DHQ: Digital Humanities Quarterly
2013
Volume 7 Number 1

Sounding for Meaning: Using Theories of Knowledge Representation to Analyze Aural Patterns in Texts

Tanya Clement <tclement_at_ischool_dot_utexas_dot_edu>, University of Texas, Austin
David Tcheng <davidtcheng_at_gmail_dot_com,>, University of Illinois, Urbana-Champaign
Loretta Auvil <lauvil_at_illinois_dot_edu>, University of Illinois, Urbana-Champaign
Boris Capitanu <capitanu_at_ncsa_dot_uiuc_dot_edu>, University of Illinois, Urbana-Champaign
Megan Monroe <madey_dot_j_at_gmail_dot_com>, University of Maryland, College Park

Abstract

Computational literary analytics that include frequency trends and collocation, topic modeling, and network analysis have relied on rapid and large-scale analysis of the word or strings of words. This essay shows that there are many other features of literary texts by which humanists make meaning other than the word, such as prosody and sound, and how computational methods allow us to do what has historically been a more difficult method of analysis — trying to understand how literary texts make meaning with these features. This paper will discuss a case study that uses theories of knowledge representation and research on phonetic and prosodic symbolism to develop analytics and visualizations that help readers discover aural and prosodic patterns in literary texts. To this end, this paper has two parts: (I) We describe the theories of knowledge representation and research into phonetic and prosodic symbolism that underpin the logics and ontologies of aurality incorporated in our project. This basic theory of aurality is reflected in our use of OpenMary, a text-to-speech application tool for extracting aural features; in the "flow" we coordinated to pre-process texts in SEASR's Meandre, a data flow environment; in the instance-based predictive modeling procedure that we developed for the project; and in *ProseVis*, the reader interface that we created to allow readers to discover aural features across literary texts. And (II), we discuss readings of several works by Gertrude Stein (the portraits "Matisse" and "Picasso" and the prose poem *Tender Buttons*) that were facilitated by this work.

Introduction

Humanities data, for which cultural institutions such as libraries and museums are becoming progressively more responsible, is like all data: increasing exponentially. Many scholars have responded to this expanded access by augmenting their fields of study with theories and practices that correspond to methodologies that use advanced computational analysis. The very popular Digging into Data challenge is a testament to the wide array of perspectives and methodologies digital projects can encompass. In particular, the first (2009) and second (2011) rounds of awards include projects that are using machine learning and visualization to provide new methods of discovery. Some analyze image files ("Digging into Image Data to Answer Authorship Related Questions") and the word ("Mapping the Republic of Letters" and "Using Zotero and TAPoR on the Old Bailey Proceedings: Data Mining with Criminal Intent"). Others provide new methods for discovery with audio files by analyzing "large amounts of music information" (the *Structural Analysis of Large Amounts of Music* and "the Electronic Locator of Vertical Interval Successions (ELVIS)" project) and "large scale data analysis of audio -- specifically the spoken word" (the "Mining a Year of Speech" and the "Harvesting Speech Datasets for Linguistic Research on the Web" projects). [1] At this time, however, none of these projects is looking at how we can analyze literary texts for patterns of prosody and sound; none are looking at the sound of text as it contributes to how we make meaning or interpret literature. 1

At a time when digital humanities scholars are enthusiastic about "Big Data" and are also struggling to make ties between theory and methodology, this paper discusses theories and research tools that allow scholars to analyze sound 2

patterns in large collections of literary texts. For the most part, researchers interested in investigating large collections of text are using analytics such as frequency trending and collocation, topic modeling, and network analysis that ultimately rely on word occurrence. The use case discussed here, which is supported by the Andrew W. Mellon Foundation through a grant titled “SEASR Services,” [2] seeks to identify other features than the “word” to analyze literary texts — specifically those features that comprise sound including parts-of-speech, accent, phoneme, stress, tone, and phrase units. To this end, this discussion includes a case study that uses theories of knowledge representation and research on phonetic and prosodic symbolism to develop analytics and visualizations that help readers of literary texts to negotiate large data sets and interpret aural and prosodic patterns in text.

3 In this piece, we describe how computational analysis, predictive modeling, and visualization facilitated our discovery process in three texts by Gertrude Stein, the word portraits “Matisse” and “Picasso” (first published in Alfred Stieglitz’s *Camera Work* in 1912 and in her collection *Geography and Plays*, 1922) and the prose poem *Tender Buttons* (1914). The following discussion focuses primarily on the theories, research, and methodologies that underpin this discovery process. First, we discuss the theories of knowledge representation and research into phonetic and prosodic symbolism that underpin the logics and ontologies of aurality incorporated in this project. This basic theory of aurality is reflected in our use of OpenMary, a text-to-speech application tool for extracting aural features; in the “flow” we coordinated to pre-process texts in SEASR’s Meandre,[3] a data flow environment; in the instance-based predictive modeling procedure that we developed for the project; and in *ProseVis*, the reader interface that we created to allow readers to discover aural features across literary texts. Second, this discussion addresses new readings of the word portraits “Matisse” and “Picasso” and the prose poem *Tender Buttons* by Gertrude Stein that have been facilitated by these modes of inquiry. This article outlines the theoretical underpinnings and the technical infrastructure that influenced our process of discovery such that humanities scholars may consider the efficacy of analyzing sound in literary texts with computational methods.

Knowledge Representation

4 Theories of knowledge representation can facilitate our ability to express how we are modeling sound in a computational environment. Before defining what we mean by “the logics and ontologies of aurality,” however, it is useful to discuss why these definitions are necessary at all. John F. Sowa writes in his seminal book on computational foundations that theories of knowledge representation are particularly useful “for anyone whose job is to analyze knowledge about the real world and map it to a computable form” [Sowa 2000, xi]. He defines knowledge representation as “the application of logic and ontology to the task of constructing computable models for some domain” [Sowa 2000, xi]. In other words, theories of knowledge representation are transparent about the fact that computers do not afford representations of “truth” but rather of how we think about the world in a certain context (the *domain*). For Sowa, *logic* is “pure form” and *ontology* is “the content that is expressed in that form” [Sowa 2000, xiii]. When developing projects that include computational analytics but lack logic, “knowledge representation is vague, with no criteria for determining whether statements are redundant or contradictory;” similarly, “without ontology” (or a clear sense of what the content represents), Sowa writes, “the terms and symbols are ill-defined, confused, and confusing” [Sowa 2000, xii]. Accordingly, if researchers and developers are unclear about *what* we mean and *how* we mean when we seek to represent “sound,” it is difficult for literary scholars to read or understand the results of any computational analytics we apply to that model.

5 In his seminal article, “What is Humanities Computing and What is not?” John Unsworth completes the very useful exercise of identifying various digital humanities projects that adhere to the aspects of knowledge representation put forth by AI scientists Davis, Shrobe, and Szolovits [Davis et al 1993]. Namely, the authors claim that knowledge representation “can best be understood in terms of five distinct roles it plays” [Davis et al 1993]. In the interest of defining and explaining our logic and ontology for this project, we will likewise map the development of our methodology project to these same parameters. Listed below is each of the five roles that knowledge representation plays in a project according to Davis, et al.

1. A knowledge representation is most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting, i.e., by

reasoning about the world rather than taking action in it.
2. It is a set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world?
3. It is a fragmentary theory of intelligent reasoning, expressed in terms of three components: (i) the representation's fundamental conception of intelligent reasoning; (ii) the set of inferences the representation sanctions; and (iii) the set of inferences it recommends.
4. It is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance a representation provides for organizing information so as to facilitate making the recommended inferences.
5. It is a medium of human expression, i.e., a language in which we say things about the world.

[Davis et al 1993]

After defining the first and second roles of knowledge representation in more detail in the first part of this piece, we aggregate a discussion of the next two aspects in the second part. Finally, part three of this piece includes the final role and a more comprehensive discussion of how all five roles are at play within our specific readings of texts written by Gertrude Stein. 6

Defining surrogates and ontologies

The first role of knowledge representation, as Davis describes it, is “most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it” [Davis et al 1993]. This surrogacy is essential to our understanding of aurality and the ways that text operates as a surrogate for sound systems. In this project, we are defining aurality as *the pre-speech potential of sound* as it is signified within the structure and syntax of text. 7

Aurality and the Subjective Nature of Sound Surrogacy

We call this surrogate an *aural* representation in order to emphasize the relationship that the written text bears on how sound contributes to meaning making practices in literary texts. While Walter Ong famously focuses on the “orality” of text as a testament to the history of oral cultures, Charles Bernstein focuses on the “aurality” of text, which he calls the “sounding of the writing” [Bernstein 1998, 13]. Bernstein explains that “orality” has an “emphasis on breath, voice, and speech ... *Aurality precedes orality*, just as language precedes speech” [Bernstein 1998, 13]. Bernstein makes further distinctions that frame pre-speech aurality as a perspective that focuses on the poem (the writing) rather than the poet (the performance). Bernstein considers the aurality of text as a kind of music in amorphous shapes of patterns and in grand sweeps. This is not to say that aurality depicts the exactness of traditional metered scansion. In contrast, when we read words, we say them differently depending on our regional dialects, our proficiency with the language, or even the physical differences of our mouths.[4] Correspondingly, sound is not represented in text as the representation of one particular utterance or scansion. In fact, textual structures present an imperfect representation of sound that inevitably incorporates the chaos, inexactitude, and confusion of aurality [Bernstein 1998]. By couching this sound surrogate within the context of *aurality* instead of *orality*, we are acknowledging that our sound surrogate gestures towards many potential utterances, to any of many noise-profuse, context-driven, reading performances. 8

Ultimately, understanding and defining sonic phenomena is a subjective practice. In a recent special issue of *differences* titled “Sound Senses,” editors Rey Chow and James Steintrager ask what is the “Object of Sound” for study and they note the slippery, diffusive nature of sound as “something not obviously divisible” [Chow and Steintrager 2012, 2]. “Objects as sonic phenomena are points of diffusion that in listening we attempt to gather,” they write. Most significant for this discussion, they articulate the work of sonic interpretation as this “work of gathering” in “an effort to unify and make cohere” [Chow and Steintrager 2012, 2]. Likewise, Walter Ong sees the work of interpreting or reading texts as a gathering of sounds: 9

Written texts all have to be related somehow, directly or indirectly, to the world of sound, the natural

> habitat of language, to yield their meanings. '"Reading"' a text means converting it to sound, aloud or in the imagination, syllable-by-syllable in slow reading or sketchily in the rapid reading common to high-technology cultures. [Ong 2002, 8]

10 While Ong essentializes the relationship between written texts and sound in terms of his study of literacy in oral cultures, Charles Bernstein points back to the difficult work of identifying the osmotic relationship between sound and meaning when interpreting poetry, arguing, "[t]he relation of sound to meaning is something like the relation of the soul (or mind) to the body. They are aspects of each other, neither prior, neither independent" [Bernstein 1998, 17]. Finally, in writing her word portraits, Gertrude Stein also notes the subjective nature of her own "gatherings" of sound and how that work influenced her creation of literary texts: "I had the habit," she writes, "of conceiving myself as completely talking and listening, listening was talking and talking was listening and in so doing I conceived what I at the time called the rhythm of anybody's personality" [Stein1988c, 174].

