelling examples of other texts and collocations of texts, Piper uses these computational “inverse document frequency ratios” (50) to situate topic
modelling as a method for the “mapping and aligning of semantic space” (140) which he feels “demonstrates some of the promise and limitations of computational models for humanities scholars” (140). I argue that his sense of limitation is only brought on by the nature of the method employed. The topic models presented here are if anything only surface level observations of word frequency, and would have to go a much longer way to be thought of as indicative of theme beyond any notion that is not already obvious. They do, although, present in their constitution potential for interpretive inference as to what their frequency implies about the philosophical orientation of the text in question, but these are merely creative inferences drawn from word count and not anything necessarily semantic or interpretive in nature. Indeed, Piper affirms my sense of the project’s limitations in his own words, saying that “Interpretive activities, many argue, take over once the distilled and computer results have been generated” (56). This is true, especially when the nature and extent of the results are dictated by the nature of the method employed.

This is where I want to draw an important distinction that will be withstanding throughout my project: that computational application can, I argue, be employed without the adoption of scientific method. I am drawing distinctions here that are meant to characterize the computational approaches in the next chapters: “computational method” in these scholarly voices, has been construed as also mean quantititative or data-based-statistical inquiry, thus dictating the nature of results drawn from the study, perhaps because of inherent assumptions about the nature of computation and its ‘numerical’ dispositions. Method, in my view, should be a disciplinary marker—quality versus
quantity; counting versus interpreting; distant or close—and can be served by computational application.

**Phase Four(?): Toward a More Feasible Interdisciplinarity**

So far these digital humanities practitioners have illustrated the ways in which computational data-based inquiry can expand the categories and types of interpretation possible when brought to bear on textual materials, and they have done work to involve written texts in the analytical frameworks of scientific, quantitative inquiry, seeing how their methods can draw unique knowledge from texts about things like semantic space and linguistic distribution. But I argue that this developmental horizon of the field has merely used computation to view texts in a scientific light, interpreting texts in scientific ways. That it has resulted in the production of fruitful knowledge (I believe that it has, and I will never suggest that it has not) is self-evident, but it still misses an important point that is perhaps only a matter of technological dexterity and creativity. Returning to my point about application and method as differential terms, Piper and others like him have covered how computational application of data-centric method can draw new quantitative interpretive knowledge from texts. I would still like to see how computerized application of a humanities method (close reading) can enhance the interpretive act of textual analysis that has been traditional to the humanities; and not just simply invent new ways of interpreting—although there is room and need for that—but innovate the ones we’ve always had.

So far in this chapter I have tried to illustrate the intellectual trajectory of the digital humanities and the influential factors from which it grew, and the field it has become today. In the beginning I outlined the digital humanities’ three conceived phases
as embodied in the corresponding scholarship: the digitization phase, the editorial/archival phase, and the current phase of literary data. I proposed that, in the spirit of my study, we conceive of a fourth phase, one in which we exert recursivity back to the enhancement of old methods in addition to experimentation with new ones.

In a broad sense, we can see how the digital humanities has developed into an outgrowth of the sciences with its embrace of technological innovation, and the ways this impulse led to the consequent obfuscation of traditional humanities methods in favor of a new ‘scientizing’ of humanities materials. I had suggested that the concern with ‘bigness’ and ‘breadth’ over ‘depth’ and ‘precision’ originated in the acts of mass digitization in the beginning and carried over to influencing the field’s desirable analytical purviews, harnessing the mass-aggregatory potential of the computer and applying it to correlative analytical purposes.

I reflected upon this disciplinary term of “method” and used it to criticize the way we think about interdisciplinarity and the feasible borrowing of knowledge and the pursuit of disciplinary goals. As these are the things at stake as the digital humanities continues to develop and outline its institutional and disciplinary imperatives, I cautioned us to consider how we think about method and outcome—the nature of knowledge drawn and the method employed as an indicator of the disciplinary contribution at hand in the project. I am courteous to the data-centric approach of the third phase acknowledging that it did indeed create new and fruitful methods of analysis and expanded the questions that we can ask and answer about literature and literary history, but I also pointed out that I was only left wanting in considering whether these new data methods serve the sciences more than the humanities. It occurred to me when viewing distant readings of textual and
literary statistical figures that it appeared as though text was being fitted into epistemological frameworks of the sciences—asking scientific questions about humanities materials through the use of data analytic methods facilitated by computer application. They viewed ‘method’ as disciplinary rather than computational, which led to a replacement of humanistic inquiry with scientific inquiry. I still want to view computational humanistic inquiry without the borrowing of scientific method and statistical output.

The main takeaway from this chapter, then, would be to answer how I would define feasible interdisciplinarity in light of what was just overviewed. Feasible interdisciplinarity that gives respect to the primary disciplinary goals of the field in question would rely on a differentiation between two terms: application and method. Method should work in service of the discipline seeking to benefit from the interdisciplinary application, in this case being literary studies and computational application. The method employed will be innate to the primary discipline, so close reading and interpretation as particular to literary studies. The interdisciplinary application, then, would be employed to serve this methodological goal. So computer application of a data analytic tool would be used to in some way modify, expedite, or enhance the deduction of a close reading from a piece of text.

When ‘method’ is construed with ‘application,’ this leads to the disciplinary disparity to which I show apprehension in my overview of data approaches to the study of text. While they contribute to knowledge about text and corpora of writing, they do so in a markedly scientific context, which appeals a particular area of the field’s practitioners and audiences but not the field in its entirety. This is why I do not discount
this third data-phase of digital humanities, but merely suggest a fourth one that I argue we have yet to come across; that maybe now it is time for a glance back at innovating old methods as we use innovation to invent new ones as well. It would be a way of rejuvenating the relevance of humanistic method that would preserve its originality rather than change its epistemological trajectory. When ‘application’ is used as the marker of disciplinary borrowing, it allows us to foreground disciplinary method and prioritize the nature of knowledge we want to produce. Computational application of data analytic method is statistical work. Computational application of close reading would be a digitally enhanced humanistic method.
Chapter 2: Methods of Computational Close Reading

“The goal of good interpretation is to understand the style completely. We presume that thought and language intertwine throughout, and the specific manner with which one understands the subject requires an understanding of the arrangement of words: i.e. the handling of language.”

- Friedrich Schleiermacher Hermeneutics: Outline of the 1819 Lectures, pp. 13

“What it communicates and how it does so and the worth of what is communicated form the subject-matter of criticism.”

- IA Richards Practical Criticism, pp. 10

“What is needed for literary satisfaction is not, ‘this is beautiful because of such and such a theory,’ but ‘this is all right; I am feeling correctly about this; I know the kind of way in which it is meant to be affecting me.’”

- William Empson Seven Types of Ambiguity, pp. 254
A Working Definition of Close Reading

If I were not so coy I would simply argue that it has been a matter of inventiveness that the methodologies of close reading have not been emulated in computational analytic processes, but that would also speak to the disservice of my previous chapter, which has attempted to prove that it was an issue of trajectory that was caused by the field’s initial ambitions and subsequent developments. Instead I advocate here for an initial, further developmental phase, one in which the question of innovation is applied to traditional methods and we consider the findings. I argue that if we are to truly establish a comprehensive definition of close reading, then computer analytic tools, in their dynamic nature, can be harnessed to operate within this methodology.

In order to argue that the coding processes I’ve written are capable of functioning as what we think of as close reading, I must first establish a definition to align their functionalities with. The first half of this chapter will be spent to that end, and the latter portion an explication of the codes in light of my precise definition.

I establish this definition—this theory—through the assemblage of particular theorists who have foundationalized the way we think about the interpretive act. I choose them in particular however because in their chronology they also represent the development of the concept, and its use today, so I derive the definition used for this study through a synthesis of their respective histories. I will first look at Friedrich Scheleirnacher’s *Hermeneutics*, the technical interpretation specifically, from the *Outline of the 1819 Lectures*, for its attention to the understanding of “style” and “the handling of language” (13) as akin to the types of literary features that can be highlighted in this coding program. Next I will look at Roland Barthes’ *The Rustle of Language*, viewing his
“reality effect” and the “collusion of a referent and a signifier” in the “concrete detail” (148) as similarly relatable to these functionalities, that they can be used to observe this textual and linguistic interplay. Moving away from this theoretical phenomenology of interpretation and into literal scholarly conceptions of close reading, I look at IA Richards’ *Practical Criticism* and William Empson’s *Seven Types of Ambiguity*. In particular, I look at Richards’ “four types of function, four types of meaning” (175)—those being sense, feeling, tone, and intention—as being similarly comparable to the coding program’s capabilities. I use Empson’s types of ambiguities in the same way, making use of the way they describe the properties of language that are analyzed in the interpretive act—the aspects of language that structure and evoke interpretation. I then apply their concepts to computational thinking, the primary task being to envision these criteria of close reading in a way that is translatable to the way a data analytic platform analyzes a dataset.

Hermeneutics, and the technical interpretation, lend useful perspective here for the way Scheleiermacher is keen to draw attention to units of language, methods of composition, and aspects of grammatical structure as part of the “unity of the work” (13) that drives the content and creates the interpretive situation. He says that “The unity of the work derives from the manner in which the grammatical constructions available in the language are composed or connected” (13). By grammatical constructions we can take this to mean the order or presence of words as a particular construction, as well as punctuation in the same which underlies grammatical structures. A simpler way of understanding “the handling of language” (13), the technical interpretation helps us to
understand that an interpretive situation in a text is derived from the formulaic ordering and structuring of language, as figures, and perceivable as graphical figures.

Here I can give an example of such a methodology, stemming from Schleiermacher’s technical interpretation, which receives interpretive content through the understanding of grammatical and linguistic structuring in the text. How can we look at “the handling of language” as indicative of “theme” in the hermeneutic sense? I would do so by first ascertaining an interpretive hypothesis that falls in-line with these aspects of hermeneutics: that theme is indicated by word prominence as part of a grammatical construction. I test this theory with the following word cloud rendered in Python, based on Edith Wharton’s *Ethan Frome*:

(Figure 3.1)

In using Schleiermacher to illustrate this point I add the concept of “graphical figures” to his conception of grammatical constructions—that they have graphic qualities in their appearance, and thus their frequency can be measured to make inferences about theme and content. On a remedial level we see this in the prominence of names: the main character’s, Ethan, is the most prominent, being that the novel is about him. The competition of the love triangle between Ethan, Mattie, and Zeena is evident in theirs
being the two second-largest in the cloud; and yet on a deeper level, if we notice that Mattie’s is slightly larger than Zeena’s, it perhaps alludes to her place in the romantic dynamic, that she is the most desirable to Ethan, and ahead of Zeena in the ostensible competition. On a deeper, more formalistic level, the word “said” in this cloud—excepting the word “night” as indicative of prominent imagery—similarly encourages an interpretive inference, suggesting that there is a predominance of dialogue in the text, leading one to consider that dialogue is the text’s central method of content delivery—as an interpretive proximity, implying that the reader receives and perceives the story and its thematic moments through observing the exchange of dialogue between characters, positioning them as an onlooker or eavesdropper on the events of the story by listening in on their conversations.

As a direct example, we could observe the exchange of what is perhaps the climax of the novel, the moment in which, while Ethan and Mattie were supposed to be on their way to the train station for Mattie to be dropped off and leave Starkfield, they decide to attempt suicide as a means of remaining together in death, avoiding her departure from the farm.

“What’s the good of either of us going anywhere without the other one now?” he said.

She remained motionless, as if she had not heard him. Then she snatched her hands from his, threw her arms about his neck, and pressed a sudden drenched cheek against his face. “Ethan! Ethan! I want you to take me down again!”

“What where?”

“The coast. Right off,” she panted. “So ‘t we’ll never come up any more.”
“Matt! What on earth do you mean?”

She put her lips close against his ear to say: “Right into the big elm. You said you could. So ‘t we’d never have to leave each other any more.”

“Why, what are you talking of? You’re crazy!”

“I’m not crazy; but I will be if I leave you.”

“Oh, Matt, Matt—” he groaned. (165)

So while the novel *Ethan Frome* is framed as a confessional monologue that speaks about the events of these people in the third person, the proximity of interpretive observation afforded to the reader is more often—and at the most important of moments—through a first-hand observation of dialogue, essentially disembodying the narrator and placing the reader in a direct selectively omniscient view of the characters’ vocal exchanges with one another. While traditionally thought of as a confessional monologue with an unreliable narrator, the rendering of these moments of dialogue leads us to suspect that the narrative at times transitions into an interpretive perspective of selective omniscience. This presumption is amplified in other moments that accompany dialogue, too, where details of the characters physical behavior and appearance imply omniscience beyond that which the narrative claims itself to be: the discussion of these events by someone who is equally receiving them through another observer, Mrs. Hale. The predominance of dialogue shows us a depth of omniscience in the narrative structure of *Ethan Frome*, that the narrative structure frames itself as a first-person confessional monologue by an unreliable narrator but does not inherently function that way.

Using computational thinking to view these grammatical constructions of word prominence as graphical units of language, we can see how its application can lead to
substantive interpretive inferences. So we will let this initial observation serve as one part of an applied data analytic capable of enacting close interpretation: that grammatical constructions such as diction and word prominence can be indicative of thematic as well as formalistic concerns.

Roland Barthes’ reality effect and the process of signification of the referential fit well into this example when focusing on the semiotics of what he calls “the concrete detail” (148) in a similar way. “The ‘concrete detail’ is constituted,” he says, “in the collusion of a referent and a signifier” (148). This idea is relative to the realism that I intend to read here, because it provides us with an understanding of its narrative structure that is useful in developing an analytical criteria within the parameters of computation. Barthes calls the realist narrative “fragmentary, erratic, confined to ‘details’” which together form its interpretive space; much like a collage, it tends to be *en medias res* and at times non-linear, leaving the loosely interrelated collage of details to interpretation by a viewer.

But the details that comprise the interpretive space of the realist narrative exist in a referential illusion that only emblematizes ‘the real.’ “[F]or just when these details are reputed [my emphasis] to denote [original emphasis] the real directly,” Barthes reminds us, “all that they do—without saying so—is signify it [original emphasis].” I choose to add emphasis to “reputed” for a specific reason, for we should remember that realism claims its truth value while also consisting of generalizations that variably navigate the bounds of the realistic and the romanticized, marking the nebulous nature of the realist aesthetic. But in the same way, the concrete details which comprise its narratives operate from within chains of signification that can be similarly traced in grammatical and
linguistic constructions, such as words as signifiers whose referents and signifying chains can be used to point to broader thematic and interpretive situations in a text.

This begins to propose that, similar to the example related to Schleiermacher, word choice, or the presence of a particular signifier in an area of a body of text, like a piece of literature, can be used to highlight that area of the text as consisting of a particular theme or interpretive situation. If word prominence is indicative of theme (in addition to, at times, formal composition), then so too is word presence, two features which are attainable from within the functionalities of computational processes ostensibly thought of as “quantitative” in nature.

I will return to the word cloud to further illustrate this point, combining the notions of word prominence and signification as a way of deducing interpretive ideas. Among the third smallest in size on the cloud is the word “face.” If I take face to signify, among other things, notions of “person,” “character,” or more formalistically, “characterization,” I can guide myself from here to passages containing this word and judge their interpretive value. One such moment from Ethan Frome comes to mind. The sentence itself, first, reads: “There was something bleak and unapproachable in his face, and he was so stiffened and grizzled that I took him for an old man and was surprised to hear that he was not more than fifty two” (3). This is already a key moment in Ethan’s characterization that foreshadows the tale’s end, but I might choose, as part of my analytical method, to move from here and examine the entire passage of which this detail sits as a constituent part, so I would expand my perspective of the language surrounding this signifying object.
From first identifying the signifying detail, then the sentence construction around it, I draw my attention from the signifying word to the entire passage around it. So from the notion of the word “face” being a prominent word, I looked for its presence in a passage as one influenced by the presence of the concrete detail. Understanding the word’s signifying chain, I arrive at a passage that is key in elucidating Ethan Frome’s characterization:

It was there that, several years ago, I saw him for the first time; and the sight pulled me up sharp. Even then he was the most striking figure in Starkfield, though he was but the ruin of a man. It was not so much his great height that marked him, for the “natives” were easily singled out by their lank longitude from the stockier foreign breed: it was the careless powerful look he had, in spite of a lameness checking each step like the jerk of a chain. There was something bleak and unapproachable in his face, and he was so stiffened and grizzled that I took him for an old man and was surprised to hear that he was not more than fifty-two. I had this from Harmon Gow, who had driven the stage from Bettsbridge to Starkfield in pre-trolley days and knew the chronicle of all the families on his line. (5).

What conclusions are to be drawn from this deduction? Following the signification we originally inquired upon, I am aiming to further understand Ethan Frome’s characterization on the assumption that passages containing the word “face” consist of
content related to this subject. This method suggests the assumption to be true in this instance. But what I see through the broader perspective provided is more telling.

