1-12-2015

Feminist Markup and Meaningful Text Analysis in Digital Literary Archives

Hannah M. Schilperoort
San Jose State University, hannahschilperoort@gmail.com

Hannah Schilperoort
Candidate for Master in Library and Information Science, San Jose State University, Dec. 2015
Master in English Literature, San Francisco State University, May 2010
hannahschilperoort@gmail.com
831-359-7069
January 12, 2015

Feminist Markup and Meaningful Text Analysis in Digital Literary Archives

Abstract

In this research paper, I examine three digital archives of women writers--University of Nebraska-Lincoln's Willa Cather Archive, Northeastern University's Women Writers Online, and University of Alberta's Orlando Project--for evidence of encoding practices and computational text analysis experimentation that supports feminist scholarship. I provide a brief overview of text encoding practices and controversies in digital literary studies, emphasizing research that suggests heavily detailed and interpretative markup results in more meaningful text analysis outcomes. I situate feminist text encoding and analysis practices and technologies within a larger argument for the use of detailed, interpretative and critical markup. I begin my research on the premise of Jacqueline Wernimont's assertion that text encoding and analysis practices and technologies are political tools that open a space for feminist intervention that can support feminist literary scholarship and reveal the integral place of women's writing within the field of digital literary studies.

After examining the three digital archives as well as any documentation of their markup and text analysis practices, I determine that feminist markup--tagsets specifically designed to support feminist inquiry--exists on a spectrum and that highly detailed and interpretative feminist markup leads to more meaningful text analysis outcomes for feminist scholarship, revealing complex social, cultural and political data pertaining to gender and literary history. I conclude that feminist-specific markup is important and necessary to support feminist scholarship, and

that detailed and interpretative markup can be leveraged to produce more meaningful and critical text analysis results in other areas of digital literary inquiry.

Introduction

In this research paper, I look closely at three digital archives of women writers for evidence of feminist encoding practices and text analysis experimentation that supports feminist scholarship. I chose to examine the University of Nebraska-Lincoln's Willa Cather Archive, Northeastern University's Women Writers Online, and University of Alberta's Orlando Project because all three archives utilize the Text Encoding Initiative (TEI) guidelines and are involved with computer based text analysis projects. The notion and practice of feminist markup has evolved out of the digitization and encoding of women's writing as well as feminist literary criticism. As I examine digital archives of women's writing, I focus on reaching an understanding of encoding practices and attempt to generate a working definition of feminist markup by looking at documentation of markup practices of women's writing. Secondly, I am looking for connections between feminist markup practices and text analysis, especially for evidence that supports or does not support a direct cause and effect relationship between interpretative and critical feminist markup and more meaningful text analysis outcomes.

Statement of Problem

Despite the undisputable interpretative nature of text encoding, traditional digital literary scholars have prioritized structural over overtly interpretive and critical markup in an attempt to produce the most objective and reliable scholarly editions as possible. However, as we will see in the following literature review, critics of this perspective reveal not only the fallacy of objectivity but also the benefits of embracing interpretative and critical markup for scholarly research and text analysis. In particular, in the following literature review, I will explore feminist theories in relation to the digitization and markup of women's writing, along with the possibility of more robust and meaningful text analysis.

Literature Review

Some TEI Basics

Digital literary studies is a subset of digital humanities concerned with digitization of literary texts, preservation and representation of digital texts, computational text analysis, and new ways of data visualization (Siemens and Susan Schreibman xix). Text encoding is a central concern of digital literary studies, contributing to the preservation of digital texts, scholarly editing, and preparation for digital display and computational text analysis. First developed in 1987, the Text Encoding Initiative (TEI), which refers both to the set of guidelines used for textual markup as well as the international consortium that maintains the guidelines, is currently the recommended standard of text encoding for digital scholarly texts in the humanities ("TEI: History").

TEI markup includes descriptive metadata about the text in the TEI Header, and structural metadata that identifies and separates the textual elements of the content (Van den Branden, Terras and Vanhoutte). Publication 5 (P5), the current phase of TEI, is expressed in XML (eXtensible Markup Language), which dictates the syntax, the structural and hierarchical layout of textual markup. The TEI guidelines are designed to be open and customizable, and the semantics of the markup is determined by the Document Type Definition (DTD), the agreed upon set of element tags and corresponding attributes, of the particular encoding project ("A Gentle Introduction to XML"; Van den Branden, Terras and Vanhoutte).

One of the primary purposes of TEI is to produce machine-readable texts, enabling a computer to perform functions, such as search retrieval, display or analysis, based on the elements of the text that are marked. Buzzetti and McGann explain: "It is through markup that textual structures show up explicitly and become processable" (64). Structural elements, such as chapters or paragraphs, for example, are marked, enveloped within opening and closing tags denoting the structure, so that they are recognized as such by computer software (64).