11 Interpreting or representing sound is also a subjective practice, however. Dwight Bolinger notes that this subjective work of gathering is a "best guess" that is based on what can be considered divisible and measurable syntactical units. "In the total absence of all phonological and visual cues," he writes, "the psychological tendency to impose an accent is so strong that it will be done as a 'best guess' from the syntax" [Bolinger 1986, 17]. In other words, when we seek to "sound out" a written word, we make "best guesses" for those sounds based on the possibilities of sound that are represented by the structural features of a word including parts-of-speech, the position of a word in a phrase (e.g., consecutive verbs or nouns), sentence type (e.g., a declaration or a question), and information structure (e.g., given and inferable information in a dependent clause is frequently de-accented) within its syntactical context. This discussion emphasizes the fact that identifying sound surrogates that represent a literary text's aurality remains subjective, especially within a computational system like ours that relies on "best guesses" for gathering syntactical units.

An Ontology of Sound in which Sound is a Meaningful Aspect of Literary Texts

12 Davis's second role for knowledge representation is as "a set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world? The commitments are in effect a strong pair of glasses that determine what we can see, bringing some part of the world into sharp focus, at the expense of blurring other parts" [Davis et al 1993]. In this project, we are committed to an ontology of sound in which sound is a meaningful aspect of literary texts.

13 The debate concerning whether or not sound contributes to how we interpret written texts has a long history.[5] While French theorist Ferdinand de Saussure argued that the relationship between sound and meaning was essentially arbitrary (1916), Socrates previously argued there was a significant (i.e. *signifying*) relationship there [Ong 2002, 78]. Roland Barthes identified two aspects of sound that contribute to meaning: the *pheno-song*, which maps to our concern with aurality in this project, and refers to "all the phenomena, all the features which belong to the structure of the language being sung, the rules of the genre, the coded form of the melisma, the composer's idiolect, the style of the interpretation: in short everything in the performance which is in the service of communication, representation, expression"; and the *geno-song*, which is the "volume of the singing and speaking voice, the space where significations germinate" [Barthes 1978, 182].

14 In order to create aural surrogates with computational modeling we include textual features of sound that research has shown correspond to how we interpret texts. Reuven Tsur, for example, works backwards from long held beliefs about specific meanings of sound in poetry to arrive at rules that can support or refute these meanings. We are also interested in those features of text that create possibilities for interpretation. For instance, Tsur notes that a reader's sense that sounds make meaning is an abstract or impressionistic regard, but Tsur seeks to "use phonetic and phonological generalizations in an attempt *to determine the rules* on which certain impressionistic generalizations are founded" based on "widespread beliefs concerning the ''aesthetic'' quality of speech sounds" [Tsur 1992, 64]. In this regard, Tsur attempts to balance "two aims": (1) to legitimate "impressions" of sound as an "integral part of criticism" and (2) to define rules that harken toward "scientific impartiality" or empirical analysis [Tsur 1992, 64]. Where we differ with Tsur is in his attempt to "claim back the largest possible areas of criticism from arbitrary impressionism" [Tsur 1992, 64]. Our

emphasis in this discussion, in contrast, is to expose the *partial* nature of determining such rules by exposing how our ontological commitments — those "aspects of the world we believe to be relevant" [Davis et al 1993] — influence our choices to expose particular sound features.

15 In particular, we adhere to the idea that all meaning making is an act of abstraction that is dependent not only on the objects of study (words *or* sound) but also on the context in which we find them (words *and* sounds). Quoting the work of Hrushovski, Tsur claims that there are four kinds of relations between sound and meaning: "(a) Onomatopoeia; (b) Expressive Sounds; (c) Focusing Sound Patterns; (d) Neutral Sound Patterns" [Tsur 1992, 2]. Benjamin Hrushovski describes as "Expressive Sounds" as "[a] sound combination [that] is grasped as expressive of the tone, mood, or some general quality of meaning" [Tsur 1992, 2]. Most important for Tsur and our project is the comparison that Hrushovski draws between the act of making meaning with sound and the act of making meaning with words: "an abstraction from the sound pattern (i.e., some kind of tone or 'quality' of the sounds is parallel to an abstraction from the meaning of the words (tone, mood etc.)" [Hrushovski 1968, 444]; [Tsur 1992, 2]. The same sounds can make meaning by reflecting different relations, and as a result, the same sounds can evoke different moods. For instance, Tsur cites the example of a line by Poe ("And the silken, sad, uncertain rustling, of each purple curtain"), which uses onomatopoeia and a quatrain by Shakespeare that begins "When to the sessions of sweet, silent thought…" that uses expressive sounds. Both of these examples use the sibilants /s/ and /š/ to evoke, respectively, "noisy potential" and "hushing potential" [Tsur 1992, 2]. Tsur denies claims that "all speech sounds are equal" since "for readers of poetry it is difficult to escape the feeling that some speech sounds are more equal … more musical, more emotional, or more beautiful than others" [Tsur 1992, 53].

16 Sound also contributes to meaning with prosodic and phonetic elements. Prosody has been defined by linguists as comprising intonation, stress, and rhythm to convey linguistic meaning through phrasing and prominence [Cole 2011]. Prosody makes meaning in part by reflecting a speaker's identity, gender, regional dialect, ethnolect, affect and emotional engagement, and cognitive process. Consequently, prosody can be used to study human behavior, culture, and society [Cole 2011]. In terms of prosody, Dwight Bolinger defines the term "intonation" to include not only accents and stress but also symbolic meaning since "it is generally used to refer to the overall landscape, the wider ups and downs that show greater or lesser degrees of excitement, boredom, curiosity, positiveness, etc." [Bolinger 1986, 11]. Moreover, there is a body of research in linguistics and psychology called phonetic symbolism that dates back to controlled studies done by Edward Sapir (1929) and his student Stanley Newman (1933). In these studies and subsequent ones, links are established between the sounds of vowels and consonants and readers' perceptions of size (big and small), volume (full or empty), speed (fast and slow), intensity (dull and sharp) and value (pleasant and unpleasant) [Shrum and Lowrey 2007, 40–47]. These ideas that the meaning of sound correlates to the structure of the text are significant to how we have chosen to develop an infrastructure for analyzing sound in text. With this project, we are identifying and modeling these potentially meaningful aspects of literary texts for interpretive analysis.

Intelligent Reasoning and Pragmatically Efficient Computation

17 The above theories in aurality and research in phonetic and prosodic symbolism undergird the choices we have made in developing a technical, computational infrastructure for analyzing the sound of literary texts. Shifting our attention to consider two more of Davis's roles of knowledge representation including knowledge representation as a "fragmentary theory of intelligent reasoning" and as a "medium for pragmatically efficient computation," this section will discuss three essential parts of the infrastructure that represents the sound of text in our project. First we consider our decision to use OpenMary, a text-to-speech application tool, that extracts aural features from literary texts; next, we discuss the data flow we developed in SEASR's data flow environment (Meandre) to produce a representation of the data for modeling as well as the predictive modeling procedure we implemented to analyze patterns across these extracted features; and finally, we introduce *ProseVis*, the reader interface we created to allow readers to discover and interpret these extracted aural features and patterns *in conversation with* (not a replacements for) the literary texts.

OpenMary

18 In this project, we use OpenMary,[6] an open source, text-to-speech system, to create a text-based surrogate of sound.

Developed as a collaborative project of Das Deutsche Forschungszentrum für Künstliche Intelligenz (German Research Center for Artificial Intelligence) Language Technology Lab and the Institute of Phonetics at Saarland University, OpenMary captures information about the structure of the text that make it possible for a computer to read text in multiple languages (German, British and American English, Telugu, Turkish, and Russian; more languages are in preparation) and create spoken text. We chose OpenMary as a useful analytic routine for analyzing these texts after first parsing our texts against the CMU (Carnegie Mellon University) Pronouncing Dictionary and then validating sections of the Mary XML Output against human-parsed sections. In a simple comparison based on analyzing Gertrude Stein's novel *The Making of Americans*, we noticed that many "unknown" words were returned in the CMU comparison. That is, many of the words that Stein used in her lexicon, though common words, were returned as "unknown" words in the results (such as "insensibility," "meekness," "well-meaning," and "slinks"). OpenMary's recommendation, on the other hand, incorporates a "best guess" model in any given prosodic situation — that is, it is based on an algorithm or a set of stringent rules that draws on the kind of research that Tsur, Bolinger and others have mapped for how we make meaning with sound, which includes part-of-speech, accent, phoneme, stress, tone, the position of a word in a phrase (e.g., consecutive verbs or multiple nouns), sentence type (e.g., a declaration or a question), and information structure (e.g., given and inferable information in a dependent clause is frequently de-accented) [Becker et al 2006].