Throughout this passage and around the central detail of his face is a subtle ambiguity to his description that marks the enigma of Ethan Frome’s personality: that he is a strong, determined, capable man, but who never had the proper opportunity to make himself so. The passage situates this through the coupling of contradicting words in its ostensibly somber descriptions of Frome, with the words connotating things like “strength” and “greatness” cast in a negative light, or one overcome by the more prominent suffering that afflicts Frome. This is evident from the beginning of the passage—calling him “the most striking figure in Starkfield, though he was but the ruin of a man”; “the careless powerful look he had, in spite of a lameness checking each step like the jerk of a chain”; and the ambiguity is gone, or rather the doubt diminished fully, by the following, central phrase that led us here: “There was something bleak and unapproachable in his face, and he was so stiffened and grizzled that I took him for an old man and was surprised to hear that he was not more than fifty two,” seeming to affirm this as Frome’s most enduring quality despite his better potentials.

All of this is to say, then, that if chains of signification can be traced through words, then ideas of theme and passages dealing with a topic of interest can similarly be traced by the presence of key words that occupy the desired signifying structure. Through what we have seen thus far, this method similarly situates an interpretive inquiry within the logical and procedural capabilities of a data analytic if we were to imagine a program meant for extracting passages of subject-based content based on the detection of key
words. It can do so if they are conceptualized in the proper context, which I intend to discuss at greater length in the second part of this chapter.

Now that we have laid a solid theoretical ground, I will begin to turn to more technical definitions of close reading from within this theoretical perspective. Looking at IA Richards and William Empson for their more specific focuses on literary quality, I will examine the same translatability of analytical concerns in their descriptions of criticism as I did in Schleiermacher and Barthes’ semiotics.

Although IA Richards speaks mainly about poetry in his Practical Criticism, I believe his work figures well into this discussion of a translatable criteria of close reading for his attentiveness to language, i.e. words. The quote by Richards which I used for this chapter’s headnote begins to illustrate what I mean. Speaking about the interpretation of text he says, “What it communicates and how it does so and the worth of what is communicated form the subject matter of criticism” (10). I happen to use this quote specifically when teaching close reading to undergraduate students for the way it situates the interpretive process into three parts. By ‘What it communicates’ we take this as the latent ‘content’ of the text or story—what happens on the surface level. By ‘how it does so’ we mean the literary technique, the handling of language that creates the interpretive situation. The ‘worth of what is communicated,’ if we take worth to align with ‘criticism,’ whose Latin root means to judge, evaluate, appreciate, refers to the critical connections that we make between the surface level content and the broader human concerns with which the content is implicated, like psychoanalysis, gender, race, history and philosophy, among others.
This short but dense quotation by Richards is a highly illuminating axiom of criticism because it signifies (symbolizes) the chain or the trail by which we arrive at our critical notions when interpreting text, a chain or trail that is tethered to the particular usages of words contained within a literary technique. This means that that same such technique, that critical worth that results from the interpretation, is traceable back to the words that construct it. The words, then, act as signposts that direct our attention to the text’s literary milieu. Richards also says that “the separate senses that a word may have are related to one another, if not as strictly as the various aspects of a building, at least to a remarkable extent” (9). The subject-matter of criticism is interrelated with the choice and positioning of key words, if we are to begin to adopt computational jargon when thinking about these established conceptualizations of close reading. I believe it prudent, at this point, that we do.

My thesis here is mainly that close reading is reducible to key words, and that using such computational methods to be attentive to the textual economy of a particular word can lead to fruitful insights akin to what happens in the close reading of a text. While we do, at times, close read sentences and passages—all the same, the keywords embedded within them are what help us to arrive at the broader passage, as I began to illustrate with the word cloud. We can further affirm this by looking at Richards’ more specific elaboration of the “four types of function, four types of meaning” (175) at hand in the interpretive situation. By this I mean his theory of sense, feeling, tone, and intention. I will read his descriptions of these chief difficulties, as he calls them, and describe how they fit into my theory of key words.
To illustrate this alignment of theories I will return back to the word cloud used in the prior examples to show how it can be used to similarly detect the aspects of human utterance which Richards is speaking about.

(Figure 3.2)

Richards says that *sense* is the surface level content, the “*something*” of which the utterance says, and that in the establishment of *sense* “We use words to direct our hearer’s attention upon some state of affairs, to present to them some items for consideration” (175). The word cloud easily helps us arrive at the sense of *Ethan Frome*, that being the content of the story. From the mundane to the more precise I can deduce firstly from this body of aggregated data that the text is ‘about’ a specific set of characters—it tells us who are the main characters of focus in the story’s thematic landscape: Ethan, Mattie, and Zeena. We have already spoken about the thematic subtlety of the words “night” and “face,” as well as “said” for the prominent role of dialogue, but considering how this cloud further illuminates the *sense* of the story I would draw attention to the words “Starkfield,” “house,” and “kitchen,” as connoting the main locales of the story, its setting, but also as objects of interpretive significance, as if *place* figures
itself as central to *Ethan Frome*’s thematic composition. These words signify this interpretive proposition.

It is easy to understand the thematic object of place and setting in *Ethan Frome* as a novel but more precisely in relation to the character of Ethan himself. “He seemed a part of the mute melancholy landscape,” the narrator says of Frome in the beginning of the novel, “an incarnation of its frozen woe, with all that was warm and sentient in him fast bound below the surface” (14). Tracing the word prominence of these locales, namely Starkfield in this example, we can begin to observe how Ethan’s characterization is composed in large part through a metaphor of place and landscape—how renderings of the environment and setting are a reflection and a parallel to Ethan’s character. The word “house” is also prominent in the cloud, and we can take this to refer to instances depicting Ethan’s house in the story, such as the taking down of the “L,” implying that the deterioration of the house is relative to Ethan’s physical degradation. Of this the narrator says “I saw then that the unusually forlorn and stunted look of the house was partly due to the loss of what is known in New England as the ‘L’: that long deep-roofed adjunct usually build at right angles to the main house” (20). In context, a failing homestead would indeed be one that lost such a luxury and dismantled its L due to a lack of productivity, as it originally allowed farmers to access the barn from within their home and not bear the climate outside.

To use the words “forlorn” and “stunted” to describe this scene of the failed New England farm is surely a personification meant to draw this thematic parallel between person and place. It functions to further align Ethan with his landscape and surroundings, as if they are a part of him just as much as he is a part of the landscape. Like the house,
Ethan too is “stiffened and grizzled” (4), a “lameness checking each step like the jerk of a chain” (3); calling him “bleak and unapproachable” (3) is interesting in that it seems, at times, that Ethan is described in unhuman ways—as if he is a feature of the landscape more so than a person—and his grim and broken surroundings—his forlorn and stunted home—are described in a more anthropomorphized way because they are parts of his lost and fractured humanity.

At this point it is easy to see how, through a similar adaptation of methodology and practices, qualitative inferences are attainable through ostensibly quantitative means. Since Richards speaks predominantly about poetry in these four aspects, I feel it sufficient to deal with sense as the central component to understanding prose, mainly the fiction realism dealt with in this example, for the way that they intertwine and essentially allude to one another; to grasp one is at times to adequately implicate the others. Or, in a more general sense, that inferences about tone, intention, and feeling are reachable when dealing with the sense of prose, as I illustrated in the close readings above. Afterall, Richards notes that “Feeling (and sometimes Tone) may take charge of and operate through Sense in another fashion… [which is] better described as sense interfering with and dominating feeling and tone” (180). If feeling, tone, and intention operate through sense, as Richards has noted, then we can understand how I have arrived at the ecological, spatial, and anthropomorphic metaphors surrounding Ethan Frome’s character when I first attempted to grasp the “sense” of places—Starkfield, house, kitchen—as a central narrative detail. This interrogation of the “sense” of these words and their place in the text’s interpretive milieu led me to the feeling (its underlying metaphorical value),
tone (the negative emotional connotations) and intention (the characterization that these
details functioned to create).

Of last relevant mention in the realm of theory would be William Empson’s *Seven
Types of Ambiguity* which similarly lays the foundation for the ways we think about
interpreting written language, serving as a final component. As I have been endeavoring
to argue that words act as signposts to broader thematic and interpretive moments, I draw
similar attention to his descriptions of the ambiguities that encourage interpretation.
Empson notes that “behind this notion of the word itself, as a solid tool rather than a
collection of meanings, must be placed a notion of the way such a word is regarded as a
member of the language” (6). He is speaking of the word’s operative function on an
individual level, how it contributes to the interpretive milieu of the whole. In more
abstract but illuminating terms he elaborates that “one may know what has been put into
the pot, and recognize the objects in the stew, but the juice in which they are sustained
must be regarded with a particular respect because they are all in there, too, somehow,
and one does not know how they are combined or held in suspension” (6)—and at this
point I would begin to suggest that further example of this would merely echo what I
have demonstrated before with my attentiveness to word choice, its prominence and
presence as indicative of interpretive atmosphere, as a pathway toward analytical focus:
whether we call it the handling of language, the interaction of *what is says* and *how it
does so*, or the fabric of ambiguity that invites interpretation, we notice that interpretive
potential is hinged to keywords that delegate larger passages which we then close read for
interpretive significance.
Among the seven types that Empson provides, I feel the first is most worth mention in the context of my discussion for the way it feasibly aligns with the examples I provided above. “First-type ambiguities arise,” Empson says of this type, “when a detail is effective in several ways at once, e.g. by comparisons with several points of likeness, antithesis with several points of difference” (22). This is metaphor, the comparison of unlike things for abstract relatability. Here I would return to the corresponding descriptions of the house and the character (Ethan) as a spatial metaphor. The metaphor is similarly hinged to the ambiguity of the choice words used to levy the corresponding descriptions. The “stunted” and “forlorn” house versus the “stiffened” and “grizzled” person. The definition of the word “forlorn” used here to describe a house, is emotive in nature, unlike a physical edifice, it is nonetheless imbied with human quality; and so too unlike its subject is the word “stunted” which refers to physical growth and development. As a verb the word “grizzled” means to complain, grumble, and “stiffened” means, obviously, to become rigid. The ambiguity here lies in the applications of these words to their respective subjects, which, in their “logical conflict” (234), which Empson notes as central to interpretive ambiguity, they situate the spatial-ecological metaphor that surrounds Ethan’s characterization—not in the way they “fit” logically with their respective subjects, but for the ways in which they do not.

I have endeavored thus far to understand close reading through a variety of important theoretical perspectives, to see it in a skeletal sense, in a basic criteria that can be adapted to computational methodology. The commonalities of these theories have allowed me to do so by understanding how, in their commonality, each seems to center around “the word”—a key word—as a signifying feature of interpretive significance.
Schleiermacher’s technical interpretation was one of identifying grammatical constructions, the handling of language, as indicative of theme, to which I added the idea of grammatical constructions as consisting of “graphical features” embodied in words that guide us to the broader interpretive space surrounding them in sentences and passages. Barthes’ worked very much in the same with his notions of the sign and chains of signification among words and throughout passages. I used these voices to first suggest that the presence and prominence of key words acts as a signposting that encourages and drives the potential for further interpretation, the layering of meaning and significance.

Moving toward more current voices, I involved IA Richards’ *Practical Criticism* and William Empson’s *Seven Types of Ambiguity*, noting that Richards’ theory of *sense* is similarly tied to the significance of *the word* as both emblematic and indicative of the text’s broader interpretive features. Empson affirms the same, but with a more direct attention to the lyrical fabric of metaphor that underpins interpretive ambiguity. In each example I have tried to support my own theory that interpretive space is reducible to the key word as an initial guide, and that the rest works more or less metonymically, identifying the part and tracing it towards a view of the larger whole. We have seen this in the way I was able to move from identifying prominent key words—“face,” “kitchen,” “said” as an action—as well as descriptives, “grizzled,” “stunted,” etc.—and using them as a means to deduce broader thematic, and even formalistic inquiries.

I mean to do this as a way to understand the act of close reading in a technical sense, but also as one that is—most importantly—*translatable* to a computational disposition; i.e. in a way that a computer analytic can mimic and carry out. The word cloud began to illustrate this adaptability when I used it to focus on word prominence, but
the further functions reserved for the latter part of this chapter will show it most vividly.

At its base, close reading has thus far been understood as “the thorough interpretation of a
text passage by the determination of central themes and the analysis of their
development” (Janicke 2); a “sustained, probing analysis… meticulously analyzing
patterns in language”; “a way of attending to the interplay of saying and meaning”
(Hinchman 443). But these are limited and broad, I would suggest, and do not present
such a precise criteria as I would feel is suitable to a computational framework. It is a
preconception, that the act of close reading is just that, more abstract than the binary and
figure-oriented disposition of a computational analytic meant purely to deal with numbers
and figures. But sifting through these theories and identifying the common emphasis on
“the word” as a signifying feature of the broader interpretive concerns that close reading
attends to has allowed us to view close reading in a computational way.

At this point I levy a definition of close reading that I argue is suited to the
analytical dispositions of computation: close reading is the study of the ways in which
words operate interpretive moments in writing. Words function within passages that we
close read, but they also guide our attention to the passage in the first place. So, close
reading begins with identifying key words, then key passages.

From this standpoint, we understand that computers are suited to the identification
of key words by command. This is self-evident from within the context of even mundane
and user-level applications such as Adobe and the way we search through PDFs. But as a
data analytic framework for computer coding it would follow two steps, beginning
essentially with the word cloud approach and then progressing toward the processes I
have reserved for the end of this chapter: using the word cloud to identify word
prominence as a thematic indicator, and then broadening the analytical perspective to observe how that word or words function and operate in larger sentences and passages. Thus, a computational program capable of facilitating the act of close reading would be one that can in one part identify significant key words that predominate a text (as did the word cloud) and then in a latter step aggregate passages containing this word or words so that a reader can observe their function in larger interpretive moments, the ways in which that piece of significant language is handled in and throughout the text. Close reading through computational means then becomes a two-step process of word clouding, to identify significant words, and then topic modelling, to observe their operation in larger quotes and passages.

Types of Computational Close Reading

At this point in the discussion we are ready (or I hope we are ready) to take a markedly computational turn as regards the topic and substance of my discussion. Now I will begin with an explication—as well as a demonstration and instruction—on the computer programming language that will output the material for our close readings of the texts dealt with in this study: the print(close reading) that I alluded to in my preface, introduction, and which serves as the title—the main focal point—of this volume. As a prefatory remark, I ask that my literary scholar readers ‘bear with me’ as I do not want to excite alarm in them with how foreign this material may look at the outset; but hopefully the concept is clear enough to them now.

I call this portion dealing with my close reading codes one part an explication, demonstration, and instruction for several reasons. Chiefly among them is that I want for
this process to be made use of by the non-user—as I said in my introduction—the literary scholar without a humanities computing background, so while I present the codes I will likewise exert great dexterity to explaining their function so that a lay user could borrow and make use of these processes with ease and precision. As a demonstration and instruction, I mean that I aim not to only present the outputs and datasets as static entities. As the next chapter will be reserved for direct discussion of the texts in question, I will also explain the literary-methodological thinking behind the codes and outputs, alluding to the ways in which I intend to make use of the output material in the next chapter as the substance of my close readings, similar to what we saw in the beginning of this chapter with the word cloud.

To reiterate the literary dispositions of this study before dispensing with the coding methodologies, my goal here is to use the data analytic capabilities of computer coding to, at one point, test the depth of analysis possible through computer tools which have as of now been considered most applicable to distant readings of large-scale textual features. These codes and the methods employed with them have already begun to suggest, through reference to the word cloud alone, that we can indeed draw in our interpretive proximity to a text just as long as the tools are harnessed with the proper intention, that we can read the subtle, precise textual details that situate interpretation on a phenomenological level within a single text, like *Ethan Frome*. On another level, by attaining a close interpretive proximity to the literary text through computational means, we are likewise empowered to consider the cognitive-interpretive proximity of the reader in the subject-matter of literary realism.
So there will ideally be two sets of codes to present in regard to the four texts I’ve chosen for this study. The first will be a word cloud composition of the novels as a way of identifying dominant and prominent textual features and details, akin to the measures of word prominence from the beginning of this chapter that I used to trace important thematic and formalistic details in the text. The next step will be in the quotation-aggregation codes which gather instances of the prominent key word in larger passages and sentences, wherein we can observe their semantic operation in larger moments of the text. In this culmination I argue that the program has performed a close reading in the computationally translatable sense: it will be shown to have extracted passages from the text based on keyword specifications, in this case passages based on thematic significance relative to prominent textual details.

In designing these codes, as I have affirmed throughout, my goal was to align the literary purpose with the computer methodology, or, phrased in another way, to align the literary methodology with the computer capability. The more overarching goal has been to preserve the institution of literary studies over that of computer science when instituting computer integration, as a prioritization of disciplinary borrowing that did not obfuscate literary studies from the analytical inquiry but foregrounded it, and positioned the extra-disciplinary integration as a secondary to the primary analytical field. So while on a surface level it may appear we are using computer application as an expedient to tasks that could otherwise be done without it, amounting to a matter of preference at that point, I argue here that the coding processes I have developed also amount to an act of viewing, a way of close reading, because of the ways it is delegated by the computer
methodological capabilities. It may, in its culmination, constitute a computational worldview on reading because of its particularity.