There is no doubt that the act of encoding is an interpretative act (Cummings 458-59; Hockey 48). Paul Eggert asserts that “texts do not have an unproblematic objective existence; they are not self-identical” (429). Thus, texts, even before digital markup is added, are open to subjective interpretation and multiple readings, meaning that there is no single objective text that would lead to a single correct marked-up version of the text. From this perspective, adding markup adds another interpretative level to the text rather than simply describing objective, inherent or static content elements. Furthermore, although digitization projects use an agreed upon DTD to identify and mark textual elements in a common way, inevitably, encoders are forced to make difficult decisions based on their own interpretations of text. For example, James Cummings explains: “It is up to the researcher applying the markup to decide what portion of the text to wrap a <title> element around” (458). Often this could be a fairly straightforward decision, but some situations might prove more problematic. For instance, where do the main title end and the subtitle begin (458)?

Despite the obvious interpretative nature of text and text encoding, traditionally, scholars working on encoding projects have attempted to maintain objectivity as much as possible by adhering to the “Ordered Hierarchy of Content Objects” (OHCO) textual model. Renear explains that the OHCO model “postulates that text consists of objects of a certain sort, structured in a certain way.” Structural elements relating to the “intellectual content” of the text, such as chapters, paragraphs, titles, stanzas, lines and so on, are marked in a hierarchical and linear fashion (Renear). XML stems from this OHCO model of a text, assuming “that a document is a single hierarchical structure and that each element nests neatly within another element” (Hockey 45). For example, paragraphs occur within chapters and lines occur within stanzas.

Although the simple hierarchical nature of XML often helps to simplify encoding decisions, texts do not always adhere to a strict hierarchical structure, resulting in scholars

encoding non-hierarchical elements in a hierarchical way or producing ill-formed XML texts that do not properly validate. Cummings points out that XML is limited "with regard to the encoding of multiple overlapping hierarchies" (460). For example, "when one structure runs concurrently with another and the encoder wishes to record both of these structures simultaneously" (460). Cummings offers a simple example of the problem of marking up "paragraphs [that] may split over pages" (460). In many cases, scholars choose to favor the element which is more important to the intellectual structure of the text rather than the physical structure. Thus, in the example given above, the paragraph would most likely be privileged over the page (461).

A Case for Interpretative and Critical Markup

As previously mentioned, XML dictates the hierarchical structure and layout of the text, but the particular DTD of the digitization project dictates the semantics, the tag elements and their attributes. In "A Case for Heavy Editing: The Example of *Race and Children's Literature in the Gilded Age*," Amanda Gailey explains that "XML provides the general rules for structuring tags," ensuring that opening tags are properly closed and tags are properly nested within other tags, but the tags used, their meaning, and their attributes are determined by the TEI DTD (130). Gailey continues to explain that as long as the text follows the hierarchical tagging rules, we can put any type of vocabulary inside the tags "as long as the terms are in brackets and we close and nest them properly" (131).

Though the openness and flexibility of TEI allows for a wide variety of markup options, scholars tend to make markup choices that denote structural textual elements rather than thematic, symbolic, or critical possibilities. As Gailey clarifies, traditionally, TEI is "primarily focused on noting the structural or formal features of a text," which, although still interpretative, is thought to be less interpretative and controversial than "overtly interpretive or critical claims" (131). For example, labeling a textual element as a "poetic line" would less controversial than labeling an element "homoerotic" (131).

Gailey posits that “there is nothing about XML that precludes using it to make interpretative claims about text” (131). However, most scholarly corpora projects choose not to employ interpretative markup. One reason for this is because editors of scholarly editions are mainly concerned with creating a reliable, somewhat “objective” text rather than offering criticism (132). Another reason has to do with producing well-formed conformant XML text. Interpretative elements, such as metaphors, for example, will often “compete hierarchically” with physical elements of the text, resulting in “technical errors” and texts that will not validate according to XML standards (132). XML’s hierarchical structure also makes it difficult, if not impossible, to “accommodate several different interpretations of the text coexisting in the same file” (132). In addition, “deep, critical markup is time-consuming” and requires literary scholars familiar with the text to do the encoding (132).

Despite the apparent difficulties, Gailey believes that employing interpretative and critical markup will greatly enhance scholarly interactions and research with digital texts. For instance, marking metaphorical interpretations would expand search results to include implicit textual possibilities in addition to referents found in explicit structural elements such as stanzas or chapters, giving more meaningful and complete search results and textual context (133-34).

According to Gailey, “heavy editing--deep markup and conspicuous editorial guidance--is arguably necessary for many readers to even make basic use” of digital editions of lesser known texts that need a great deal of historical and ideological context to be understood (141). Though one could definitely argue that providing too much interpretative guidance could be perceived as privileging one reading of a text over another, Gailey asserts that ideally scholars and students would be able to “‘turn on’ and ‘turn off’ various commentaries on the texts” (141). The biggest road block for heavy editing is the funding needed for time and labor, but Gailey believes interpretative and critical editing is necessary to make new technologies more useful by

encouraging close reading with more opportunities for deeper, more meaningful text analysis (140-142).