The documentation explains OpenMary's system for Natural Language Processing (NLP): 19

> In a first NLP step, part of speech labelling [sic] and shallow parsing (chunking) is performed. Then, a lexicon lookup is performed in the pronounciation [sic] lexicon; unknown tokens are morphologically decomposed and phonemised by grapheme to phoneme (letter to sound) rules. Independently from the lexicon lookup, symbols for the intonation and phrase structure are assigned by rule, using punctuation, part of speech info, and the local syntactic info provided by the chunker. Finally, postlexical phonological rules are applied, modifying the phone symbols and/or the intonation symbols as a function of their context. [MARY TTS]

Further intelligent reasoning is reflected in OpenMary's folksonomic technique for representing words that are not in the CMU Pronouncing dictionary lexicon; this technique involves generating a lexicon of known pronunciations from the most common words in Wikipedia and allowing developers to enter new words manually ("Adding support for a new language to MARY TTS"). OpenMary will make a "best guess" at words that are not part of the CMU lexicon because its rule set or algorithm — its "intelligent reasoning" — for how it generates audio files is based on the research of both linguists and computer scientists. As such, this highly technical description speaks to the deeply interdisciplinary work that has formed the rules by which OpenMary represents the sound of literary texts in a digital file — as an interface between human-perceived rules for reading and methods for machine processing. 20

As a byproduct of this process, OpenMary outputs a representation of the sound of text in XML that reflects a set of possibilities for speech that are important indicators of how the text could potentially be read aloud by a reader. Specifically, OpenMary accepts text input and creates an XML document (MaryXML) as output with attributes like those shown in Figure 1. This example represents the phrase "A kind in glass and a cousin, a spectacle and nothing strange" from Gertrude Stein's text *Tender Buttons*. 21

```
<s>
    <prosody pitch="+5%" range="+20%">
        <phrase>
            <t g2p_method="lexicon" ph="@" pos="DT">  A</t>
            <t accent="L+H*" g2p_method="lexicon" ph="' k AI n d" pos="NN"> kind</t>
            <t g2p_method="lexicon" ph="' I n" pos="IN">  in</t>
            <t accent="L+H*" g2p_method="lexicon" ph="' g l { s" pos="NN">  glass</t>
            <t g2p_method="lexicon" ph="' { n d" pos="CC">  and</t>
            <t g2p_method="lexicon" ph="@" pos="DT">  a</t>
            <t accent="L+H*" g2p_method="lexicon" ph="' k V - z @ n" pos="NN"> cousin</t>
            <t pos=","> ,</t>
            <boundary breakindex="4" tone="H-L%"/>
        </phrase>
    </prosody>
    <prosody pitch="+2%" range="+10%">
        <phrase>
            <t g2p_method="lexicon" ph="@" pos="DT">   a</t>
            <t accent="L+H*" g2p_method="lexicon" ph="s - ' p E k - t @ - k @ l" pos="NN">   spectacle</t>
            <t g2p_method="lexicon" ph="' { n d" pos="CC">  and</t>
            <t accent="L+H*" g2p_method="lexicon" ph="' n V - T I N" pos="NN">  nothing</t>
            <t accent="L+H*" g2p_method="lexicon" ph="s - ' t r EI n dZ" pos="JJ">    strange</t>
            [...]
        </phrase>
    </prosody>
</s>
```

Figure 1.

As shown above sentences (<s>) are broken into prosodic units and then phrases (<prosody> and <phrase>), which are, in turn, broken into words or tokens (<t>). These word elements hold the attributes that mark “accent”, part of speech (“pos”), and “ph” — phonetic spellings (transcribed in SAMPA),) broken into what we refer to as “sounds” separated by “–”, with an apostrophe (“ ' ”) preceding stressed syllables. Other information is included at the phrase level such as “tone” and “breakindex.” [7]

22

Meandre Data Flow Environment

The SEASR (Software Environment for the Advancement of Scholarly Research) team at the University of Illinois at Urbana-Champaign has been working on creating a computational environment in which users who are interested in analyzing large data sets can develop data flows that push these data sets through various textual analytics and visualizations.[8] This environment, called Meandre, provides tools for assembling and executing data flows. A data flow is a software application consisting of software components that process data. Processing can include, for example, an application that accesses a data store, one that transforms the data from that store and analyzes it with textual analysis, and one that visualizes the transformed results. Within Meandre, each flow is represented as a graph that shows components as icons linked through their input and output connections (see Figure 2). Based on the inputs and properties of a component, an output is generated upon execution. Meandre provides basic infrastructure for data-intensive computation by providing tools for creating, linking, and executing components and flows. As such, Meandre facilitates a user’s ability to choose how her information will be organized and ultimately the kinds of inferences that can be made from the resulting data.

23

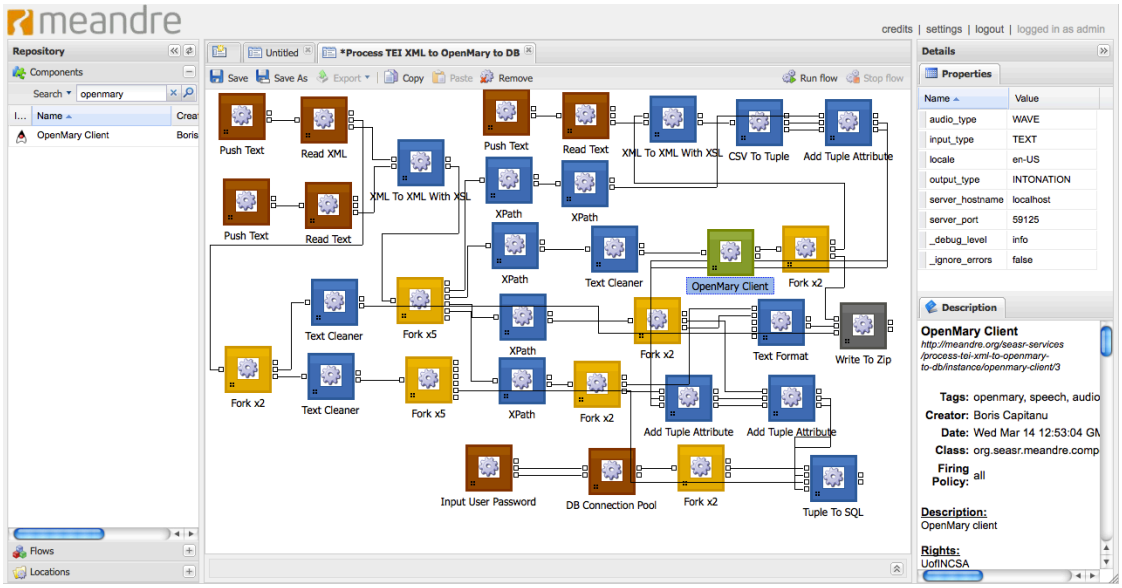

Figure 2.

The ability to explore a text's aurality was not represented within SEASR until we added a Meandre component to use OpenMary (shown as the green box module in Figure 2). Meandre components were used to segment the book into smaller chunks of text before passing it to OpenMary for feature extraction, because sending large amounts of text to OpenMary created memory problems associated with processing the complete document. Consequently, the flow processes each document in our collection through the OpenMary web service at a paragraph level. Meandre is also used to create a tabular representation of the data (see Figure 3). The features represented from the MaryXML are part of speech, accent, phoneme, stress, tone, and break index, because research shows that these features have a significant impact on how we make meaning with sound.[9] We also include information that is useful in terms of framing the context of the sounds within the document's structure (chapter id, section id, paragraph id, sentence id, phrase id, and word id). This allows words to be associated with accent, phoneme, and part-of-speech within the context of the phrase, sentence and paragraph boundary. Figure 2 shows the flow with the components that are used for executing OpenMary and for post-processing the data to create the database tables. Green components are for computing (i.e. the OpenMary processing component), the blue components are transformation components (i.e. XSL transformation), the red components are input components (i.e. loading the xml file), the dark gray component is an output component (i.e. writing a file) and the yellow component are control flow components, (i.e. forking - duplicating an output). Another benefit of creating this flow in Meandre is that readers who wish to analyze these results or who wish to produce data for their documents will have access to the same flow[10].

24

tei_chapter_id	tei_section_id	tei_paragraph_id	sentence_id	phrase_id	word	part_of_speech	accent	phoneme	stress	tone	break_index	s
1	1	1	1	1	A	DT	NULL	@	0	H-L%	4	
1	1	1	1	1	kind	NN	L+H*	k Al n d	1	H-L%	4	
1	1	1	1	1	in	IN	NULL	I n	1	H-L%	4	
1	1	1	1	1	glass	NN	L+H*	g l { s	1	H-L%	4	
1	1	1	1	1	and	CC	NULL	{ n d	1	H-L%	4	
1	1	1	1	1	a	DT	NULL	@	0	H-L%	4	
1	1	1	1	1	cousin	NN	L+H*	k V	1	H-L%	4	
1	1	1	1	1	cousin	NN	L+H*	z @ n	0	H-L%	4	
1	1	1	1	1	,	,	NULL	NULL	NULL	H-L%	4	
1	1	1	1	2	a	DT	NULL	@	0	L-L%	5	
1	1	1	1	2	spectacle	NN	L+H*	s	0	L-L%	5	
1	1	1	1	2	spectacle	NN	L+H*	p E k	1	L-L%	5	
1	1	1	1	2	spectacle	NN	L+H*	t @	0	L-L%	5	
1	1	1	1	2	spectacle	NN	L+H*	k @ l	0	L-L%	5	
1	1	1	1	2	and	CC	NULL	{ n d	1	L-L%	5	
1	1	1	1	2	nothing	NN	L+H*	n V	1	L-L%	5	
1	1	1	1	2	nothing	NN	L+H*	T I N	0	L-L%	5	
1	1	1	1	2	strange	JJ	L+H*	s	0	L-L%	5	
1	1	1	1	2	strange	JJ	L+H*	t r EI n dZ	1	L-L%	5	

Figure 3.

