Metatation: Annotation as Implicit Interaction to Bridge Close and Distant Reading

HRIM MEHTA, University of Ontario Institute of Technology
ADAM BRADLEY and MARK HANCOCK, University of Waterloo
CHRISTOPHER COLLINS, University of Ontario Institute of Technology

In the domain of literary criticism, many critics practice *close reading*, annotating by hand while performing a detailed analysis of a single text. Often this process employs the use of external resources to aid analysis. In this article, we present a study and subsequent tool design focused on leveraging a critic's annotations as implicit interactions for initiating context-specific computational support that automatically searches external resources. We observed 14 poetry critics performing a close reading, revealing a set of cognitive practices supported through free-form annotation that have not previously been discussed in this context. We used guidelines derived from our study to design a tool, Metatation, which uses a pen-and-paper system with a peripheral display to utilize reader annotations as underspecified interactions to augment close reading. By turning paper-based annotations into implicit queries, Metatation provides relevant supplemental information in a just-in-time manner and acts as a bridge between close and distant reading.

CCS Concepts: • **Human-centered computing → Empirical studies in visualization**; *Gestural input*; *Information visualization*; User centered design;

Additional Key Words and Phrases: Digital humanities, close reading, distant reading, ink annotations, pen-based interfaces, implicit interaction

ACM Reference format:
Hrim Mehta, Adam Bradley, Mark Hancock, and Christopher Collins. 2017. Metatation: Annotation as Implicit Interaction to Bridge Close and Distant Reading. *ACM Trans. Comput.-Hum. Interact.* 24, 5, Article 35 (November 2017), 41 pages.
https://doi.org/10.1145/3131609

1 INTRODUCTION

Literary criticism is conducted through a multitude of approaches. Close reading, one of the traditional and primary approaches, involves analyses of spatial and temporal interactions between the syntactic, semantic, structural, rhetorical, and phonetic features within a text. These analyses lead to the generation of a "reading" or interpretation of the text, which then informs the next reading by its discoveries. From our study, we found that this process can repeat upwards of 30 times in a feedback loop that operates between discovery and re-reading. In contrast, distant reading, one of the more modern approaches to text analysis, leverages computational tools for detecting patterns in a corpus as a whole. A lack of understanding of the close reading practice in the human-computer interaction (HCI) community and skepticism in the literary community regarding the

Authors' addresses: H. Mehta, A. Bradley, and C. Collins, 2000 Simcoe Street North, Oshawa, Ontario, Canada L1H 7K4; emails: {hrim.mehta, adam.bradley, christopher.collins}@uoit.ca; M. Hancock, 200 University Avenue West Waterloo, Ontario, Canada N2L 3G1; email: mark.hancock@uwaterloo.ca.

https://doi.org/10.1145/3131609

value computation brings to literary criticism prompted us to delve into the practices of poetry analysts.

At first glance, the practice of close reading for literary analysis might appear to be straightforward. However, conversations with our domain expert, who holds a PhD in English Literature and is also a co-author of this article, revealed that when analyzing a poem a literary critic's practice encompasses varying styles of analysis with shifting points of entry into the text. This practice attempts to make sense of the many layers of structure within a text and tries to figure out how literary devices interact to create meaning. Each reading builds on the previous version, slowly building up a larger and larger network of how the text generates this meaning. Despite many available digital tools, investigations within literary studies are still most often conducted on paper. The challenges that arise when designing for this domain are centered on both the nature of the work flow and the nature of how literary critics build knowledge. Humanities epistemologies are built rhetorically, meaning that each individual user must have the room within the system to think creatively and the system must be open enough to accommodate this work flow. Often, this is about supporting intuition and not only is each user different in their requirements in regards to this intuition, but after each "reading" of a text these requirements change. This is a unique set of constraints for designing a system. Existing digital tools for literary analysis create a different work flow and thus seem to be discordant with the critic's needs as described by our domain expert. To assess whether this is the case and to gain insight into what comprises the close reading practice of the critics, we undertook an observational study.

In this article, we present the results of the observational study of literary critics' work flow when performing close readings of poetry and the subsequent design implications for the development of digital literary analysis tools. When asked to analyze a poem, we observed that literary critics participate in a very complex cognitive "shorthand" for trying to make sense of the many layers of structure within a text. We extend a previous work on external cognition and the role of annotations therein to the domain of literary criticism and provide evidence that the act of close reading for the purpose of interpretation embodies the cognitive processes essential to making sense of a single text. From our study of literary critics' practice, we note that a large part of the act of close reading is *thinking by doing* through annotation; that is, the practice of annotation is not merely an exercise in note-taking, but is rather a means for sensemaking. These findings demonstrate the need to support close reading processes as a starting point when designing digital literary analysis tools followed by the now common distant-reading-based corpora analysis and comparison techniques.

While the existing work flow with pen-and-paper has been adequate for the task of close reading, the critics also manually engage with external resources while performing the readings. This is not only time-consuming but also disruptive to the analysis process. We introduce computational support to the existing feedback loop of reading, hypothesis generation, and re-reading. We do this to enable easy access to external resources and to augment the existing workflow by providing analysis support. This support comes in the form of new literary connections/interactions within a text grounded in the ones identified by a critic.

To demonstrate this digital augmentation, we also present the design and implementation of a tool, Metatation, based on the design guidelines derived from our observational study of literary critics. This tool addresses the need for supporting the practice of annotation when performing a close reading of a single text and guides distant reading of literary corpora on the basis of these sensemaking artifacts (annotations). Bridging these two in-conflict but complementary approaches to literary analysis, Metatation allows literary critics to interact with a text using pen-and-paper, and processes the free-form annotations they make during close reading as *implicit cues* for identifying and generating supplementary metadata relevant to their current thought process

in real-time (see Figure 1). Metatation then helps situate these findings in the context of a corpus of interest, supporting verification/exploration of a critics' line of investigation.

Existing tools in this domain [1, 8, 29, 33] do an excellent job of presenting meta-information in the form of highly stylized visualizations, but end up neglecting the critics' *experience* of the poetic language. These tools do not support the practice of annotation and present supplemental information up front, before any work has been done by the analyst, limiting the possibilities for the types of cognitive engagement required by the sensemaking task. This is further complicated by the fact that literary critics often perform multiple readings of a single text with a shifting focus. While these tools support the exploration of patterns either in a single text or in large collections of texts, the exploration is generally prompted by top-down research questions/hypotheses and there is limited scope for modifying the objective facts presented by the tools based on the subjective position of the critic.

We found that by instituting a subtle shift in design from offering information before the active workflow has started to after the critic has already conveyed components of interest, not only allows for augmenting the close reading process with minimal disruption, but also helps constrain the distant-reading-based support to be more relevant to the critic's current line of investigation. This shift in the moment of interaction facilitates the individual subjective experience of a critic and offers objective facts in relation to that subjectivity creating a feedback loop not present in previous tools. Thus, while existing tools provide meta-information through an automated analysis of the text being read, Metatation interprets the free-form annotations being made on the poem by the critic. Using Anoto[1] pen-and-paper, in real time, Metatation presents information relevant to only those features of the text that are being currently noticed and analyzed by the critic.

The contributions of this article are as follows: First, we present the results of an observational study of expert poetry critics as they analyze poems, providing insights into the sensemaking process involved when conducting a close reading. Second, we provide design guidelines for the development of digital tools for supporting literary analysis through both close and distant reading by inferring reader intent based on underspecified interactions in the form of reader annotations. Last, we designed and implemented a prototype system, Metatation, that provides context-specific meta-information to a critic in real-time based on the derived design guidelines.

Our work builds upon research in digital humanities and HCI to provide digital tools for literary analysis that bridge the gap between close and distant reading. By focusing on integrating this technology into the existing work practices of literary critics, we can support sensemaking while augmenting available information with little disruption.

2 RELATED WORK

There are several areas of work that are most related to our own: tools to support literary analysis, studies of annotations, use of annotations for computation, and the use of pen-and-paper vs. digital documents in active reading.

2.1 Literary Analysis Tools

Close reading is the practice of comprehensive analysis and interpretation of a single text. It focuses on paying close attention to linguistic, spatial, and structural features of a text and their interactions with each other, followed by reasoning through the observations made to produce a reading of the text. This process is inherently experiential—interpretations of the text vary from one person to another as well as from one context to another. Close reading can be contrasted with the approach of distant reading [20, 31] that harnesses the power of computation to aggregate

[1] http://www.anoto.com.

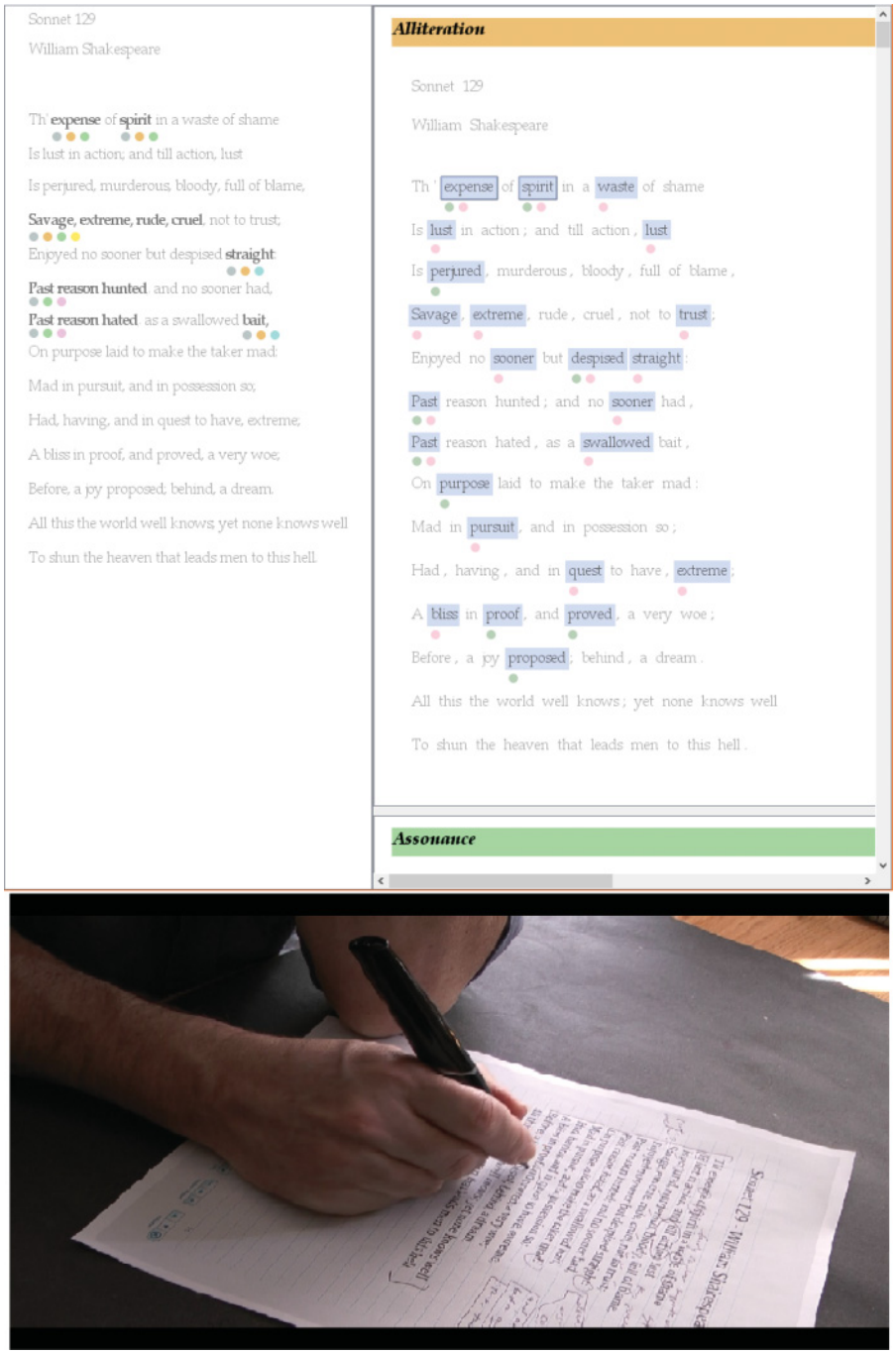

Fig. 1. The Metatation interface. As a reader annotates on paper while reading, the worksheet viewer panel (top-left), presented on a peripheral display, updates in real-time to highlight the annotated words in bold and adds color-coded dots below these words to indicate availability of implicitly generated query results pertaining to the words. The query results are presented on the peripheral display as a stream of metadata tiles (top-right).

and quantitatively analyze text corpora for the purpose of literary historical analysis. Tools for distant-reading-based literary analysis present broad overviews of collections of text to support exploration of trends across documents. Jänicke et al. [21] present an in-depth survey of existing literary analysis tools based on both close and distant reading.

Tools focusing primarily on supporting literary analysis of prose often look to demonstrate patterns across collections of texts. Text visualization tools like TextArc [38], Docuburst [12], WordSeer [33], and Compus [15] aim to support exploratory analysis of literary texts through distant reading. These tools present broad overviews of text corpora based on quantitative linguistic analyses while also supporting exploration of finer connections that reveal patterns within and between documents. Our work presents an alternative approach, focusing on close reading and the relationships between parts of a single text, which are then extended to guide distant reading based exploration of text corpora.

Poetic language often defies the general rules of syntax and semantics observed in prose, resulting in a complex dynamic system that exhibits a variety of highly charged features. Several visualization tools have thus been developed specifically for supporting the analysis of poetry. PoemViewer [1] supports close reading of poetry by allowing an exploration of semantic and phonetic relations in a poem through the use of rule-based visual mapping techniques. It visualizes semantic relations, such as word repetitions, parts-of-speech, and sentiment, and presents various types of phonetic relations, such as assonance, consonance, alliteration, end rhyme, and so on, in addition to providing information about the physiology of sound production. Myopia [8], a visualization tool for close reading of poetry, presents poetic elements, such as meter, sound, syntax, metaphor, and personification to support literary analysis of TEI (Text Encoding Initiative)—encoded[2] texts. ProseVis [11] allows interactive exploration and visualization of sonic features in a text, such as phrase boundaries, parts-of-speech, phonetic spelling, stress, and rhythm markings, to aid the discovery of phonetic patterns within a text at different levels of granularity. Poemage [29] is another such tool that supports visualization of the interaction of complex sonic patterns across a poem.

While these tools offer highly detailed meta-information through visualizations of poetry, they do not provide support for annotation, which our study reveals to be intrinsic to the sensemaking process. We build on the work of these previous systems by supporting the close reading process through the interpretation of annotations followed by their extension to distant reading based queries.

2.2 Studies of Annotations

There is a subset of previous work that specifically investigates ink annotations and has implications for the design of digital annotation systems. Marshall [26] and later Marshall and Brush [27] have investigated annotations made by students in textbooks and when reading academic papers. They provide observations that are largely consistent with our own study of annotation practices in poetry analysis, having noticed the importance of spatial arrangement, the need to support idiosyncratic annotations, and the need to minimize distraction from the reading task [26]. Moreover, they provide a coding of annotations into anchor-only (e.g., underline, circle, margin bar, etc.), content-only (e.g., notes, marks (such as *), doodles, etc.) and compound (anchor + content) marks, and evidence that most annotations in their study were anchor-only.

Our work extends this literature by considering annotations made in the domain of poetics, with the object of understanding how literary critics dismantle the linguistic structures as they read and re-read poems continuously to make sense of the texts in question. Our focus on analyzing

[2] http://www.tei-c.org/index.xml.

the annotations originates not from the need for developing annotation systems for this domain but rather as a tangible means of understanding the thought process involved when performing a literary analysis and to investigate the possibility of augmenting the close reading practice with computation. We employed the external cognition framework [39] to explain the role of annotations in close reading and attempted to infer the features being investigated by a critic based on the properties of their annotations and the text being marked.

2.3 Annotations for Computation

Most of the work on reading practices has so far focused on moving from paper to digital, and thus many tools have been developed to support paper-like annotations on digital devices [37, 53]. There are fewer examples of research that explore the idea of processing physical/digital annotations for the purpose of computation.

XLibris [41], a pen-enabled tablet display with paper-like affordances for supporting active reading, was the first to make an attempt at using annotations made by a reader to generate additional relevant content. Words annotated by a single pen stroke were used as keywords to retrieve research papers relevant to the one being studied. These papers were then presented to the reader as links in the margin close to the pen stroke that invoked this retrieval. Our system, Metatation, also employs annotations made by a reader during the analysis process as cues to generate real-time context-specific supplementary data to augment close reading. However, our system differs from this work in that we focus on physical pen strokes on paper, and process information specific to the words on the page, rather than higher level concepts like related literature to the entire document or context.