Another goal, however, to further the preservation of literary studies in this interdisciplinary framework, was to maintain the desire for recourse to the hardcopy text, regardless of the data analytic gleanable through the coding functionality. The use of a coding interface should not capitulate the very things I bemoaned in the previous chapter, and replace the text—instead, as an act of viewing, it should illuminate the text for us and act as a guide in navigating it, analyzing it. The coding process then should amount to an act of navigation in addition to one of viewing, because it illuminates details, patterns, which we can then go back and observe. It becomes a mode of dissection where we can view the text’s anatomical parts and bring that perspective to bear on our consideration of the larger whole.

A final caveat for the coding exhibition that is to follow would be this: I will exert great energy, in the presentation of these codes, to explain them in a way that a non-user, after observing me ‘unpack’ these lines, would be confident in using this program and experimenting with it themselves. I have conceded before that introductory study and training in coding is inherently necessary to articulate one’s self at this level of intermediacy, but at the very least I am to hope that after my explanation a reader would feel that a knowledge of coding is at the least more accessible to them now, that they have coherent questions that they could feel confident bringing to such a learning environment and letting it serve as the first beginning step of their new training.

In what follows I will demonstrate and explain the Python coding processes that together create a computational close reading methodology.
To reorient ourselves, I will recount the analytical procedures that the Python codes open for us, envisioning it as an interpretive process in and of itself. I would suggest that when we read, we are looking for significance, or at other times, patterns. Being textual and therefore linguistic in nature, word—through their arrangement, quantity, and quality—create their significance and embody meanings and themes that predominate and thus characterize the text in question. A logical way then to identify significance in the realm of a text’s meaning is by ascertaining the prominence of particular words and then, by considering their type and quality, understanding the theme and meaning that they bring to the moment of representation. The notion can be summarized in a short adage that, I would like to affirm, should mark the nature of computational interpretation: word prevalence is indicative of theme because diction constitutes the text’s formal and semantic composition.

The first computational task is then to ascertain word prominence, and then, observing their operative function on interpretive levels. As I showed with my examples of Ethan Frome, this can give way to both formalistic and thematic questions. I use the word ‘prominence’ in the first stage rather than ‘quantity’ because quantity would entail a complete manifest of the text’s lexicon. This I believe is irrelevant because I am not interested in a rote datafication of the text’s linguistic properties, but am vested more in the interpretive concern with significance, which is connoted by prominence—the recurrence of a word, the textual space it occupies and why, how it operates in the text, the ways the language is handled.

For the purpose of variety, I will base the next example on Kate Chopin’s The Awakening, and use it to introduce questions that I will treat in-depth in the next chapter.
To begin, I will measure word prominence looking for significance on the levels of form and theme. The word cloud used to measure these factors appears relatively different than what we see in *Ethan Frome*. In the case of *The Awakening*, we see other particularities stand out:

(Figure 3.3)

At first glance of this frequency measurement I can discern several things of interpretive substance, while others appear too obscure for immediate conjecture. The first anomaly I notice is the predominate size of “did” and “is,” the past and present conjunction of the phrase “to be” in English. This leads me to hypothesize that there is a tension, parallel, or conjunction between the use of past and present verbal tense in this story and its narrative structure, that there could be an interpretive duality to the rendering of tense as the story’s details are received by the reader in the interpretive moment. More inquiry into this initial notion will be applied in the second analytical stage.

Aiming to identify more word-objects for examination, I would move from a formalist consideration of verbal tense and look for something thematic, one that could be coupled with or not coupled with this initial observation. I would next arrive at the words “upon” and “no,” prevalent in the cloud, and consider what their frequency might suggest
along the lines of thematic analysis. The word “no” makes me arrive at my notions about
the word “upon,” but at any rate, I interpret the word “no” as a verbal declaration and a
speech act, the declining of a request or question asked by a speaker. This makes me want
to consider the conversational nature of requests and questions in the text of The
Awakening, that analyzing them specifically could elucidate thematic questions about the
dynamics of power and authority in the story (who is saying no—why and to whom? And
what thematic sentiments are conveyed), if we are to be drawn now to passages where
requests or questions by a speaker are being responded to with a negative, and
considering their thematic effect. The prevalence of the word “no” among so many
others, suggests that this is a dominating compositional feature that contributes to the
text’s thematic milieu, and so is worth understanding with precision.

In regard to the significance behind the prevalence of the word “upon,” it
similarly makes me envision a speech act of exchange but also the exchanging of
physical things, as if we were to ask the question of when and in what nature things are
placed upon other things in this story (such as money upon the countertop, head upon
one’s bosom, and other instances we are likely to find when analyzing direct word-
ocurrence passages) and how do these moments additionally elucidate an interpretive
perspective.

We can further investigate the significance behind the occurrence of these words
through additional Python processes. In another step closer to these interpretive
properties of the text (peering into the cloud), we can use a frequency analysis to tell us
where and in what quantity these words appear. The module, in plain words, returns a list
of the pages on which a word occurs. I will start with the word “no” to continue to pursue
my interest in the significance of speech acts of declination in *The Awakening*. Using a process known as a PDF file parser, I can retrieve passages, in five-line chunks, of every occurrence of this word throughout the text, as a way of taking a cursory glance at its appearance, this time abbreviated into the specific moments themselves. While some are bound to be more substantive and important than others, I am able to arrive at this particular instance on page seventeen:

(Figure 3.4)

A few words on pragmatics, we can see here how the output operates in the form of a list, where one is enabled to analyze the otherwise disorganized corpus of text with focus on a particular object of inquiry, and how it molds a quantitative operation to assist with a qualitative question. As I searched through the sample sets, this occurrence on page seventeen struck me as being exactly what I had initially hypothesized: a declination through dialogue. The line itself, to me, is fitting for an epigraph on a comprehensive discussion of *The Awakening*.

“And have you been running away from prayers ever since, mach’re?” asked Madame Ratignolle, amused.
“No! oh, no!” Edna hastened to say. “I was a little unthinking child in those days, just following a misleading impulse without question.”

I have arrived at my first analysis passage—or, to be more literal, these coding processes have brought me ‘close enough.’ Noting that there is much more to observe in this moment as regards a broader understanding of the story, I can use an additional PDF file reader to return me the entire contents of a particular page and read the broader moment there, or I can refer to the text itself on my own. It would be a matter of one’s desire, or need, to lift their hands from their computer and shift their focus to another object. I will not belabor with its presentation here, but move now into my close reading of the passage in question.

This moment depicting the speech act of conversational declination on the part of Edna’s response to Madam Ratignolle’s question captures an important sentiment of Edna’s psychological characterization, one that foregrounds the major themes of the story as well as sets them in motion toward the culmination that we receive at the end in Edna’s literal suicide, and its symbolic connotations.

This exchange occurs fairly early in the story, which is why I feel it important in foreshadowing and setting in motion themes that persist throughout the story. At this moment, Edna is sitting with Madam Ratignolle, and Edna is visibly deep in thought amid the conversation of the rest of their companions. Madam Ratignolle asks Edna “of what are you thinking?” (17). Edna’s first response, a declination, appears interesting in that it implies hesitance on the part of Edna revealing her true emotions. She says, perhaps curtly, “Nothing,” but continues, just as abruptly, “How stupid! But it seems to me it is the reply we make instinctively to such a question” (17). What is the story of The
Awakening, but one about an individual’s inner tension between conflicting instincts: of mother, of woman, and of individual. As in much of the story, Edna is hesitant to speak the truth. But, as in the culmination of the story, she tends to ultimately give way to her true feelings nonetheless in spite of this conflict.

As Edna divulges what she was thinking about, we receive the first foreshadowing of the ominous ocean by the way she describes the first scene at mind, and the second that comes after it. She says firstly, that she was thinking of:

the sight of the water stretching so far away, those motionless sails against the blue sky, made a delicious picture that I just wanted to sit and look at. The hot wind beating in my face made me think without any connection that I can trace of a summer day in Kentucky, of a meadow that seemed as big as the ocean to the very little girl walking through the grass, which was higher than her waist. She threw out her arms as if swimming when she walked, beating the tall grass as one strikes out in the water (17).

So much of The Awakening is about physical space, and Edna physically escaping spaces wherein social laws that she struggles with and against are imbibed. In the novel itself, the ocean as a thematic object is the main vehicle for this metaphor. Its tenor is physical space and the proximity it extends—or affords Edna—from the places in which she feels irrevocably oppressed. As this is a competition between social law and individual passions—competing instincts—Edna alludes to this escapist impulse subtly and subliminally throughout the story, connoted here in her fascination with and mental attraction to “the sight of the water stretching far away” [my emphasis] (17). This moment begins to underscore Edna’s unconscious desire to get away.
Edna notes to Madam Ratignolle that this scenery reminded her also of a similar sensation during a day in Kentucky when she was younger. She elaborates upon this as well when asked by Madam Ratignolle, and the scene is just as telling in capitulating this thematic metaphor. Madam Ratignolle asks rather plainly, and in a trivializing way, “Where were you going that day in Kentucky, walking through the grass?” (17). Edna does not respond in kind to this act of trivialization, but relates the scene sentimentally as if she were recounting it to herself at this point moreso than Madam Ratignolle. She says: I don’t remember now. I was just walking diagonally across a big field. My sunbonnet obstructed the view. I could see only the stretch of green before me, and I felt as if I must walk on forever, without coming to the end of it. I don’t remember whether I was frightened or pleased. I must have been entertained. Likely as not it was Sunday, she laughed: and I was running away from prayers, from the Presbyterian service, read in a spirit of gloom by my father than chills me yet to think of (17).

Again Edna turns a pastoral, scenic depiction into a rumination on the unconsciously freeing, liberating emotions afforded by great physical expanse between one’s self and others. What Edna says lastly in this moment makes significant what she says first. She was “running away from prayers”—or more specifically, escaping a Presbyterian service, “read in a spirit of gloom” by her father, and which “chills me [her] yet to think of.” Edna’s characteristic mode of escape from oppressive social spaces is in that of attaining and extending space between herself and those social environments. On a philosophical level, she escapes her community by fleeing to the pastoral, which brings physical as well as emotional relief, escaping social law by seeking physical space. On a literal level, she
says “I could see only the stretch of green before me, and I felt as if I must walk on forever” (17), positioning the act of leaving the community for a pastoral expanse as both a physical as well as an emotional escape.

This thematic moment essentially culminates in the conversational declination that I first identified in my word frequency analysis, when Madam Ratignolle keys on Edna’s phrase of “running away from prayers.” The phrase that elicits Edna’s declension is significant in itself: “And have you been running away from prayers ever since, mach’re?” (17). The moment of response by Edna reads specifically as “‘No! oh, no!’ Edna hastened to say” (17). This narrative commentary after the response is key in setting the context of Edna’s response and dressing it in a particular sentiment that marks Edna’s characterization as embodied in moments of dialogue such as this. Edna “hastened” to say “no” as if she were correcting herself, or, correcting her response before the ears of her social community members. It marks a sublimation of Edna’s true emotions, as much of the story does. Following this inference, we can infer that Edna is lying in her response to Madam Ratignolle, that she indeed had been running away from prayers ever since, and that, eventually, we will see her swim away, too.

The rest of Edna’s response to this pivotal question is dressed in similar language, where Edna lies about her true sentiments to save face before her social companion, making the scene more ironic than anything—ironic in the sense that, while on one level she is lying to Madam Ratignolle, she is also lying to herself, signifying the grapple and struggle with competing desires that characterizes the story. Edna says, in reference to her escape out into the field away from the Presbyterian sermon, that “I was a little unthinking child in those days, just following a misleading impulse without
thinking” (17). Edna exerts what she feels is a proper deference to what is “right” in this moment, that the unthinking impulse was the one that led her away from this social space that made her so uncomfortable in favor of the open pastoral space where she felt individuality and freedom. But the truth is, perhaps unbeknownst to her, that the true unthinking impulse is the one that led her to stay in these oppressive social spaces, and the true, thinking impulse is the one encouraging her to get away. The emotional repression, as we can see, is latent in the language when viewed conversationally.

Edna continues to unconsciously elaborate her emotional oppression through irony until the end of this scene. She continues, trying to redact her Freudian slip: “On the contrary, during one period of my life religion took a firm hold upon me; after I was twelve and until—until—why, I suppose until now, though I never thought much about it” (17). That Edna says religion “took a firm hold” on her in her younger years conveys sublimated feelings of oppression of which she desires not to admit—to her listener or, at this point, herself.

The moment culminates in physical cues that further convey Edna’s indecision about her emotional motivations, and continue to set in motion the escapist themes of the story through the desire for pastoral physical space. As Edna speaks about the presence of religion in her life, she stops abruptly saying “But you know..” and the text surrounds this utterance with physical cues to convey Edna’s emotional tribulation: “she broke off, turning her quick eyes upon Madam Ratignolle and leaning forward a little so as to bring her face quite close to that of her companion” (17). Edna now feels compelled to speak in confidence, almost secret, to Madam Ratignolle about what she has next to say, proclaiming that “Sometimes I feel this summer as if I were walking through the green
meadow again; idly, aimlessly, unthinking and unguided” (17). Edna is reaching a breaking point that is slowly but surely progressed toward throughout the story and in the seemingly most mundane of narrative moments. That progression is to be observed in the subtleties of Edna’s conversations with others, perhaps even more so than the literal moments of physical action on Edna’s part—in her acquiring a house of her own, for example, seen as a pivotal step in her character development and the themes of physical proximity that permeate the story, Edna can be observed pleading for more space in some of the very first pages of the story, even conveying that she had been striving for that space long before the story even began.

Hopefully now we feel that consideration of the use of the word “no” in *The Awakening* can teach us much of its interpretive atmosphere. In this single instance of many throughout the text, we have been able to delve deeply into the essence of Edna’s character and see how this is integral to a broader understanding of the text, its themes, narrative form, and its fictive motivations. Hopefully, too, we can begin to see how the human and computational interpreters must work synergistically for a truly logical collaboration of methods. The computer identified prominent words, their quantitative significance, and, through the consideration of the literary theories employed to justify this chapter, I was able to deduce its interpretive significance, to understand its prominence as a unit within a larger interpretive situation. It affirms my initial maxim on (computational) close reading: that word prominence is indicative of theme and form. Conversational declinations, connoted in the use of the word “no,” convey aspects of Edna’s emotional repression that both foreshadow and set in motion the major themes of escape that comprise the novel *The Awakening*. 
Introducing the Python Close Reading Bot

What I have supplied here in this chapter is but one initial exhibition of computational close reading in action, and a full treatment of the four texts—*Ethan Frome*, *The Awakening*, *The Country of the Pointed Firs*, and “The Yellow Wallpaper” will follow in the next chapter as my primary literary contribution and computational demonstration. But, before that, it would be fitting here to present the codes themselves that perform the functions that I used to compose this analysis.

I will warn non-users to please trust me, as these codes will appear at first glance a nonsense of periods, parentheses, dashes, and incomplete words; but I will also encourage (caution) us to understand that this is a language, too—a command language, to speak computationally—but it is a language at its base, and society’s authorities of everything linguistic are still—and should be, I argue—the best equipped to adopt and understand it. Coding is—to put it into words our field would be accustomed to—administering commands to the computer to handle, aggregate, or organize data. The nature of the commands lies in what you want the computer to do with your data. In my case, it was, in a series of steps, firstly to create a word cloud detailing word prominence, and secondly to retrieve passages and pages of text based on a keyword of my choosing. I deduce my analysis from the desired organization that I elicited from my dataset. The organizational method encoded in the computer is, essentially, a construction of my analytical interest, combined with the computer’s capability to identify significance in instances like the word cloud.
To illustrate my first analytical step, and to reiterate in a concluding sense, I use the program first to identify and measure the degrees of word prominence throughout the text (organize it into perceivable data) because of my first theoretical notion that theme and form are connoted by the presence of significant words that act as signposts for particular themes and of particular forms. The word cloud allows me to perceive word prominence, and make inferences on how to carry out the next steps involved. The word cloud process is quite simple, and I would even suggest accessible as a beginner-level function.

The word cloud function is a short one save some adjustments a user can make based on their desired output. It involves first processing the page-as-text making it readable to the program, and then setting the text into a “string”—literally so, a straight line, word by word—and using that string the generate the cloud, which produces a visualization (the cloud) organizing the words by size based on their respective quantity.

The initial code looks like this:

```python
[1] import textract
[2] import difflib
[3] import nltk
[4] import re
[7] from wordcloud import WordCloud
[8] long_string = (book1)
[9] from nltk.corpus import stopwords
[10] filtered_words = [stopwords.words('english')]
[11] wordcloud.generate(long_string)
[12] wordcloud.to_image()
```

(Figure 3.5)

Running this command in a Jupyter cell will produce the word cloud visualization that I used in the discussions above. A bit of what these phrases mean on a practical level is in
order. The first four lines, the imports, are a series of process oriented specifications that orient the program for a specific function. The module “textract” is invoked for text processing, “difflib,” “nltk,” and “re” are all similar text readers, making the program aware of things like pagination and file format.