Once the features for aurality were extracted for a collection of documents, we wanted to compare the aurality between the documents and identify the documents that had similar prosody patterns. This comparison was framed as a

25

predictive problem, where we used the features from one document to predict similar documents. We developed an instance-based, machine-learning algorithm for the predictive analysis that can be broken into the following steps:

1. **Defining a prediction problem**:

   Our hypothesis is that several books in our collection have similar prosody patterns and should “sound” more alike.

2. **Defining examples for machine learning**:

   Figure 4 shows the process we follow to create “examples” for machine learning, starting with the OpenMary output, and transformation to a database table in Meandre. Next, we use our predictive analysis algorithms to derive a “symbol” from the OpenMary output at the sound level (i.e., each row of the tabular data). This symbol is an id that represents a unique combination of *just* those features we associate with prosody including part of speech, accent, stress, tone, and break index. There are over six thousand symbols because there are over six thousand combinations of these attribute values. Once symbols are defined, we create a moving window — a phrase window — across the sounds to create the examples we use for comparison. We define the window size of this phrase window to be the average phrase length produced by the OpenMary analysis. We select the average length of a phrase in the data set, not in order to maximize classification accuracy, but in order to best simulate how readers perceive sound at the phrase level (Soderstrom, et. al). Shorter or longer phrase windows are possible and window size does affect accuracy — these choices, again, reflect the “intelligent reasoning” and “pragmatically efficient computation” aspects of knowledge representation that Davis, et. al have identified.

   For our collection, the size of the phrase window is fourteen so the set of input features are the fourteen symbol ids for the given phrase. In total, there are 1,434,588 phrase windows of fourteen symbols from nine books. Finally, we added the “class” attribute, which is an id for the book in which the phrase window exists. The class attribute (the book) is the attribute that we predict.

3. **Modeling**:

   For the predictive analysis we use an instance-based approach, which is based on learning algorithms that systemically compare new problem instances with training instances. In this project, we use a full, leave-one-out cross validation. That is, for each prediction, the phrase window is compared to each phrase window from all other books.[11] The prediction is the probability that the phrase window is in each class (or book), so the probability distribution over all classes sums to 1.0 (as seen by the row values of the bottom table in Figure 4). In our collection, there are nine class attributes, one for each book.

   To predict which book a given phrase window exists, this phrase window is compared to all other phrase windows from all other books by computing a distance function.[12] In order to build the best prosody model, one must systematically optimize the control parameters of the machine-learning algorithm to maximize accuracy. There are over one million phrase windows that need to be compared with each other requiring twenty-eight trillion window comparisons. This amount of computation needs to be done for each bias parameter setting considered during bias optimization. Each bias is a new experiment to run with different parameters so adding a book is a new parameter and changing the phrase window from 14 to 15 is a new bias.[13]

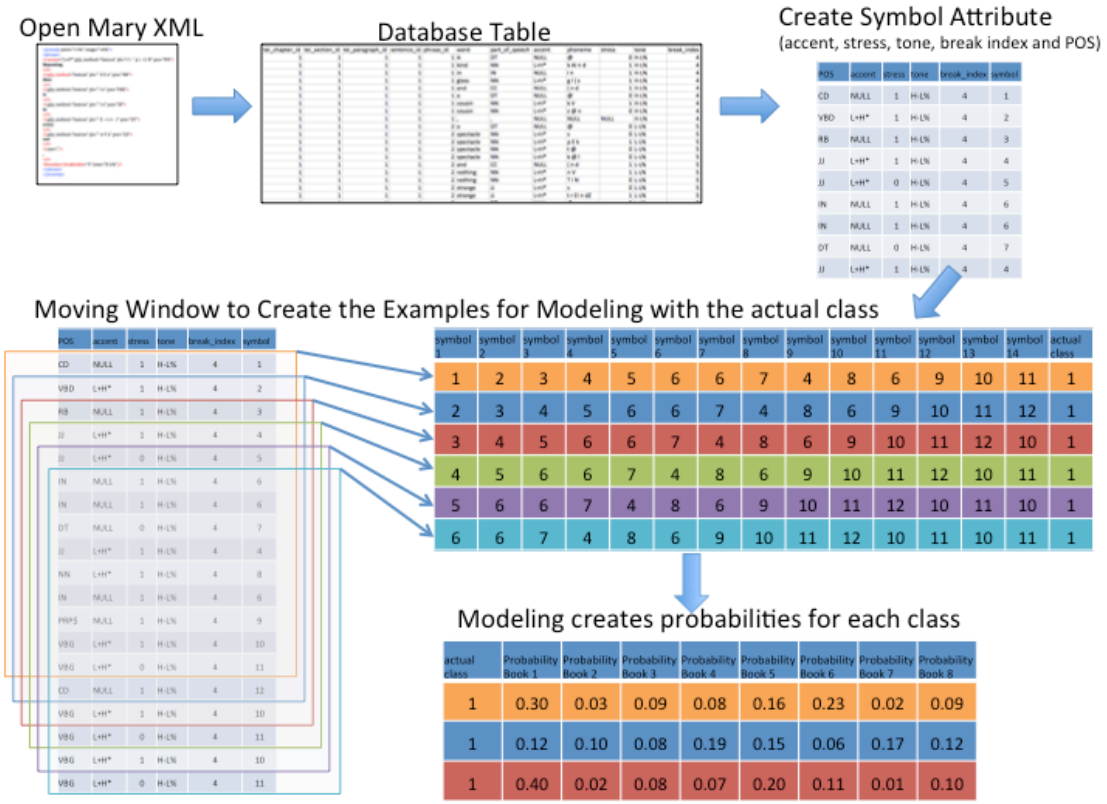

Figure 4.

We describe these extensive processes to show that intelligent reasoning and pragmatically efficient computation require extensive amounts of processing power. As such, these are not experiments that can be run on a home computer. Changing the way we analyze text (moving away from the grapheme and towards the phoneme) is complicated by the need to collaborate across disciplinary realms (such as an English Department or School of Information collaborating with a Supercomputing Center and Visualization Lab). Further, the results produced by these processes comprise another set of large amounts of data that must be made comprehensible to readers or scholars interested in analyzing sound patterns in text. 31

ProseVis

An essential aspect of this project is ProseVis, a visualization tool we developed to allow a reader to map the features extracted from OpenMary and the predictive classification data to the words in the contexts to which readers are familiar.[14] We developed this project with the ultimate goal of facilitating a reader's ability to analyze sonic features of text and research has shown that mapping the data to the text in its original form allows for the kind of reading that literary scholars engage: they read words and features of language situated within the contexts of phrases, sentences, lines, stanzas, and paragraphs [Clement 2008]. Recreating the contexts of the word not only allows for the simultaneous consideration of multiple representations of knowledge or readings (since every reader's perspective on the context will be different) but it also allows for a more transparent view of the underlying data. If a reader can see the data (such as sounds and parts of speech) within the contexts of the text with which they are familiar and well-versed, then the reader is empowered within this familiar context to read what might otherwise be an unfamiliar, tabular representation of the text. 32

Using the data produced by Meandre, ProseVis highlights features of a text. Figure 5, for example, shows two short prose pieces by Gertrude Stein called "word portraits" and titled "Matisse" and "Picasso." Stein's word portraits were writing projects in which character development progresses without narrative, much like still-life portraits of a person that also "do not tell a story" [Stein1988c, 184]. Rather, portraits provide a telling snapshot in time. Stein draws the comparison to portraits because her attempt to create written portraits was much like what she considered a painter's 33

ought to be — to create “a picture that exists for and in itself” using “objects landscapes and people” without being “deeply conscious of these things as a subject” [Stein 1990, 497]. In this first ProseVis example, we see Stein’s portraits “Matisse” and “Picasso” rendered as a series of rows with colored blocks. Because ProseVis maintains a list of unique occurrences of each attribute extracted by OpenMary, the reader can choose to color the visualization by any of these attributes such as part-of-speech, tone, accent, word, and phonetic sound. (Figure 6 shows the same information, zoomed to illustrate how the words are legible beneath the blocks of color.) The panel on the right is the control panel where the reader can choose how to display the text and prosody features. Options include the ability to show lines by phrases or sentences or chapter or stanza group. In these ways, the reader can examine prosodic patterns as they occur at the beginning or end of these text divisions.

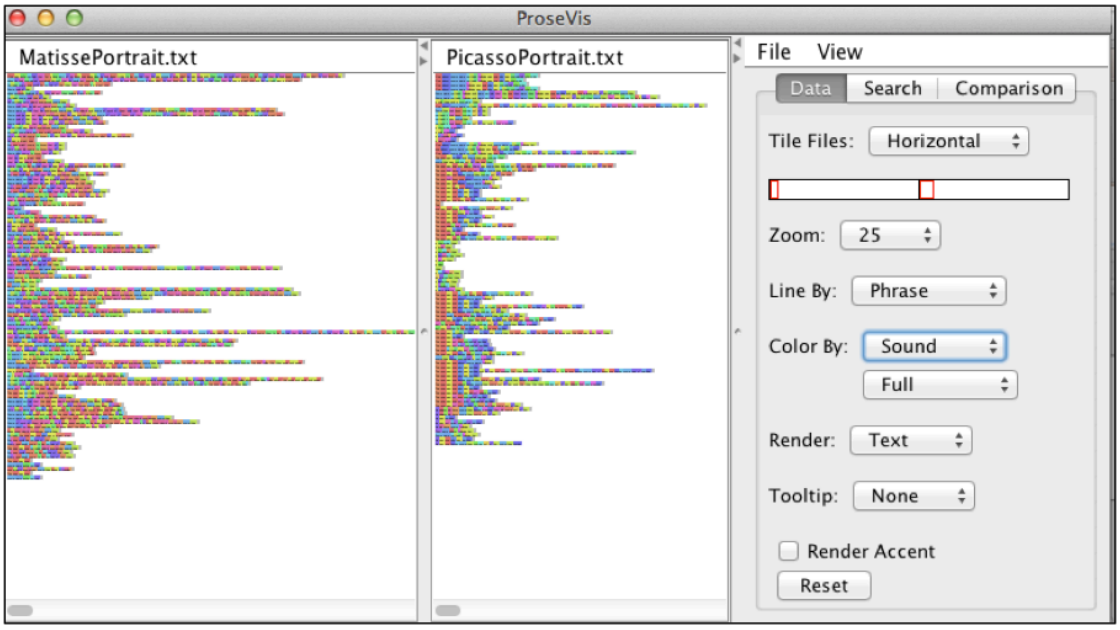

Figure 5.