Text Analysis and the TEI

Text analysis is a form of analysis that uses computer software to quantifiably analyze digital text. In the field of digital literary studies, text analysis is often used as a supplement or extension to traditional methods of literary analysis, such as close reading, by applying automated ways of providing quantitative evidence or support (or lack of) for existing or working interpretations (Hockey 66; Jockers 6).

Concordance text retrieval programs have been and continue to be the basic tools for computational text analysis (Hockey 49, Jockers 4, 15). Hockey explains that these tools are "designed to carry out detailed analysis of words, phrases, or other components of a text or collection of texts" (50). Concordance programs are used to create word-frequency lists, word indexes, and keyword-in-context (KWIC) lists (Hockey 49-65). Word indexes created by concordance and text analysis programs are used by scholars to investigate "questions of authorship and style" as well as "larger interpretative issues like plot, theme, genre, period, tone, and modality" (Hoover 517). For example, someone could examine the use of a particular word or phrase in the dialogue of female and male protagonists to determine gender assumptions apparent during the cultural and historical time period in which the text was written.

Text analysis tools are now easier to access and use than ever before. As Jockers notes: "With the spread of broadband and the accessibility of the Internet, many tools that were once platform dependent and command line in nature have been 'reinvented' for the web so that scholars may now do small-scale text processing and analysis on remote web servers using any number of web-based applications" (15). In addition, new types of visualization tools create word clouds or other types of charts that present word-frequency results in esthetically pleasing visual ways rather than simply displaying a basic word index (Jockers 15).

Textual markup is important to text analysis because it "can be used to control analysis" (Hockey 57). Hockey argues that TEI "provides a much better mechanism than any other encoding scheme for handling the complexities of scholarly texts in the humanities" (45). The content of the text is encoded so that researchers can use digital text analysis tools to identify any desired aspect of the text, including paragraphs, speakers, emphasis, margins, and so on, allowing researchers to analyze the text in a myriad of complex ways (Cohen and Rosenzweig). Besides basic elemental structures, such as paragraphs, the structural elements chosen to be marked up depend on the analytical or research goals of the given project. With highly detailed TEI markup, XML-aware tools could potentially make "standard types of vocabulary analysis in novels," such as "word frequency distributors," into "smart data" by also including information about the narrative voice, such as the gender or social class of the narrator and/or characters (Flanders and Jockers 19-20).

Feminist Markup and Analysis

Along the same lines as Gailey's argument for interpretative and critical markup, feminist digital humanities scholars have also acknowledged a need for heavy editing for scholarly digital editions of women's writing. A significant number of digital archives of women's writing have emerged over the last couple of decades, including, but not limited to those I explore in this paper: the Orlando Project, Women Writer's Project, and Willa Cather Archive. As Jacqueline Wernimont explains in her article "Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives," digital archives are important for recovering and making visible women's writing that has historically been marginalized, but mere presence is only the beginning (para. 2-5). A more complete feminist intervention of the text, Wernimont argues, would include feminist modes of production and technologies, such as text encoding, markup and text analysis (para. 2-5).

Wernimont argues that the use of interpretative markup can produce blurred boundaries that highlight feminist literary criticism by not only providing digital access to marginalized women's writing but also using that writing to illustrate inconsistencies in dominate male definitions of genre and other literary elements and research goals (para. 6). For example, instead of forcing women's writing into patriarchal or canonical definitions of genre, the interpretative nature of markup can be mobilized to create new definitions of genre and new genre-tags that more accurately describe women's writing (para. 6). In addition, when DTDs are made public rather than invisible, they can be read as a "generative . . . model of a text" rather than the "sole authoritative model" (para 13). This opens up a space for feminist intervention, for understanding markup as interpretation, for new modes of production that do not privilege pre-established patriarchal definitions of text. From this perspective, TEI, as it is expressed in XML, and its predecessor SGML, can be seen as "political rather than neutral tools" that allow for the space of feminist intervention (para. 11).

Wernimont points to text analysis tools as a possible site of feminist intervention. Such text analysis tools could include tools that are customized with the specific intention of analyzing women's writing or tools that use digital archives of women's writing as a "testbed" for experimentation for the purposes of future enhancement (para. 15). It is easy to understand how tools customized for the analysis of women's writing could be considered feminist, but what about when women's writing is being used as a testbed for tools designed for general text analysis purposes? Wernimont suggest that such tools might develop, either consciously or unconsciously, a feminist aesthetic due to the nature of the testbed (women's writing). Even so, this does not guarantee a feminist aesthetic or that such tools will be utilized by the greater digital humanities for feminist or other critical analysis purposes (para. 15). It is up to feminists and other critical theorist to "leverage digital tools to transform literary scholarship in meaningful ways" that do not "continue old patriarchal privileges of expertise and authority" (para. 12).