Lines five and six perform a process known as “tokenization,” which processes the otherwise visual document—the novel—into lines of text. It is the default for the program to receive the document as rote images of pages, and processes like textract essentially “code” or “process” the document in a way that is readable as textual data. The sixth line creates a variable, “book1,” that contains this newly processed body of text. It is then invoked in lines seven and eight when generating the cloud. After we “import word cloud,” note our desired function, we then convert the text into a long string: long_string = (book1).

I will return to lines nine and ten, as the final lines are more important for an initial understanding. Lines eleven and twelve take the long_string that we created from the tokenized text and then converts that into a visual cloud. Perhaps we can sense the syntax of the commands at this moment, as what we have done is taken a document: theawakening.pdf, converted it into text: (text = textract.process('theawakening.pdf')), then named it in its new form: (book1), converted that text into a string: (long_string), and converted that string into a cloud: (wordcloud.generate(long_string), with the final line: wordcloud.to_image(), producing the visualization on which I based my previous discussion.

I said I would return to lines nine and ten to describe their specific role. I do so because they are inherently optional but also important for a number of reasons.
Including stopwords in a text processing function can be used to basically ignore words that would otherwise overcomplicate or obscure the desired data output. In the case of the word cloud, this would involve words with little to no interpretive value, like articles, but that might appear frequently and take up space in the cloud but not be representative of the more meaningful lexicon that one is looking to measure. In looking for more substantive, meaning bearing words, I use stopwords as a way to prioritize the most substantive words in the text, such as nouns, adjectives, and other descriptors that one feels would be relative to theme and interpretation. In other scenarios, a user may want to include rather than omit stopwords—for example, I would not put it past a talent linguist out there, who may be interested in interpreting the quantity and occurrence of the indefinite article.

The next process is another similar act of quantity approximation. To be feasible, I will continue to walk back through my example of *The Awakening*. Since the word cloud produced by the above code helped us ascertain the prominence of the word “no,” I next wanted to measure its quantity per-page. The following code allows me to do so as a way of identifying the areas of the text most worth my attention in this regard. This is more of an optional step, as the *repeated* occurrence of a word may or may not be significant in various circumstances.

The process itself is a short one, involving a brief series of commands with some that are repeated, such as the file upload.

```python
[1] import PyPDF2
[2] import re
[3] fileName = 'theawakening.pdf'
[5] numPages = object.getNumPages()
[6] print(numPages)
```
pattern = "[Nn]o"

for I in range(0, numPages):
    pageObj = object.getPage(i)
    text = pageObj.extractText()
    for match in re.findall(pattern, text):
        print(f'Page no: {i+1} | Match: {match}')

(Figure 3.6)

Lines one through four import the necessary modules and read the pdf document, converting it into an object that can be manipulated as in the previous codes. Lines five and six instruct the program to notify us of the page number containing our desired data output, for easy reference. Line seven identifies the pattern we are looking for from within the larger dataset (the text of the novel). We will notice an obscurity in the pattern specification: “[Nn]o” being the way I phrase the word “no.” I do this because I am interested in capitalized as well as uncapitalized occurrences, and the brackets mark this specification. Lines eight through twelve are what essentially perform the quantification process, identifying instances of the word “no” by the page of their occurrences and their quantity. The phrase “for I in range(0, numPages):” means, in plain words, “for instances (of the above pattern) in the page range—with 0 in Python connoting all pages—then the following lines: pageObj.extractText, meaning, give us the text when it matches this pattern. But we do not want the text just yet, so the final two lines are essential. We say “for match in re.findall(pattern, text)—meaning, for all matches of this pattern in this text (then the next line) we want to “print(fPage no) or, give us the page number.

The output is a fairly decipherable list that one could use as a point of reference, noting how many times a particular pattern occurs on a particular page. An example of the output looks like this:
After understanding the prominence and then the frequency of a particular keyword, the next process will allow us to retrieve passages containing any key word we desire, as sample sets on which to base a close reading of the thematic or formal object. It is a bit longer than the others, and can be adapted for further precision, especially considering, for example, how my interest in the word “no” developed into one of conversational declination—we can apply added specificities to this effect.

```python
import PyPDF2
import re

fileName = 'theawakening.pdf'
object = PyPDF2.PdfFileReader(fileName)
numPages = object.getNumPages()
print(numPages)

pattern = r'\[Ss\]aid|[Nn]o'

for i in range(0, numPages):
    pageObj = object.getPage(i)
    text = pageObj.extractText()
    for match in re.findall(pattern, text):
        print(f'Page no: {i+1} Match: {match}')
```

(Figure 3.7)
for line in text.split('n'):
    lines.append(line)
lines.append('n')
lnum = 0
for line in text.split('n'):
    lnun = lnun + 1
    match = re.search(pattern, line)
    if match:
        line = line.replace('n', ' ')
        line = line.encode()
        print(f'Page no: {i+1}')
        print(lines[lnum-3])
        print(lines[lnum-2])
        print(lines[lnum-1])
        print(line)
        print(lines[lnum+1])
        print(lines[lnum+2])
        print(lines[lnum+3])
        print('--------------------------------')

(Figure 3.8)

By now, hopefully, aspects of the code seem familiar and we can further understand the syntax of computational close reading. The import modules occur in the first two lines of code, while lines three, four, and five upload the document of text. Line five, in particular, instructs the program to differentiate page numbers. Lines six and seven institute the pattern of interest, which I’ve emended to include instances where the words “said” as well as “no” occur in proximity of one another, connoting occurrences of conversational nature, as per my analytical interest. All lines after line seven are indented because they are specifications of the original command to “identify these patterns.” For this instance in range 0, means, to reiterate, for all of these instances in all pages, identify this pattern. Lines eight and nine instruct the program to do both things: to identify the page number of the instance and retrieve the text in question.
The phrases “lines = []” and “lines.append("")” instruct the program to output these results into lists, isolating them from the rest of the dataset. Lines fifteen through eighteen set the pattern specification: “for line in text.split("n"):” essentially means to differentiate these patterned instances from the rest of the text, and the phrase “match = re.search(pattern, line)” means to search for these patterns within lines of text. The final lines, nineteen through thirty, set the parameters for the output we are to receive. In plain words, we say “if match:” output the following details. The final print commands may seem repetitive, but they mean to output the line of the pattern instance (print(line)) and the three surrounding lines (-1,-2,-3 and +1,+2,+3), so the pattern instance will appear in the middle. So the program identifies a line of text in which our keyword appears, and then outputs sample sets of that line with three additional lines on each side. A sample of the output I used for my discussion of *The Awakening* looks like this:

![Example output](image-url)
Difficult to accurately illustrate here, but this image, within the program, is a scrollable list of all instances of the key word with three surrounding lines on each side. In the example from page seventeen, we can see the moment of dialogue that I identified, where Edna says “no” to Madam Ratignolle’s question of her running from prayers, which I used to bring my attention to the more general moment of the story.

When I desired to view the text before and after this specific instance, to grasp the moment in its broader narrative context, I used a much more simple command for just outputting the entire page text of a page of my choosing, allowing me to read and conduct my analysis solely from within the Jupyter notebook. This process repeats commands from the other scripts and is much shorter, having little need for such acute precision as the others.

```
[1] import PyPDF2
[4] print(pdfReader.numPages)
[6] print(pageObj.extractText())
```

(Figure 3.10)

Line five is most pivotal to this function because it instructs the program which page of text we desire to see. In it I write “pageObj = pdfReader.getPage(17)” so that it gives me the full text of page seventeen. This is basically how I conducted my close readings of *The Awakening* that I discussed in this chapter, and will be the method by which I dispense with my full readings of the four texts in chapter three, using Python to measure word prominence as a way to ascertain the subtle interpretive elements of these literary texts. Close reading, in this computational sense, appears a process of aggregating significant details and patterns for focused interrogation.
Coming Attractions

My third and final former chapter will enact this style of analysis—computational close reading—to the upmost, applied to the texts I have named here. The task will be to attend to the entire textual economy of a particular word and observe the interpretive deductions possible in regard to our understanding of regionalism and the novels themselves. *The Awakening* will be read in much the same way but expansively related to the act of conversational declination, so attending not just to the word “no” but its other linguistic permutations, gleanable from the outputs of the keyword identifier function of the Python bot; so similar words such as “not,” “never,” “must not,” “will not,” etc., to interpret the ways in which acts of refusal function to oscillate the story’s various thematic and character based tensions.

My readings of *Ethan Frome* will be slightly more multifaceted. While I will follow on one level the same pattern of significant word identification via wordclouding and then keyword passage retrieval, I will also introduce another bot capable of reading the text in its developmental compositional progress, i.e. variant iterations of the text that represent its revision overtime, shedding light on the author’s own creative development, made possible at a speed and precision that makes the address of such a question feasible: a Python process capable of comparing texts and identifying differences.

Apart from that one “special feature” in this exhibition of computational close reading, “The Yellow Wallpaper” and *The Country of the Pointed Firs* will be similarly approached through computational close reading as I did *The Awakening*: identifying key
words through word prominence and then observing their operative nature in distinct passages.
Chapter 3: Examples of Computational Close Reading

“Why?” asked her companion. "Why do you love him when you ought not to?"

Edna, with a motion or two, dragged herself on her knees before Mademoiselle Reisz, who took the glowing face between her two hands.

"Why? Because his hair is brown and grows away from his temples; because he opens and shuts his eyes, and his nose is a little out of drawing; because he has two lips and a square chin, and a little finger which he can't straighten from having played baseball too energetically in his youth. Because -"

"Because you do, in short," laughed Mademoiselle.”

- Kate Chopin, *The Awakening*

“It was a long time since any one had spoken to him as kindly as Mrs Hale. Most people were either indifferent to his troubles, or disposed to think it natural that a young fellow of his age should have carried without repining the burden of three crippled lives. But Mrs Hale had said ‘You’ve had an awful mean time, Ethan Frome,’ and he felt less alone with his misery.”

- Edith Wharton, *Ethan Frome*

“It does seem so pleasant to talk with an old acquaintance that knows what you know. I see so many of these new folks nowadays, that seem to have neither past nor future. Conversation's got to have some root in the past, or else you've got to explain every remark you make, an' it wears a person out.”
- Sarah Orne Jewett, *Country of the Pointed Firs*

“But I MUST say what I feel and think in some way”

- Charlotte Perkins Gilman, “The Yellow Wallpaper”
Introduction: Dialogue and Dialect in Literary Regionalism; Examples of Computational Close Reading

Chapter two has established thus far that one of the ways the regionalist women’s novel delivers content to the reader for thematic consideration is by way of placing them into moments of direct conversation between characters, viewing the exchange through the cognition of a focal character and coupling the depiction of their particular dialogue with corresponding emotional and psychological underpinnings. I will return to this notion of dialogue and conversation being more central to regionalism than other factors yet considered, as a mode traditionally read for its imagery and visual aesthetic and its onus on space and place; and although some critics have noted regionalism for its use of regional dialect, I set out to argue that conversation is both hinged to thematic development in regionalism but that it is also and oftentimes the forum by which the thematic climax is executed in the regionalist novel.

I arrive at this point primarily in response to Stephanie Foote, whose *Regional Fictions: Culture and Identity in Nineteenth-Century American Fiction*, studies regionalism for its “elaborate, lyrical constructions” (17) which posits that “regional writing gave strangers with accents literary recognition at exactly the same moment that accented strangers in the form of immigrants were clamoring for recognition and representation in the political arena” (5). While hers focuses fixedly on how and why regionalists recreated the phonetic vernacular of regional peoples in writing (such as the southern Appalachian “yourn” of Steinbeck’s novels, for example) I would like to move a step further past the particularities and peculiarities of their speech and observe the emotional dynamics of their conversational habits. Truly representative and moreover
normalizing, I intent to look past the particularities of their dialect and instead observe the particularities of their conversational interactions.

I intend to pursue this notion through the same methods applied to *The Awakening*, examining passages of dialogue exchanges that contain key thematic moments, deduced through measures of word prominence via word clouding, and then the aggregating analysis passages using the Python functions I detailed in chapter two.

For the purpose of variety however, and because it is the technical and methodological aim of this work to demonstrate varieties of computational close reading, I will also model other forms of computational close reading that at times overlap with this central thesis, suggesting other interpretive aims suited to this analytical method. I will consider, for example, using applied computation to observe the development of a regionalist work in process as a way of speaking about these tools to archivists and critical editors—viewing variants of a text across two iterations, in distinct moments of its textual lifespan. Like most novels published in multiple editions across time, variant readings occur that imply change but also development in the text’s overall composition. These can often be read to then consider the development of the text’s thematic impulses and fictive intentions, bringing to the foreground interpretive situations that can illuminate a broader understanding of the text itself. I would posit this as a form of computational close reading interested in and suited for reading the compositional process embodied in the documented revision of a literary text. Or more simply, reading revision computationally.

*Reading Revision through Automated Collation*
In the case of Edith Wharton’s *Ethan Frome*, there are two important Scribner’s editions published ten years apart from one another that consist of variants that invite interpretation: the first 1911 and the second 1922 American editions. Lending itself to my regionalist thesis, these variants markedly restructure key moments in the story that, while are not exchanges of dialogue themselves, depict actions happening during or before or after moments of dialogue, which recast the sentiments surrounding the characters’ interactions, evidence of the importance of such moments to the development of the story.

The Python functions used to arrive at this study of revision were different than the one used to arrive at my close readings of word prominence. This time I used a text comparison process known as “difflib,” a program language aimed at detecting differences between two bodies of text. I include the script and a sample output below for reference and use by others.

```python
import textract
text = textract.process('ethanfrome11.pdf', method = 'tesseract')
book1 = (text)
from nltk.tokenize import sent_tokenize, word_tokenize
text = textract.process('ethanfrome22.pdf', method = 'tesseract')
book2 = (text)
from nltk.tokenize import sent_tokenize, word_tokenize
import difflib
import nltk
import re
text = textract.process('ethanfrome11.pdf', method = 'tesseract')
book1 = text.decode()
text = textract.process('ethanfrome22.pdf', method = 'tesseract')
book2 = text.decode()
from nltk.tokenize import sent_tokenize, word_tokenize
from IPython.core.display import HTML
words1 = nltk.word_tokenize(book1)
words2 = nltk.word_tokenize(book2)
htmldiff = difflib.HtmlDiff()
tbl = htmldiff.make_table(words1, words2, context = True)
HTML(tbl)
```

(Figure 4.1)
For the purpose of brevity in this discussion, I have consulted the output table in its entirety and composed an abbreviated manifest of variants that I will discuss here. In the case of variants across editions of *Ethan Frome*, the most substantive I observed were these, save many others involving mainly punctuation and hyphenation, and other lower-order aspects of its composition.

1922 (pg. 3) although he came // though he came 1911 (pg. 3)
1922 (pg. 6) his shoulders // his strong shoulders 1911 (pg. 6)
1922 (pg. 28) its wakening life // its waking life 1911 (pg. 30)
1922 (pg. 44) abreast of her // abreast of her retreating figure 1911 (pg. 47)
1922 (pg. 54) wrist-bone of the hand // wrist of the hand 1911 (pg. 58)

The most striking variant in this observation, I would argue, is to be found in the omission in the ’22 of the word “strong” used to describe the look of Ethan Frome’s shoulders as he approached the post-office in the opening scene of the novel, implying revision to Frome’s characterization across the two editions. Much narrative effort is exerted to depict the character of Ethan in corresponding physical and emotional ways: that there was “something bleak and unapproachable” in his face (4-5); that Ethan
“seemed a part of the mute melancholy landscape” (14-15); and that his failed farm, the “diminished dwelling” was correspondingly “the image of his own shrunken body” (21), so there is little wonder why, in the revision of the first edition for the publication of a second, that this subtle description should be changed to omit the word “strong”—the act of revision here was an act of enforcement of Ethan’s characterization.

While there were many, perhaps consistently throughout, instances of variants in punctuation, owed perhaps in part to accidental formatting and line-breaking rather than substantive revision, some more distinctive changes in punctuation were also present, involving mainly these:

1922 (pg. 3) white colonnade: // white colonnade; 1911 (pg. 3)

1922 (pg. 7) moral reach permitted, there // moral reach permitted there 1911 (pg. 7)

1922 (pg. 9) devoted village, and the wild cavalry // devoted village and the wild cavalry 1911 (pg. 9)

1922 (pg. 35) from hand to hand, he wondered // from hand to hand he wondered 1911 (pg. 38)

And finally, some involving the use or lack thereof of pronouns versus character names:

1922 (pg. 37) his wife’s disfavour, // her disfavour 1911 (pg. 40)

1922 (pg. 42) Denis’s cheap banter // his cheap banter 1911 (pg. 46)

The most substantive variants observed here, from the first group, imply significant and substantive revision in a number of textual and interpretive scenarios in the reading of *Ethan Frome*. Two from the first group are subtle yet important: “though he came” in 1911 was revised to “although” in the ’22, and “its waking life” changed to “its wakening
life” in the same. While not necessarily effecting meaning and presentation in a hermeneutic sense, I would encourage readers to consider that these slight phrasal adjustments do indeed affect the tone, meter, and cadence of the narrative lyricism, and should therefore be noted as a substantive change. These types of clippings and addings of suffixes and prefixes persist variably throughout the two editions, and ought to be considered substantive rather than accidental for the way they effect the syllabic structure of the narrative. Such revisions would be done authorially for one of two reasons, in my opinion: to instill brevity or expand breadth in the orality of the narrative voice.