Figure 6.

When visualizing the text at the sound level, we encountered three primary issues: (1) The set of unique sounds in a given text is too large to assign each one an easily discernible color in the display; (2) When doing a string-based comparison of one complete sound to another, it is not possible to detect subtler, and potentially critical similarities that

form patterns such as alliteration and rhyming.[15] To address the second issue, we broke each syllable down into three primary constituents, and allowed for the display to target each of these constituents individually. The constituents that we identified as the most informative were the leading consonant sound, the vowel sound, and the ending consonant sound:[16]

Word	Sound	Lead Consonant	Vowel Sound	End Consonant
Strike	s tr I ke	S	AI	k

Table 1.

This breakdown provides the reader with a finer-grained level of analysis, as well as a simplified coloring scheme. As a result, if the reader chooses to color the visualization by the sound, they have the additional option of coloring by the full sound, or by a component of sound such as a leading or ending consonant or a vowel sound. Figure 7, Figure 8, and Figure 9 show these alternate views. Further, a reader can render all the words as phonetic spellings (“sound”), parts-of-speech (“POS”), or take out the underlying information altogether (to leave just color) instead of text.

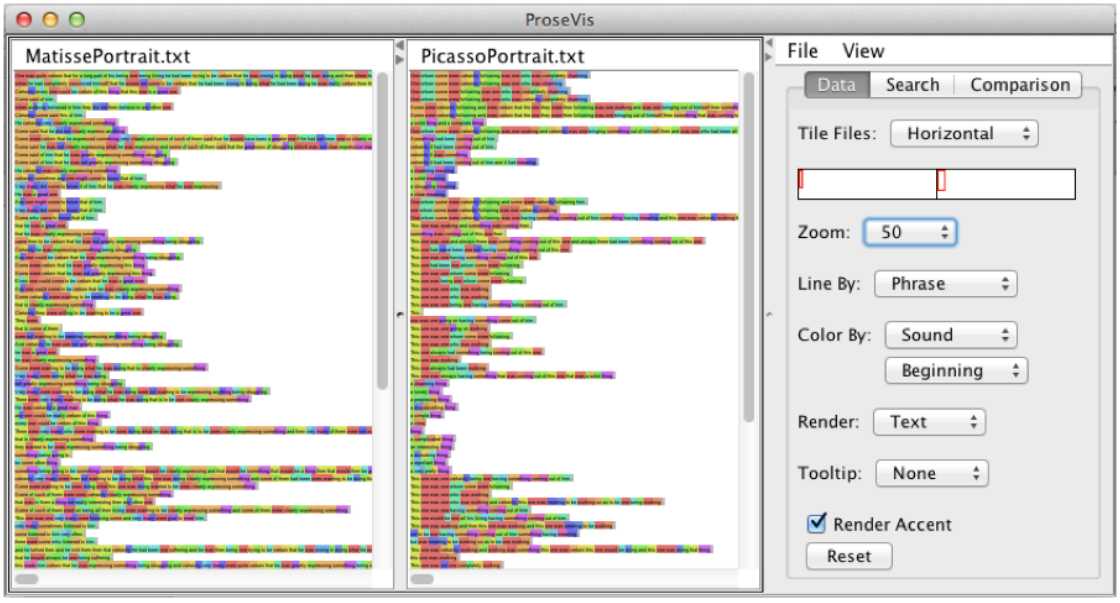

Figure 7.

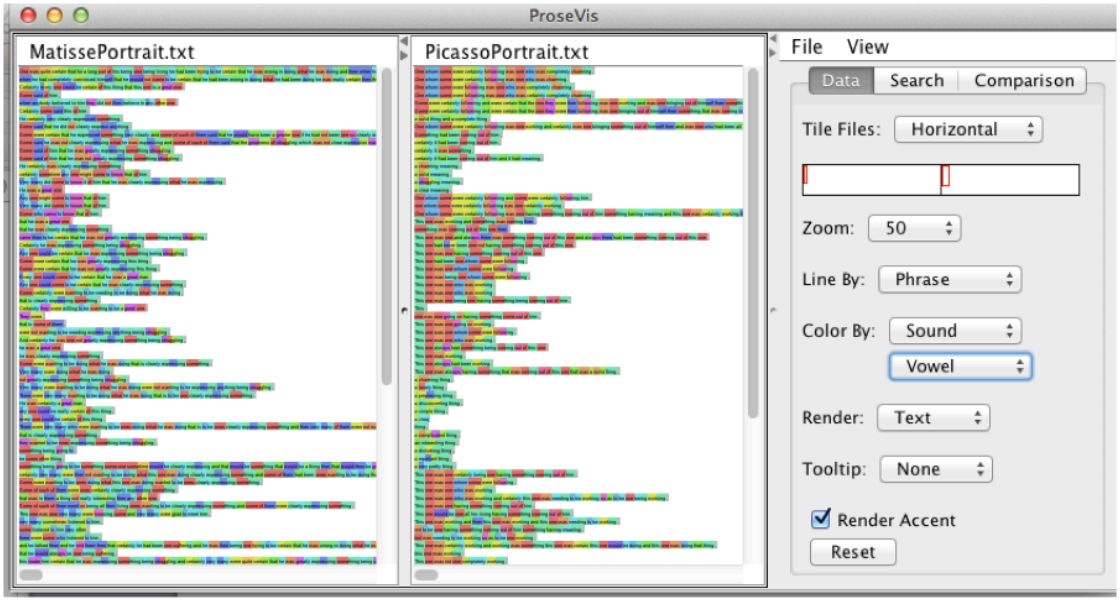

Figure 8.

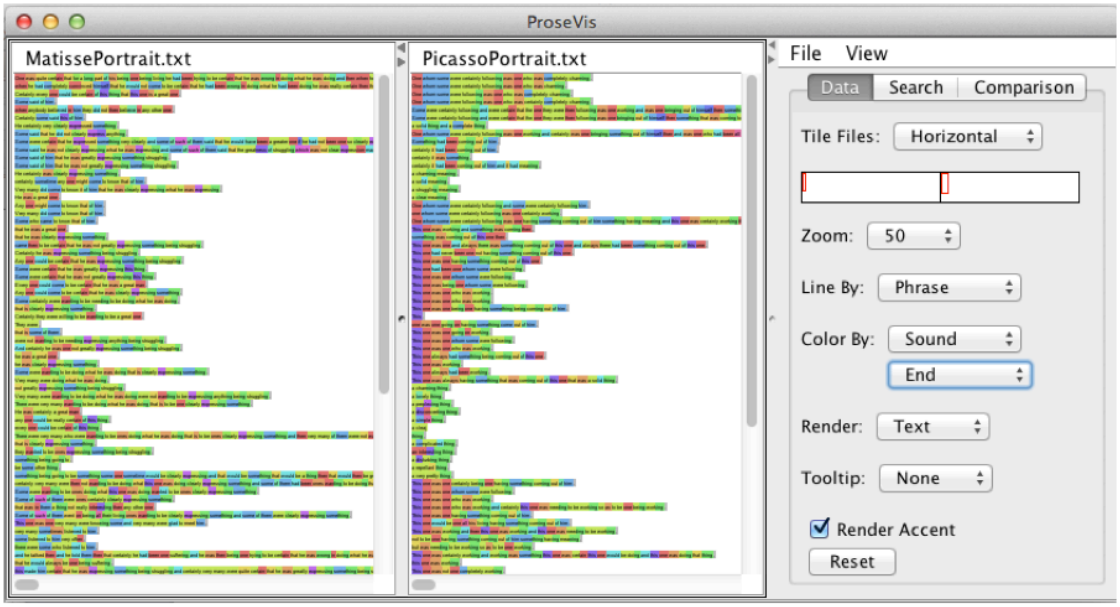

Figure 9.