Shipman et al. [43] analyzed annotations made by students when preparing briefs on legal cases for a moot court session and characterized high-value annotations that can be used to identify important passages from a document. While their tool is similar to our work in that their system also extracts meaning from annotations made by hand on documents, their tool is intended to automate the process of summary and extraction when reading based on annotation characteristics, rather than augmenting the processes involved in sensemaking when reading.

2.4 Pen-and-Paper vs. Digital Documents in Active Reading

Active reading is the practice of engaging with the content of a text for the purpose of extracting its meaning. It differs from close reading, practiced in the domain of literary criticism, in that the purpose of close reading is not limited only to the deduction of the meaning of a text but also focuses on how various low-level linguistic features of the text generate meaning when combined into a larger network. Close reading is concerned with the system that leads to meaning, where active reading is interested in a particular meaning derived from a text, such as a summary. Active reading has been studied by previous literature with the aim of understanding the effects of substitution of pen-and-paper by digital documents on the process of text comprehension.

O'Hara and Sellen [36] presented a comparison of active reading practices for performing a text summarization task when using physical and digital documents. It was noted that annotating while reading aids text comprehension by facilitating the development of an internal representation of the text through re-structuring and collation of information, and that annotations also serve as sources for later reference. Physical paper was reported to have better affordances to support active reading practices than digital documents, such as flexibility of free-form in-context ink annotations, ease of switching between annotating and reading, ease of cross-referencing of multiple documents, and bimanual interactions, among others. Morris et al. [32] employed a similar study design that used paper or digital documents on tablets, and that also compared a stylus-enabled horizontal display to a dual-monitor desktop setup. It was observed that paper and tablets were

preferred overall for both reading and annotating, with dual-monitor desktop setups performing the worst. Though experience of using tablets and horizontal displays was comparable to that of paper, insufficient margin space and mode switching while inking were noted as challenges. However, no significant difference in terms of the number of annotations made and annotation forms was observed. Similarly, our observations of poetry critics annotation strategies when performing close reading also reflected the preference for using paper over digital documents and indicated the need for providing support for free-form annotations.

The benefits of coupling physical and digital documents were explored as early as DigitalDesk [51], a physical desk that allowed interaction with digital documents via a projected display. Since this work, many systems have explored the augmentation of physical or digital paper in connection with handwriting [24, 47, 54]. Many researchers have also specifically investigated the use of predefined pen gestures [23, 35, 46, 49] to control a digital system, as well as predefined regions [23], hotspots [44], or document structure [35] to interpret annotations digitally in a variety of domains, including academic publications [35, 46], slide handouts [44], and music composition [49]. MouseLight [45], a tool that explores bimanual interaction of a mobile projector and digital pen to augment physical paper with virtual information, is an example of subsequent research focused on closing the feedback loop from digital to physical documents. Research into the design of pen-and-paper digital systems has also lead to investigation of digital alternatives to paper for supporting reading and writing. For example, Chen et al. [9] developed a multi-surface digital slate-based system and performed a longitudinal study to test the affordances such a system offered for graduate students' active reading practices.

These systems are related to our own work, but focus on active reading of materials such as, scientific publications and lecture slides, where the primary goal is to understand the arguments being made by the text than to analyse how the linguistic features work together to generate meanings. While some of our design decisions use known technology (e.g., Anoto and stroke recognition), our focus is on utilizing underspecified interactions in the form of free-form reader annotations in combination with the semantics of the annotated text to augment the practice of close reading. The choice of achieving the same through the use of a paper-digital system is prompted by the findings of our observational study of literary critics' practices.

3 OBSERVATIONAL STUDY: POETRY ANNOTATION

In order to better understand how literary critics unravel the structures within a poem through close reading, we designed an observational study. Based on the guidance of our domain expert in regards to the primacy of annotation, our observations mainly focused on annotations as sense-making artifacts. We were also interested in what, if any, external tools (digital or otherwise) these critics already used and how they were integrated into this process.

3.1 Participants

We recruited 14 participants (11 male and 3 female), comprising 3 PhD students and 11 university professors with varied areas of expertise, who publish on literary criticism and/or teach poetics, from three different universities in Canada.

Literary critics generate insight about poetry through close reading. In previous literature, close reading of poetry is often confused with active reading tasks. Upon further examination, guided by our domain expert, the two practices are actually quite different. Active reading focuses on analyzing the content of a text to understand the arguments being made, typically with the goal of generating a summary of the text. Close reading on the other hand focuses on analyzing how the various linguistic features interact to generate different possible interpretations. Close reading

tends to be iterative, sometimes being repeated upwards of 30 times to try to continually build up a view of how the network of meanings in a poem are ordered and structured.

A typical close reading session would begin with a reading out loud to establish intuitions based on sound and rhythm and would then lead to questions that help to establish expectations. An example of this would be on the structure and layout of the poem: a sonnet has 14 lines but not all 14 line poems are sonnets. A poetry analyst would first count the lines of the poem and then see if there were other factors that suggested the sonnet form. Namely, these would be a poem set up in an octave (8 lines) and a sestet (6 lines) split by a volta (or turn) where the content of the poem would in some way change directions (often metaphorically) at this structural point. There are many different forms of poems including formless varieties making this step often quite difficult. A second pass of the poem could look for punctuation (how many sentences are there? Are there periods or semi-colons? Is punctuation used at all?) Any and/or all of these types of questions help to uncover the way that the poem unfolds in time. Each answer to one of these questions builds a different view of how the poem is creating meaning. The next pass could be based on individual words and their meanings (Are their synonyms and/or antonyms present? Do all the words relate etymologically? Are words being built up with prefixes and suffixes or are they simple in terms of diction?). Next, an analyst could look for repeating patterns in sound, rhythm, or visuals. This process continues with the analyst making notes in the margins based on an evolving set of assumptions and facts that lead to what literary critics call a reading. These readings are then compared to other readings of the same poem or often compared to other poems to try to establish an argument. These arguments are thought of as conversations and the process continually repeats itself to push the conversation forward. This is only one way this knowledge-building process unfolds—there are many types of literary criticism that focus on more cultural than structural engagements with texts. This literary analysis process bears resemblance to the practice of reflection in action observed by Schön [42]

Literary critics are formally trained to identify varied literary devices and to understand how these devices interact to generate different meanings for a given text through close reading. The training, however, does not enforce a standard set of notations to mark these findings. Close reading tends to be performed using pen-and-paper with a few rare exceptions where critics work with a touch tablet. Dictionaries and anthologies are some of the most commonly used physical resources used in the analysis process. Digital resources such as the online Oxford English Dictionary, previous scholarly work, and search are also common.

3.2 Choice of Poetry

To ensure that our study data represented a breadth of poetic styles and time periods and that the poems we used were of recognized scholarly importance, we drew on the Norton Anthology of English Poetry [16] to create our study dataset. Literary critics, as a practice, generally specialize in specific time periods of literature. Thus, someone who is an expert in Modernist literature would not have comparable proficiency in Elizabethan drama. If a literary critic is assigned a poem from their area of expertise, there is a high probability of the poem having been previously analyzed. In this case, asking the critic to do a close reading (analysis) would end up simply being a performance/teaching of the poem and no longer reflective of their sensemaking practice. Consequently, ensuring the inclusion of works of diverse poetic styles from different eras was essential to mitigate the effects of expertise bias in our study. Dividing literary history into appropriate epochs is a complex problem, and so we deferred to the well-respected anthology to supply the time constraints for our selection process as well as the texts themselves. We randomly selected 14 poems, 2 from each of the 7 different time periods (1510–1620, 1620–1690, 1690–1780, 1780–1830, 1830–1880, 1880–1920, and 1920–). These 14 poems were then randomly grouped into 7 pairs, such that

each pair included poems from 2 different time periods. To avoid expertise bias, such as having a Shakespearean scholar annotate a Shakespearean sonnet, pairs of poems were counterbalanced and randomly assigned to the participants for analysis so that each pair was analyzed by two participants, once in each ordering. As a result, none of our participants undertook the study with poems that fell within their area of expertise. Having two poems per participant and common poems between two participants permitted us to observe how the annotation practices varied based on both the reader as well as the work being analyzed.

3.3 Task and Procedure

For our study of poetry critics, each participant was asked to perform a close reading (analysis) on one pair of poems. Each poem was provided in an assigned order on a separate sheet of Anoto paper, printed in Times New Roman with 1.5 line spacing. Participants were provided with a LiveScribe[3] Anoto pen to use for annotations. The Anoto pen tracks pen position on the Anoto dot pattern permitting us to record pen strokes made by our participants as they annotate the printed sheets of paper.

Observational sessions were conducted in the regular work environment of the participants and were video and audio recorded. The video was recorded from two separate angles: one from directly above the desk to capture how the participant analyzed the poem, and another facing the participant to note any other external physical cues of the sensemaking process. We also logged pen strokes from the Anoto pen and paper using the Anoto LiveScribe software and collected the physical paper at the end of the session.

Participants were instructed to perform a close reading on each poem in accordance with their own work practices. The participant was given a printed copy of the poem and instructed to conduct an analysis until they were finished or until 15 minutes had passed. In our pilot studies, we discovered that this analysis process was a highly personal experience and the presence of the experimenter was found to be distracting. Therefore, the participants were left alone to work during each of these 15 minute periods.

Participants were not required to annotate the poem, but if they did, they were requested to use the provided LiveScribe Anoto pen. Participants were also permitted to access any form of available external resources that they would normally use. Following the analysis session, participants were asked to explain what they had annotated and why through a retrospective think aloud process using the annotated poem as a guide. The same procedure was then repeated for the second poem assigned to a participant.

The observational sessions were followed by a brief interview (non-audio-video-recorded) regarding participants' current practices when performing a close reading. We asked them how often they perform close readings as a part of their work, how many readings they perform per poem and what are the differences between multiple readings, what tools they use for annotating, do they follow a personal system for annotations, do they use color when annotating, do they mark directly in books or on photocopied sheets, what they do with the annotated sheets, do they ever go back over their annotations and for what purpose, and do they recognize what the annotations were for if they go through previously annotated sheets.

3.4 Data Coding Process

We used open coding for qualitative analysis of the recorded sessions. We first created an interactive tool to synchronize, replay, and assign codes to the recorded pen stroke log files, scans of the final annotated poem pages, videos of the analysis sessions, and videos of the retrospective think

[3] http://www.livescribe.com.

Table 1. Categories of Codes in the Final Code Set and Subcategories of Form Codes

Code category	Code subcategory
Form	Ellipses, underlines, polygons, connectors, brackets, text, miscellaneous notations
Cognitive purpose	CO, EML, EML+CO, A

Form code category	Form code subcategory
Ellipse	Circles, spirals, concentric circles
Underlines	Straight single, straight double, wavy
Polygons	Squares, rectangles, triangles, others
Connectors	Arcs, angular arcs, single-headed arrows, double-headed arrows, cross-headed arrows, arrow-headed wavy arcs, circle-begin and underline-end arcs, circle-headed arcs, lines
Brackets	Regular square, half-square, round, double round, curly, vertical line
Text	Letters, numbers, special characters
Miscellaneous notations	Squares, rectangles, triangles, others

Table 2. Initial Code Categories for Examining Annotation Function

Purpose code category	Purpose code subcategory
Poetic elements	Connections, grammar, syntax, semantics, diction, sound, rhyme, alliteration, consonance, assonance, repetition, syllable counts, stress, rhetoric, allusions, bibliographical information
Reading habits	Reading the poem out loud, reading the poem silently, pausing to look up information, idiosyncratic habits while reading

aloud sessions. This tool also allowed us to query the coded data to see, for example, all the codes for annotations in a particular area of the page or for a particular time frame.

Annotations serve as a tangible proxy for the cognitive processes involved when conducting a close reading of a poem. We thus oriented our analysis toward categorizing annotations to understand what the participants were marking, for what reason, and how. Two researchers iteratively worked on randomly selected samples of data to develop a mutually agreeable code set and guidelines for categorization of the annotations. The final code set comprised of two categories of codes, *form* and *cognitive purpose*, each of which consist of several subcategories of codes (see Table 1).

The *form* code category was defined for analyzing how participants were identifying points of interest during their readings. Considering that literary critics undergo similar training for performing analyses of texts, we initially thought that there might be a systematic notation being used for marking specific features of interest as a critic analyzed a poem and thus the possibility of leveraging the form of annotations to identify their purpose. In the initial iteration of the coding process, the *form* code had 29 subcategories of codes (see Table 1) that were grouped and refined in subsequent iterations to generate the final form code categories listed in Table 1.

To investigate what features the participants were annotating and to serve what purpose, we started by categorizing annotations based on the poetic elements they identified in addition to noting participants' reading habits, generating the code categories listed in Table 2. We realized, however, that while these codes helped us identify the types of linguistic features and thus supplementary data of possible interest to poetry critics, they were inadequate at indicating the higher level cognitive processes involved that explained the purpose of these annotations. We then reviewed frameworks that explain the role of external representations in supporting cognition. The switch to an external cognition based code set was driven by the fact that while participants were consistent in terms of what linguistic features they were identifying, the annotation forms used to mark these features were idiosyncratic, leading to annotation form-function ambiguity within and between participants. Categorization of the annotations based on the external cognition

framework, which is more high level and abstract, permitted the intent of the annotations to be discerned consistently across all participants despite this ambiguity. In addition, the coarse categorization by this code set enabled us to generalize highly specific actions performed by the participants to the more abstract cognitive tasks of hypothesis generation and hypothesis verification.

Codes based on the *cognitive purpose*, served by the process of annotation in text comprehension, included: computational offloading (CO), externalizing to reduce memory load (EML), both computational offloading and externalizing to reduce memory load (EML+CO) and ambiguous (A). CO and EML codes have been derived, in the context of annotations and close reading, based on the main cognitive benefits of using external representations identified by Preece et al. [39]. CO represents the act of thinking through the content for identifying patterns of interest by annotating, whereas EML refers to the act of tracking and assimilating intermediate hypotheses formed based on the patterns found and a critic's existing literary knowledge through the use of annotation. EML+CO is used to denote those annotations for which both CO and EML are being performed as a contiguous unit, while A marks those annotations whose purpose is unclear. Descriptions of the purpose of annotations, elicited through the retrospective think aloud sessions, were referenced to guide this *cognitive purpose* based categorization. Figure 2 shows an example of the result of the coding process, based on *cognitive purpose*, when applied to the final annotated poem sheet for participant P11.

Annotations were also noted as being located either on the text itself (word space), between the lines and stanzas of the poem (white space) or in the margins (margin space). At the outset, we manually coded annotations based on the space they occupied. However, our inability to achieve coder agreement due to frequent cross-overs of annotations across the different spaces on the page led to our decision to use automatic categorization of annotations based on their location. Annotations with more than 50% of the area of their bounding box falling under one or more bounding boxes of the lines of the poem were categorized as being in the word space. Similarly, annotations with more than 50% of the area of their bounding box falling under one or more bounding boxes of the spaces between the lines and stanzas of the poem were said to occupy the white space. All other annotations were categorized as belonging to the margin space.

Finally, all annotations were manually grouped into sets that represented a semantic unit. For example, three consecutive underlines of rhyming words. The start and end times of annotation events were also recorded.

Following the code set development, 2 of the researchers independently coded 4 randomly chosen sessions out of the 28 analysis sessions and Cohen's κ was calculated to measure inter-rater reliability. The measure revealed good agreement on coders' judgments for both annotation form, $\kappa = 0.75$, and annotation function, $\kappa = 0.66$. The remaining 12 analysis sessions were then independently coded by each of the 2 researchers.

4 RESULTS AND DISCUSSION

The results of our study highlighted several themes common to many of our participants. Firstly, free form annotation is an important part of the close reading process in literary criticism (Section 4.1). We also observed that annotation use was highly idiosyncratic and polymorphic (Section 4.2). The function of annotations can be disambiguated through patterns in space, time and forms used and the cognitive purpose served by annotation (Section 4.3). Sometimes annotation serves as a placeholder for intuition, and annotation becomes a form of experiential cognition [34] (Section 4.4). Participants made use of external resources to aid their analysis, and usually deferred their use of tools until they have finished a complete reading, reducing interruptions (Section 4.5).

These themes will be explored in depth in the following subsections, and inform the design of our subsequent system, Metatation.