Two of the other substantive variants I mentioned involve the removal of a word that strongly effects the thematic presentation in the text, and a third involves the addition of a word effecting the interpretive moment in the same way. I’ve already covered the removal in the ’22 of the word “strong” to describe Ethan Frome’s shoulders as he approached the post-office in the opening of the novel, which was evidently an important detail framing his characterization early on in the novel, but equally peculiar is the removal of the word “retreating” from the ’22 edition in the scene early on in the novel where Denis Eady approaches Mattie to offer her a ride home from the church social. Approaching Mattie to offer her a ride on his buggy, the text of the 1911 edition says that he “gave the horse a cut that brough him quickly abreast of her retreating figure” (47). In the 1922 however, “retreating” is removed, and Denis approaches merely her “figure” without this additional behavioral description.

Again, this subtle variant detail effects the interpretive moment in a significant way. As it’s presented in the 1911 edition, the retreating figure of Mattie upon Denis’ approach implies indecision at best or reluctance at most on part of Mattie’s responding
to Denis’ act of courtship, his offering her a ride home as a way to spend time with her. We should remember that Frome is watching this exchange take place from a secluded position in the shadows nearby, so this moment depicting Mattie’s internal motivations is key to the text’s establishment of character chemistry. For Mattie not to have been retreating from Denis’ approach in the 1922 edition suggests revision to Mattie’s initial characterization. Retreating would have implied her intended refusal of Denis’ invitation; removing it leaves the notion of her desires—her romantic desires—more ambiguous, deepening the central thematic effect about Mattie’s romantic agency that persists throughout the story. Perhaps by describing Mattie as not retreating, but just being there, perhaps waiting hesitantly, the narrative leaves more interpretive space to suppose that Mattie would have been open to Denis’ invitation, but decided not to acquiesce out of hesitation or reluctance, perhaps after consideration that Frome would be nearby waiting for her, as we know him to be. The action and emotion connoted in the word “retreating” is hinged to Mattie’s depiction, as the two variants are not one and the same, but rather constitute significantly different interpretive moments in regard to Mattie’s characterization.

The last substantive of importance occurs on pages fifty-four and fifty-eight, in the 1922 and the 1911 editions respectively. It is slight, but still relative to image. Midway through the novel Frome is described as, in an act of affection, grabbing Mattie by the “wrist-bone of the hand” (1922) versus the “wrist of the hand” (1911). I find the addition of the word “bone” to modify the imagery in this scene significantly by infusing a type of intimacy about the image through anatomical descriptors. A metaphorical and imagistic “depth” is achieved by rendering the image in thought objects that are literally
inside of the body, as the image performs an act of penetration that is as much physical as it is emotive. The tender sentimentality that radiates between Ethan and Mattie is emblematized in the way he touches her—and them each other—at an innermost place of her being, rather than from a physical and metaphorical outside of the material body. Touching Mattie at her wrist-bone imbibes in the moment a feeling of the depth of their connection to one another, so while similarly subtle in nature compared to the other substantives, the varying space for interpretation that the variant offers is significant.

*The Country of the Pointed Firs, The Significance of Place in Regionalism, the Significance of Titles, and Keyword Passage Collection*

On the importance of titles SJ Wilsmore writes that “Titles are the means by which an individualist tradition makes clear its literary wares, their owners, and their resulting nature” (404). I propose the question of what critical insight could an interpretive study of titles provide when applied in such a way, a consideration of its embodiment of a works literary wares and their resulting nature, and its implications in the text’s overall hermeneutic. I align this particular study not in the reading of the title on the cover of the novel, or the tops of its pages, per se, but in the study of passages in which the words of the title appear, hypothesizing that they function as a sort of call-back or allusion to an overarching sentiment conveyed in its diction, and its relation to the broader story. Of topic here will be Sarah Orne Jewett’s *The Country of the Pointed Firs*. My argument is that the title is connotative of the text’s primary aesthetic object: place. While this may appear a banal observation of a piece of regionalism, I locate this connection not through the line of the title but from an object foregrounded within it: the
fir tree and its appearance in moments of the story. These appearances of the fir tree and
mentions of it both metaphorical and literal, I argue, are central to an understanding of
the aesthetic of place that drives this novel and which serves as much of the interpretive
ground of regionalism. Adherent to my original thesis, I will read depictions of place in
contention with moments of conversation, and compare their interaction with this central
object of the fir tree to understand the weight each mode of depiction carries in the text
and to what extents my aesthetic observation of the firs serves as a ground for
interpreting and understanding each facet of the novel’s methods of representation.

While the regionalist mode, I argue, seems centered on place setting in its
aesthetic and depicting community values of unique, at times remote regions, this
observation I made about the fir tree and its appearance as a connotative thematic object
works to underscore an oscillation of impulses in the way the novel depicts its region of
interest, a tension that underscores the depiction, of one between these communal values
of the remote region and the individualism that a person within such a place still
nonetheless carries. This can be observed in a number of moments working subtly
throughout the novel, but can be seen most apparently in the surface-level plot of the
novel: the narrator’s retreat to Dunnet Landing was necessitated not so entirely much as a
vacation, but a writing retreat—she is meant to be working on her own writing as she
tries to experience and appreciate the pastoral beauty and sense of community of the
people and place of Dunnet Landing.

It is my argument that this enigmatic tension between the individual and the
community in Country of the Pointed Firs is signaled in-text by the appearance of the fir
tree in some varying way. To make this clear, I bold the appearance of the word in each
of my analysis passages, because its function is at the same time subtle as it is important. At the outset, situating this tension to come, the appearance of the fir tree signals the thematic and overarching philosophical sentiments with which the novel functions to dress the regional depiction, observable here:

After a first brief visit made two or three summers before in the course of a yachting cruise, a lover of Dunnet Landing returned to find the unchanged shores of the pointed firs, the same quaintness of the village with its elaborate conventionalities; all that mixture of remoteness, and childish certainty of being the centre of civilization of which her affectionate dreams had told. (3)

Over the course of a yachting cruise one is seeing and visiting myriad places, yet this quote seems to attempt to affirm that none are like the remote, uniquely-isolated, rural place of Dunnet Landing, here called the “unchanged shores of the pointed firs.” To elaborate, I would argue that this moment imbibes the regional place with a type of pre-lapsarian quality, setting it against the more popular or metropolized places that one may visit on a yachting cruise, like some coastal cities depicted in a novel of manners, meeting-places for members of a more social community. But there is no overarching social structure that penetrates Dunnet Landing but the one community contained within it, and it preserves for itself this sense of uniquely-isolated individuality of community, with its “childish certainty of being the centre of civilization.” The shores of Dunnet Landing are at the same time the end of its geographic boundary as well as its social-political boundary.

But even within the isolated social economy of Dunnet Landing tensions between the individual and the collective persist, gleanable in the narrator’s interactions with the
native people while staying there. Different from the residents of Dunnet Landing, the narrator did bring an ostensible “outside” social commitment with her: the development of her writing. There is then a tug-and-pull that goes on in the narrator’s contemplations about her priorities while living there, whether to maintain her industrial commitments to her writing or immerse herself deeper in the conventionalities of the island’s lifestyle. Attending to the broader linguistic economy of the aesthetic object of the “pointed firs” however, I similarly located this passage through observation of nature imagery, in this case the mention of “wild herbs” which appear more subliminally in this moment:

Literary employments are so vexed with uncertainties at best, and it was not until the voice of conscience sounded louder in my ears than the sea on the nearest pebble beach that I said unkind words of withdrawal to Mrs. Todd. She only became more wistfully affectionate than ever in her expressions, and looked as disappointed as I expected when I frankly told her that I could no longer enjoy the pleasure of what we called “seein’ folks.” I felt that I was cruel to a whole neighborhood in curtailing her liberty in this most important season for harvesting the different wild herbs that were so much counted upon to ease their winter (5)

“Seein’ folks,” in this passage and in much of the novel’s description of the island residents’ daily habits, is synonymous with the work one must do to maintain their daily lifestyle, in this case gathering herbs in the fall that will help one remain healthy during the long winter. It is at the same time an act of self care and an act of immersion and community with one’s fellow residents. Being an outsider and a visitor, the narrator carries a tension about her motivations unrealized by those regularly living at Dunnet Landing: “the voice of conscience sounded louder in my ears than the sea on the nearest
“At these ‘unkind words of withdrawal’ to Mrs. Todd by the narrator, her companion is understanding yet disappointed, nonetheless affectionate in this understanding, at that, at the narrator’s inability to take part in a native habit of herb gathering. This reaction by Mrs. Todd at the narrator—the outsider—signals further social knowledge about regional place-setting, that perhaps Mrs. Todd is disappointed yet nonetheless understanding, because she knows the narrator to be an outsider with social commitments beyond the shores of their remote region, that she knows, perhaps, that an outsider such as the narrator cannot be expected to fully immerse and commit themselves to the “elaborate conventionalities” of the region of which she herself is not a fully native part, that the community of the regionalist place has its limits when it comes to outside, visiting individuals. In the narrator’s inability to gather medicinal herbs for the winter, a chore of the natives, this moment supposes that a visitor can visit a place like Dunnet Landing, but can never truly know what it means to live there.

There are other passages that seem, peculiarly, to suggest the inverse, that a community so vested in collectiveness and a common lifestyle is at the same time equipped and attuned to intense isolation during the winters, but that it is a type of collective isolation which each and every individual understands in part of one another, markedly different from Mrs. Todd’s reluctant acceptance of the narrator’s recusal of herself from the chore of herb gathering. At the close of the summer season the narrator contemplates the following:

They did not expect to see one another again very soon; the steady, hard work on the farms, the difficulty of getting from place to place, especially in winter when boats were laid up, gave double value to any occasion which could bring a large
number of families together. Even funerals in this country of the **pointed firs** were not without their social advantages and satisfactions. I heard the words “next summer” repeated many times, though summer was still ours and the leaves were green (49)

So while on the surface it would seem that the community of Dunnet Landing is marked by its collective, communal lifestyle, prolonged isolation is also engrained in their lifestyle, working to make moments of collective gathering all the more special for that very reason. It reveals an inherent ambiguity of the regional mentality: whether they are indeed driven by community or individualism. Here it seems they are driven to cherish community *because of* individual isolation that marks the island lifestyle. I observe this mainly in the contradictory way this moment casts the event of a funeral, a somber event, but one not without its “satisfactions” because it brings members together and into meeting.

So while there is “double meaning to any occasion which could bring a large number of families together,” like a funeral that is at one point a scene of mourning but also a reason for communal gathering, there is also double meaning, in this novel at least, to what makes a place a *region* in regionalism: is it indeed marked by community or isolation? It seems, in moments like this, that while the regionalist place in question is marked by its isolation from a broader metropolized body politic, there is also isolation within the isolated region—oftentimes a rural, agrarian place—which precipitates, necessitates the sense of community within the place. The people of the region garner, seek, and cherish a sense of community because of their collective sense of individual isolation.
In addition to the themes embodied by regionalism, there is much to be understood through this method of analysis, the tracing of the appearance of this aesthetic object of the fir trees, about regionalism’s methods of content delivery—those methods of narrative immersion in which I was interested in my introduction. Perhaps by now we can note that The Country of the Pointed Firs is a highly imagistic novel, consisting primarily of the narrator’s contemplations of the regional landscape and the sentiments evoked in their image, constructing its narrative effects. In my keyword passage gathering, moreover, I was at a loss to find the fir tree mentioned in any pivotal moment of dialogue, but I was also pressed to find any pivotal moment of dialogue in The Country of the Pointed Firs, much less a predominance of it.

Instead The Country of the Pointed Firs, in contrast to the other novels dealt with thus far, is an experience of viewing whereas Ethan Frome and The Awakening are, at times, experiences of listening in. In The Country of the Pointed Firs, the reader views the regional landscape through the narrator’s eyes and registers the experience through the narrator’s emotional viewpoint. This is evident as its mode of narrative immersion in the final two passages I choose to read here. The first:

We were standing where there was a fine view of the harbor and its long stretches of shore all covered by the great army of the pointed firs, darkly cloaked and standing as if they waited to embark. As we looked far seaward among the outer islands, the trees seemed to march seaward still, going steadily over the heights and down to the water’s edge. (14)

The image is conveyed in the first person, “We were standing…”, and the language dresses it in aesthetic felt by the narrator and viewer who becomes mirrored in the reader.
The “long stretches of shore” are covered by “the great army of the pointed firs,” the metaphorical usage of “great army” functioning to convey their multitude, showing the large breadth of the view of the coastal landscape. They are “darkly cloaked and standing” as sentinels waiting at attention, “waiting to embark,” conveying the position they are in, perhaps leaning forward in the wind but at such a great unison that, from afar, they could appear the image of a standing army lording over the landscape holding it in place. The military metaphor of the coast lined with pointed firs persists in the same way in this passage, that they “march seaward still” as one part an emblem but also a defining authority of the island.

In such acts of viewing in the novel we see other metaphors that dress the image of the regional landscape in emotional sentiments with philosophical connotations, this time a metaphor of grandiosity, enormity, and anthropomorphication, again with the fir tree as its aesthetic center piece.

Through this piece of rough pasture ran a huge shape of stone like the great backbone of an enormous creature. At the end, near the woods, we could climb on it and walk along to the highest point; there above the circle of pointed firs we could look down over all the island, and could see the ocean that circled this and a hundred other bits of island ground, the mainland shore and all the far horizons. It gave a sudden sense of space, for nothing stopped the eye or hedged one in, that sense of liberty in space and time which great prospects always give. (21)

This “great backbone of an enormous creature” is nothing more than a large boulder at the top of a mountainside onto which travelers could climb and observe the island landscape from afar, such is why the narrator likens the enormity of the sight to being
atop the back of an enormous creature that is larger that life, a pastoral figure that towers over the landscape and grasps it, in viewing, in its entirety all at once. They visually grasp the totality of the island from “above the circle of pointed firs” as if it were a sacrilegious throne or podium which they are mortally privileged to occupy at that moment. The image that follows seems to proceed around this point as if it were the center of the island and thus the world they inhabit: they could “see the ocean that circled this and a hundred other bits of island ground.” The grandiosity is conveyed in the effect that this viewing point is conceived by the narrator, subtly, as if it indeed were the center of the world—that coveted regionalist point of freeing isolation that is immune to any governing social body similar to what one feels in a bustling metropolis, the thematic heart of regionalism, as the view “gave a sudden sense of space. … that sense of liberty in space and time which great prospects always give.”

The Nature of Dialogue in First-Person Regionalist Narratives: the Word “We”

Keyword passage gathering via Python has thus far taught us much about the nature of regionalism and narrative content delivery in the case of The Country of the Pointed Firs and has figured well into the larger arguments that I have been trying to make about regionalism and its methods of content delivery. These sprung from an initial assumption I made about the use of my computational analytic tool: that a data-based study of the frequency of the use of titles in-text can reveal myriad substantive opportunities for close reading and broader thematic understandings. To my slight dismay, it has admittedly revealed The Country of the Pointed Firs to be an outlier in my discussion as compared with what was said about The Awakening and Ethan Frome: that
The Country of the Pointed Firs, in contrast, is a more imagistic novel than Ethan Frome and The Awakening, which contrastingly depend more on dialogue for the delivery or their climaxes and the progression of their plot. But this is merely the nebulous nature of a literary mode: each in their variety work to place the reader within the fabric of a cognitive and aesthetic experience, emulate the emotions of the characters interacting on the page, and elicit the reader’s empathy and consideration. Regionalism functions, then, primarily on imagery and dialogue, varying by author.

Despite this notion about dialogue and image noticeable at a preliminary glance, I am still not ready to concede my thesis entirely, in part due to a particular finding as a result of the word clouding function in Python. How could we, for example, suppose that image is the predominant aesthetic mode of The Country of the Pointed Firs when word clouding reveals the word “we” to be the most common word in the novel? We can observe this in the image pasted below:

(Figure 4.3)

To continue to observe the balance of weight between dialogue and image in the regionalist novel, I will next read the significance of the appearance of this evidently most common word in the text, the word “we.” There are a number of moments worth
observing to help understand its function, and thus the corresponding function of
dialogue with that of imagery in the novel.

The passage I’ve chosen to explore this notion is long yet telling, and contains
dialogue that comes after a previous passage I analyzed here, the one where the narrator,
balancing her need to complete her writing and her desire to immerse herself in the
simplicity of the island lifestyle, tells Mrs. Todd that she can no longer take part in the
labor of herb gathering. The exchange that takes place is telling of the ways characters
interact with each other and the ways explicit as well as unspoken dialogue reveal things
about the nature of regionalist “isolation” which, evident in this instance, can be both
emotionally liberating as well as painfully lonely, an enigma of the regionalist onus on
place and remoteness. In a technical sense, my attempt at passage gathering regarding the
occurrence of the word “we” has yielded passages containing not necessarily its usage in
dialogue but more so its usage after dialogue has happened and that is then being
contemplated by the narrator, conveying further thematic consideration. It reveals to us
that the narrator oftentimes considers her relationship with characters, particularly Mrs.
Todd, after having a conversation with them.