Finally, under the “Comparison” menu, readers can see the predictive modeling data layered on top of the text. Here, each color represents a different book (listed on the right) and each sound is highlighted according to which book it is most like. When all the books are selected, the color reflects which book has the highest probability or comparison for a given sound.[17] 36

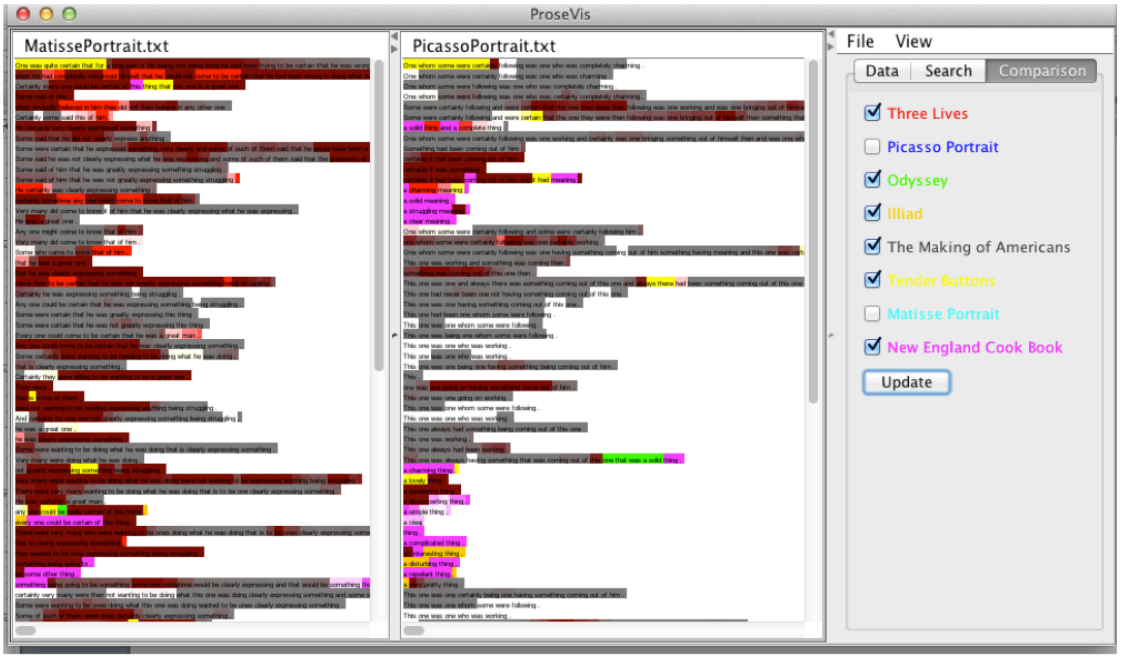

Figure 10.

Reading the portraits “Matisse” and “Picasso” and *Tender Buttons* in ProseVis

As discussed, one of Gertrude Stein’s early modes of experimentation was to create word portraits in the modernist mode. At the same time, she sensed an immediate connection between the acts of speech (talking and listening) and her work creating portraits of people in words. Derrida minimizes the distinction between writing and speech or voice (and therefore sound) by showing how both are perceived by the *différance* that is signification. In reading Gertrude 37

Stein's work, however, Scott Pound argues that "Derrida obscures a distinction between written and spoken language that a discussion of poetics cannot do without. Poststructuralism's demonstration of the difference writing makes must therefore be set in relation to the difference sound makes" [Pound 2007, 26–27]. As discussed in the first part of this essay, in order to investigate the difference that sound makes, we are transparent about the fact that a representation of sound is subjective. What is most significant for this discussion, therefore, is Derrida's claim that "[i]n order to function, that is, to be readable, a signature must have a repeatable, iterable, imitable form; it must be able to be detached from the present and singular intention of its production" [Derrida 1991, 106]. In this project, the form of the "signature" or sound of text is the iterable, repeatable data that is produced by computational analysis.

38 Further, we can imagine this imitable form as a layer of data (a reading or another "text") that we are using as an overlay on the "originary" text as a means or a lens to read the literary text differently. This "new" perspective on Stein's texts is not only important for understanding her creative work; it is important for reconsidering what we have learned not to consider. For instance, Craig Monk argues that Gertrude Stein lost favor with Eugene Jolas, founding editor of *transition*, for political and personal reasons. Yet, the history can be and has been read differently: that Jolas preferred James Joyce's writing to Stein's because Joyce was held up as the revolutionary writer of his time. According to Monk, Jolas laid down a gauntlet in 1929 when he published his "The Revolution of the Word Proclamation" (issue 16/17 of *transition*). This revolution, Jolas writes, requires "[t]he literary creator" or writer "to disintegrate the primal matter of words imposed on him by textbooks and dictionaries" ("Introduction"). Joyce, argues Monk, epitomized Jolas's revolution with his "neologisms and portmanteau words" [Monk 1998, 29] while Stein's "little household words so dear to Sherwood Anderson never impressed [Jolas]" ([unpublished autobiography], 201 qtd. in Monk 32). As a result, while Jolas would publish much of Joyce's work including a serialization of his 'Work in Progress' (which subsequently became *Finnegans Wake*) as well as essays by prominent authors who wrote about Joyce's work, he only published one more piece of Stein's after his 1929 manifesto. Finally, in 1935, Jolas publicly denounced her writing in a *Testimony Against Gertrude Stein* (a supplement to *transition*, July 23). Perhaps the most salient observation Monk makes for this discussion is his conclusion that "it was only as Jola's preference for the verbal in poetry began to emerge clearly that that the discussion of the visual analogies used often to describe Stein's works might be read, in hindsight, as implicitly derogatory" [Monk 1998, 30].

39 In fact, the idea that James Joyce's mode of experimentation incorporated elements from music while Stein's works, in contrast, reflected influences from the visual arts has been debunked and explored and complicated by too many scholars to rehearse again in this space, but the fact remains that as a culture, we are not far removed from the situation in which *transition*'s audience found itself: we have been summarily prohibited from reading sound patterns by a system of production that favors one mode of interpretation over another: the grapheme over the phoneme. Sound patterns are difficult to discern. Using computational analysis to mark (to make imitable and repeatable and *visual*) expressions that correspond to sound is a step in attempting to discern the relationships between all the various features of text that contribute to meaning. Stein describes this confluence of features this way: "I began to wonder at about this time just what one saw when one looked at anything really looked at anything. Did one see sound, and what was the relation between color and sound, did it make itself by description by a word that meant it or did it make itself by a word in itself" [Stein1988c, 91].

40 Primarily, the last part of this discussion is an exploration, using ProseVis and the data from OpenMary and Meandre's processes in reading sound patterns in Gertrude Stein's portraits "Matisse" and "Picasso" and her prose poem *Tender Buttons*. In this exploration, we are concerned above all with Davis's final role of knowledge representation, namely as "a medium of human expression, i.e., a language in which we say things about the world" [Davis et al 1993]. What is at stake in this section is not to create new readings of Stein's texts (this would take much more deliberation and space) as much as it is to demonstrate how we have come to analyze literary texts in digital environments as visual texts that are divorced, quite often, from attributes of sound. In computational environments, productive and critical representations of knowledge should show a consideration for the multiplicity of ways that humans express and understand themselves through how we say things about the world with literature. Considering how to represent and analyze the sound of text in these readings represents a step towards pushing computational discovery practices past singular representations of the word and, thus, singular modes of interpretation.

“Matisse” and “Picasso” (1912)

41

The relative success of Stein’s methods for creating the rhythm of a character is evident in the response of scholars. Wendy Salkind argues that with her portraits “Matisse” and “Picasso”, Stein expresses a “disenchant[ment] with Matisse and his painting” and a sustained “belief in the genius of Picasso.” In particular, Salkind notes the ways in which sounds and rhythms work to create these readings:

> We can *hear* that adulation and disappointment in the phrase repetitions she uses in both pieces. She writes about the effort of creating art, the struggle to be constantly working, to be consistently expressing something, and to find greatness among your followers… When the *Picasso* description above is *spoken aloud*, the repetition of the “w” sound continuously brings your lips forward, as if in a kiss. The monosyllabic sentence flows and arrives on the emphasis of the *double syllable* resolution of the final word. Although also *monosyllabic*, the *Matisse* description is pedestrian, lacking fluidity. When *spoken*, her words describing him don't feel nearly as good in your mouth.

42

These same patterns are evident in the ProseVis visualization in Figure 7 in which the beginning consonant sound “w” of words like “was,” “one,” and “were” is represented in red. Clearly, there are fewer concentrated patterns in the “Matisse” portrait on the left than in the “Picasso” portrait on the right but “Matisse” has 283 “w” sounds out of 2129 total sounds (7.5%), which is actually more than “Picasso,” which has 271 “w” sounds out of 1246 sounds (4.5%). The visualization suggests that rather than volume of sounds, Salkind’s reading may have more to do with the close repetition of the “w” sounds in “Picasso” — the successive opening and closing of the lips to make these sounds could mimic kissing more readily than the sporadic lone “w” sounds used in “Matisse.” Further, if we color the text according to the “accent” data (see Figure 11 and Figure 12), we see the dark blue areas that indicate high pitch or accented words that are more prevalent in “Picasso” than in Matisse. These representations invite us to ask more questions: what is accent doing in the text to contribute to readings like the one Salkind proposes? What is the role of sound and prosody in this text?

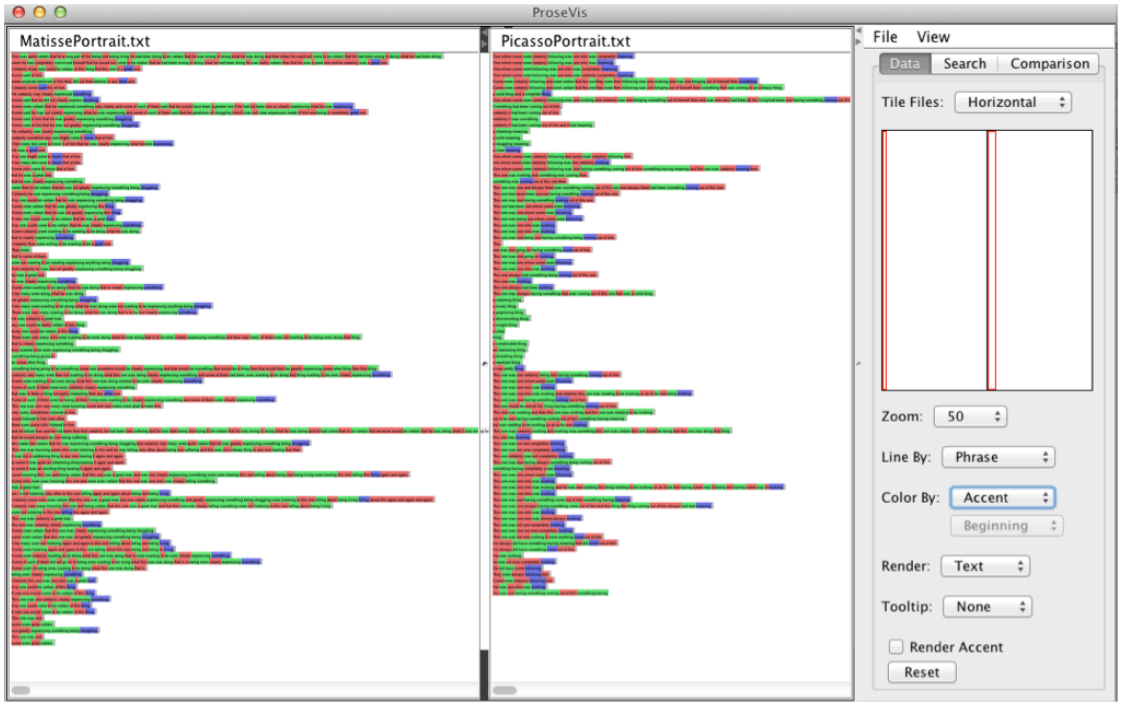

Figure 11.