Method

Setting out, my primary research goal was to examine digital archives of women's writing for evidence of feminist intervention in terms of markup and text analysis. I explored three digital archives of women's writing: Women Writer's Online, the Willa Cather Archive, and the Orlando Project. I chose not to include other digital archives of women's writing, such as the Victorian Women Writers Project and the Dickinson Electronic Archives, because they do not experiment with text analysis tools. I looked at the websites of all three digital archives for documentation concerning markup choices and text analysis. I also searched for other available documentations about these three archives in library databases, Google Scholar, and Open Access journals. I looked specifically for explicit mention of feminist interpretative and critical markup and implicit implications of a feminist mindset.

As I had already read Jacqueline Wernimont's article "Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives," in which she situates Women Writers Online and the Orlando Project as central to feminist scholarship and digital humanities, I already expected to find signs of feminist markup in those two digital archives, but I wanted to see the evidence for myself and to look deeper into experimentation with text analysis tools. As far as the Willa Cather Archive, I was unsure of what I would find. As I began my research, I was especially interested in looking for connections between interpretative and critical feminist markup and more meaningful text analysis for feminist scholarship and for digital literary studies as a whole, to which Wernimont alludes.

Wernimont argues that there is "an impact on technological development when the test cases are exclusively the writing of women," when "women's writing is the 'testbed' for digital development" of new encoding practices and text analysis tools, but that this impact is not always acknowledged as a fundamental contribution to the field of digital humanities (para 15). Wernimont maintains that relationship between "women's writing-as-testbed" and the

development of new encoding practices and tools "raises the possibility that there might be a feminist basis for tools and methods, even if those are not themselves feminist" (para. 15). Wernimont asserts that the "particular effects" of the this relationship "deserve essays of their own," which is exactly what I am attempting to do in this research paper: look for a direct cause and effect relationship between women's writing, feminist or interpretative encoding practices, and more meaningful text analysis outcomes for feminist scholarship.

Results

Women Writers Online

According to their website, "Women Writers Online is a full-text collection of early women's writing in English, published by the Women Writers Project at Northeastern University." The digital texts, originally published in print between 1526 and 1850, are available with a paid subscription. A free one-month trial subscription, which I was easily able to acquire, is available for both institutions and individuals. The Women Writers Project is directed by Julia Flanders, head of the Digital Scholarship Group and Professor of the Practice of English at Northeastern University.

As reported in the "Methodology for Transcription and Editing" page, accessible through the "Frequently Asked Questions" page, texts are treated as documents rather than works of literature, as cultural artifacts "whose historical specificity is part of [the] value." Encoders "do not emend the text or create critical or synthetic editions; each encoded text is a transcription of a particular physical object." Thus, the encoding practices of Women Writers Online consist of traditional formal and structural markup and avoids overt interpretative and critical markup.

On the other hand, in their collaborative paper "Feminism in the Age of Digital Archives: The Women Writer's Project," Wernimont and Flanders explain that the early modern women's writing digitized for Women Writers Online often defied the "process of categorization, explication, and description central to digital text markup" (428). In order to remain true to the

women's writing being encoded, they chose to forego traditional genre categories, replacing them with new categories, and the markup tags that represented them, in order to more accurately describe and classify women's writing (429).

For example, digitized texts that might have formally been categorized according to the traditional women's genre of "romance" were, upon closer examination, found to defy such limited categorization. Signifiers of "romance" in early women's writing might refer to a "romance" with history, writers, books, plays and sometimes even the more traditional, lovers (428). Wernimont and Flanders explain: "The texts we digitize regularly confound easy categorization (and we would argue this is not unique to women's writing but that women's writing has given us occasion to make this clear). They thus require finer descriptions of form and in some cases, implementation of new terminology or hybridized terms" (429). So, although the encoding practices of Women Writers Online do not engage in critical markup in relation to the body of the texts, the did reinvent traditional markers of genre to classify texts in the TEI Header, an act which can definitely be considered a feminist intervention.

On the "Current Projects and Plans" page of the website, the Women Writers Online Lab is described as space to "test out experimental interface ideas" and "give users a chance to work with beta versions." The goal is to create an "experimental annex or 'sandbox' connected to the main WWO interface." At the time of writing this paper, the "small set of prototypes currently visible" to non-subscribers can be explored on the Women Writers Online Lab portal page. The portal offers a demonstration of what kind of analysis these tools are capable, and are not representative of the tools' only analytical use. To be clear, I analyze Women Writers Online particular *use* of the tools rather than examine the entirety of the tools' technical specifications and capabilities.