With the word “we” occurring most commonly after instances of dialogue and
during moments of the narrator’s consideration of the exchange, a useful example to
observe would be in the following, a significant moment where the narrator ruminates on
the balance between maintaining her industrial duty of writing and continuing her easy
and pleasurable assimilation, mentally and emotionally, into the mundane cultural
conventionalities of Dunnet Landing:
In taking an occasional wisdom-giving stroll in Mrs. Todd's company, and
in acting as business partner during her frequent absences, I found the July days
fly fast, and it was not until I felt myself confronted with too great pride and
pleasure in the display, one night, of two dollars and twenty-seven cents which I
had taken in during the day, that I remembered a long piece of writing, sadly
belated now, which I was bound to do. To have been patted kindly on the
shoulder and called “darlin',” to have been offered a surprise of early mushrooms
for supper, to have had all the glory of making two dollars and twenty-seven cents
in a single day, and then to renounce it all and withdraw from these pleasant
successes, needed much resolution. Literary employments are so vexed with
uncertainties at best, and it was not until the voice of conscience sounded louder
in my ears than the sea on the nearest pebble beach that I said unkind words of
withdrawal to Mrs. Todd. She only became more wistfully affectionate than ever
in her expressions, and looked as disappointed as I expected when I frankly told
her that I could no longer enjoy the pleasure of what we called “seein' folks.” I felt
that I was cruel to a whole neighborhood in curtailing her liberty in this most
important season for harvesting the different wild herbs that were so much
counted upon to ease their winter ails.

“Well, dear,” she said sorrowfully, “I've took great advantage o' your bein' here.
I ain't had such a season for years, but I have never had nobody I could so trust.
All you lack is a few qualities, but with time you'd gain judgment an' experience,
an' be very able in the business. I'd stand right here an' say it to anybody.”
Mrs. Todd and I were not separated or estranged by the change in our business relations; on the contrary, a deeper intimacy seemed to begin. I do not know what herb of the night it was that used sometimes to send out a penetrating odor late in the evening, after the dew had fallen, and the moon was high, and the cool air came up from the sea. Then Mrs. Todd would feel that she must talk to somebody, and I was only too glad to listen. We both fell under the spell, and she either stood outside the window, or made an errand to my sitting-room, and told, it might be very commonplace news of the day, or, as happened one misty summer night, all that lay deepest in her heart. It was in this way that I came to know that she had loved one who was far above her. (5)

While this quote is rather large, viewing this moment in its entirety is important to understand the semantic and interpretive significance of the word “we,” especially to be able to appreciate its proximity in this passage—at the end—and the role it serves in being there, at the end of this conversation, within a moment of the narrator’s contemplating the exchange between her and Mrs. Todd. To preface my point, I will note again Stephanie Foote’s words on the fictive, representative purposes of regionalism: that “regional writing gave strangers with accents literary recognition at exactly the same moment that accented strangers were clamoring for recognition … mediating the average urban citizen’s sense of self estrangement in a word filled with increasing numbers of strangers” (5). A more expanded way of saying this is that regionalism depicted interactions between urban people and metaphorical foreigners from rural regions to facilitate social understanding and empathy for cultural differences. This is exactly what is happening in this passage, a borrowing of understanding of cultural difference is
happening between the narrator and Mrs. Todd, in their surface-level tug-and-pull over how to spend one’s time: in the regional conventionality or the urban-industrial.

But the word “we” has special importance here because of its semantic place in the passage. The passage is a contemplation by the narrator of the tension between her individualistic responsibility and her newfound place in the collective society of Dunnet Landing. And only after contemplating and perhaps by result verbally registering this tension does she begin using the collective plural pronoun of “we” to speak about herself and those around her, connoting her sense of belonging and togetherness, and this pattern persists throughout the novel in conversational exchanges just as this one, where the narrator contemplates a tension of responsibility, ruminates upon its effect on her individual self, speaking with more “I’s” and “she’s” to connote her-versus-them, and then transitioning into what often appears a compromise or resolution which yields peace between herself and the person in question, be-it Mrs. Todd or someone else, where she joins them together in a plural reference. I will illustrate this in a close reading of the passage.

“In taking an occasional wisdom-giving stroll in Mrs. Todd's company,” the narrator says, “I felt myself confronted with too great pride and pleasure in the display, one night, of two dollars and twentyseven cents which I had taken in during the day, that I remembered a long piece of writing, sadly belated now, which I was bound to do.” Amid taking part in the habits of living there, the narrator has forgotten the conventionalities of her previous region of place, and the social expectations that were vested in her being there. Instead, she finds herself immersed behaviorally in this place. Yet she is self-aware of the assimilative act taking place, and goes on to verbalize the
tension in this contemplative moment. While decided in her dedication, she concedes that in the opportunity to “have been patted kindly on the shoulder and called darlin’, to have been offered a surprise of early mushrooms for supper” amid other pleasantries, to resist them still “needed much resolution.” Yet she does, nonetheless, say “unkind words of withdrawal to Mrs. Todd, once “the voice of conscience sounded louder in my ears than the sea on the nearest pebble beach.” The narrator chooses to maintain her native-regional conventionality.

Mrs. Todd, in her habitual grace, “only became more wistfully affectionate than ever in her expressions,” though no doubt disappointed. But it becomes interesting in that the narrator is apparently more disappointed, saying that she felt “cruel to a whole neighborhood … in this most important season for harvesting the different wild herbs that were so much counted upon to ease the winter ails.” She empathizes more with Dunnet Landing than herself and the plight of her individual responsibilities, she feels one with and a part of this new community more so than the one she came from. They compromised in this situation of tension between cultural mentalities and priorities. “Mrs. Todd and I were not separated or estranged by the change in our business relations” the narrator says, “on the contrary, a deeper intimacy began” suggesting, quite virtuously of the regionalist mode, that their social bond and friendship had deepened through an understanding of their differences. The pivotal usage of the collective “we” seems then to come as a result of their established understanding, and embodies the essence of their compromise: to resign to their respective duties but to also preserve time, when the time was right, to speak with one another: “We both fell under the spell,” the narrator says, of making this interaction a habit, of maintaining their respective individualist cultural
conventionalities—the narrator’s commitment to her writing, Mrs. Todd’s to the foraging and gathering necessary to their sustenance on the island—and making time for congenial interaction nonetheless.

The moment models the social-didacticness that Stephanie Foote and others read as a mark of the realist genre. The reading I have been able to provide through my attentiveness to the semantic economy of the word “we”—the most common word in the novel—has helped us understand the grammatical and semantic subtleties of the process when depicted in writing, that the narrator feels herself a part of the collective after contemplating and then compromising the social tension, having transitioned from using individual, first person pronouns to describe herself and her counterpart, noting the conditions in a unilateral sense—this happened to me and this happened to them—and then, after acquiring a compromise, referencing herself and her counterpart together (as “we”) in a collective sense, one that connotes the acceptance established by the two individuals, and metonymically, the two ways of living, the two regions.

So it begins to seem as though conversation is still important to the development of a regionalist novel and the development of regionalism’s fictive intentions. Even in a novel terse in its dialogue and more deeply vested in imagery and aesthetic, the act of conversation serves in advancing the novel’s thematic tensions. Moving toward “The Yellow Wallpaper” as an additional example I would argue that semantic conversation such as this moment can be read even outside of quotation marks, when such an exchange is taking place in moments of narrator or character contemplation. In moments such as these, the narrator or speaker in question is having a conversation with themselves, and can be read conversationally for the same semantic effect, appearing a type of refrain or
chorus that ruminates on the plot-action that happens throughout the story. The language of conversation inhabits contemplation when interpreted in this way, and the predominance of a plural pronoun over any other word in the text, I argue, is testament to the markedly conversational and contemplative nature of *The Country of the Pointed Firs*, and ought to influence how we think of a tale that on the surface seems so deeply dominated by imagery and scene setting.

**Self Referentiality in “The Yellow Wallpaper”**

In the attempt to make comprehensive observations about the nature of literary regionalism I thought it fitting to include a short story in my selection of texts as a way of grasping regionalism in its various permutations, that of the novel and the short story being its most prominent forms. Charlotte Perkins Gilman being an exemplar of regionalism, I also choose her text for the way it exemplifies my theory about the markedly conversational nature of regionalism. This may seem a paradox when applied to “The Yellow Wallpaper” but only if we forget the notion on which I ended my previous section, that reading text conversationally need not be bound to text between the confines of quotation marks, but that conversation can be read as an act of locution in literature as a way of contemplating and handling the development of the story’s central themes. In this case, as I set out to read “The Yellow Wallpaper” I read the stream of consciousness that occupies much if not all of the story as an act that splits the story and its embodied character along two lines of thought: conformity and non-conformity to the marital oppression that galvanizes the tale. As the narrator speaks in her journal through which we receive the story, she is contemplating the advice of her husband in tension
with her own internal impulses—what he believes is best for her versus what she believes is best for her. In these moments, we can observe the mental and emotional tug-and-pull between individual passion and social expectation that marks the regionalist genre. I refer to this phenomenon as self-referentiality, connoting the ways in which the narrator speaks to herself, about herself, to others, and about others, a study of which that can be revealing of much of the story’s thematic composition.

The lexical frequency distribution of the text lends itself to this analysis of self-referentiality and the overall referentiality that structures the story. Illustrated in the word cloud below, we can observe how the most prominent words are “Me,” “My,” and “John,” the narrator’s husband and the story’s antagonist.

(Figure 4.4)

While the words “is,” “have,” and “can” are equally prominent, I would like to focus on the self-referential pronouns and their implications with mentions of the character John. the word “me,” firstly, is an accusative first person objective usage of the word “I,” but as a self-referential, “I” is notably absent from this word cloud of a story so bent on the narrator’s ruminations of self. Why could this be? I argue that it is a mark of the character’s internal conflicts: as an objective usage, it signifies the ways in which the
narrator spends much time speaking about herself but outside of herself; perhaps, at
times, embodying the phenomena that she likewise views herself with scrutiny much as
the antagonist husband does, adapting his sense of viewing scrutiny so much so that she
applies it to herself. We can illustrate this loss of agency in sample passages to come.
Being attentive to the textual economy of the word “me,” we can observe how the
narrator speaks about herself as illustrative of the text’s concern with conflict between
individual passion and social law. The word “my” presents equally insightful
opportunities for reading, as an indicator of acts of possession—what does this narrator
feel she is in firm, secure possession of? How does it mark her character development, or
more latently, her degradation into madness? The prominent occurrence of “John” is
more self-evident than these conversational subtleties embodied in moments of
contemplation. As the antagonist, he is often the cause or source of the tension that
ruminates in the narrator’s contemplations, moments where the ruminates over the
conflict between her husband’s edicts and her own individual desires.

A conversational reading of pronouns in conjunction with the appearance of the
antagonist John in “The Yellow Wallpaper” can bring to the foreground much of the
gendered tension and dynamics of agency and power that mark the story. “John laughs at
me, of course,” the narrator says, describing her ailment and impulses, “but one expects
that in marriage” (2), alluding to the normalization of this conflicted power dynamic, of
the husband’s dominant agency over the narrator so much so that she not only doubts
herself, but feels her ostensible abuse is normal. Much of the story details such emotional
oppression disguised as normalcy, in another line describing her husband’s oppressive
care she says “John is very careful and loving, and hardly lets me stir without special
direction” (3). There is double meaning to the ways in which the narrator discusses her situation—manifestly, the narrator believes that the way she is treated is right and good, product of her social conditioning and her supposition of the husband’s role, a caretaker; but the reader is—or is encouraged to be—fully aware of this disparity in the narrator’s speech, for latently, we know these acts to be wrong, and for much of the story find ourselves in favor of the wife’s need for “congenial work, with excitement and change” (3) despite these things being against the testaments of her husband. The narrator however does not possess this clarity—the reader is more omniscient than she—and so we observe the tug-and-pull of her desires against her beliefs as they shift toward her eventual psychotic episode. As this change develops, so does the nature of her self-referentiality.

As the narrator struggles with this tension between individual desire and social expectation, she is wont to empathize with herself emotionally but not morally. The following moment illustrates this issue well:

My brother is also a physician, and also of high standing, and he says the same thing. So I take phosphates or phosphites, whichever it is, and tonics, and journeys, and air and exercise, and am absolutely forbidden to work until I am well again.

Personally, I disagree with their ideas.

Personally, I believe that congenial work, with excitement and change, would do me good.

But what is one to do?
I did write for a while in spite of them; but it does exhaust me a good deal having to be so sly about it, or else meet with heavy opposition (3).

The narrator’s brother, as a family member, supports the edicts of the husband because of his kinship but also because of his social role as a doctor, imbibing him with authority that ostensibly supersedes the validity of the narrator’s impulses for personal desire. The social mores overpower the individuality of the narrator’s regard for herself, so much so that she views her disobedience as morally wrong, saying that “it does exhaust me a good deal having to be so sly about it,” about her desire to write as an emotional outlet, construing the act of writing with an act of work which is forbidden to her because of the doctors’ opinions about what is best for her, rather than her own. “But what is one to do?” is in this context a redundant question because that decision is made for her: she is to listen because it is right to do so, or so she is conditioned to believe, much to the reader’s chagrin, who in their continued moral omniscience over the situation knows this not to be true.

Viewed in this way, the story appears a continuous stream of consciousness by the narrator, ruminating over what she conditionally believes is right versus what she feels is right for herself, and it is the constant weight of oppression placed upon her intrinsic desires that causes her psychotic break. Over the course of the story, we can see the narrator begin to hesitate to empathize with herself, in a moment such as this one:

I sometimes fancy that in my condition if I had less opposition and more society and stimulus—but John says the very worst thing I can do is to think about my condition, and I confess it always makes me feel bad.
So I will let it alone and talk about the house (3).

As the narrator speaks about her condition in this passage, the use of the dash represents a significant moment of emotional development. She is speaking about her desires: “I fancy that in my condition if I had less opposition and more society and stimulus”—but then she cuts herself off, recounting John’s desires in spite of her own: “but John says.” A figurative refrain, it signifies the nature of their collective power dynamic: in spite of what she wants for herself—what she says—John wants something different—John says. The line afterwards, “I confess it always makes me feel bad,” becomes ambiguous at this point, as to whether she feels bad to think about her condition, as John encourages her not to, or that she feels bad to have her desires for society and stimulus dismissed by her caretaker. It is ambiguous whether she feels bad to obey John, or whether she feels bad about the desire to disobey him.

The moral double entendre in the narrator’s speech is moreover most prevalent when John is the subject of discussion.

I get unreasonably angry with John sometimes. I’m sure I never used to be so sensitive. I think it is due to this nervous condition. But John says if I feel so I shall neglect proper self-control; so I take pains to control myself, before him, at least, and that makes me very tired (3).

The irony is the scene is that the reader knows full well by now that the narrator is being perfectly reasonable, and that she is only “so sensitive” because of his acts of emotional and physical oppression. She is conditioned to believe, on the other hand, that “it is due to this nervous condition,” while the reader knows that it is not. The arc of the story
however, as we know, is of the narrator’s breaking away from these oppressive acts, all
the while it leads to a psychical degradation rather than a semblance of personal victory.
A point of development toward this end occurs when she begins to maintain her personal
beliefs, yet views the act as perverse when she says “I take pains to control myself, before
him, at least, and that makes me very tired.” This moment signifies the fact that she has
lost faith in the truth of her husband’s edicts, and has devolved now to viewing it as an
act of mandatory obedience rather than care from a loved one. As the narrator views these
edicts as mandates however, she also strives to disobey, or desires to do so; but even this
impulse is perceived through a fractured morality: her psychosis is essentially capitulated
by her perceived desire to break from the social norm, or, in her view, to do the wrong
thing. The resulting conflation of her internal desires with this perceived immorality are
what pushes the narrator toward the culmination of her psychosis that we see at the end.

As a reflection of her emotional and psychological tribulations and her effort to
cope with her situation, the narrator’s language of self-reference changes markedly in the
second half or final portion of the story, demarcated in my view by the appearance of the
strange woman—or spectre of a woman—that the narrator begins to see or hallucinate in
the garden as she looks out the window and eventually behind the yellow wallpaper. As a
coping mechanism depicted in writing, the narrator ceases to speak about herself and
refer to herself—a hopeless cause, at this point—and instead ascribes her troubles to this
symbolic figure, endeavoring to “save her” as the narrator herself wishes she could be
saved. In these passages and through to the culmination of the story, self-referential
words such as “me” and “my” are replaced with third-person references to this figure
who is a metaphor for the narrator’s own self, using “she” and “her” to reference the symbolic person and embodied self.