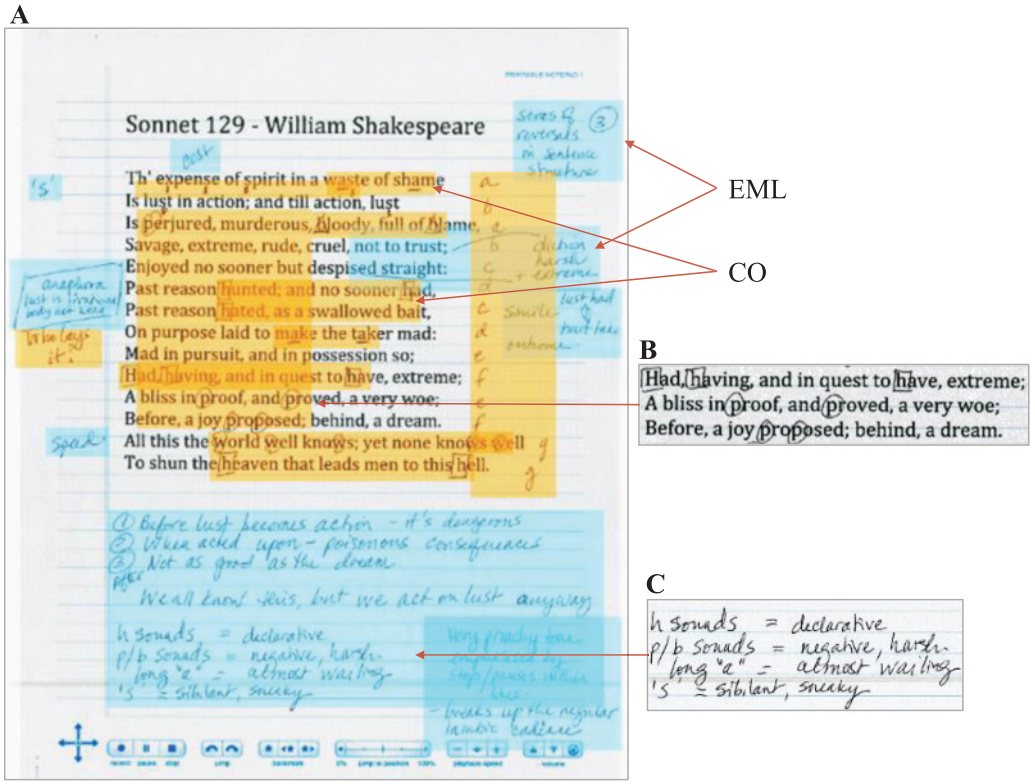

Fig. 2. Annotations made during close reading of poetry. Panel (a) shows *cognitive purpose* codes assigned to annotation units for one of the participants' annotations during the study, as visualized by our coding tool. Orange and blue bounding boxes represent annotation units categorized as CO and EML respectively. Panel (b) is an example of an annotation unit, identifying the repetition of sounds, categorized as CO. Panel (c) shows an example of an annotation unit, noting observations about repetitions of sound across the poem, coded as EML.

4.1 Role of Free-Form Annotations in Sensemaking

The practice of annotating while analysing poetry was reported as being intrinsic to the current work practices of 9 of our 14 participants. For the remaining five participants, this practice was noted to be more prevalent when working on poems that they were planning to teach, but was nonetheless an important part of their routine. All of our participants reported that they usually use a pen/pencil to annotate printed sheets of paper when performing a close reading to analyze poetry. Only two of the participants stated that they have previously used a stylus/touch-enabled tablet for the same purpose. Two of the participants stated that they occasionally use colored pens to visually separate annotations made on the current reading of a poem from those made on previous readings of the same poem or to visually separate different themes emerging from the poem.

A majority of participants reported using physical tools (pen/pencil) over digital tools (stylus, computer) for annotation in close reading. The reason for this preference could be that paper supports free-form annotations well [36]. Free-form ink annotations are highly suited to the close reading process since they afford the flexibility needed for the idiosyncratic manner of annotating

with minimal attentional overload, in contrast with digital ink annotations that require explicit switching between the available markup tools. Paper also provides additional affordances such as quick navigation, portability, rotation and physical manipulation of pages, and the ease of cross-referencing multiple documents side-by-side, among others, to support the close reading task [36]. Whether or not substituting physical paper with a stylus-enabled tablet is appropriate and provides equivalent support for the cognitive processes involved is unclear, based on our current study, and further investigation is warranted. Other forms of annotations such as speech annotations as explored in RichReview [55] might be interesting to consider but we have no evidence of our participants having used speech recordings for the purpose of analysis; further investigation is needed.

On the basis of our observational study, it is clear that the experience of language contributes to the understanding of poetry and is reflected through the practice of annotation. That is, people "think by doing" through annotation in poetics. For example, P7, whose annotations comprised of pictograms and other common annotation forms, explained that through the use of annotations he was "working through the progress in parts [of the poem] visually; it's an abstraction of what is being seen in the poem." When asked about what they do with the annotated sheets and if and for what purposes would they generally refer back to previously annotated sheets, a majority of our participants responded that the annotations are transient and they do not usually preserve these annotated sheets other than for teaching purposes. For example, P6 stated that the practice of annotating is more a technique for brainstorming. Two of the participants stated that they do preserve these sheets since they are a record of work but rarely do they refer back to the sheets. These findings are in sync with Dix's [14] observations regarding the role of externalization in supporting internal processing of the content being analyzed leading to sensemaking and how "the marks on the paper are only important while they are being actively considered." The practice of annotation in the domain of literary criticism is not simply a means for note taking but rather forms the basis for sensemaking by supporting reflection in action [42]. Free-form annotations are thus a required part of the analysis process of literary critics.

Two participants stated that following a reading, they use the annotations made while reading to synthesize the questions raised/hypotheses into main themes which then prompt further readings. All of our participants noted that they undertake multiple readings of a poem, ranging anywhere between 5 and 30, with shifting focus to characterize the interactions between various features of the poem. One of our participants explained how multiple readings of a poem lead to a cumulative buildup while another participant compared the process of unraveling a text through close reading to building a structure with Lego blocks. All of our participants do agree that when referring back to their annotated sheets, they are able to recognize a majority of their annotations and the annotations act as a prompt for recalling their thoughts about the poem.

4.2 Polymorphism of Annotations

All of our participants reported that they do not follow a formalized personal system of annotations. From our analysis and observations of the annotation process, it was clear that annotations made by participants were highly idiosyncratic (specific to their process alone) and polymorphic (annotation form did not consistently serve the same function for different participants, or even throughout the annotation process of a single poem for an individual participant leading to form-function ambiguity).

Figure 3(a) shows an example of how underlines are being used by P14 for highlighting both words that are indicative of volition as well as words that exhibit alliteration and internal rhyme. In Figure 3(b), P4 has overloaded the use of circles to indicate repetition of sound patterns, peculiarity of punctuation use, and words employed as qualifiers. Similarly, Figure 3(c) shows how circles and

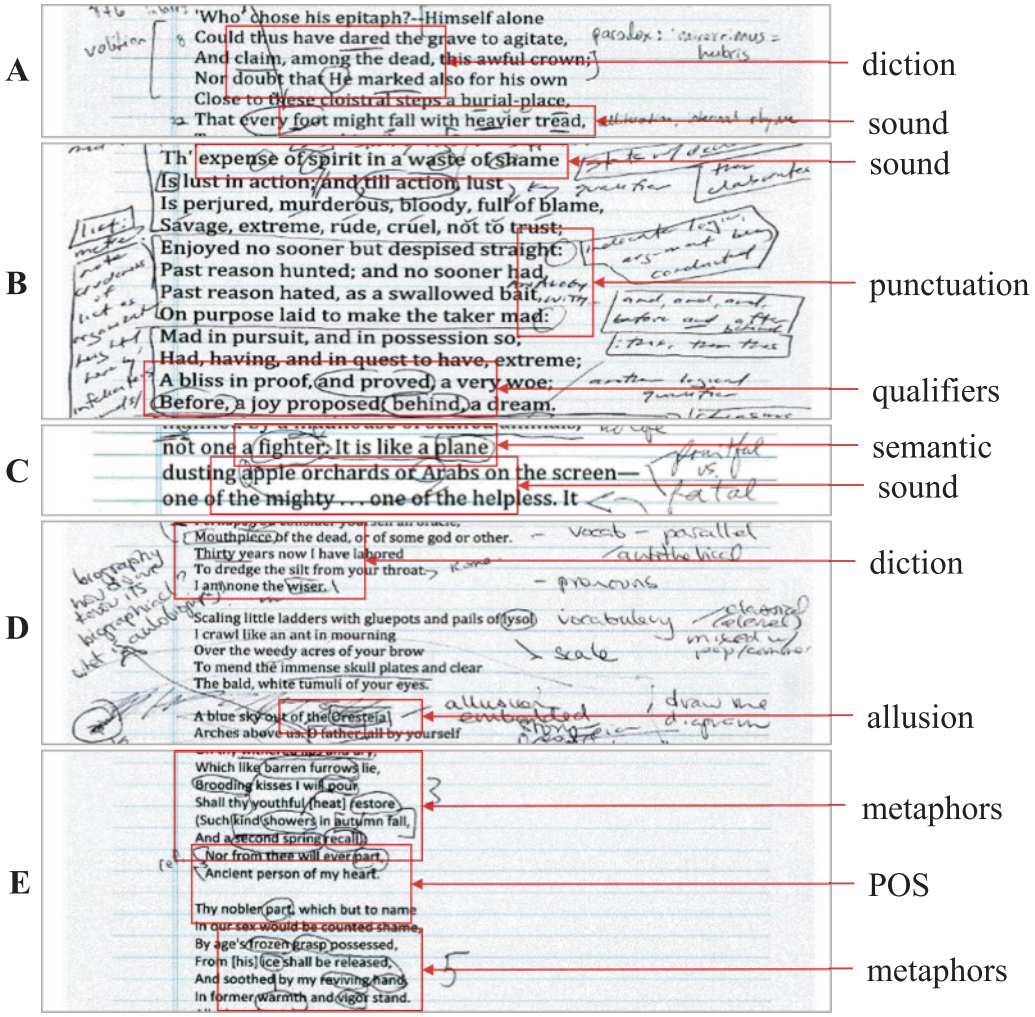

Fig. 3. Snippets of participants' final annotated sheets exhibiting annotation polymorphism. Annotation form and function do not hold a one-to-one relation as can be seen in the examples. Panel (a) shows an example of how underlines are being used by P14 for highlighting both words that are indicative of volition as well as words that exhibit alliteration and internal rhyme. In (b), P4 has overloaded the use of circles to indicate repetition of sound patterns, peculiarity of punctuation use, and words employed as qualifiers. Similarly, panel (c) shows how circles and connectors have been used to convey synonymy as well as phonetic relations between words in P1's work. Panel (d) is another example of how the meaning associated with an annotation form (square) changes from stanza to stanza for P9. In (e), P8 has identified both the change in part-of-speech of "part" and a set of metaphorical words through the use of the same annotation form, circle.

connectors have been used to convey synonymy as well as phonetic relations between words in P1's work. Figure 3(d) is another example of how the meaning associated with an annotation form (square) changes from stanza to stanza for P9. Last, in Figure 3(e), P8 has identified both the change in part-of-speech of "part" and a set of metaphorical words through the use of the same annotation form, circle.

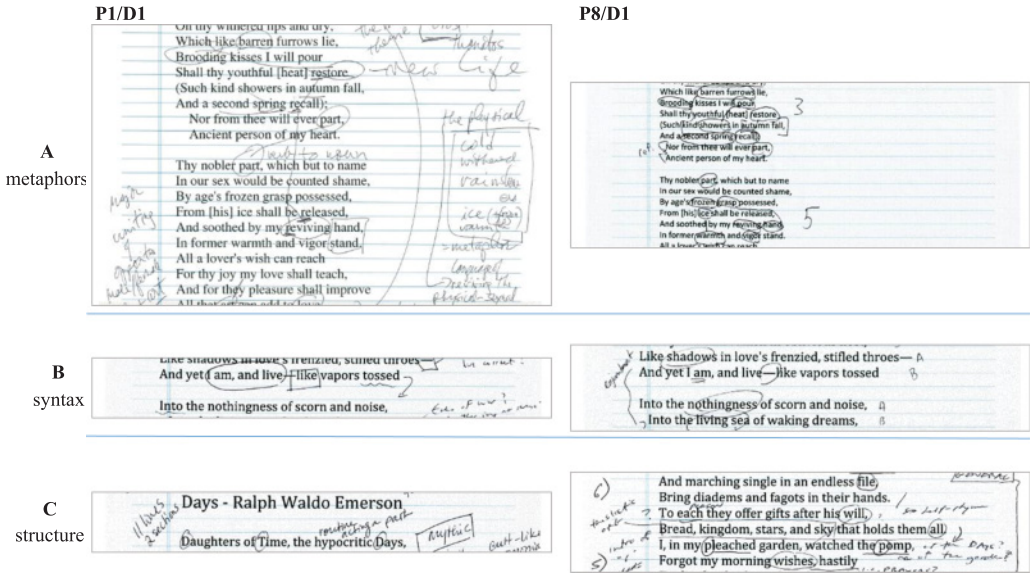

Fig. 4. Examples of consistency in the types of activities that participants analysing the same pairs of poems engage in. (a) Both P1 and P8 have identified the use of metaphors in their analysis of D1. (b) P6 and P13 both are indicating enjambment between stanzas of D12. Panel (c) shows notes on poem structure made by P4 and P11 for D8.

However, we noticed that all of our participants were engaging in similar activities, although the places on the page and the times in the process varied across participants. For example, both P1 and P8 have identified the use of metaphors in their analysis of D1 (see Figure 4(a)). P6 and P13 both indicating enjambment between stanzas of D12 as shown in Figure 4(b) is another example of inter-participant consistency. Similarly, Figure 4(c) shows notes on poem structure made by P4 and P11 for D8.

Annotation polymorphism is indicative of the constant shift in the meaning associated with the different annotation forms during the analysis and thus a need for avoiding constraints on the location or function of readers' annotations in a digital literary analysis tool. What this means for design is that the tool needs to be robust enough to work with incomplete information since the attribution of meanings to different annotation forms can greatly vary depending on a variety of factors (explained in more detail in the following section). Gesture-based systems that associate an annotation form to a single purpose are inappropriate and restrictive for literary critics due to this polymorphism since the community does not follow a standard set of notations.

4.3 Annotation Function Disambiguation and Use of Computation

We observed that our participants were actively grouping written thoughts and ideas into three main areas of the page: on the text itself (word space), in between the lines of text (white space), and in the margins. Looking into what function individual annotations served on each area of the page, a clear pattern began to emerge. Annotations that served as a means of computational offloading (CO) were observed to operate mainly in the word space and the white space whereas those that served as a means of externalizing to reduce memory load (EML) were noted to occupy mainly the margin space (see Figure 5).