Two out of the four the examples of text analysis tools, labeled Parallel Coordinates and Mean and Deviation, give various visualizations of the corpus as a whole. Neither use of these

tools, at least in this particular variant, seems particularly feminist in nature, as the purpose is geared toward a visual comparison of frequencies rather than critical analytics. However, the data generated could very well be used for a feminist scholarship. The use of the third tool, Dramatic Speakers by Gender is decidedly feminist in nature. The tool is used to compare the percentage of male and female speakers in each scene of two plays written by two women writers, allowing for an exploration of female voice and subjectivity in early women's writing. The complex outcomes of this text analysis tool rely on detailed markup that denotes the gender of each speaker in the text. Markup denoting gender is more structural than interpretative, but it does represent a significant attention to textual detail that exceeded typical formal markup practices and can be considered a feminist intervention when used to support feminist inquiry. The use of the fourth tool, Fictional Correspondence Networks, compares the number of letters, along with writer and recipient, written by characters in a novel, and organizes the visualization by gender. Although, simply offering the ability to organize by gender does not alone constitute a feminist critique, such ability could certainly support feminist inquiry, and also requires detailed markup that denotes the gender of each letter writer in the text.

The four tools were "created using the free open-source Protovis visualization framework developed at Stanford University." According to the Protovis page on TAPoR, "Protovis is no longer under active construction [but] many of the concepts used in its development have been implemented in its successor, D3.js." This supports Wernimont's observation that women's writing is being used as a testbed for the development of tools for the digital humanities. As evidenced by the use of the tools carried out by Women Writers Online, text analysis that relies on detailed markup leads to outcomes that are meaningful to feminist scholarship. Detailed markup that denotes the gender of narrator, speaker or character can give insight on female voice and subjectivity in a historical context of women's writing. As Wernimont alludes, when feminist inquiry is used to test the development of text analysis tools like Protovis and its

successor D3.js, a greater possibility occurs for the creation and perpetuation of digital humanities tools that are influenced by women's writing and support feminist scholarship.

The Willa Cather Archive

The Willa Cather Archive is published and maintained by University of Nebraska-Lincoln and edited by Andrew Jewell. The archive contains scholarly and first editions of Cather's novels and collections, short fiction, interviews, speeches, letters, nonfiction, and journalism, all of which are encoded in XML using TEI standards. The archival and accompanying resources are publically available free of charge. I could not find much documentation about the markup practices of the archive, besides the fact that all texts are XML/TEI conformant. I did not find any explicit mention of feminist practice or influence in regards to markup practices or the archive in general.

However, I did find documentation of the use of somewhat interpretative markup in Jewell's article "The Professor's Mouse: Cather Scholarship in the Digital Age," available on the "About the Cather Archive: Introduction" page. Jewell states:

> We are marking up every person, title, character, and group name in a way that will allow highly sophisticated searches. If a scholar is writing a piece on Cather and Shakespeare, that person will be able to do a search that locates every mention of Shakespeare in Cather's journalistic writings, even if the actual works she used were something like "the great bard" or "the author of Macbeth." (para. 9).

I call this only somewhat interpretative markup because regularizing names is not quite as interpretative or controversial as marking conceptual signifiers such as metaphors, for example.

Jewell asserts that one of the "foremost 'experiments' on the Willa Cather Archive is TokenX, text analysis, visualization, and play tool created and customized for the Cather Archive by Brian Pytlik Zillig" ("The Willa Cather Archive" para. 8). According to the archive's "Introduction to TokenX" page, TokenX allows scholars to visualize, analyze, and understand

the text in new ways, including word frequencies, concordances, and patterns of usage. The tool and texts are freely available on the archive for anyone to use.

Due to their general analytical nature, none of these functions can really be perceived as specifically feminist. Nor does the Willa Cather Archive claim to have any specific feminist aim. The data generated by the TokenX tool could be much more meaningful if, for instance, word frequencies could be displayed along with more complex data about the narrator, speaker or other contextual information, which would require more detailed markup of Cather's texts as well as more robust analysis tools. However, Willa Cather is a woman writer. Although she is already a canonical figure in American literary history, free digital access to any woman writer, even an established canonical figure like Cather, can be seen as a feminist recovery of women's writing.

The Orlando Project

According to the Orlando Project website, unlike most digital archives of women's writing, the Orlando Project is not a "digital edition of an existing text" ("What is Orlando" para. 4). *Orlando: Women's Writing in the British Isles from the Beginnings to the Present*, published online by Cambridge University Press in 2006, is continuously updated digital literary scholarship consisting of "collaboratively authored biographical, critical and historical material on over 1000 writers" ("About the textbase"). Orlando, directed by Susan Brown, Patricia Clements and Isobel Grundy, is available by paid subscription. Orlando is not an archive of digitized historical texts, but is a born-digital resource about women writers and literary history. I was able the access the full resource through the King Library at San Jose State University.

Orlando explicitly situates their agenda within feminist scholarship concerned with women's writing and literary history ("About Us: Project" para. 1). As reported by the "About Us: Project" page: "The Orlando history focuses on gender and other aspects of cultural formation, and it emphasizes the intellectual, material, political, and social conditions, including

writing by men, that have, over time, helped to shape writing by women. These, and many other considerations, have determined the Orlando Project's tagsets" (para. 4).