Much of the conclusion of the story fixates on the narrator’s obsession with the patterns in the wallpaper as a type of depiction of a cabin fever of sorts, but this reading of self-referential pronouns also helps us understand the centrality of this apparition that the narrator perceives as being trapped behind it. Coupled with the narrator’s hallucinations are descriptions of action by this apparition that are now embodiments of what the narrator wishes she could do. Instead of stating her desires, what she would like to do, she perceives the apparition as doing it, since her own motivation has been repressed beyond repair. “The pattern does more—and no wonder! The woman shakes it!”—this is a moment wherein the narrator, when of sound mind, may have ruminated on grasping the bars on her windows and trying to break free, but she no longer has the mental space to consider her own desires and motivations, so she has transferred them to an outside figure that possibly can, and so the apparition shakes the pattern on the wallpaper. It is a metaphor for the claustrophobia of her social constrictions, marked by referentiality that pivots from the self to an other.

The creeping that the woman-apparition does, or that the narrator sees her doing, here becomes a metaphor for acting for one’s self or by one’s own motivations—doing as one pleases; or more closer to the substance of the metaphor, moving freely. Amid her fractured sense of morality, to move freely is cast in a negative light, hence referred to as creeping. In one moment the narrator says, “I see her on that long road under the trees, creeping along, and when a carriage comes she hides under the blackberry vines” (13). The narrator has been, since her internment at the estate, confined
to her room, discouraged from going outside and among society. The apparition creeps outside at night because “It must be very humiliating to be caught creeping by daylight” (13), and when a carriage appears, a figure of observation, she hides.

In the final lines of the story however, the narrator’s psychotic images essentially arrive at the point in which she started but through a different linguistic lens. The narrator, subtly and without strong signaling, equates herself with the apparition of the woman behind the wallpaper. “[T]here are so many of those creeping women” she says, “and they creep so fast. I wonder if they all come out of that wallpaper as I did?” (15). The final climax of the final lines moreover are the first instance in which the narrator refuses a demand by her husband, not opening the door and instead creeping around the room. At this act of refusal comes also the narrator’s conflation of the apparition with herself, her psychotic episode constituting an attempted act of refusal and reclaiming of her personal agency by imagining that she is the woman freed from behind the wallpaper: “‘I’ve got out at last’ said I, ‘in spite of your and Jane! And I’ve pulled off most of the paper, so you can’t put me back!’” (15). This is the latent content of the metaphor, that the narrator freed the woman from behind the wallpaper so that she can creep—move freely once again. Manifestly, we know that the narrator suffered so deeply from her mental and physical isolation and confinement—mental in the sense that every utterance of her own desire was repressed by her husband, physical in that she was confined to the bedroom—that she proceeds to pace mindlessly in circles around the room as the last fragment of personal volition still afforded to her.

Conclusion: Regionalism on the Level of Diction
Using computational close reading in the form of a word frequency analysis has taught us much about the nature of regionalism, and using computation to study the occurrence of key words has allowed us to answer significant questions about regionalism regarding its central formal properties of imagery and the depiction of conversation. In the case of Edith Wharton’s *Ethan Frome*, we used computation to observe the revision that took place across two editions, viewing the different ways the author handled depictions of character interactions on the level firstly of image, where we see Ethan’s description by the narrator change from appearing to have “strong shoulders” to the new rendering bereft of this description, enforcing his characterization as “the ruin of a man.” In character interactions we encountered similar subtleties relative to the story’s central thematic concerns, with varying depictions of Mattie’s interaction and reaction to certain characters to a different metaphor cast around her interaction with Ethan, grabbing her by the wrist-bone of the hand rather than just the hand. In this instance the observations teach us that regionalism is not purely conversational in nature, that while thought of as a form interested in depicting regional dialects, it is just as dependent on the depiction of imagery and action framing conversational exchanges in order to advance its fictive and thematic notions.

Considering the balance between image and conversation that structures the literary form of regionalism, we moved into discussion of *The Country of the Pointed Firs* toward various ends. For the purpose of general interest, I posited a preliminary hypothesis that I felt suited to computation, considering the interpretive insights it could garner, and found that they were many. We found that titles in and of themselves, at times, embody subtle yet significant thematic and aesthetic features that are influential to
broader understandings if they are willing to be seen and interpreted as such, using this assumption to notice how the object of the fir tree often appears in the story as a signal or a metaphorical tenor for important moments of thematic development in the story. A following measure of word prominence in the story, apart from this initial object of study, helped us arrive at the importance of the word “we” in the story as connoting in conversation a central aspect of regionalism: empathy with others from dissimilar social backgrounds. Reading the conversational nature of the regionalist tale, we observed moments in which the narrator of The Country of the Pointed Firs at first set herself apart from the natives of Dunnet Landing and then, through continued interaction embodying moments of social understanding, came to liken herself to them, feeling a belonging part of that community rather than an outsider, emblematic of regionalism’s social didacticness that sought to bring recognition to regional peoples.

At this point we arrived at a significant analytical question about regionalism. As we teetered between analyzing aspects of imagery and conversation in a type of supplementary confluence, I questioned which was the primary formal feature of the literary mode. Does imagery or conversation form the foundation of literary regionalism? In order to answer this question, I responded to relevant scholarship to restructure how we think about conversation in regionalism, viewing it not just as the graphical depiction of the phonetics of a regional dialect, but rather reading conversation for its diction and speech acts, not just within the confines of quotation marks but also in moments of character contemplation, leading us toward an analysis of pronouns and language of self-reference in “The Yellow Wallpaper.”
Regarding the “Yellow Wallpaper” I would argue that it has presented us with our most useful exemplar in answering this question about conversation and imagery. Being freed, so to speak, from looking at actual dialogue and instead reading the dialogue of character contemplation as an act of conversing with one’s self, as well as the story’s psychological framing, has allowed us to observe the real confluence of image and conversation in literary regionalism. Much of “The Yellow Wallpaper” consisted of contemplation, and the conversational act of self-reference and its change in nature throughout the story provided us with a forum by which of observe and understand the story’s thematic culmination and the development of the narrator’s psychosis. Yet much of her contemplations that we used to understand the depiction of her internal psychology were also hinged to her hallucinations—images—and their metaphorical effect, whereas the beginning of the story positioned these contemplations as responses to literal conversations that the narrator had with her oppressive husband. All the same, imagery and conversation seemed to sit in equal balance in their advancement of theme in this regional tale.

I still feel, nonetheless, equipped to argue an answer to this literary enigma, of whether imagery or conversation are the driving force in the thematic development and formal composition of regionalist literature. In answer I posit, as we have seen in these selections, that imagery often frames the broader and surrounding aesthetic in which the conversation takes place. Imagery, as an aesthetic frame, enforces the themes and interpretive substance that is to be drawn from moments of conversation. In Ethan Frome, for instance, we see Ethan’s fractured body before we hear him speak; we see how Mattie interacts with others compared with how she interacts with Ethan (she
retracted from Ned Hale’s approaching figure, and Ethan grasps her by the wrist-bone, leaving deep emotional sentiments unspoken because the image suffices. In *The Country of the Pointed Firs*, similarly, the fir tree is an emblem of the landscape, imbuing the narrator with feeling that influences her interaction with others—feelings of understanding and compromise, that pastoral beauty naturally influences those that live within it. “The Yellow Wallpaper” more subtly navigated this balance of influential factors. At the outset much of the story’s development was delegated by the narrator’s conversations with her husband set in tension with contemplations she had about herself, but her moment of psychosis and ostensible “breaking free from the wallpaper” was capitulated through acts of viewing after acts of conversation became superfluous in her eyes. Culminating not in a conversational moment, it culminated rather in the depiction of her hallucinations, of images of the creeping woman outside, the shaking of the wallpaper, and the narrator’s creeping circles around the room. Literary regionalism, then, renders images of regional place in an aesthetic that frames and emblematizes moments of interaction and conversation, which then together advance notions about the human condition on various social levels.
Conclusion: Computational Close Reading

“Close Reading: spec. in Literary Criticism, close criticism, close reading, etc., critical and detailed analysis of a text; an example of this. Also applied to the analysis of other works of art.”

- Oxford English Dictionary

Distant Reading: “aims to generate an abstract view by shifting from observing textual content to visualizing global features of one or several texts”

- Janicke “‘On Close and Distant Reading in Digital Humanities: A Survey of Future Challenges.”

Computational Close Reading: Using computational analytic tools and processes to study the textual, linguistic, and/or semantic economy of words in a body of text, artistic or otherwise.

- Damiano Consilvio

***

In literary scholarship augmented by digital humanities application, the digital and the human exist in a symbiotic relationship: the human is an interpreter, and the digital tool is an aggregator. In my experience studying interdisciplinarity between computer science and literary studies—the theoretical, methodological, and disciplinary proclivities of the digital humanities—I have found that neither one can do the work of the other. Computer technology is algorithmic in nature, based in a language of
recognizing numbers and symbols exclusively, and the linguistic elements of textual data are likewise recognized as figures, not interpreted, but merely aggregated based on human input—commands—that teach the machine to recognize and retrieve the language in question. The computer merely prides itself on the speed and precision of aggregation possible when language is reduced to computationally recognizable figures, but post-aggregation it is the human mind that is essential because while it cannot retrieve innumerable data instantaneously, it is sensitive to metaphor, figurative language, the things that make literature literary and which serve as the fodder for interpretation and critical analysis. So while the human cannot perform instantaneous bulk data aggregation as can the computer, the computer cannot interpret the abstractions of artwork as intimately as can the human. This essay will illustrate this notion as the central crux of the digital humanities and the conjoining thread of computer science and literary studies: as one aggregates the data better than the other, and one interprets the data better than the other.

My theory of the Digital Humanities as a disciplinary symbiosis of literary studies and computer science grounded in aggregation and interpretation falls in line with the established history of the Digital Humanities recounted by prominent scholars like N. Katherine Hayles and Jerome McGann. In terms of the aggregatory nature of digital technologies, Hayles reminds us of the “quantitative” nature of the “first wave of the digital humanities” which saw scholars, in the service of curation, “mobilizing the search and retrieval powers of the database, automating corpus linguistics, stacking hypercards into critical arrays” (26). Jerome McGann interprets this first wave as a late age of print imperative for “the reediting, for online environments, the entirety of our cultural
Inheritance” (157). In large part, early Digital Humanities was concerned with using digital environments for the preservation of the historical textual record. While a digital cloud is a secure place for its immateriality, its greater purpose was founded in the configuration of databases, which expedited the retrieval of textual material in a way never possible with hard-copy print, as evidenced in prominent database projects such as McGann’s own Rosetti Archive, and the Rutgers Whitman Archive, among others. Each of these projects and others like them are able to house, organize, and make available a greater breadth of material than could be done feasibly—in both a physical and economic sense—in a print edition.

But viewing the Digital Humanities from within the extent of its first stage is dangerous as it begins to ostensibly place competing value between the print and the digital, and hence the technological and the literary, giving way to an apprehension that critical interpretation will in a later stage be undervalued in favor of mere tool making, and that the institutional proclivities of literary studies—intimate sensibilities for the abstractions of language, critical thinking and literary analysis—will fall to the wayside as literary studies becomes less interpretive and more curatorial, and interpretation is replaced by invention. That is not an argument I would like to put forth in this essay, but rather one that my original theory of symbiosis will push back upon. A richer Digital Humanities rather would account for how each field—computer science and literary studies—use their native institutional methods to complement the analytical intentions of the other, or how methods of data management can be used as tools for performing interpretation.
Hayles, McGann, and others have likewise been attentive to this institutional disparity and take measures to prevent against it. Hayles firstly acknowledges the important turn that DH must take in order for the two fields not to fall into a disciplinary quagmire. “The second wave is qualitative,” Hayles writes “interpretive, experiential, emotive, generative in character. IT harnesses digital toolkits in the service of the Humanities’ core methodological strengths: attention to complexity, medium specificity, historical context, analytical depth, critique and interpretation” (26). McGann finds this pivotal turn in the “design and construction of statistical models for studying language formalities of many kinds, ranging from social and historical linguistics to the study of literary forms” (3). The Digital Humanities was purely a quantitative aggregator in its first phase—the database and digital edition phase—and its tools were curatorial in nature: large scale databases with sophisticated spot-light search functionalities that could bring before a reader vast materials like the entirety of the *Leaves of Grass* variations or otherwise unpublished texts contained only in manuscript, all without the physical and material work involved in retrieving these texts in hardcopy form. Now while this first phase does mark a point in which the computer performs a task in a way that the human cannot—as a ‘mass aggregator’—the texts themselves were still left in the database in their immediacy to be interpreted by a human reader and not a machine. It therefore ostensibly left unanswered the ways in which Digital Humanities application could serve literary interpretation, or how computer science could truly inform literary practice. At this point all the digital project accomplished was to house the texts in a database, to construct a tool.
Beyond acknowledging the tool-making caveat, DH scholars have suggested ways to embrace integration of the fields while also preserving their individuality. Taylor Arnold in *Debates in the Digital Humanities* responds to this specifically through the context of data analytics, exploring “how statistics—the organization, analysis, interpretation, and presentation of data—is a fundamental interlocutor and methodological approach for the digital humanities” (293). She strikes a chord in the methodological correlations she draws between statistics and literary studies, envisioning within the literary scholar a sense for literary data similar to that of the statistician that deals with primarily numerical figures. “DHers can concentrate” she writes, “on developing new modifications for extracting [my emphasis] humanities knowledge” (298). Extraction is a key word here because it brushes closely with my theory on aggregation—that DH application can find its strength in its later stage in much the same place it found it in the beginning: in the aggregation of textual material. It would seem, then, that the disciplinary tension between the two is vested more-so in what a responsible scholar does with the data: whether they house it in a database (first phase) or actively interpret it (second phase) after the computer aggregates it.

If we are to follow Taylor Arnold’s direction for symbiotic integration of disciplinary methods, we must understand how the use of method must go beyond discipline and apply to the individual scholar, as well. “To fully utilize statistical analysis” Arnold continues, “one needs to be proficient in programming with a language such as R or Python” (298). Whereas institutional method is innate in a field, skill must be adapted to secure feasible integration. The literary scholar attempting to perform Digital Humanities studies must be attuned to computer methods if they are to be
harnessed properly. In this instance literary studies preserves itself and takes the lead over computer science, as it would seem that only a literary scholar can use computerization to perform interpretation, but a computer scientist would be far harder tasked to perform a literary study on their own. It does mark a significant learning curve for the literary-scholar-turned-DH practitioner, but it also marks the stewardship, so to speak, of the literary scholar in the digital humanities. Viewed in this way, we can envision a Digital Humanities in which DH makes tools for the literary scholar, and not tools to replace them.

To return to my original notion, that in the Digital Humanities the computer and the human exist in a symbiotic relationship, and that the methodological strengths of each can complement each other in the pursuit of their individual disciplinary goals, is to begin to understand what a nuanced, third phase Digital Humanities should look like: it should be a field in which the scholar adopts the methods of an outside discipline as a tool to pursue the goals of their own where particular skills are lacking in contrast to the other. In the Digital Humanities this axiom rests on the binary of aggregation and interpretation: only the computer can aggregate and organize large breadths of material instantaneously, and only a human mind can feasibly and logically interpret language, metaphor, figurative language and the like. If the aggregated material finds its final resting place in a database or digital edition, then the ostensible DH project relegates itself to that of mere tool making. If, on the other hand, the textual data is retrieved computationally in order to serve as the basis or material for interpretation, and is therefore aggregated and then interpreted, only then can it constitute a truly interdisciplinary project.
To follow this theoretical synthesis, I will illustrate these notions about equivocal interdisciplinary application through reference to on-going research that is meant to inform my dissertation. Earlier in my career I composed and independently published a digital scholarly edition of Edith Wharton’s *Ethan Frome*. It is currently housed online via a URL and is in essence a multimodal hypertext, but nothing more. I prided it for its usage of media annotations, or explanatory notes similar to print critical editions but that include other forms of media in their explanation, at times featuring an image of a hollow-backed bay or a livery stable, and at other times including an audio clip of a song characters’ were mentioned to be dancing to in a particular scene. I also accounted for variants across editions by including variant lines in hyperlinks on the page text. In this project I felt I identified a strong locus of the interdisciplinary connections between computer science and literary studies, as I had used an online publishing and composing platform—Scalar—to create a digital edition capable of doing things—of aggregating data—in ways a print edition could not. But, identifying this project within the phases of the Digital Humanities, I quickly found that I had not, but that I had rather fallen into the toolmaking dilemma of DH that was the cause of interdisciplinary tension rather than rich collaboration.

The shortfall of this project was that I had composed a tool for vivid reading, but I had left critical interpretation to the wayside. The issue was that the project was concerned with textuality and not textual interpretation. In harnessing the capabilities of computerization, I had merely done something that a print book could not, and by its ostensible conclusion I found that I had not yet engaged with literary practice at all—I did not interpret the text, I only aggregated it with the help of computational tools. Still
the project provided me with a foundational notion that would inform the rest of my Digital Humanities knowledge: I explored the ways in which literary scholars could use computers to do things they could not otherwise accomplish with print materials and in print practices. In this initial exploration of interdisciplinarity it was my premise which was misguided: I began with questioning how digital composition could achieve things that print could not. Still, it brought me to the cusp of nuanced understandings of the purposes of Digital Humanities: it taught me that the computer was no more than an aggregator, and that the human was still essential to execute interpretation.