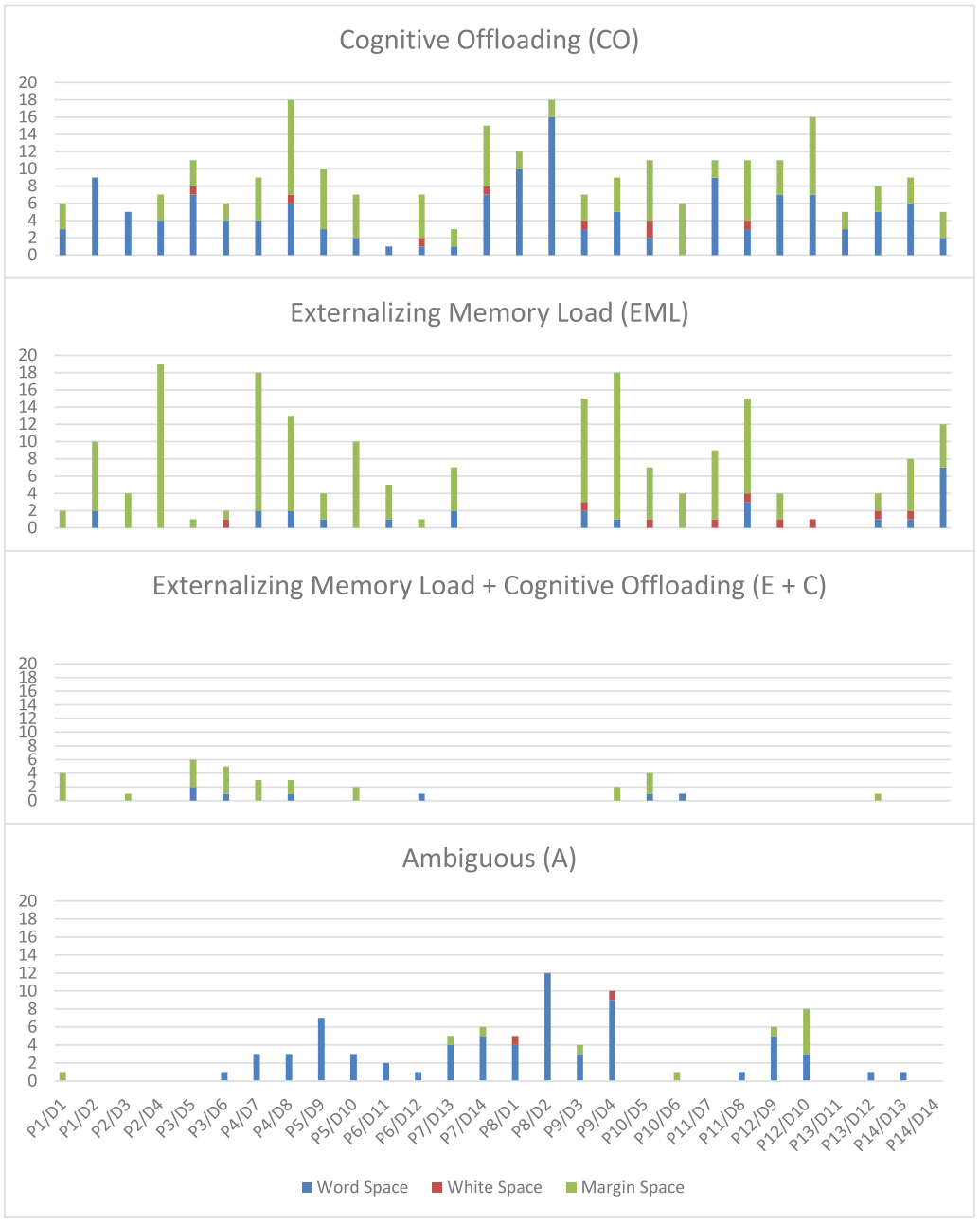

Fig. 5. Cognitive purpose codes by participant (P1–P14) and poem (D1–D14) and space (colors). Poems are aligned in columns.

Implicit connections of points of interest through the use of similar forms (frequently ellipses and underlines) in spatial and temporal proximity, in addition to explicit connection through the use of connectors, were prevalent with CO annotations. Figure 6 constitutes of examples of CO annotations made by the participants. In Figure 6(a), P1 has implicitly linked the repetitions of

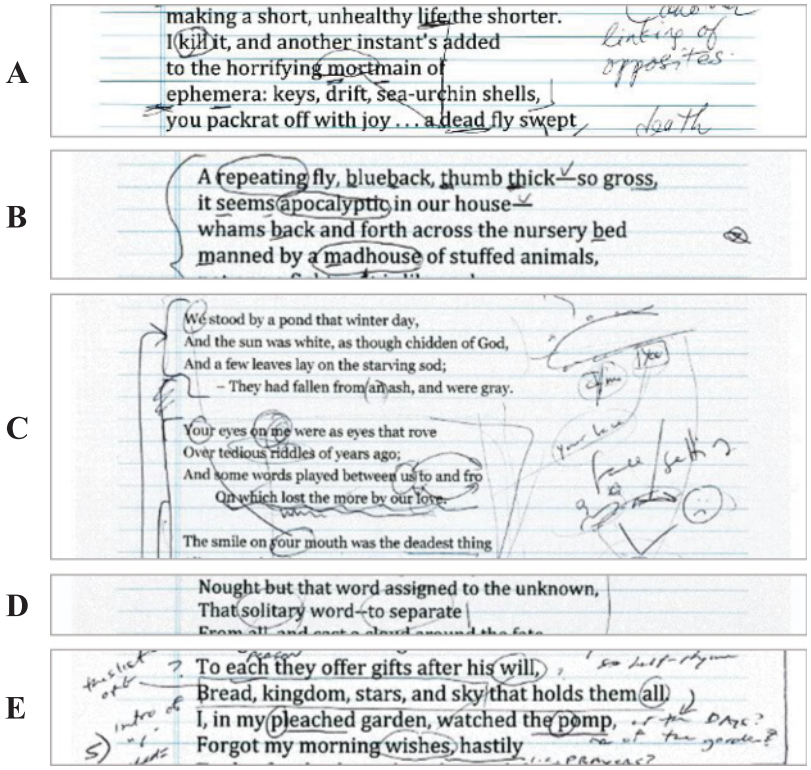

Fig. 6. CO annotation snippets of participants' final annotated sheets. As can be seen in the examples, CO annotations tend to occupy the word space and are characterized through the use of common forms in spatial and temporal proximity (implicit connection) as well as explicit connectors for linking words of interest.

"m" in "mortmain" and "ephemera" through the use of underlines, whereas synonymy between "kill" and "mort" has been implied via circles. He has also explicitly connected "mort" and "dead" and "dead" and "life" to point out the synonymy and antonymy relation respectively between the word pairs. Similarly, Figure 6(b) shows how the repetitions of "b," "th," "s," and "m," and those of "—" have been implicitly linked using underlines and check-marks respectively in addition to an emphasis on diction through the use of circles by P8. Figure 6(c) is a snippet from P7's sheet showing how he explicitly connected the occurrences of the pronouns "we," "your," and "me," whereas Figure 6(d) is an example of implicit connection of words of interest through the use of common forms in proximity from P7/D13. Similarly, Figure 6(e) presents an example of implicit linking of the sound in "pleached" and "pomp" and explicit connection of synonyms "will" and "wishes" in the work of P4.

EML annotations mainly comprised notes about the participants' interpretations of patterns and themes observed in the poem as well as reminders to look up supplementary information, such as word usage or alluded works. These were linked to the source text through the use of explicit connectors or through spatial proximity to relevant parts of the text. Notes providing general commentary about the poem usually occupied the space towards the bottom of the page. Figure 7(a), (c), and (d) shows examples of EML annotations that are notes about patterns observed in the poems D2, D8, and D4 by P1, P4, and P9 respectively whereas Figure 7(b) shows EML annotations commenting about the themes of the stanzas of D4 by P2. Notes by P11 and P1

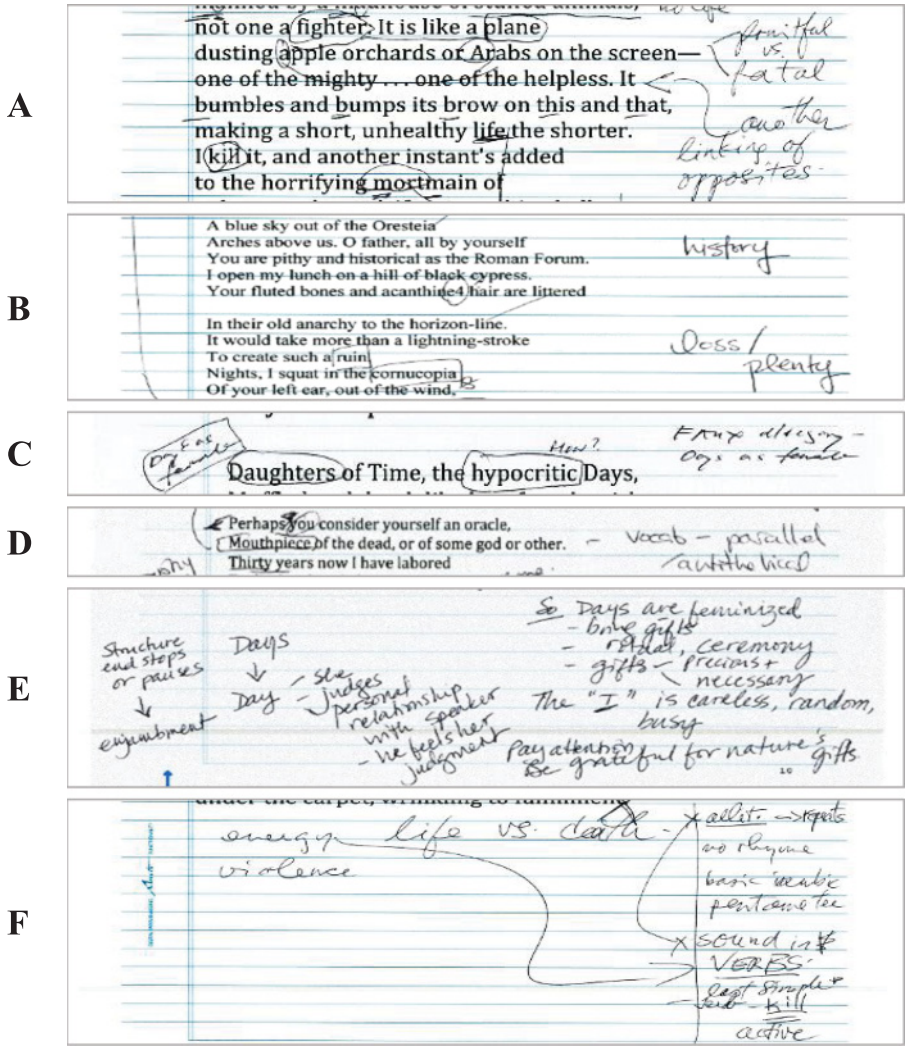

Fig. 7. EML annotation snippets of participants' final annotated sheets. As can be seen in the examples, EML annotations tend to occupy the margin space and generally comprise of notes regarding observed patterns of interest or general commentary about the poem.

synthesizing the poem content in D8 and D2, placed in the margins at the end of the poem text, are shown in Figure 7(e) and (f), respectively.

EML+CO annotations are characterized by the occurrence of an EML annotation, commenting upon the intermediate conclusion about an observed pattern, immediately following a CO annotation, identifying and working through a pattern of interest in the text, and a linking of the two types of annotations through the use of explicit connectors or spatial proximity. For example, Figure 8(a) shows how P1 identified the change in part-of-speech of the words "art" and "part" across stanza breaks and accompanied this with a note explaining that the two words have assumed noun form following their usage as verbs. Similarly, in Figure 8(b), P4 has marked the occurrence of chiasmus in "well knows" and "knows well" along with a note detailing the effect of

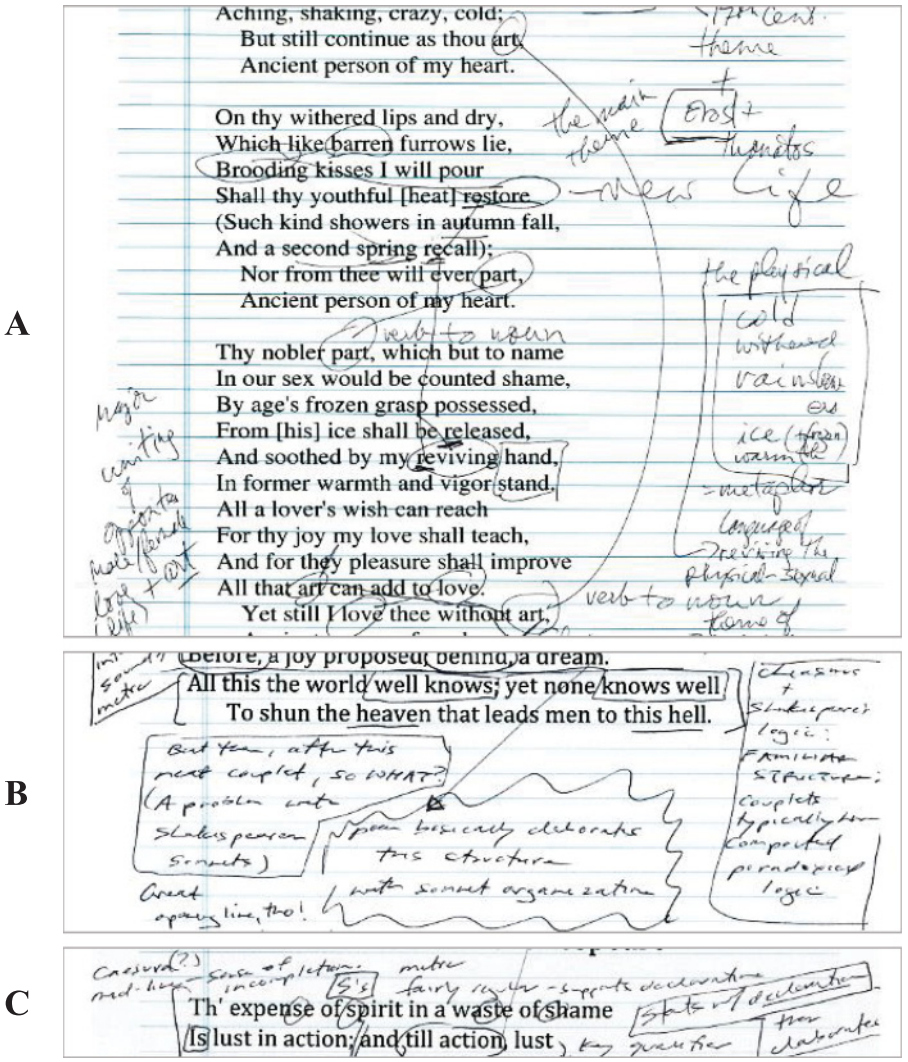

Fig. 8. EML+CO annotation snippets of participants' final annotated sheets. As can be seen in the examples, EML+CO annotations comprise of consequent occurrence of CO and EML annotations.

its usage in the last couplet. Figure 8(c) is another example of EML+CO annotations where the use of caesura has been highlighted in the word space and discussed in a note placed in the margin.

Both CO and EML annotations were consistently interleaved throughout an analysis session for all of our participants whereas EML+CO annotations were relatively rare.

Thus, even though annotations made by the participants are polymorphic in nature, the use of common annotation forms in spatial and/or temporal proximity often indicated related points of interest in the poem pertaining to the current thought process of the reader. That is, differences in space, time, and form of annotations are used to convey the differences associated with what the participants are marking and for what reason. Thus, the form of annotations, coupled with the spatial and temporal distance between them as well as the cognitive purpose they serve, could be

used to infer the type of computational support that could be of value in augmenting a reader's analysis process at a given moment in time during close reading. Determining a universal threshold for grouping strokes based on spatial and temporal proximity, however, is difficult due to inter-participant diversity; further research is warranted.

The dependence of the purpose of annotations on space is also indicative of the importance of maintaining the spatial and textual context of the poem and how the topology of a poem functions and restricts the ways in which the features of a poem operate to generate meaning. This need for maintaining the poem space has also been noted by poetry critics in previous works [29]. This enforces restrictions on manipulating the poem space as explored in LiquidText [48] and RichReview [55] for regular text documents.

CO annotations serve as a visible trace of a reader's contemplation on patterns of interest. Relations between the sets of words identified as points of interest by the reader through the use of CO annotations can be used for deducing the type of supplementary data to be generated. CO annotations can be identified through the space on the page that the reader is annotating in as well as through the presence of similar forms in spatial and/or temporal proximity. Similarly, EML annotations assimilate intermediate hypotheses based on patterns of interest found and a critic's existing literary knowledge. Processing this handwritten textual content could help provide more detailed insight into what the current hypothesis is instead of relying solely on relations between the words marked by CO annotations for inference. The inclusion of EML annotations in the inference process would particularly strengthen distant reading based exploration.

We believe that this form of user-initiated data generation would work well in the domain of literary criticism. A follow up with our participants suggests that they would be amenable to being informed of things that they might have overlooked or simply did not know. Inferring the features of interest based on annotations being made and the words being marked during a close reading could lead to ambiguous results. However, this ambiguity, instead of being a handicap, serves more as a strength with this domain in particular. It is the ambiguities in linguistic features that lead to multiple interpretations of a poem and make the spatial and temporal interactions interesting [29].

By basing suggestions on existing marks, we can find new connections similar to the patterns explicitly noted by the critics. Presenting these new connections in real-time allows for the consideration of alternative interpretations and helps augment or modify their subjective position. The use of a critic's annotations for inference also makes it possible to shift the focus of their analysis based on the interactions between the annotated parts of the text. If these close-reading-based annotations are then extended to distant-reading-based analyses of a corpus of interest, one could situate the findings/hypothesis in a broader literary context. This mimics the existing practice of how literary critics form an argument in relation to a poem. In addition, these extensions should be made available to the critic during and/or after a reading depending on the analysts approach. This supports reflection after the act but also supports play [29].

In addition, the use of reader annotations as a means of implicit interaction for requesting supplementary data could minimize the disengagement from the analysis process and interruption to the flow of reading that usually results from a switch to the process of retrieval of additional material from external resources.