The "Going Electric" section on the "Help" page, accessible at orlando.cambridge.org, provides a wealth of information about the markup of Orlando. Orlando overtly states that their DTD tagsets were specifically developed to facilitate feminist inquiry and scholarship. In their words: "The *Orlando* tagset is designed to identify elements of writers' lives and writing that are important to an understanding of women's literary history: it attempts to map the diverse and changing literary conditions under which women's writing has been shaped and received" (para. 19).

Orlando explains: "Although literary history is heavily dependent on quantifiable and factual statements, it is most interested in the unquantifiable. Biographical accounts of authors consist not only of birth, death and publication dates, but also of information (and sometimes speculation) about such matters as political allegiances, religious beliefs, race, class, and sexuality" (para. 7). The tagsets "represent and make searchable both relatively straightforward and highly interpreted information" (para. 7). Thus, "it is possible to search such fairly simple concepts as Anthologization, Contract, Copyright, or Earnings, and also ones, such as Motifs, Techniques, or Narrative Voice, on which the views of its readers, like those of its subjects, are bound to be various" (para. 7). Thus, Orlando's markup, although it does not ignore structural details, is highly interpretative, conceptual and grounded in feminist scholarship.

Orlando's markup consists of four DTDs, of which "the 'formal' tags associated with sections, paragraphs, and the like" are based on the TEI Guidelines and "the 'content' tags associated with the knowledge domain of literary history," which concentrate on the author's lives, writing and literary careers, are home-grown by the Orlando Project ("About Us: Tagsets"

para. 2). The highly detailed tagsets allow users to perform a number of complex searches, follow interrelated links, and make meaningful connections between writers, culture and history.

In "Text Mining and Visualization for Digital Literary History," Susan Brown insists: "The extensive tagging makes Orlando a unique resource for experimentation with data mining, machine learning, and visualization techniques to investigate the impact that interpretive markup has on the data mining and the visualization of results" (para. 2). Brown and her colleagues have been experimenting with text analysis on the Orlando textbase and have found that indeed the complex interpretative nature of the markup is producing meaningful analytic outcomes.

In "Visualizing Varieties of Association in Orlando," Brown and colleagues document their experience with a "prototype for accessing linkages" between authors, ranging from many tags in their tagset, from "politics, plots, or relations with publishers" (1). Even in prototype form, the Orlando Degrees of Separation tool, now updated and accessible through TAPoR, "reveal[ed] a high degree of underlying complexity in the associations between people in the textbase" (2). These complex associations would not be possible without Orlando's complex and highly interpretative tagset.

Also available on TAPoR is Orlando Breadboard, a tool that "allows one to easily find and view authors, with a compact representation of their connections to other authors, places, literary references, organizations and miscellaneous." As suggested on the tool's general information page, the connections between authors made using Orlando Breadboard give evidence of "how women's writing changed the world." Another tool, Orlando Vision, accessible through TAPoR, is "an application for visualizing a specific collection of authors, and the links or associations between them," which "are determined by co-occurrence in the Orlando dataset." Concerning these tools developed for Orlando, there is no mention of a specific feminist analytic goal in TAPoR, but the tools were in fact developed to take advantage of Orlando's interpretative and critical markup, which is based in feminist literary scholarship.

Discussion

I believe that all three digital archives can be understood as participating in feminist scholarship, whether or not feminist scholarship is an explicit goal. The Willa Cather Archive does not explicitly situate itself within feminist scholarship, nor does it employ highly interpretive or critical markup, but the fact that Cather is a canonical female figure in literary history establishes the archive within the larger recovery of women's writing. Because of the archive, most of Cather's work is freely available to anyone with the means and the motivation to study her work, no doubt resulting in meaningful feminist scholarship.

Women Writers Online and the Orlando Project explicitly situate themselves within feminist scholarship but their encoding practices differ greatly. Women Writers Online adheres to the more traditional path, emphasizing formal and structural markup. However, they do keep feminist scholarship in mind, creating new genre classification tags in order to more accurately describe women's writing and facilitate feminist inquiry. I did not find documentation explicitly acknowledging the markup of gender, but the text analysis outcomes of the two tools, "Dramatic Speakers by Gender" and "Fictional Correspondence Networks," suggest that such detailed markup of gender was employed with the specific purpose of generating a gendered analysis of the texts involved. The Orlando Project is the most explicitly feminist of the three archives. Orlando's primary objective is to support feminist scholarship, creating unique tagsets with a wide range of highly interpretative and critical markup, covering a wide range of signifiers that make social, cultural and political connections between women writers, their work and literary history, to support feminist inquiry.