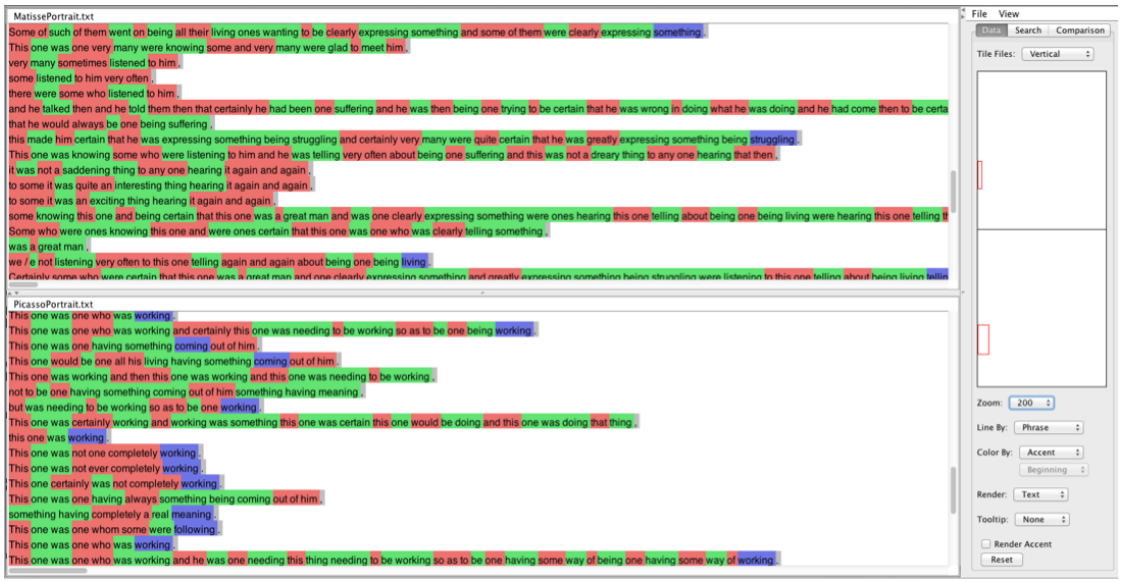

Figure 12.

Other comparative patterns are clear as well. In Figure 5, in which full sounds are represented, and each vertical line is a phrase, there is an inversion of patterns between the two pieces. In “Picasso,” phrases (represented in each line) begin with the yellow/red pattern and evolve into the blue/green pattern at the end of the phrase (or line). The reverse is true of the color sequences in “Matisse.” A closer look at these patterns in Figure 6 shows that Stein starts phrases about Picasso with specific referents to him such as “This one,” while phrases about Matisse begin with more general referents such as “Some” as in “Some of a few.” Conversely, while the “Picasso” phrases evolve into expressing an abstract notion of a thing as in “something” and end again with the specific reference to him again in “one,” the “Matisse” phrases start with vague language (“Some”) and get more specific in the middle of the phrase (referring to “he”) and ending with vague terms referring to an abstract “thing”. These patterns or expression are highlighted in this visualization because they are emphasized by certain sounds. In the “Picasso” phrases, ending sounds are ones created by first opening and then closing your lips such as the “o” and “m” and “n” sounds in “some” and “one” — nonetheless with the “om” sound, your throat remains open. The prevalent Matisse sounds are one would make by beginning with closed lips and ending with widened or more opened lips such as the “i” and “e” sounds in “thing” and “he” — in this case the reader is closing off the breath, squeezing it with her mouth. One could argue that Stein’s play with sounds shows how she represents these artists: the “Picasso” sounds are open, contributing to the sense of “fluidity” upon which Salkind remarks; the “Matisse” sounds, on the other hand, shorten the breath and restrict the mouth’s movement into the next sound. The visualizations facilitate our ability to examine how the words in context correspond to these sound patterns. 43

Tender Buttons (1914)

For our predictive modeling study, we compared the sounds of Gertrude Stein’s *Tender Buttons* to that of *The New England Cook Book* [Turner 1905]. Margueritte S. Murphy hypothesizes that *Tender Buttons* “takes the form of domestic guides to living: cookbooks, housekeeping guides, books of etiquette, guides to entertaining, maxims of interior design, fashion advice” [Murphy 1991, 389]. By writing in this style, Murphy argues, Stein “exploits the vocabulary, syntax, rhythms, and cadences of conventional women's prose and talk” to “[explain] her own idiosyncratic domestic arrangement by using and displacing the authoritative discourse of the conventional woman's world” [Murphy 1991, 383–384]. Murphy sites *The New England Cook Book* (*NECB*) as a possible source with which to compare the prosody of *Tender Buttons*: 44

> Toklas, of course, collected recipes, and she later published two cookbooks, *The Alice B. Toklas Cookbook* (1954) and *Aromas of Past and Present* (1958). Through Toklas then, at least, Stein was familiar with the genre of the cookbook or recipe collection and would appropriately “adopt” and parody that genre in writing of their growing intimacy. Significantly, Toklas’s name as “alas” appears repeatedly in

"Cooking" as well as elsewhere in *Tender Buttons*. [Murphy 1991, 391]

45 It is immediately clear from a simple frequency analysis that the word "Alas" only appears in the one "Cooking" section in *Tender Buttons*, albeit it appears there thirteen times. In order to analyze whether the texts had similar prosodic elements, however, we attempted to make this comparison evident with predictive modeling.

46 To focus the machine learning on this hypothesis, we chose nine texts for comparison and only used features that research has shown reflect prosody such as part of speech, accent, stress, tone, and break index. The nine texts we chose were "Picasso" (1912), "Matisse" (1912), *Three Lives* (1909), *The Making of Americans* (1923), *Ulysses* by James Joyce (based on the pre-1923 print editions), *The Iliad*, translated by Andrew Lang, Walter Leaf, and Ernest Myers (1882), *The Odyssey*, translated by S.H. Butcher and Andrew Lang (1882), and of course, *Tender Buttons* (1914) and *The New England Cook Book* (1905).[18] The features we included from the OpenMary data do not include the word or the sound. The break index, which marks the boundaries of syntactic units such as an intermediate phrase break, an intra-sentential phrase break, a sentence-final boundary, and a paragraph-final boundary, is particularly important because readers (and correspondingly the OpenMary system) use phrasal boundaries to determine the rise and fall or emphases of particular words based on their context within the phrase (Soderstrom, et. al). As mentioned, to further bias the system towards the manner in which readers make decisions on sound, we selected a window size that represented the average size of a phrase across the nine texts. We also hypothesized, in order to measure the tool's efficacy, that *The Odyssey* and *The Iliad* are most similar.

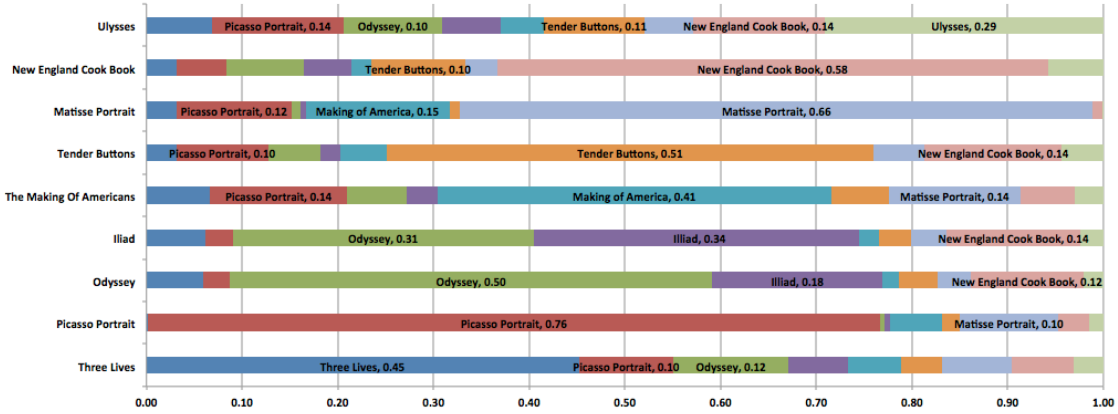

Figure 13.

47 First, we defined a prediction problem for machine learning to solve: Predict from which book the window of prosody features comes. Figure 13 visualizes the results of our predictive analysis. The analysis results show that machine learning makes the same similarity judgment that Murphy had made: *Tender Buttons* and *NECB* sound more similar to each other than they do to any of the other books in the set. In the visualization, row four shows the results for *Tender Buttons* in which the analysis has chosen *NECB* as the matching text more often (at 14%) than any other book including others by Stein ("Picasso" is chosen 10% of the time). As well, row two, which shows the results for *NECB*, shows that *Tender Buttons* is chosen more often (10%) than any other book when trying to predict the actual class or the book itself. Another prediction that shows the algorithm's success is expressed in rows six and seven, which indicate that the computer confuses *The Odyssey* and *The Iliad* – texts that are known to be very similar in terms of prosody — the highest percentage of times. Interestingly enough, while both the results for the *Iliad* and the results for the *Odyssey* show a high correlation with the cookbook (14% and 12% respectively), neither the results for *Tender Buttons* or for the cookbook show a high correlation with Homer's texts. This seems to indicate that the aspects of *Tender Buttons* and the cookbook that make them sound like each other are not those that make the cookbook sound like Homer's texts. Further, the fact that *Tender Buttons* has very little correlation with texts that are seen as similar to the *NECB*, shows how strongly *Tender Buttons* is correlated with the other texts in the set. Consequently, its consistent correlation with the cookbook is a much stronger match than might otherwise be indicated.