The requirements of annotation may be a limitation of physical paper and pen, and the physical space available for EML, but nonetheless there is a pattern to cognitive sensemaking in terms of where it happens within the space given. Current tools created for this domain disrupt both time (in terms of interaction) and space (in terms of available annotation space) constricting the analyst's ability to think through the text. For example, a system like PoemViewer does an excellent job of presenting information about linguistic features of a given text. However, we have found that a necessary part of the process for our participants is uncovering this information themselves. If

all of the meta-information is presented first and annotation is not supported, then the workflow and sensemaking are not allowed to happen over time. Several of our participants describe the "experience" of the poem over time changing as they build up more and more connections to a single thought. This is not to say that existing tools do not provide opportunities for analysis, they simply create a different workflow than the one we have observed in our participants' close reading practices.

4.4 Annotations as a Reflection of Experience

Some of our participants stated that there were times when they were unsure as to why certain words of the poem seemed important to them, but these words were nevertheless marked and, at a later time, associated with other relevant points of interest once this understanding was solidified. We frequently observed that the expert readers would have some intuition and identify something as being important, before they knew how or why. For example, both P1 and P8 had initially annotated the adjective "repeating" (D2) for its peculiarity with later realization that the adjective was used by the poet as a literal indicator of the repeating sound patterns in the poem. These observations resonate with Schön's [42] description of the role intuition or previous knowledge play in sensemaking when observing reflective practices of professionals across varied fields.

Several participants would also make notes in the margins as placeholders for themes to return to. The simple act of writing "death?," or "rebirth?," within the margins acted as a reminder to check new findings or ideas with previous intuitions. Another example is marking expected structures within the texts. The addition of line numbers and rhyme schemes was prevalent within our participants and act as a visual shorthand to test expectations against. For example, if the expectation of a Shakespearean sonnet is that the poem has 14 lines and follows the rhyme scheme `abab cdcd efef gg`, then the poem follows expectations. Any deviation from these expectations can add a layer of meaning to the search. Deviations from expectations were another example of marginal notes that were used as "search terms" by our participants.

This behavior is important to consider when it comes to designing an interaction with a system, because it implies that analysts must be allowed the ability to think and interact with their own intuition. Current tools for poetics, which bulk process poems and present results all at once, do not allow for this thoughtful exploration. Slowing down the provision of analytic assistance, through just-in-time support integrates better with the work flow we observed.

4.5 Use of External Resources and Minimal Interruption

Three of our participants accessed external resources (both digital and physical documents) to look up metadata such as word definitions, usage histories, poetic forms, and terminology. In addition, information about the poet, such as era, their other works, and that of their contemporaries, was referenced during the analysis process. For example, both P4 and P11 annotated "pleached" (D8) and then looked up senses of the word in a physical dictionary at the end of their reading. Similarly, P6 referenced an on-line illustration of a spinning wheel to analyse the previously annotated words in D11 describing its various parts. One of our participants had a tablet with him throughout the analysis sessions for quick web look-up. It is also important to note that external resources were usually accessed at the end of a single reading of the poem, with the exception of those times when supplementary information had to be promptly retrieved in order to proceed further with the analysis process, e.g., when coming across archaic usage of words in the poem.

This behavior also indicates that the metadata must be available on demand and in relation to what has been annotated by the participants during the close reading process. Because the analysis

is time dependent, and may be started again once a given reading is completed, a robust system will allow for analyst-initiated metadata and the ability to access that data when needed in a non-intrusive manner.

5 DESIGN OF METATATION

Our observations of poetry critics' sensemaking process when performing close reading revealed the necessity of sustaining the experience of interacting with poetic language, critical to the process of interpretation, by providing support for free-form annotations (Section 4.1). We also noted the need to permit critics to reflect upon the content on their own and thus the necessity of slowing down the provision of assistance in the design of a literary analysis tool (Section 4.4). Based on these design guidelines, we developed a tool, Metatation, that couples a desktop-based literary support application with physical paper (augmented by the Anoto dot pattern) through the Anoto pen. The Anoto pen is a digital pen that records its position on a physical sheet of paper, augmented with the Anoto dot pattern, by processing digital snapshots of the pattern generated by an optical lens embedded beneath the pen-tip. Annotations made by a critic while reading initiate interaction with the system. The annotations are processed in real time, taking into account their space on the page, time and order of occurrence and form, to infer their function (Sections 4.2 and 4.3) and are treated as implicit interactions for requesting context-specific supplementary metadata. Specifically, Metatation interprets CO annotations to deduce reader intent and generates apt assistance based on this interpretation to better support the critic's current thought process. Identification of CO annotation units by Metatation is guided by findings of the observational study that annotations belonging to the same unit tend to be marked through the use of the same shape and are spatially and temporally close to each other. Thus, despite form-function ambiguity, Metatation can derive CO annotations units based on the properties of those annotations. Words marked by the annotations in a single CO unit are then analyzed by Metatation to deduce the literary devices being identified by the critic to generate suggestions. Our system is thus subtle in that critics are not required to create specific pen gestures, such as "rhyme" or "alliteration" annotations, to access analytical support. While poetry critics do not have specific goals when performing an analysis and look for a wide variety of linguistic features, such as metaphor, visual and phonetic imagery, and rhetoric, and their spatial and temporal interactions, we have focused only on generating supplementary semantic and phonetic data due to the limitations of existing natural language processing techniques at identifying higher level poetic devices.

Based on the observations of our study, we realize the need to minimize interruptions to the critic's reading process (Section 4.5) that could arise as a result of presentation of the generated supplementary reference materials. Video recordings of the analysis sessions from our study also indicated that critics, when performing close reading of a poem, tend to focus on the document on their desk with their head bent down. Metatation thus presents the analytic assistance on a peripheral display to minimize interruptions to the flow of reading, as the data is outside of the critic's field of view and the critic has the choice to decide when to look at the computational support provided. We considered alternatives where the generated data is projected onto the desk or on the document itself, but decided against them to avoid distraction from even simple visual cues in the reader's field of view.

In the following sections, we describe the design and implementation of our system in detail.

5.1 Example Use and System Description

Once a reader specifies a poem to work with, the desktop application generates a printable image file that is a composite of the Anoto dot pattern and the poem. As the reader annotates the poem using an Anoto pen, the pen captures and communicates the pen position on the paper in real

time, via Bluetooth, to our application running on a nearby computer. The application groups the received pen points into pen strokes and then clusters these pen strokes based on spatial and temporal proximity. Pen strokes are categorized as ellipses, underlines, or connectors by a geometric recognizer. Words associated with pen strokes are then extracted from the poem.

The query framework then processes the sets of annotated words, either implicitly grouped by pen strokes having a common form in a cluster or explicitly grouped through connector strokes, to identify possible semantic and phonetic relations between these words. On finding a relation, the query framework looks not only for other words in the poem that are similarly related to the annotated words, but also for all other sets of words in the poem that share the same relation. The query framework also extends this search to available relevant corpora to find all sets of words that exhibit these relations. The semantic relations considered are synonymy, antonymy, and word repetitions. The phonetic relations considered are alliteration, consonance, assonance, and rhyme. The query framework also fetches word definitions, usage history, usage examples, and etymology. Acts of annotation thus become *implicit queries*. For example, through circling two words that rhyme, our system detects the rhyme and propagates a rhyming query across the entire poem. The results generated by the query framework are then formatted into a structured presentation or visualization called a *query tile* and provided to the reader in a non-interruptive manner. The reader can then choose when to pause the process of annotation and reference the fetched metadata, which is provided on a nearby screen for easy glancing. The architecture of Metatation is illustrated in Figure 9.

5.2 Preprocessing

An initial calibration of the Anoto pen with respect to the display device being used is essential to ensure reception of accurate position information from the pen by the system. Extents of the Anoto dot patterns used by the system are manually recorded and stored to facilitate a mapping of the raw pen points captured by the pen to locations on a composite of a poem and the dot pattern. Raw text file of a poem to be analyzed using the system should be formatted such that the poem title and name of the poet are followed by the content of the poem stanzas with each of the stanzas separated by a newline. An input poem file specified by the user is processed by the system and the poem content is rendered over the Anoto dot pattern image file to generate a printable composite of the dot pattern and the poem, as shown in Figure 10. Software architectures for the design of pen-and-paper-based digital systems described in Print-n-link [35] and PADD [17] were useful in the design of the system components handling generation of augmented paper documents and the Anoto pen and document content calibration process.

The system also stores information about the relative position of a word in the poem (stanza, line, and word index), extents of the word's bounding box printed on the physical sheet of paper, the part-of-speech (POS) tag assigned to the word by the POS tagger from the Stanford NLP Toolkit [25], word pronunciations from the syllabified CMU phoneme dictionary [2], synonyms and antonyms from Merriam-Webster Thesaurus [13] and WordNet [30] and word etymology, senses, definitions, usage history, and usage examples from Merriam-Webster Dictionary [13] into a JSON file for quick access by the query framework, as shown in Figures 9 and 10.

5.3 Stroke Clustering and Shape Recognition

As a reader writes on physical paper augmented with the Anoto dot pattern, the Anoto pen generates TUIO[4] events that are received by the desktop client. We group pen points between consecutive pen down and pen up events as pen strokes. These pen strokes are then clustered using

[4]http://www.tuio.org.

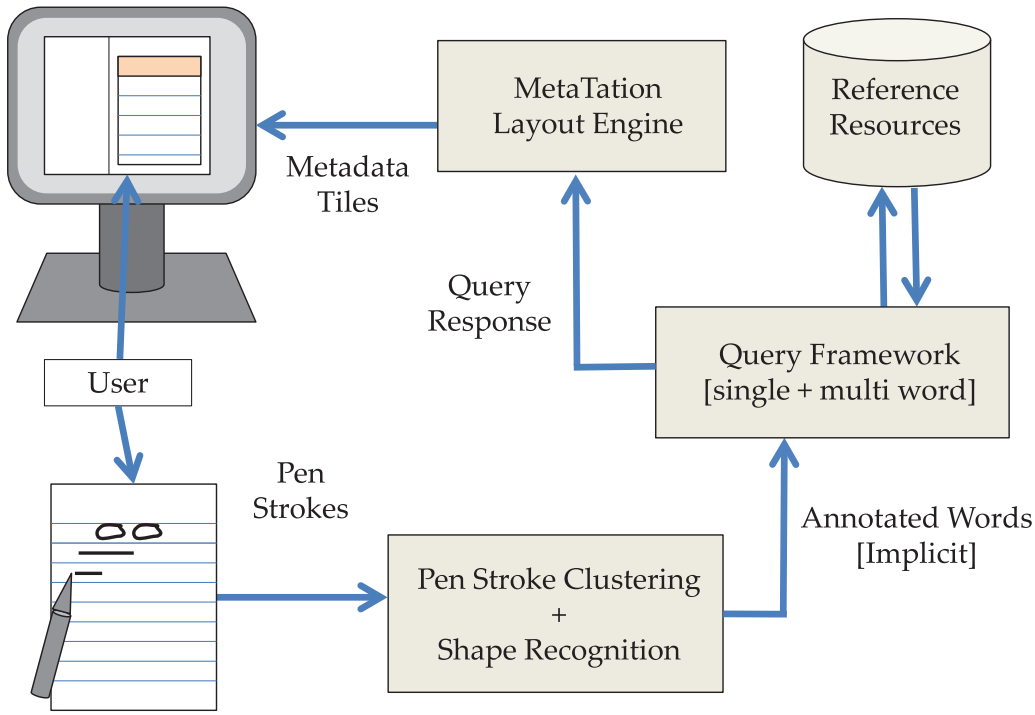

Fig. 9. Metatation system architecture. As a reader annotates a printed sheet of poem, the recorded pen strokes are clustered and categorized in real-time based on their space, time and form to generate "pen-stroke words." A query framework analyzes the pen-stroke words to find semantic and phonetic relations between them. On finding one or more of such relations, the query framework extends the search to similar relations of interest in the poem as well as in other corpora. The Metatation layout engine then presents these results returned by the query framework in the form of metadata query tiles on a peripheral display.

hierarchical agglomerative clustering as presented by Chiu and Wilcox [10]. We empirically observed that the single-linkage criterion outperformed both complete-linkage and mean-linkage criteria in terms of clustering accuracy when tested using participant data gathered during the observation study. Thus, we defined the linkage criterion that determines the distance between clusters as the minimum of the pairwise distances between the pen strokes in the clusters. The distance between two pen strokes is defined as a weighted function of the spatial and temporal distance between them. The temporal distance between two pen strokes s_1 and s_2, is defined as

$$time = \begin{cases} 1 \text{ if } t_{start}(s_2) - t_{end}(s_1) > 60 \text{ seconds} \\ (t_{start}(s_2) - t_{end}(s_1))/60 \text{ otherwise,} \end{cases}$$

where $t_{start}(s_2)$ is the start time of a pen stroke and $t_{end}(s_1)$ is the end time of another pen stroke, assuming that s_1 precedes s_2. The time metric has an empirically determined upper bound of 60 seconds to facilitate its normalization across all pen stroke pairs.

The spatial distance, *space*, between two pen strokes is defined as the minimum distance between the two bounding boxes for s_1 and s_2. The x and y values of bounding box locations are normalized by the extents of the dot pattern printed on the paper prior to their use in the spatial distance

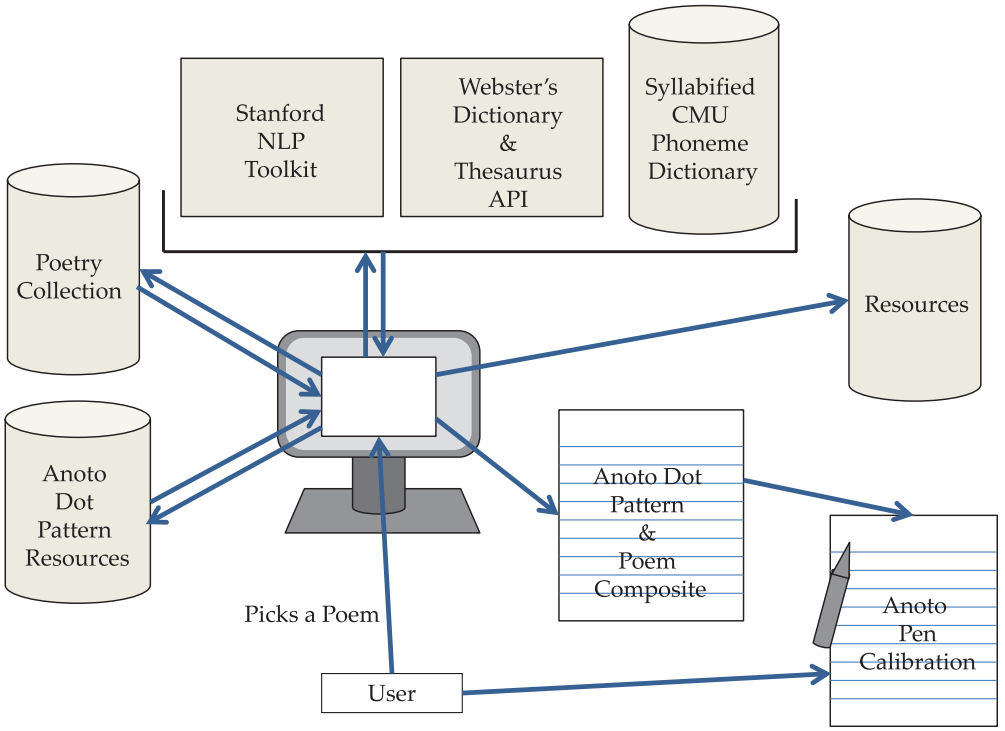

Fig. 10. Metatation preprocessing pipeline. Once a reader specifies a poem to work with from a given poetry collection, the system generates and prints a composite of the poem and one of the available Anoto dot patterns. In parallel, the system extracts metadata about the poem content from a variety of external resources and stores it in a format accessible by the query framework. If the Anoto pen to be used for annotating has not previously been calibrated with the peripheral display being used, a pen calibration step has to be performed prior to the start of the analysis session.

calculation. The spatiotemporal distance between two pen strokes is then given by

$$d = \sqrt{space^2 + time^2}.$$

On every iteration, the distance between clusters increases. The clustering process stops at an iteration i for which the ratio $r_{i+1} < r_i$ where r is given by

$$r = \frac{(d_{i+1} - d_i)^2}{d_i - d_{i-1}}$$

as proposed by Kara et al. [22].