Examining these archives for evidence of feminist markup has led me to believe that feminist markup can be understood as a spectrum instead of in terms of a concrete definition. Out of the three archives, Orlando definitely exhibits the greatest degree of feminist markup. Their encoding practices and tagsets, although containing formal and structural elements,

emphasize interpretative and critical markup rooted directly in the interests and goals of feminist scholarship. Women Writers Online has employed feminist markup to a lesser extent, focusing on formal rather than critical markup, but their creation of new genre tags denotation of gender does demonstrate a leaning towards markup choices grounded in feminist scholarship.

I believe I did find a direct connection between feminist markup and more meaningful text analysis outcomes. The Cather archive's customized TokenX tool is useful for finding and visualizing basic word frequencies, concordances and statistics, but does not include any cultural, historical, or literary context within the data. No doubt, the data generated could be used to support feminist scholarship but there is nothing inherently feminist, or even interpretive or critical about the data.

On the other hand, when applied to the texbase of Women Writers Online and the Orlando Project, text analysis provides more meaningful outcomes. This is both the result of the use of detailed markup as well as participation in the development of new, more complex text analysis tools. Women Writers Online is experimenting with text analysis tools developed by Stanford University; the way the tools are used to analyze aspects of the textbase, in terms of gender and narrative voice, for example, provide data that is far more complex and meaningful than that generated by the Cather archive's TokenX tool. Not only do the outcomes offer more complex data, but in the case of gender and narrative voice, the data generated supports the interests of feminist scholarship. However, whether this outcome is the result of feminist-specific markup or simply detailed markup is not completely clear. Perhaps any detailed or interpretative markup could be considered feminist if it is leveraged for feminist scholarship. Perhaps, in some cases, it is not the markup itself that is feminist but the means of inquiry and how scholars leverage text analysis tools.

However, outcomes of text analysis experiments with the Orlando Project *are* the direct result of feminist-specific markup. Since the Orlando Project is more explicit about the

connection between tagsets and feminist scholarship, it is easier to see the connection between feminist markup and meaningful text analysis outcomes and to argue for a direct cause of and effect relationship between the two variables. In the case of Orlando, both the markup and the text analysis tools are grounded, created, and utilized for feminist scholarship. Thus, being rooted in feminist scholarship, Orlando's text analysis outcomes have a decidedly feminist slant.

The tools offer new ways to see the complex relationships between writers and entities and situate these relationships in cultural, political and historical context. Such complex and contextual outcomes are possible because of Orlando's complex and highly interpretative markup. Furthermore, because Orlando's highly interpretative tagsets were generated with feminist inquiry in mind, text analysis outcomes include contextual data that emphasizes women writers' contributions to literary history. On the other hand, the tools developed by Orlando could also take advantage of markup that is not directly rooted in feminist scholarship, demonstrating the value of highly interpretative and critical markup for text analysis within and also beyond feminist scholarship.

Conclusion

When talking about literature, even when talking about quantitative methods of analyzing literature, it is difficult to make concrete conclusions. Literature is interpretative and so are the methods of encoding digital scholarly editions. Thus, so-called quantitative methods of computer based text analysis are also interpretative, considering that the texts being analyzed are encoded using interpretative methods.

I have determined that feminist markup exists on a spectrum and that highly detailed, interpretative and critical feminist markup does in fact lead to more meaningful text analysis outcomes. When markup and text analysis tools are rooted in the interests of feminist scholarship, the outcomes will have a decidedly feminist slant and will better support feminist inquiry and the study of women's writing.

I have also determined that heavily detailed or interpretative and critical markup in general will yield more meaningful text analysis outcomes. Tools created to take advantage of highly interpretative and critical markup will provide more complex contextual information needed for meaningful literary analysis. Thus, interpretative and critical markup is not necessarily a feminist-specific practice nor does feminist analysis have to be grounded in women's writing. Interpretative markup could support feminist analysis of texts by male authors as well as other types of critical analysis, surrounding issues of race, class or sexuality, for instance.

Although I see feminist markup as part of a larger move in digital literary studies towards the recognition of the benefits of highly interpretative and critical markup for text analysis, I do not believe this perspective detracts from the importance of feminist-specific markup. I believe that markup rooted in the interests of feminist scholarship can be applied to the writing of men as well as women, just as feminist analysis is applied to all types of literary texts, despite the gender of the author, in traditional literary studies.

More work needs to be done analyzing text analysis tools that are developed with feminist scholarship in mind or that use women's writing as a testbed to gauge outcomes of interpretative markup and text analysis. As a result, we may gain a better understanding of how to produce more meaningful computational text analysis as well as the integral role women's writing has played and continues to play in developing technologies central to the field of digital humanities and digital literary studies.

Works Cited

"A Gentle Introduction to XML." *TEI*. TEI Consortium. 16 Sep. 2014. Web. 14 Oct. 2014.

Brown, Susan, Patricia Clements and Isobel Grundy. *Orlando: Women's Writing in the British Isles from the Beginning to the Present*. Cambridge University Press, 2006. Web 9 Dec. 2014.