As I continued, my premise for exploring the Digital Humanities changed. From “how can the computer achieve things not possible in print,” I began to arrive at “how can the computer assist in literary scholarly goals.” The onus was upon preserving the innate disciplinary proclivities of literary studies while embracing the comparable capabilities of digital computer methods. I became more aware of DH practitioners’ caveat that “the creation of tools drawn from outside the humanities do not simply supersede the theoretical principles that … should inform DH practice” (David Berry 65), and I found that the initial phase-one turn in the digital humanities away from interpretation and into database was succeeded by a turn directly back in the same direction, back to literary studies as toolmakers like myself came to realize what we left behind. So with the crux of literary studies being the scholar’s sensitive interpretation of language, I shifted my focus from outdoing print processes to exploring how computer aggregation methods could be used to expedite the interpretive process.

My ideas about a feasible interdisciplinary methodology reached a fuller maturity as I studied Data Science and Data Analytics—subdisciplines of computer science but
that work similarly as does the Digital Humanities in the context of literary studies: it is methodological in focus, and concerned with assisting in the goals of an individual discipline through use of computer technologies. As a field, Data Science defines itself as “a methodology by which actionable insights can be inferred from data” or “the production of beliefs informed by data and to be used as the basis of decision making” (Laura Igual 2). Apart from the Digital Humanities as it is conceived by scholars like N. Katherine Hayles, Jerome McGann, and others, Data Science seemed to be the voice I needed for nuance because of the ways it is strictly concerned with application methods—positioning itself as a tool rather than a maker of tools. Studies in Data Science, notably Brian Kokensparger’s Guide to Programming for the Digital Humanities and Laura Igual’s Introduction to Data Science: A Python Approach to Concepts, often center around the usage of data analytic programming code in the pursuit of various interdisciplinary projects. As such, they put forth a theory of textual data that has been useful to me both practically as well as conceptually when thinking about the purpose of Digital Humanities application. Whereas previously, in my phase one digital edition, I had set out to surpass the capabilities of the print edition and suggest new editorial methods for hypertext scholarly editions, data science urged me to use the tools they provided—programming code for data analytics—to perform my original literary tasks.

Data Science then provided me with an invaluable tool, Python Programming Code, and it became my research endeavor to experiment with using this tool to perform literary tasks, and in essence, to perform literary criticism. Only when literary critique was the end goal and not in the peripheral of my project would it constitute a feasible
interdisciplinary connection, as one that embraces and accommodates both disciplines without superseding the other: a reciprocal connection. Python as a tool, most of all, represents that aggregation capability that cannot be matched by the human cognition, literary scholar or otherwise. All the same, I have found that in the realm of interpretation itself the human cognition is still essential, and that Python merely served, as computer technology does, as an aggregator. But still it allowed me to move away from the toolmaking dilemma and into a richer form of interdisciplinary collaboration: I began to discover ways that Python could extract text from a document and aggregate the material necessary for interpretation.

I saw this new premise enacted in a collation or text-comparison project, which supported well original ideas in regard to the human eye’s ability to collate two texts for variants versus the capabilities of Python’s string comparisons. It is evident that Python’s for loops for string comparison did in an instant a task which by eye would demand six to twelve hours. As a result, through Python, I was able to develop a single-code collation script applicable to any portable document format file:

```python
import textract
text = textract.process('man48.pdf', method = 'tesseract')
book1 = (text)
from nltk.tokenize import sent_tokenize, word_tokenize
text = textract.process('man52.pdf', method = 'tesseract')
book2 = (text)
from nltk.tokenize import sent_tokenize, word_tokenize
import difflib
import nltk
import re
text = textract.process('man48.pdf', method = 'tesseract')
book1 = text.decode()
text = textract.process('man52.pdf', method = 'tesseract')
book2 = text.decode()
from nltk.tokenize import sent_tokenize, word_tokenize
from IPython.core.display import HTML
words1 = nltk.word_tokenize(book1)
words2 = nltk.word_tokenize(book2)
htmldiff = difflib.HtmlDiff()
tbl = htmldiff.make_table(words1, words2, context = True)
```
The script performs a collation on two editions of the same text: man48.pdf and man42.pdf, which are files of first edition facsimiles of the Battle Royale chapter from Ralph Ellison’s *Invisible Man*, documents I was parsing for differences that I thought could mark thematic development from the initial magazine version to its eventual adaptation into the second chapter of the novel. This code is important because of what it automates in service to the human interpreter, but also for what the human operator must teach it first. Its composition marks the ways in which the computer cannot ever replace the human, but must rather be directed to do what it wants.

An understanding of the script’s operating procedure will elucidate how the human and the computer work complementarily in this type of digitally augmented literary project. The textract module, for example, teaches Python that the documents in the host directory, although image-based PDFs, are documents of linguistic data, therefore noting for Python that it is dealing with an image that contains language. The textract.process command then reads the entirety of these texts into Python so that their individual figures can be recognized and retrieved. The nltk.tokenize module on the following lines continues to process the language for Python. With textract initially, Python is recognizing the document as a large breadth of letter-figures, but not yet words. Tokenizing the text after textract teaches Python to assemble the letters into coherent words and thus sentences and paragraphs. At this point, nearly half of the code is based in teaching the computer to do something that a human can already do, realize that this picture is written language; hence their symbiotic relationship: Python could not perform this on its own without a scholar composing the proper prerequisites.
The next module in the code, difflib, represents the passing of proficiency from the literary scholar and back to the computer, initiating the process which the human could not do and for which the computer is ultimately necessary. The module difflib is essentially the collator, and combining it with html commands allows the program to complete a machine collation. The line `tbl = htmldiff.make_table(words1, words2, context = True)` is a key confluence of efforts between the human and the computer. It instructs Python to do several things: first is to identify differences in text, performed by difflib. The `tbl =` initiates a compiler with the rest of the line, combining two commands, the variant parser and the table builder, resulting in htmldiff.make_table. The parentheses after this line denotes the two texts meant to be parsed and collated, words1 and words2, which are variables containing the two pdf documents. The final portion, the context note (context=true), instructs Python to retrieve instances of variation between the texts and mark them with an ‘n’ symbol (in laymen’s terms, to detect ‘true’ ‘differences’). In this case true is relative to ‘different’ when thinking about collation: true data is, to Python, data that is different across the two texts. HTML(tbl) executes the cell, which after about two minutes of load time, outputs a tabulation of varying figures across the two texts, surrounded by five additional words on each side so that they can be easily identified in the primary documents for verification if needed. The final product looks like this, a scrollable table of every word in the text aligned side by side:
Looking at this table from the right to the left (the 48 version on the left and the 52 on the right) allows me to identify text present in one document but not in the other. In this table, I was able to locate a variant line absent in the magazine edition but present in the 52: “it took me a long time and much painful boomeranging of my expectations to achieve a realization everyone else appears to have been born with: that I am nobody but myself. But first I had to discover that I am an invisible man!” (’52 pp. 13; ’48 pp. 15)

But, as the persistent thesis of this essay, we should remember how the computer and the scholar worked together in this collation endeavor. The program merely aggregated the variants for me, and I interpreted them. Neither could have done the work of the other as efficiently, and I could not have as quickly used these variants to make inferences about Ellison’s compositional method or the thematic development of his novel chapter had the program not first collected them for me.

Using Python as a command-based aggregator, we can envision how this same script process could apply to other forms of literary interpretation. What this code does is
process the lines of text into what Python reads as ‘strings’—linear data, identified by the context specificities, in this case the ‘true’ command. As the code stands now, it can merely be revised for its context to retrieve different things from the data in question. For example, as literary scholars are apt to do, they would in preparation to compose a paper retrieve passages and quotes from a primary text to use as analytical examples and readings. These quotes and passages used in discussion are usually hinged or act in reference to the subject of the scholar’s analytical essay: for example, representations of space and place. Here again it becomes a difference between doing things by hand or having them computationally automated. In a paper attempting to explore topics like space, place, and distance, the context of the code can be revised to identify particular, relevant key words, and tabulate them with the surrounding ten to fifty figures or words to show the quotes themselves; so, it would key words relative to space, place, and distance, with the surrounding words of the paragraph, aggregating for the scholar a table of relevant quotes and passages for their discussion. The line, revised, may look like this: `tbl = htmldiff.make_table(words1, words2, context = 'space', 'place', 'distance'). But it can be made more particular, more sensitive, based on the scholar’s own cognizance of the author’s writing style and awareness of the ways that these themes might be mentioned, so they could include other key words: `tbl = htmldiff.make_table(words1, words2, context = 'space', 'place', 'distance', 'house', 'work', 'building', 'furniture', 'window', 'drawingroom'). Working in this way, being guided by the literary scholar’s necessary prior knowledge, but who makes use of the program’s speed and precision, Python would output tables of passages for the scholar to use in their discussion without having to read and highlight manually.
It is necessary to reflect upon this project in the context of the phases of the Digital Humanities recounted by other scholars. In its first phase, this project created a tool, a machine collator. Useful in its own rite, it served a particular purpose of performing a machine collation in record time, alleviating the need to collate by hand and allowing for the expedition of such text-comparison-based projects. But it is still just a tool in and of itself, and only when the aggregated material is interpreted (or, to compare jargon, processed) by a human scholar does it become a piece of literary analysis. The computer, in its numerical, mathematical language, cannot interpret, make inference, synthesize abstractions in the data, only the human mind can do that, so they meet each other in the middle. This is where the DH project moves into the second, higher phase, when the data that is aggregated is used to make qualitative inferences rather than just quantitative. The most the program is capable of doing is aggregating and numbering—retrieving the desired data, organizing it into a table, and making it available to be read, essentially a quantitative output. The quantitative work of the machine becomes qualitative, interpretive, critical and analytical, through the help of the human scholar who consults it and draws conclusions from it. Here the project enters the second phase capable of synergistically combining disciplinary methodologies in pursuit of the scholar’s desired goal, a symbiotic relationship between the human and the program.

We have been able to see this new analytical framework for which I advocate in the body of this study, in the examples I provided regarding collation but also in my more central endeavor to test the interpretive closeness possible through the use of computer tools, a thesis which, in its composition, has also allowed us to implicitly study textual methods of narrative immersion in regionalism. As I detailed in my last chapter, the
computational methods I devised helped us greatly to deduce significant notions about the nature of the regionalist narrative, of its use of conversation and imagery to levy its thematic and fictive effects.

But rather than tailoring the literature for an ostensibly computational study—viewing it for large scale, global features, fitting the text into computational and data frameworks—I instead tailored the computer method to the literary discipline. Recognizing the analytical proclivities of the computer in its ability to recognize and aggregate choice textual details, I directed this methodology to measure the variable usages of words in their prominence and occurrence, and then analyzed the output to deduce their significance. I used the computation to measure depth rather than breadth, in an effort to prove that it was possible. The computer program was key in helping me compose the close readings that followed throughout the chapters. It automated the retrieval of my analysis passages at a rate and precision that both galvanized the project and helped set it in motion. Instantaneous aggregation was its primary strength in this instance, as it was clearly suited to aiding in the close and critical analysis that I was hoping to achieve.

Automation certainly helps to the extents possible, but only to the extents possible, and there will always be the other side of the dividing line which relegates tasks which the computer is capable of and those which can only yield to humans. This binary is marked by the differences I introduced at the outset of this essay: the aggregator and the interpreter, the program and the scholar. Each equally possess particular practical abilities which the other does not, and DH projects that combine these abilities in a confluence will position themselves better in the realm of a true interdisciplinary project,
as one that adapts interdisciplinary methodologies in the pursuit of individual disciplinary goals. It is by this way of thinking that the Digital Humanities application to literary studies can ground itself. They should still be literary projects based in the proclivities of the field, projects that pursue readings of literary text to deduce historical, social, or philosophical significance, and not those that merely seek to construct and produce tools in the service of literary studies. A collation tool in and of itself is, I would argue now, only just a computer science project because its application is variable. A more overarching project that uses the digital tool to perform and then deliver literary critique, with its central onus being on the criticism itself and not the innovative quality of the tool, is more centralized in the literary discipline and therefore a literary project consisting of digital application, and not the other way around.

There is however, and of course, need for improvement of existing technology for the ways in which it inhibits the human from full usage and application. To be anecdotal in this regard, I will attest that as a literary scholar attempting to use data analytics and programming code to aggregate and study literature, the learning curve was immense, and I have yet to reach a point where I can freely and fluently compose code without being in close contact with Python specialists who are themselves computer scientists—at a loss to understand the literary aspects of my work, but aware enough of the Python processes to help me with what I am trying to do. But even still I am tasked with phrasing my question in the correct jargon. I could not ask, per se, “how to collate texts in Python,” but rather, to adapt to the discipline, I must ask an attuned Python user “how to process documents into strings and output a comparison,” so there is still a disciplinary barrier inherent in a data science tool meant ostensibly to be interdisciplinary. Such
technologies hailing from a field that prides itself for its interdisciplinarity (be-it Data Science or the Digital Humanities) should adopt a cognizance of user-friendliness in the creation and dissemination of their tools to avoid the shock-value of the learning curve involved in using them. So it would seem again that discussions of DH return back to literary studies: the issue is with the language each discipline is used to speaking.

Regardless, it is unlikely that, within the Digital Humanities or without, the computer will ever fully replace the human. It’s proficiency only goes so far, and it endures to argue that although the program can do things more quickly than the human can, the human can still and always will do things cognitively with data that the computer will likely never be able to do, simply because of the binary-numerical operating nature of the computer as a machine: it does not understand or recognize literature, only figures, and the scholar uses the code needed to make it recognizable. My discussion of the Python string-comparator illustrated this concept well, as it points to the ways in which the program can really do nothing on its own, but that it must be directed by a human mind. This is how Digital Humanities application to literary studies is meant to work: literary scholars merely use digital technologies to expeditiously perform the same literary criticism they would have otherwise.

—Damiano
APPENDICES

Appendix 1: Python Code

The following are screen captures of the coding scripts used in my computational analysis.

```python
In [1]: import PyPDF2
   ...: import re
   ...: # filename = input('path_to_pdf')
   ...: filename = 'ethanfrome.pdf'
   ...: object = PyPDF2.PdfFileReader(filename)
   ...: numPages = object.getNumPages() # print(numPages)
   ...: 79

In [7]: import tesseract
   ...: import nltk
   ...: import re
   ...: text = tesseract.process('ethanfrome.pdf', method = 'tesseract')
   ...: book1 = text.decode()
   ...: from wordcloud import WordCloud
   ...: long_string = book1
   ...: wordcloud = WordCloud(background_color='white', max_words=5000, contour_width=1, contour_color='steelblue',
   ...: stopwords = ['Gutenberg', 'gutenberg', 'project', 'and', 'that',
   ...: 'the', 'a', 'of', 'it', 'to', 'his', 'it', 'in', 'with', 'for', 'her', 'was',
   ...: 'he', 'you', 'had', 'on', 'she', 'as', 'from', 'ebook', 'him', 'when', 'which',
   ...: 'there', 'at', 'down', 'you', 'there', 'on', 'up', 'they', 'out', 'if', 'not',
   ...: 'about', 'or', 'https', 'what', 'then', 'if', 'word', 'if', 'Mattie', 'Ethan',
   ...: 'Zeena', 'Edith', 'Mr titan', 'which', 'this', 'which', 'out', 'www', 'Ethan',
   ...: 'Frome', 'org', 'files', 'by', 'all', 'be', 'back', 'into', 'to', 'one',
   ...: 'him')
   ...: from nltk.corpus import stopwords
   ...: filtered_words = [stopwords.words('english')]
   ...: wordcloud.generate(long_string)
   ...: wordcloud.to_image()
```
the arid acres of his farm yielded scarcely enough to keep his household through the winter; but I had not supposed him to be in such want as Harmon’s words implied, and I expressed my wonder. Well, matters ain’t gone any too well with him, D Harmon said. When a man’s been setting round like a hulk for twenty years or more, seeing things that want doing, it eats into him, and he loses his grit. That farm from was always ‘bout as bare’s a milkpan.

Page no: 7

know more. Once I happened to speak of an engineering job I had been on the previous year in Florida, and of the contrast between the winter landscape about us and that in which I had found myself the year before, and to my surprise Frome said suddenly: Yes! I was down there once, and for a good while afterward I could call up the sight of it in winter. But now it’s all snowed under. He said no more, and I had to guess the rest from the induction.
BIBLIOGRAPHY


19-36.


Kuhn, Jonas. “Computational Text Analysis within the Humanities: How to Combine Working Practices from the Contributing Fields?” Language Resources and


North, Joseph. “What’s ‘New Critical’ about ‘Close Reading’?: I. A. Richards and His


Totten, Gary. “Spaces of Consumption in American Realism,” *The Oxford Handbook of American Realism*, (Keith Newlin ed. and introd.), Oxford University