Prior to stroke-type recognition, pen strokes for which more than 50% of the area of their bounding box falls under the margin space are eliminated from further processing since we decided to focus solely on CO annotations [T3] (Section 4.3). The shape recognizer of our prototype system has been designed to support categorization of pen strokes into a small set of predetermined geometric primitives, namely, ellipse, underline, and connectors, as these forms were the most common and frequent among our participants' CO annotations. The recognizer estimates the fit of the pen points of a pen stroke to an ideal version of the predetermined geometric primitives through geometric tests. We recognize ellipses by comparing them to an expected shape and pattern, underlines by

testing whether the stroke is primarily horizontal and fits between lines of the poem, and connectors by testing whether the stroke follows an arc pattern. Existing stroke recognizers such as $1 [52] and $P [50] can be used instead to support identification of a larger set of stroke types.

Following pen stroke clustering and shape recognition, words annotated by each of the pen strokes are extracted. Words annotated by pen strokes having a common form (ellipse or underline) in the same cluster are then grouped together as a single query to be issued to the query framework. Similarly, words annotated by pen strokes explicitly connected through the use of a connector are sent to the query framework as a single unit. This approach allows us to group together sets of close reading observations by the analyst with high confidence that the words are related to each other. We treat these groups as hypotheses generated by the user's interaction with the poem.

5.4 Query Framework

First, the framework retrieves the definitions, usage history, usage examples, and etymology for each of the words in an issued query and sends the results to be displayed. It then inspects the set of query words to identify the presence of semantic and phonetic relations between them. If the query framework identifies some relation between the annotated words, it searches for other words in the poem that share the same relation with the annotated words, as well as all other sets of words in the poem that also hold this relation.

Detection of semantic relations between query words is supported by the synonyms and antonyms of the words (for all senses) retrieved from the Merriam-Webster Thesaurus [13] and WordNet [30]. The syllabified CMU phoneme dictionary [2] is referenced for accessing phonemes of the query words (for all listed pronunciations) for an investigation of phonetic relations. To check for *perfect rhyme*, the framework first looks for the presence of common phonemes in the stressed syllables of the query words. On finding one or more such phonemes, the framework then checks to see whether all phonemes following the matched ones are identical in both diction and preceding words. *Alliteration* and *consonance* in the annotated words are detected through the presence of common consonant phonemes in the stressed syllables, whereas *Assonance* is identified by the presence of common vowel phonemes in the query words. The results generated by the query framework are then communicated to the desktop application for display.

5.5 Metatation Interface

The desktop application interface is comprised of two panels, namely the worksheet viewer panel and the metadata tile stream panel, as shown in Figure 1. Interface components are described in detail in the following sections.

5.5.1 Worksheet Viewer Panel. The worksheet viewer panel displays the poem being analyzed and conveys the availability of relevant metadata for reference. Before a reader starts annotating the physical sheet of paper, poem content in this panel is grayed out.

As the reader annotates, the system processes the pen strokes and highlights the annotated words creating "pen-stroke words." Upon hovering over any of these pen-stroke words, other words marked by pen strokes in the same cluster/annotation unit as the currently selected word are made apparent. As metadata generated by the query framework are received by the application, color-coded dots representing the different types of available query results appear below the respective pen-stroke words. Both pen-stroke words and query result dots can be selected to filter the metadata tile stream.

5.5.2 Metadata Tile Stream Panel. The metadata tile stream panel displays the query results as they are generated by the system in a scrollable view. Query results are displayed in interactive tiles that have a layout and visual design appropriate to the query type, as shown in Figure 11.

Fig. 11. Four representative Metatation tile types. Phonetic relation tiles, such as consonance, are interactive. Here, "hunted" has two consonant phonemes, as indicated by the green and pink dots. Hovering over a word reveals a tooltip of its pronunciations with the relevant consonant phonemes highlighted. Clicking on the phoneme dots highlights all words that share the same consonant phoneme. Word pairs that rhyme are highlighted using colors and lines. Synonyms are grouped into sets through the use of color with one-to-one relations being shown by lines. Word detail tiles are titled with the query word and detailed information is provided in a structured layout.

Tiles in the metadata tile stream panel are ordered by creation time to ensure that the tiles generated by the most recent annotations are visible to a critic at a glance without the need to interact with the system.

We wanted literary critics to be able to use Metatation during their analyses, so we decided to display the results in a view such that the tiles generated by the most recent strokes could be readily visible to the critic without the need for interacting with the system. However, to ensure that the original poem structure was maintained, the query tiles had to mimic the original layout of the text. This resulted in a tile that was much larger than expected, requiring the addition of a scroll bar so that the critic could easily navigate within and between tiles. We considered displaying the query results in tabs, with one tab per tile type as an alternative. The concern with that design, however, was that only one of the tile tabs could be set to active at any given point. This would force the critic to interact with the system in order to switch tabs limiting a more comprehensive view of all the tiles generated by the most recent set of strokes. Another alternative design that we had considered was that of a card-based layout, one per stroke cluster, comprising of tiny snippets of the tiles generated. However, with this design as well, interaction with the cards became essential undermining the original design constraint. In the end, we decided to implement a scrollable view for the metadata tile stream panel as it requires simple (e.g., mouse wheel) interaction to examine the stream.

Assonance, consonance and alliteration tiles: The query framework examines an annotation unit's pen-stroke words to detect the presence of assonance, consonance, and alliteration relation between word pairs. Detection of one such relation for any given word pair triggers the search for other word pairs in the poem that also exhibit that relation. For example, if a literary critic consecutively circles two words on the physical paper with the Anoto pen that share alliterative qualities such as the words "heaven" and "hell," an alliteration tile will be generated that shows other words within the poem that also demonstrate similar properties in their opening sounds. The consonance and assonance tiles operate in similar ways with their individual sound patterns and have identical visual and interaction design.

In these tile, each assonant, consonant, or alliterating phoneme is mapped to a unique color and its occurrence in a word presented through the use of colored dots below the word. A faint border highlights the pen-stroke words that triggered tile generation. Clicking one of the phoneme dots brings forth words that contain the selected phoneme and helps see the spatial distribution of the phoneme in the poem. Hovering over a word in these tiles reveals its phonetic transcription with the phonemes of interest highlighted.

Synonyms and antonyms tile: On detecting any of the pen-stroke word pairs as being synonyms or antonyms, the query framework searches for other occurrences of synonym or antonym pairs, respectively, in the poem.

A synonyms tile highlights synonymous word pairs in the poem by explicitly connecting the words in a pair. In addition, each pair of synonyms is assigned a unique color. To better reveal the presence and spatial distribution of synonymous word sets in the poem, word pairs comprising of identical words are mapped to a single color instead. An antonyms tile is identical to the synonyms tile in terms of its visual design but presents word pairs that are antonyms. Both tiles differentiate pen-stroke words that prompted tile generation through the addition of faint border in the words highlights in the tile. These tiles were designed with the intention of adding to the set of synonyms and antonyms identified and annotated by a critic.

Perfect rhyme tile: The query framework analyzes phonetic transcriptions of pairs of pen-stroke words to check whether they form a perfect rhyme. If any of the pen-stroke word pairs exhibit perfect rhyme, the query framework then searches for other instances of rhyming word pairs in the poem and presents discovered word pairs in a perfect rhyme tile.

The perfect rhyme tile is designed to convey how the spread of rhyming phonetic patterns in the poem. To this end, rhyming word pairs are explicitly connected through links in the perfect rhyme tile. Additionally, each rhyming word pair is mapped to a unique color. To visually indicate sets of rhyming words, if two pairs of rhyming words have a word in common, both are instead assigned a single color.

Although human analysts are very skilled at identifying rhymes, this tile becomes most valuable to the analyst when it is extended out into the distant reading scale. If a critic annotates two rhyming words that hold particular significance to their reading, the query framework will search for similar rhymes in the user-supplied distant reading corpus. This approach can aid specifically in detection of allusions from other texts, but also can aid the critic in uncovering new and interesting relationships between the text being studied and literary history.

Word detail tile: For every pen-stroke word, the query framework generates a word detail tile that presents the word's etymology, senses, definitions, usage history, and usage examples. To reduce the amount of information presented to the critic, the word detail tile was originally configured to display details pertaining solely to the word's POS in the poem. However, we later changed this to present details for all word senses irrespective of the POS observed in the poem guided by the observation that the intended meaning of the word and the POS in the poem do not always coincide in poetic expression.

The word detail tile was designed to permit a critic to quickly access the dictionary entry for an annotated word as several of our participants consulted this type of meta-information during and/or after their reading. A specific example of this occurred when P11 did not know the definition of the word "fagot," which occurs in the poem *Days* by Ralph Waldo Emerson (D8) in the line:

> Daughters of Time, the hypocritic Days,
> Muffled and dumb like barefoot dervishes,

And marching single in an endless file,
Bring diadems and fagots in their hands.

The participant wrote in the margin space of the page the question "wood?." In fact, the word means a bundle of sticks, twigs, or branches bound together and used as fuel. The word detail tile would have been able to confirm the critic's intuition without having to search through a physical dictionary.

6 DESIGN ITERATION

Following the development of our prototype system, Metatation, we conducted a cooperative design session for an initial investigation into the effectiveness of the tool in leveraging free-form annotations for augmenting the process of close reading for literary criticism. We recruited two of the literary scholars that had previously participated in our observational study for gathering qualitative feedback on the design of the Metatation system. Both of our participants have been teaching poetics for the past 5 (P6) and 20 (P9) years, respectively.

Each participant was given a description of how Metatation was created using the results of the study that they had previously participated in. They were then given a demonstration of how the system works followed by a session where each participant was allowed to explore the tool while discussing their experience as well as concerns out loud with the researcher as they performed a reading of an assigned poem. We decided to assign the poem Sonnet 129 by William Shakespeare (D7), from the poetry dataset created for the observational study (see the appendix), to both of our participants since the poem exhibits all the semantic and phonetic relations that our system is capable of identifying and providing support for. In addition, neither of the participants had previously analyzed this poem during the observational study.

After the participants had a chance to sufficiently interact with the various components of the system, they were requested to elaborate upon their general impressions about the tool in a brief interview guided by the following questions:

Q1. What are your general impressions of the tool?
Q2. Did the addition of technology interfere with your process in any way?
Q3. Could you envision using a tool such as this in your daily work? What are some of the problems that you think would crop up when using this tool?
Q4. Can you think of any other modules that could be helpful in supporting the kind of work that you do?
Q5. Do you have any other suggestions for improving the tool?

The cooperative design session revealed helpful insights about the current design and directions for future work.

6.1 Design Iteration: Observations

P6 had a very telling interaction with the tool as he chose to focus on and analyze the following lines in the sestet:

A bliss in proof, and proved, a very woe;
Before, a joy proposed; behind, a dream.
All this the world well knows; yet none knows well
To shun the heaven that leads to this hell.

The participant underlined the words "bliss," "dream," "heaven," and "hell" and referenced the metadata tiles generated by the tool. He noticed that Metatation had added "joy" to the set of {bliss,

heaven} in the synonyms tile and "woe" to the set of {bliss, joy} in the antonyms tile. The participant appeared positively surprised to see the additions that Metatation had provided. He realized that he had missed the connotation of the word "joy," highlighted by the tiles, which when taken together with the other annotated words lead the lines to convey a sense of parallelism. He was genuinely excited about the possibility of using a tool such as this in his daily work that would help him catch insights that he might possibly have overlooked. He added that while he would have generally performed a single pass of reading first followed by a look up of supplementary information, he would not mind concurrently using our system while analyzing and was receptive of the results altering or enhancing his thoughts.

This is exactly the type of interaction the system was intended for. Understanding how these domain experts experience a text, the addition of one extra connection to a set that has been initiated by the user can have far reaching consequences for sensemaking within the text. We have observed our participants conducting similar actions during the study and have noticed how the addition of a single term to a cluster of annotated connections multiplicatively affects all of the other connections in the poem related to the new set. In this way, even simple additions, like an extra synonym or antonym, can end up having a profound effect on interpretation and thus, this reaction by P6 was very crucial.

P6 commented that the ease of use of the interface was paramount at ensuring seamless integration of the tool into the readers work flow and added that he was concerned by the influence of constraining the types of marks he can use when annotating on his existing work practices. As noted in the results of the observational study, a variety of annotation forms were being employed by the readers when annotating while analysing poetry. The shape recognizer of our system, however, is currently able to process only a limited set of shapes, namely, ellipse, underline, and connectors. The reason for picking such a small subset was that these forms were not only consistently prevalent across all of our participants but were also simple enough to be reliably recognized by the system. We recognize that it is essential to further expand the geometric recognizer to be able to process other annotation forms so as to not constrain the readers and negatively influence their analysis process.

The other participant (P9) started by underlining the words "waste," "shame" and "blame" followed by marking the words "perjured," "murderous," and "bloody." She was particularly impressed by how well the visual design of the tiles, while being very simple, effortlessly revealed and made apparent the patterns of semantic and phonetic relations in the poem. P9's reaction was more focused on a particular use case as she was highly interested in the possibility of using our tool, in its current form, within a teaching context, "especially in a first year course, like intro to literary genres." This was an unintended, but exciting reaction. She went on to discuss how the visual groupings of word meanings would add greatly to the process of describing the complex functions of words when working within a poem and was very excited about the possibilities the tool held for teaching. She commented that the results generated by the tool seemed serendipitous in a sense since her own annotations were being used to guide metadata generation and added that "if this was meant to be a proof of concept, consider it proved." While P9 was very excited about the potential of the tool for teaching poetics, she wanted extensions to the metadata tiles in the research context and suggested that if the tool could be expanded to include things like comparisons across multiple works of a poet or to identify allusions to other authors and literary works, it would greatly improve the task of linguistic inquiry. She recommended an addition of tiles suggesting the different poetic forms exhibited by a poem. She was also interested in the possibility of conveying the multiplicity of senses for a rhyming word (that is, a combination of the word detail and perfect rhyme tiles), to help indicate the presence of a homonym rhyme.

When asked if the use of Metatation would interfere with their work flow, P6 said that it would "not interfere" and P9 "liked how you could simply keep the machine with the tool separate from the workspace and reference it after a single pass." P9 commented that having the system "present, pinging you like social media" would completely interfere with the analysis process but liked how our tool provided her with the ability to choose when to interact and thus well supported her current work practice of referencing meta-information after a single pass of reading. Further discussion in regards to these responses highlighted for us that there are two different modes of use that can be employed with Metatation. P6 wanted to play with the poem, looking up to see what the tool could tell him during the process, whereas P9 wanted to do her work and then reference the tool. Both use states are already possible within our design and the ability to engage within a user-defined time frame ensures that interruptions to work flow can be minimized.

7 EXTENSIONS TO METATATION

While Metatation provides a platform that allows the markup of poetry to be augmented in real time by connections within a poem for close reading, our participants expressed the desire to use our system for distant reading as well; that is, they wished to transition from the close reading of an individual poem to the distant reading of it within a historical context. This also mimics the current practices of hypothesis generation and verification for argumentation. The following use case takes the specific example, identified by P6, of the usage of "heaven" and "hell" in Shakespeare's sonnets as a starting point and expands upon the analysis based on our own domain expert's observations of how a literary critic would explore such an analysis. This use case also serves as a demonstration of how a literary critic could make use of the distant-reading-based extensions to Metatation, namely our Corpus Viewer tile.

7.1 Use Case: "Heaven" and "Hell" in Shakespeare's Sonnets

To give a better sense of how a close reading actually works in practice, we had our domain expert describe what a typical close reading "pass" on a text would be using Metatation. What follows is that reading:

Imagine a literary critic performing a reading of Shakespeare's Sonnet 129 (Figure 12) using Metatation by marking up words of interest on a copy printed on Anoto paper. He notices that the poem ends with the couplet:

> All this the world well knows; yet none knows well
> To shun the heaven that leads men to this hell.

He recognizes the importance of the words "heaven" and "hell," as they are both on the last line of the entire poem—his prior knowledge of sonnets, specifically from Shakespeare, meant he understood well that these last two lines often represent an answer to a question posed in the first 12 lines—and they are antonyms, so he circles these two words. Upon completing this reading, he turns to his computer to see that Metatation has also highlighted the words "bliss," "woe," and "joy" in the Antonym tile, as these words are related to the pair he highlighted, which he made note of as reinforcing the importance of the words he circled.

Based on his prior expectation that this pairing is likely to be an "answer" to a "question" posed earlier in the sonnet, he revisits the earlier part of the poem and notes that one of its themes is lust, and by extension sexuality. He surmises, therefore, that the final line "heaven that leads men to this hell" is a reference to sexual activity and how its pursuit is all-encompassing. While he feels he has gathered some evidence for this interpretation, the presence of this pairing in other sonnets allows him to explore whether there is further evidence.