Brown, Susan, Patricia Clements and Isobel Grundy. *The Orlando Project*. University of Alberta. Web 9 Dec. 2014.

Brown, Susan. "Text Mining and Visualization for Digital Literary History." *The Canadian Writing Research Collaboratory*. Canadian Foundation for Innovation. 1 Dec. 2010. Web 9 Dec. 2014.

Brown, Susan, et al. "Visualizing Varieties of Association in Orlando." *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1.1 (2009): 1-5. Web 9 Dec. 2014.

Buzzetti, Dino and McGann Jerome. "Critical Editing in a Digital Horizon." *Electronic Textual Editing*. Ed. Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth. New York: Modern Language Association, 2006. Print.

Cohen, Daniel Jared, and Roy Rosenzweig. "To Mark Up, Or Not to Mark Up." *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Fairfax: Center for History and New Media, n. d. Web. 14 Oct. 2014.

Cummings, James. "The Text Encoding Initiative and the Study of Literature." A Companion to Digital Literary Studies. Ed. Ray Siemens and Susan Schreibman. Oxford: Blackwell, 2007. Print.

Eggert, Paul. "Text-Encoding, Theories of the Text, and the `Work-Site." *Literary & Linguistic Computing* 20.4 (2005): 425-435. *Academic Search Premier*. Web. 9 Dec. 2014.

Flanders, Julia, and Matthew L. Jockers. "A Matter of Scale." Northeastern University, Boston, MA. 18 Mar. 2013. Keynote Lecture from the Boston Area Days of Digital Humanities Conference.

Flanders, Julia. "Women Writers Online." *Women Writers Project*. Northeastern University. Web. 14 Oct. 2014.

Gailey, Amanda. *A Case for Heavy Editing: The Example of* Race and Children's Literature in the Gilded Age. *The American Digital Scholar in the Digital Age.* Ed. Amy E. Earhart and Andrew Jewell. Ann Arbor: University of Michigan Press, 2011. 124- 205. *Digital Culture Books*. Web 9 Dec. 2014.

Hockey, Susan M. *Electronic Texts in the Humanities: Principles and Practice*. Oxford: Oxford University Press, 2000. Print.

Hoover, David L. "Quantitative Analysis and Literary Studies." *A Companion to Digital Literary Studies*. Ed. Ray Siemens and Susan Schreibman. Oxford: Blackwell, 2007. Print.

Jewell, Andrew. "The Professor's Mouse: Cather Scholarship in the Digital Age," presented at the 2005 International Willa Cather Seminar, Lincoln, Nebraska. Web 9 Dec. 2014.

Jewell, Andrew. "The Willa Cather Archive," Literature Compass Blog. Mar. 2008. Web 9 Dec. 2014.

Jewell, Andrew. Willa Cather Archive. Center of Digital Research in the Humanities. University of Nebraska-Lincoln. 2004-2014. Web 9 Dec. 2014.

Jockers, Matthew L. *Macroanalysis: Digital Methods & Literary History*. Urbana: University of Illinois Press, 2013. Print.

Orlando Project. "Orlando Breadboard." TAPoR. 6 Sep. 2013. Web. 9 Dec. 2014.

Orlando Project. "Orlando Degrees of Separation." TAPoR. 8 Feb. 2012. Web. 9 Dec. 2014.

Orlando Project. "Orlando Vision (OVis)." TAPoR. 6 Sep. 2013. Web. 9 Dec. 2014.

Ramsey, Stephen. "Algorithmic Criticism." *A Companion to Digital Literary Studies*. Ed. Ray Siemens and Susan Schreibman. Oxford: Blackwell, 2007. Print.

Renear, Allen H. "Text Encoding." *A Companion to Digital Humanities*. Ed. Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell, 2004. Web. 14 Oct. 2014.

Rockwell, Geoffrey, Kirsten C. Uszkalo, and Stefan Sinclair. *TAPoR*. 2014. Web. 11 Nov. 2014.

Siemens, Ray, and Susan Schreibman. Editor's Introduction. *A Companion to Digital Literary Studies*. By Siemens and Schreibman. Oxford: Blackwell, 2007. Print.

Stanford Visualization Group. "Stanford Vis Group: Protovis." TAPoR. 20 Nov. 2014. Web. 10 Jan. 2015.

"TEI: History." *TEI*. TEI Consortium. 20 Jan. 2011. Web. 14 Oct. 2014.

Van den Branden, Ron, Terras, Melissa and Edward Vanhoutte. *TEI by Example*. 23 Dec. 2012. Web. 14 Oct. 2014.

Wernimont, Jacqueline, and Julia Flanders. "Feminism in the Age of Digital Archives: The Women Writer's Project." *Tulsa Studies in Women's Literature* 29.2 (2010): 425-435. *Project Muse*. Web 9 Dec. 2014.

Wernimont, Jacqueline. "Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives." Digital Humanities Quarterly 7.1 (2013). Web 9 Dec. 2014.