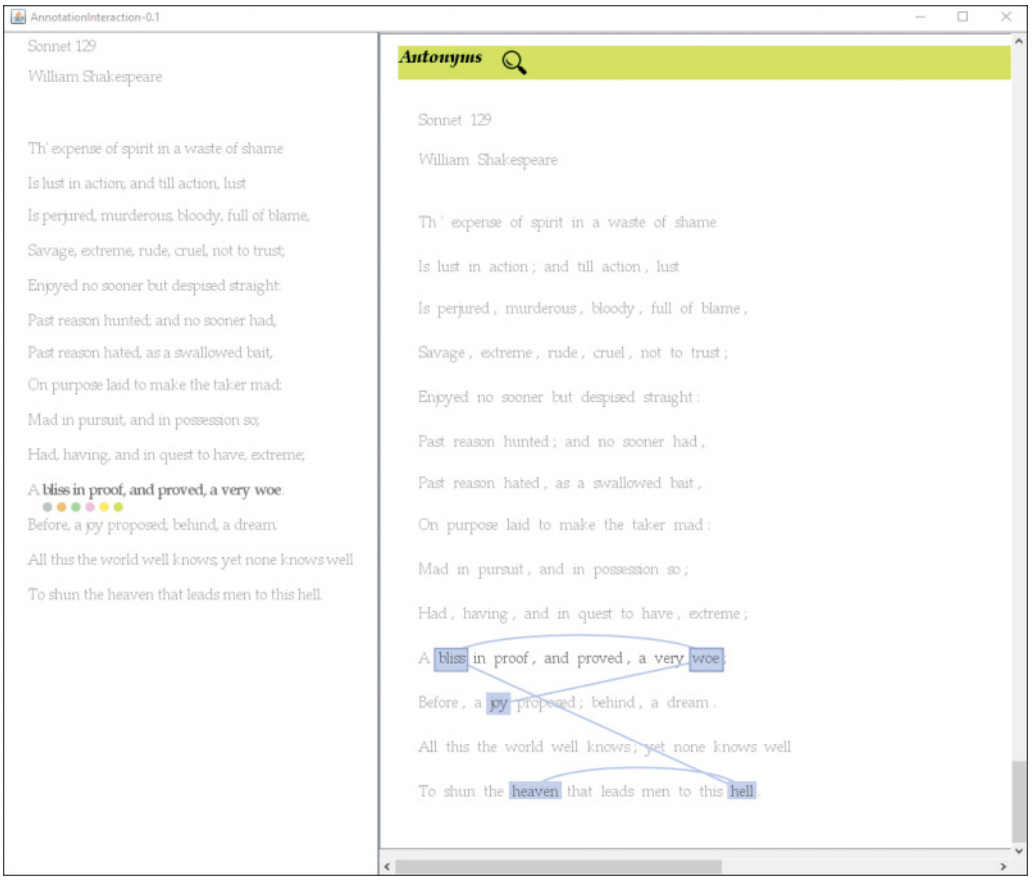

Fig. 12. Antonyms tile generated by Metatation's query framework when a reader underlines "bliss in proof, and proved, a very woe" on a printed sheet of the poem due to the identification of an antonyms relation between "bliss" and "woe."

He therefore chooses to do a distant reading by investigating what other of Shakespeare's sonnets contains this specific pairing by bringing up the Corpus Viewer tile containing all antonym pairs in all of Shakespeare's sonnets (Figure 13). In the antonym tile, he clicks on the link between "heaven" and "hell," which filters these sonnets to only those containing this pair (Figure 14). Through this distant reading, he is able to investigate another sonnet, Sonnet 145, which use the same "heaven"/"hell" pair (Figure 15). Upon opening this new sonnet, he makes note of the context that this same pair is used:

> 'I hate' she alter'd with an end,
> That follow'd it as gentle day
> Doth follow night, who like a fiend
> From heaven to hell is flown away;

This quote indicates that the poet is saved from "hell" by the love interest denying his sexual advances. The literary critic now has mounting evidence for the connection between this pairing and sexuality, due to the sexual interpretation of the antonym pair "heaven" and "hell" in Sonnet

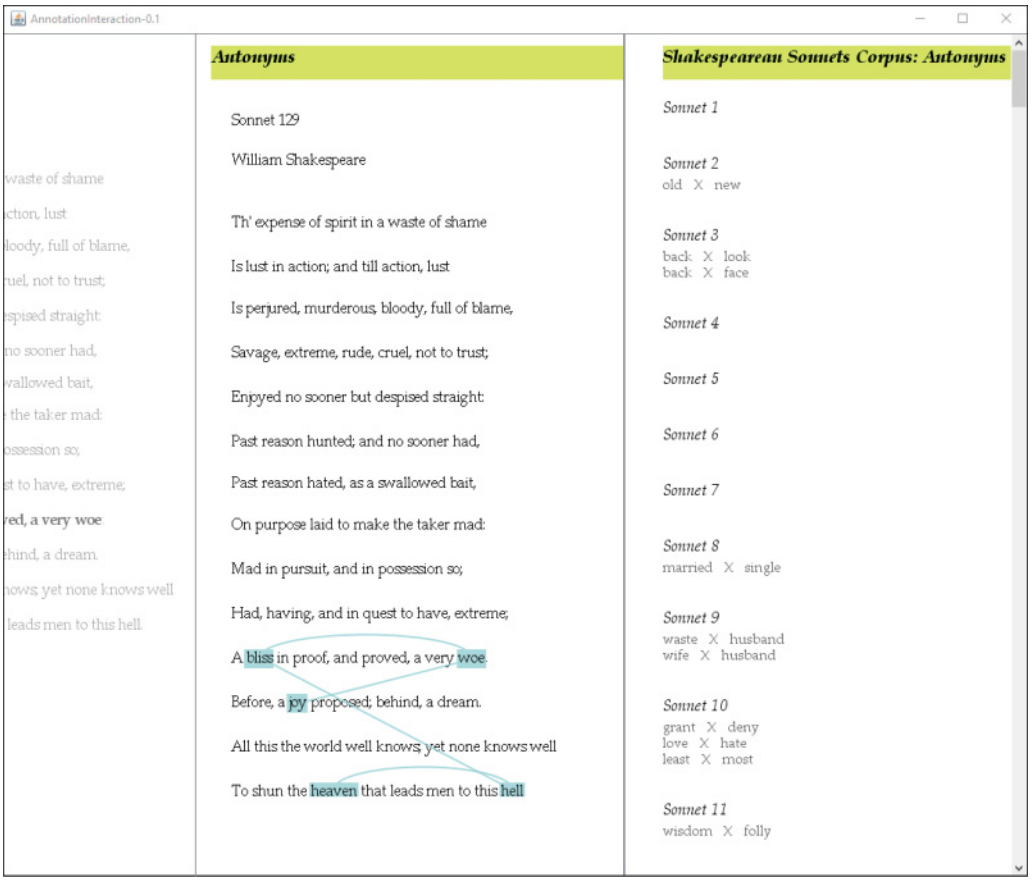

Fig. 13. Clicking on the search icon on the antonyms tile (see Figure 12) in the metadata tile stream panel, the reader can switch to distant reading-based analysis of the sonnet by accessing the Corpus Viewer tile that lists all antonym pairs across the Shakespearean sonnets corpus. Clicking on the peeking Worksheet Viewer panel on the left, hides the Corpus Viewer tile and switches the analysis mode back to close reading.

129, and the sexual denial in sonnet 145 that allows the poet to be "flown away" from the idea of heaven and hell (sex and its repercussions). The critic then begins to understand "heaven" to be the dream of the act, as in Sonnet 129, and that the poet is saved from the torture of love by being denied sexual intercourse.

7.2 System Support

As this use case highlights, it became evident after our cooperative design session that there would be value for the analyst to be able to extend their close readings queries to not only search for other words in the poem that exhibit similar relationships to those identified by the reader, but also to search for relationships of interest across user-supplied corpora.

As a proof-of-concept, the Antonyms query tile was then updated to include information on how the critic's observations in a single poem relate to texts in a large corpus of English poetry (here, the Shakespearean sonnets corpus). It is this ability to take annotations made on a single poem (close reading) and extend them out to larger corpora (distant reading) that bridges these

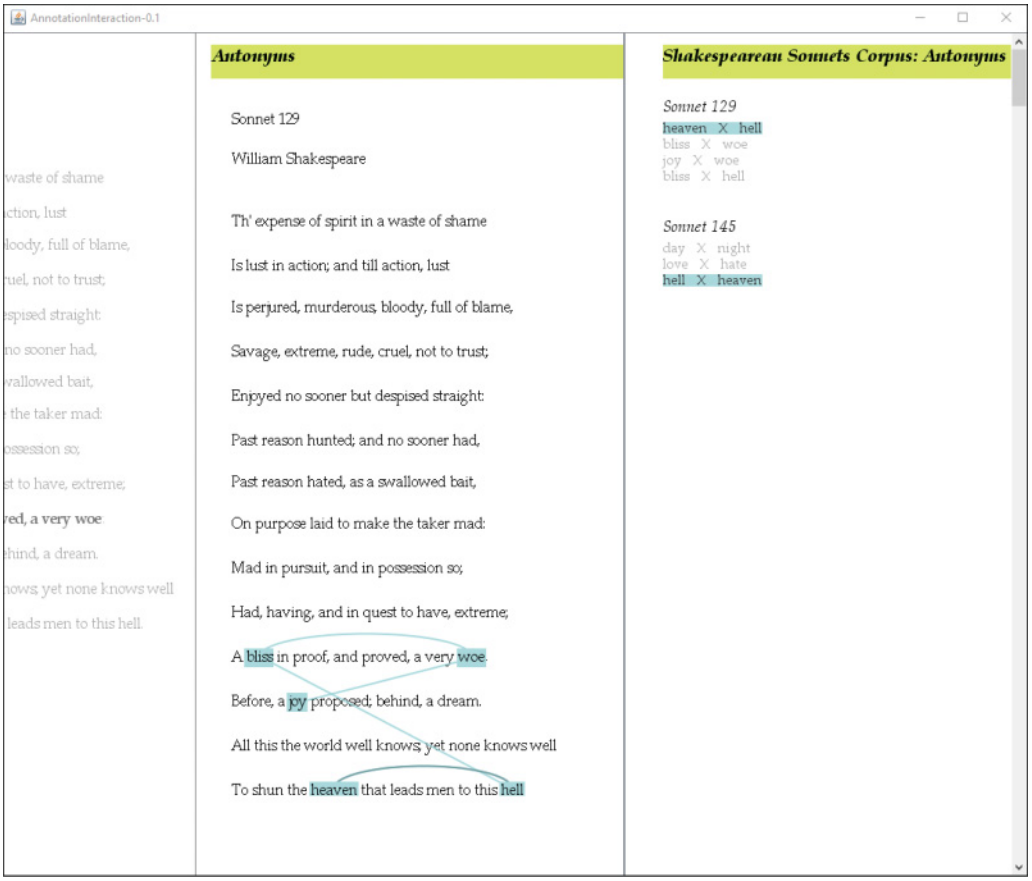

Fig. 14. Clicking on the link between "heaven" and "hell" on the antonyms tile filters the list of all antonym pairs across the Shakespearean sonnets corpus in the Corpus Viewer tile to only show those sonnets that contain this particular antonym pair.

common approaches to literary criticism. The system leverages computational power to identify relevant connections of possible interest in large corpora that would otherwise be hard to track manually. For example, if a participant annotates two antonyms in a poem but misses a third, the system will not only highlight the third antonym in that poem, but also trigger a search for pairs of antonyms in the provided corpora. Thus, annotations made while performing close reading could be used to spur distant-reading-based exploration of relevant text corpora, and both approaches to literary criticism could be seamlessly supported. This approach challenges the current trend in distant reading, which is to generate automated pattern searches within large corpora using topic models [3–7, 18, 19, 28, 40]. Metatation inverts that process, relying on the close reading work of an analyst and generates distant reading queries by extending those local questions to larger sets of texts.

In our use case (Section 7.1), the distant-reading-based Corpus Viewer tile allowed for the extension of close reading techniques out into larger corpora. While a critic could (and likely would) read all 154 Sonnets in a day or two, Metatation allows them to load a corpus of all English poetry and perform this analysis in seconds as they perform a close reading. Once a line of reasoning is

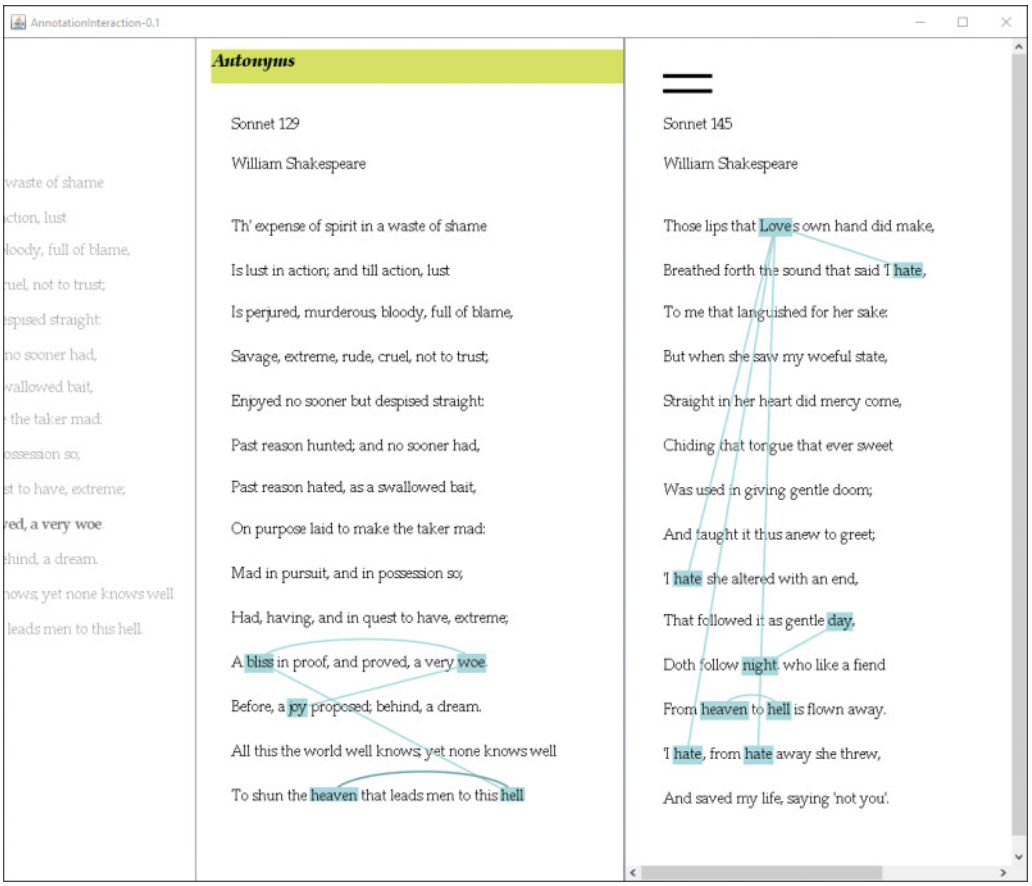

Fig. 15. Clicking on the sonnet title in the Corpus Viewer tile, the reader can investigate a specific sonnet of interest further, supporting comparisons with the original sonnet. Clicking on the icon at the top of the newly opened sonnet, permits the reader to switch back to the full list of sonnets.

followed through and noted, the analyst then has the ability to zoom in back to the original poem and continue the reading. This example is but one of possibly hundreds of threads that remain to be pulled for a full close reading to be executed, but the ability to zoom out to larger corpora and zoom back in to the original work dramatically aids the sensemaking process without replacing the cognitive requirements of the practice with computational processing.

8 FUTURE WORK

In the following sections, we discuss the avenues for future work that we want to explore in relation to the work presented here.

8.1 Metadata Query Tiles

There is a great potential for the development of new meta-information tiles for supporting research-based analysis of poetry. While we have developed the distant-reading-based analysis support for only the antonyms tile in the current prototype, other query tiles could be similarly extended to enable comparisons of semantic as well as syntactic context of usage of annotated

words across different corpora such as, all works of a poet, works of other contemporaries of the poet, and poems of the same genre across literary time periods. Metadata tiles hinting at the possible poetic forms exhibited by a poem based on the rhyming patterns and structure observed in the poem as suggested by one of our participants (P6) would also be a useful addition to the system. Identification of alluded works or poets in a poem has also been recommended as one of the possible additions for supporting research-based analysis (P9). Since the current prototype allows exploration of only basic sonic and semantic ambiguities in the poem, computational augmentation for the varied points of focus/entry during a literary analysis is fairly limited and could lead to user frustration as the system repeatedly offers similar types of meta-information. We are thus interested in expanding the limited set of query types with support for identification of more complex rhetorical devices such as metaphor and irony. Relating the themes of a poem to major historical events during the time of writing the poem to better situate the work in the context of sociocultural effects on literary development is another interesting aspect to explore. To address the repetitiveness of system-generated recommendations, a machine learning model could be introduced in the system that would allow a critic to teach the system about poetic devices it is currently unable to process. This, however, comes with its own challenges in the form of the critic being interrupted from his analysis process to teach the system what poetic device he has identified and how the device functions. Thus, further investigation is needed to determine how to balance user interaction with the system and interruptions to the reading process caused by these interactions.

8.2 Augmenting Different Types of Annotations

We made a decision to augment only CO annotations with digital metadata for our proof-of-concept prototype. The reason for not doing so with EML annotations was that they comprised mainly of handwritten notes and while a lot of current handwriting recognition systems provide fairly accurate digitization of the input text, further exploration of how this content could be used to infer the type of relevant metadata and guide the distant reading process is warranted. Augmenting both CO and EML annotations would help better integrate the tool with the existing analysis process of literary critics by allowing the system to be more robust in terms of the range of computational support it provides. An exploration of the possibility of taking into account the semantic context of the annotated words in addition to the spatial and temporal distance between the respective annotations and their forms could also help refine the metadata being generated by the system. We have also considered the possibility of overloading idiosyncratic reading habits, such as tapping on the desk for counting syllables or when figuring out the stress and sound patterns in a line, as implicit interactions for requesting analytic support and would like to explore the efficiency and usability of these interactions.

8.3 System Evaluation and Further Studies

We would also like to further evaluate our system through a longitudinal study to identify how the use of annotations, made while reading, as implicit interactions for requesting meta-information influences their annotation practices and to investigate how well the system integrates with their existing work flow. Since digital slates [9] have shown promise as a substitute for paper-based textual analysis in the context of graduate student use, we are also interested in exploring the possibility of substituting pen-and-paper with stylus-enabled digital tools to further investigate how crucial a role paper plays in the close reading process of literary critics and to determine whether and how the annotation process gets affected by this change in medium. The transition from Anoto pen and physical paper to stylus-enabled digital tools would not pose any challenges for our system since the system has been developed to function with TUIO events.

8.4 Generalization to Other Domains

While our system was informed by the specific domain of poetry analysis, other domains could easily benefit from its use, including the analysis of legal documents to connect cases and provide background on involved parties, the analysis of patient data in health care, and in newer forms of pedagogical practice, such as computational rhetoric, where the focus is on exploring artistic writing through computer-supported exploration. Metatation employs annotations made while reading a poem as implicit interactions to infer a critic's task requirements in real-time. The system's shape recognition and stroke clustering modules in tandem with an analysis of the literary relations between underlying words permit this inference. While the design of these modules in Metatation is guided by our observations of literary critics' annotation practices, modularity of our software architecture allows for substitution by domain-specific stroke clustering, shape recognition, and querying modules per target user's practices and needs. The output of these modules that identify user interest can then be fed into domain-specific analytical/learning models to generate relevant metadata/recommendations. Thus, the software architecture can be used as a reference when designing a system that uses annotations as implicit interactions for inferring user needs in real-time. Additionally, Metatation's design also serves as an example of a low-impact system that minimizes interruptions to users' existing workflow, permitting them to reference system-generated metadata if and when needed. Separation of the user's workspace (paper) and that of the system's (display) ensures that system recommendations do not interfere with a user's current task and permits collaboration between the user and the system to be initiated and controlled by the user.

9 CONCLUSION

There has been a tremendous growth in the field of digital humanities in recent years in the context of developing computational tools that support distant reading. However, the practice of close reading is still relevant as a method of understanding how artful language works together to create meaning. In this work, we explored the idea of overloading the act of annotation to connect close and distant reading.

To this end, we presented the results of an observational study probing into the annotation strategies employed by literary critics as they analyze poetry as close readings. Based on the results of the study, we derived guidelines for the design and development of tools for supporting literary inquiry and analysis, and also for connecting those analyses into the domain of distant reading. We also reported on the design and implementation of Metatation, a free-form annotation-based literary analysis tool that generates and presents just-in-time, context-specific metadata, as a reader annotates a poem printed on a physical sheet of paper, in a non-interruptive manner. We then presented a discussion with domain experts that lead to an iteration in the design of our tool. We concluded with a description of the directions for future development of our work.

Our tool, Metatation, has the potential to augment existing work practices in the domain of literary criticism while connecting close reading to larger corpora and literary history. In particular, it provides the advantage of supporting an already familiar method of annotating text with a physical pen, and presenting meta-information in a separate space to aid in the sensemaking process unique to this domain.

APPENDIX

POETRY DATASET FOR OBSERVATIONAL STUDY

Table A1. Poems Assigned Per Participant for the Observational Study

Participant	Poem ID	Title	Poet
P1	D1	A Song of a Young Lady to Her Ancient Lover	John Wilmot
	D2	Harriet	Robert Lowell
P2	D3	Adlestrop	Edward Thomas
	D4	Colossus	Sylvia Plath
P3	D5	Ode, Written in the Beginning of the Year 1746	William Collins
	D6	To the Virgins to Make Much of Time	Robert Herrick
P4	D7	Sonnet 129	William Shakespeare
	D8	Days	Ralph Waldo Emerson
P5	D9	The Painter	John Ashberry
	D10	Spring	Edna St. Vincent Millay
P6	D11	Huswifery	Edward Taylor
	D12	I Am	John Clare
P7	D13	A Gravestone Upon the Floor in the Cloisters of Worcester Cathedral	William Wordsworth
	D14	Neutral Tones	Thomas Hardy
P8	D2	Harriet	Robert Lowell
	D1	A Song of a Young Lady to Her Ancient Lover	John Wilmot
P9	D4	Colossus	Sylvia Plath
	D3	Adlestrop	Edward Thomas
P10	D6	To the Virgins to Make Much of Time	Robert Herrick
	D5	Ode, Written in the Beginning of the Year 1746	William Collins
P11	D8	Days	Ralph Waldo Emerson
	D7	Sonnet 129	William Shakespeare
P12	D10	Spring	Edna St. Vincent Millay
	D9	The Painter	John Ashberry
P13	D12	I Am	John Clare
	D11	Huswifery	Edward Taylor
P14	D14	Neutral Tones	Thomas Hardy
	D13	A Gravestone Upon the Floor in the Cloisters of Worcester Cathedral	William Wordsworth

REFERENCES

[1] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. 2013. Rule-based visual mappings – With a case study on poetry visualization. *Comput. Graph. Forum* 32, 3 (2013), 381–390.

[2] Susan Bartlett, Grzegorz Kondrak, and Colin Cherry. 2009. On the syllabification of phonemes. In *Proceedings of Human Language Technologies*. Association for Computational Linguistics, Boulder, CO, 308–316.

[3] David M. Blei. 2012. Probabilistic topic models. *Commun. ACM* 55, 4 (April 2012), 77–84. DOI: http://dx.doi.org/10.1145/2133806.2133826

[4] David M. Blei and John D. Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning (ICML'06)*. ACM, New York, NY, 113–120. DOI: http://dx.doi.org/10.1145/1143844.1143859

[5] Jordan L. Boyd-Graber and David M. Blei. 2009. Syntactic topic models. In *Advances in Neural Information Processing Systems: Volume 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.). Curran Associates, Inc., Red Hook, NY, 185–192.

[6] Jordan L. Boyd-Graber, David M. Blei, and Xiaojin Zhu. 2007. A Topic Model for Word Sense Disambiguation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

[7] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L. Boyd-graber, and David M. Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems: Volume 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Eds.). Curran Associates, Inc., Red Hook, NY, 288–296.

[8] M. Chaturvedi, G. Gannod, L. Mandell, H. Armstrong, and E. Hodgson. 2012. Myopia: A Visualization Tool in Support of Close Reading. In *Proceedings of the Conference on Digital Humanities (DH'12)*.

[9] Nicholas Chen, François Guimbretière, and Abigail Sellen. 2013. Graduate student use of a multi-slate reading system. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, 1799–1808.

[10] Patrick Chiu and Lynn Wilcox. 1998. A dynamic grouping technique for ink and audio notes. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology (UIST'98)*. ACM, New York, NY, 195–202. DOI: http://dx.doi.org/10.1145/288392.288605

[11] Tanya Clement, Loretta Auvil, David Tcheng, Boris Capitanu, Megan Monroe, and Ankita Goel. 2012. Sounding for meaning: Analyzing aural patterns across large digital collections. In *Proceedings of the Conference on Digital Humanities (DH'12)*.

[12] Christopher Collins, Sheelagh Carpendale, and Gerald Penn. 2009. Docuburst: Visualizing document content using language structure. *Comput. Graph. Forum* 28, 3 (2009), 1039–1046.

[13] Merriam-Webster Inc. 2006. *The Merriam-Webster Dictionary*. Merriam-Webster, Inc., Springfield, MA.

[14] Alan Dix. 2008. Externalisation – How writing changes thinking. *Interfaces* 76 (2008), 18–19.

[15] Jean-Daniel Fekete and Nicole Dufournaud. 2000. Compus: Visualization and analysis of structured documents for understanding social life in the 16th century. In *Proceedings of the ACM Conference on Digital Libraries*. ACM, San Antonio, TX, 47–55.

[16] Margaret Ferguson, Mary Jo Salter, and Jon Stallworthy. 2005. *The Norton Anthology of Poetry*. WW Norton & Company, New York, NY.

[17] François Guimbretière. 2003. Paper augmented digital documents. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*. ACM, 51–60.

[18] Gregor Heinrich. 2009. A generic approach to topic models. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2009, Bled, Slovenia, September 7–11, 2009, Proceedings, Part I*. Springer, Berlin, Heidelberg, 517–532. DOI: http://dx.doi.org/10.1007/978-3-642-04180-8_51

[19] Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99)*. ACM, New York, NY, 50–57. DOI: http://dx.doi.org/10.1145/312624.312649

[20] Matthew Jockers. 2013. *Macroanalysis, Digital Methods and Literary History*. University of Illinois Press, Champaign, IL.

[21] Stefan Jnicke, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. 2015. On close and distant reading in digital humanities: A survey and future challenges. In *Proceedings of the Eurographics Conference on Visualization (Euro Vis) – STARs*. DOI: http://dx.doi.org/10.2312/eurovisstar.20151113

[22] Levent Burak Kara, Leslie Gennari, and Thomas F. Stahovich. 2008. A sketch-based tool for analyzing vibratory mechanical systems. *J. Mech. Des.* 130, 10 (2008), Article 101101, 11 pages.

[23] Chunyuan Liao, François Guimbretière, Ken Hinckley, and Jim Hollan. 2008. Papiercraft: A gesture-based command system for interactive paper. *ACM Trans. Comput.-Hum. Interact.* 14, 4 (2008), 18.

[24] Wendy E. Mackay, Guillaume Pothier, Catherine Letondal, Kaare Bøegh, and Hans Erik Sørensen. 2002. The missing link: Augmenting biology laboratory notebooks. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM, 41–50.

[25] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

[26] Catherine C. Marshall. 1997. Annotation: From paper books to the digital library. In *Proceedings of the 2nd ACM International Conference on Digital Libraries (DL'97)*. ACM, New York, NY, 131–140. DOI: http://dx.doi.org/10.1145/263690.263806

[27] Catherine C. Marshall and A. J. Bernheim Brush. 2004. Exploring the relationship between personal and public annotations. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'04)*. ACM, New York, NY, 349–357. DOI: http://dx.doi.org/10.1145/996350.996432

[28] Jon D. Mcauliffe and David M. Blei. 2008. Supervised topic models. In *Advances in Neural Information Processing Systems: Volume 20*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis (Eds.). Curran Associates, Inc., Red Hook, NY, 121–128. http://papers.nips.cc/paper/3328-supervised-topic-models.pdf.

[29] N. McCurdy, J. Lein, K. Coles, and M. Meyer. 2016. Poemage: Visualizing the sonic topology of a poem. *IEEE Trans. Visual. Comput. Graph.* 22, 1 (January 2016), 439–448. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467811
[30] George A. Miller. 1995. WordNet: A lexical database for English. *Commun. ACM* 38, 11 (1995), 39–41.
[31] Franco Moretti. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, Brooklyn, NY.
[32] Meredith Ringel Morris, A. J. Bernheim Brush, and Brian R. Meyers. 2007. Reading revisited: Evaluating the usability of digital display surfaces for active reading tasks. In *Proceedings of the IEEE Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*. ACM, 79–86.
[33] Aditi Muralidharan, Marti A. Hearst, and Christopher Fan. 2013. WordSeer: A knowledge synthesis environment for textual data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management (CIKM'13)*.
[34] Donald A. Norman. 1994. *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*. Basic Books, New York, NY.
[35] Moira C. Norrie, Beat Signer, and Nadir Weibel. 2006. Print-n-link: Weaving the paper web. In *Proceedings of the ACM Sympsium on Document Engineering*. ACM.
[36] Kenton O'Hara and Abigail Sellen. 1997. A comparison of reading paper and on-line documents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'97)*. ACM, 335–342.
[37] Ilia A. Ovsiannikov, Michael A. Arbib, and Thomas H. McNeill. 1999. Annotation technology. *Int. J. Hum.-Comput. Stud.* 50, 4 (1999), 329–362.
[38] W. Bradford Paley. 2002. TextArc: Showing word frequency and distribution in text. In *Poster presented at the IEEE Symposium on Information Visualization (INFOVIS'02)*.
[39] Jenny Preece, Helen Sharp, and Yvonne Rogers. 2015. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley and Sons, London.
[40] Radim Řehůřek and Petr Sojka. 2010. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.
[41] Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price. 1998. Beyond paper: Supporting active reading with free form digital ink annotations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'98)*.
[42] Donald A. Schön. 1983. The reflective practitioner: How professionals think in action. Basic Books, New York, NY.
[43] Frank Shipman, Morgan Price, Catherine C. Marshall, and Gene Golovchinsky. 2003. Identifying useful passages in documents based on annotation patterns. In *Research and Advanced Technology for Digital Libraries*. Springer, New York, NY, 101–112.
[44] Beat Signer and Moira C. Norrie. 2007. PaperPoint: A paper-based presentation and interactive paper prototyping tool. In *Proceedings of the ACM Conference on Tangible and Embedded Interaction (TEI'07)*. ACM.
[45] Hyunyoung Song, François Guimbretière, Tovi Grossman, and George Fitzmaurice. 2010. MouseLight: Bimanual interactions on digital paper using a pen and a spatially-aware mobile projector. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'10)*. ACM.
[46] Jürgen Steimle, Oliver Brdiczka, and Max Mühlhäuser. 2009. CoScribe: Integrating paper and digital documents for collaborative knowledge work. *IEEE Transactions on Learning Technologies* 2, 3 (2009), 174–188.
[47] Lisa Stifelman, Barry Arons, and Chris Schmandt. 2001. The audio notebook: Paper and pen interaction with structured speech. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'01)*. ACM, 182–189.
[48] Craig S. Tashman and W. Keith Edwards. 2011. LiquidText: A flexible, multitouch environment to support active reading. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, 3285–3294.
[49] Theophanis Tsandilas, Catherine Letondal, and Wendy E. Mackay. 2009. Mus ink: Composing music through augmented drawing. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM, Boston, MA, 819–828.
[50] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2012. Gestures as point clouds: A $ P recognizer for user interface prototypes. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. ACM, 273–280.
[51] Pierre Wellner. 1993. Interacting with paper on the digitaldesk. *Commun. ACM* 36, 7 (1993), 87–96.
[52] Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM, 159–168.
[53] Joanna Wolfe. 2002. Annotation technologies: A software and research review. *Comput. Compos.* 19, 4 (2002), 471–497.

[54] Ron Yeh, Chunyuan Liao, Scott Klemmer, François Guimbretière, Brian Lee, Boyko Kakaradov, Jeannie Stamberger, and Andreas Paepcke. 2006. ButterflyNet: A mobile capture and access system for field biology research. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*. ACM, Montreal, Canada, 571–580.

[55] Dongwook Yoon, Nicholas Chen, François Guimbretière, and Abigail Sellen. 2014. RichReview: Blending ink, speech, and gesture to support collaborative document review. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. ACM, 481–490.

Received June 2016; revised July 2017; accepted August 2017