48 Using the ProseVis interface, we can see within the context of the text where these associations have been made.

Figure 14 shows *Tender Buttons* and *NECB* in ProseVis. In both panes, each sound is highlighted according to which book it is most like. When all the books are selected, the color of the book that has the highest probability for a given sound is shown. As well, sounds are brighter or less so depending on the level of probability. Figure 15 shows *Tender Buttons* compared to “Picasso” with only a subset of texts “turned on” including NECB, *The Making of Americans*, “Matisse,” and *Three Lives*. In this view, *Tender Buttons* again shows a comparatively higher number *NECB* (pink) matches than the shorter “Picasso” text. In the close-up view in Figure 16, it is easier to see the lighter and darker shades of pink (in the line “alas the back shape of mussle”) and yellow (in the line “alas a dirty third alas a dirty third”). The darker shades indicate that the probability that the sound is more like *Tender Buttons* or *NECB* is greater.

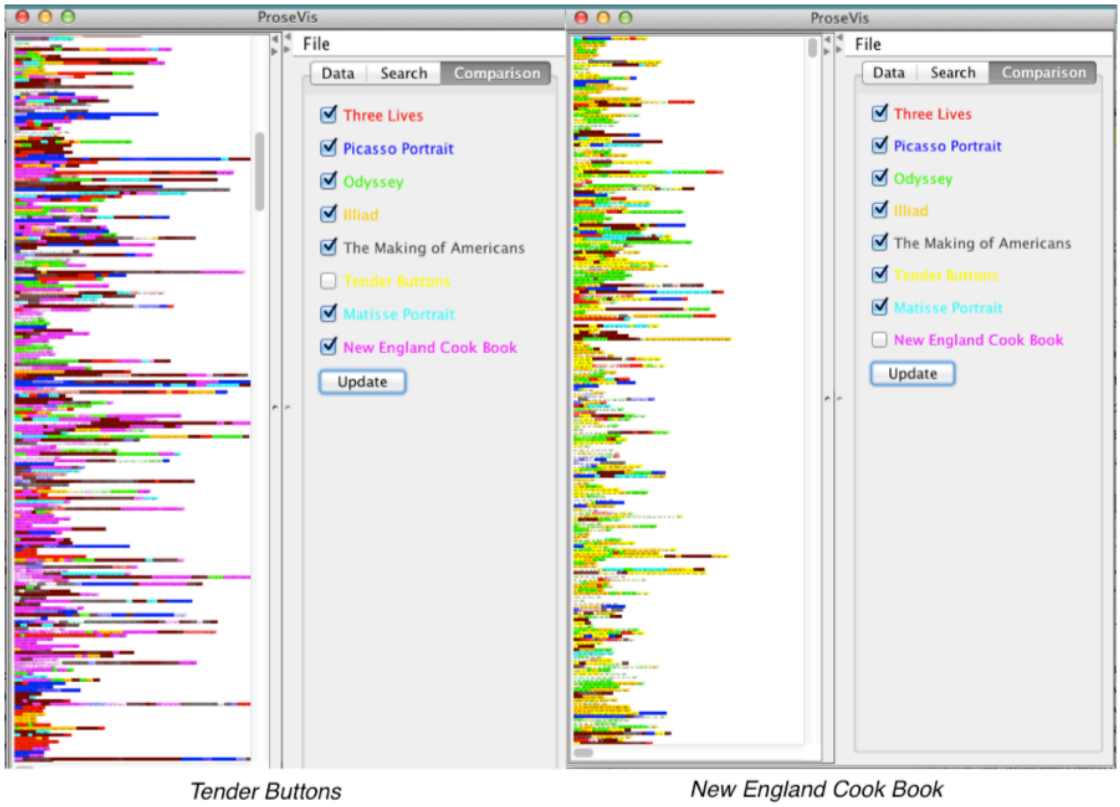

Figure 14.

[19]

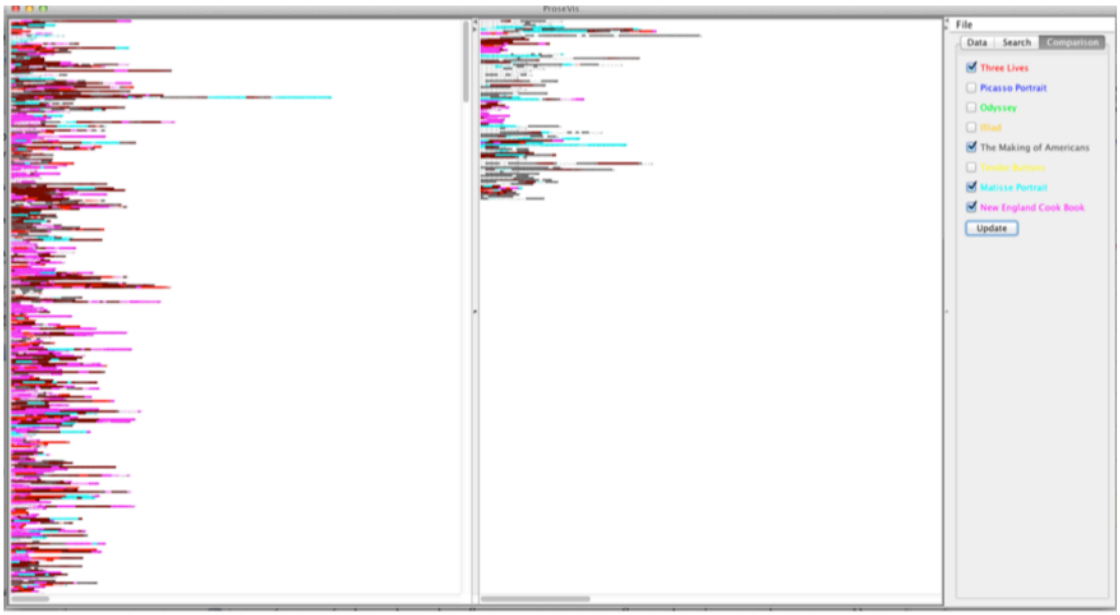

Figure 15.

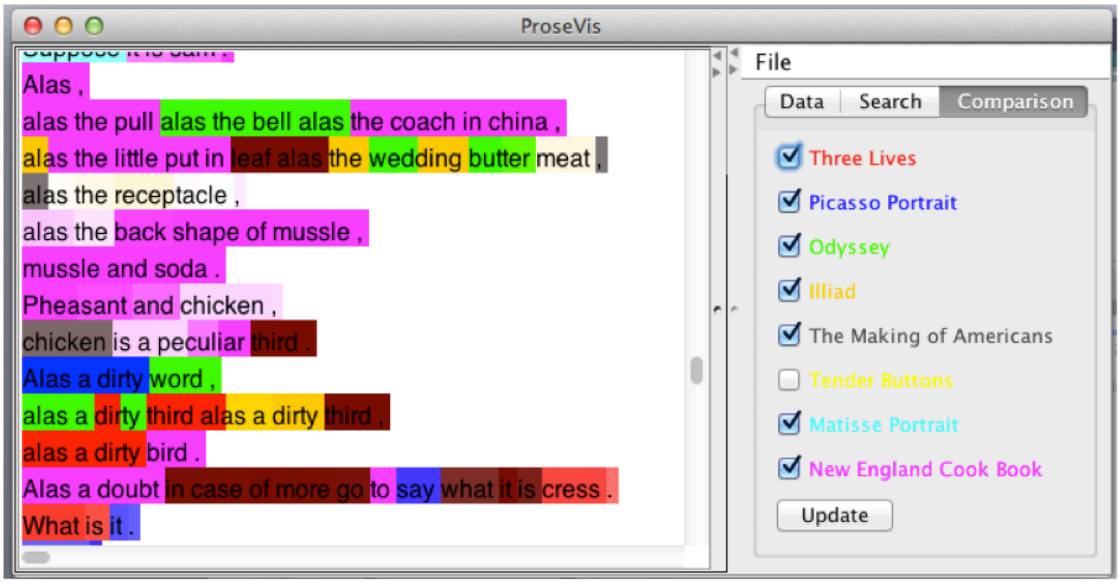

Figure 16.

The story the visualizations tell is two-fold. First, these visualizations are useful in allowing us to test or generate hypotheses about prosody and sound in the texts. Figure 16 shows us that the section surrounding “alas” is, in fact, more like *NECB* than the other books, supporting Murphy’s hypothesis that the area around “alas” has the rhythm of *NECB*. At the same time, we can see in Figure 15 that all of the sections of *Tender Buttons* are *not* most like *NECB*. In this figure, the top of the view is colored red, grey, and light blue, indicating that this area is more like *Three Lives*, *The Making of Americans*, and “Matisse” respectively. A subsequent research question could concern the nature of this difference. Further, Figure 17 shows two views of *Tender Buttons* visualized in ProseVis. On the left, the same list is divided into two colors showing that half the list is correlated with *NECB* and half the list is more strongly correlated with *The Making of Americans*. On the right, there is another list with half of the list correlating more strongly to the *Odyssey* and the bottom half correlating to “Picasso.” This visualization immediately engenders questions concerning why the first part of the list is different than the second half when the two halves seem remarkably similar.

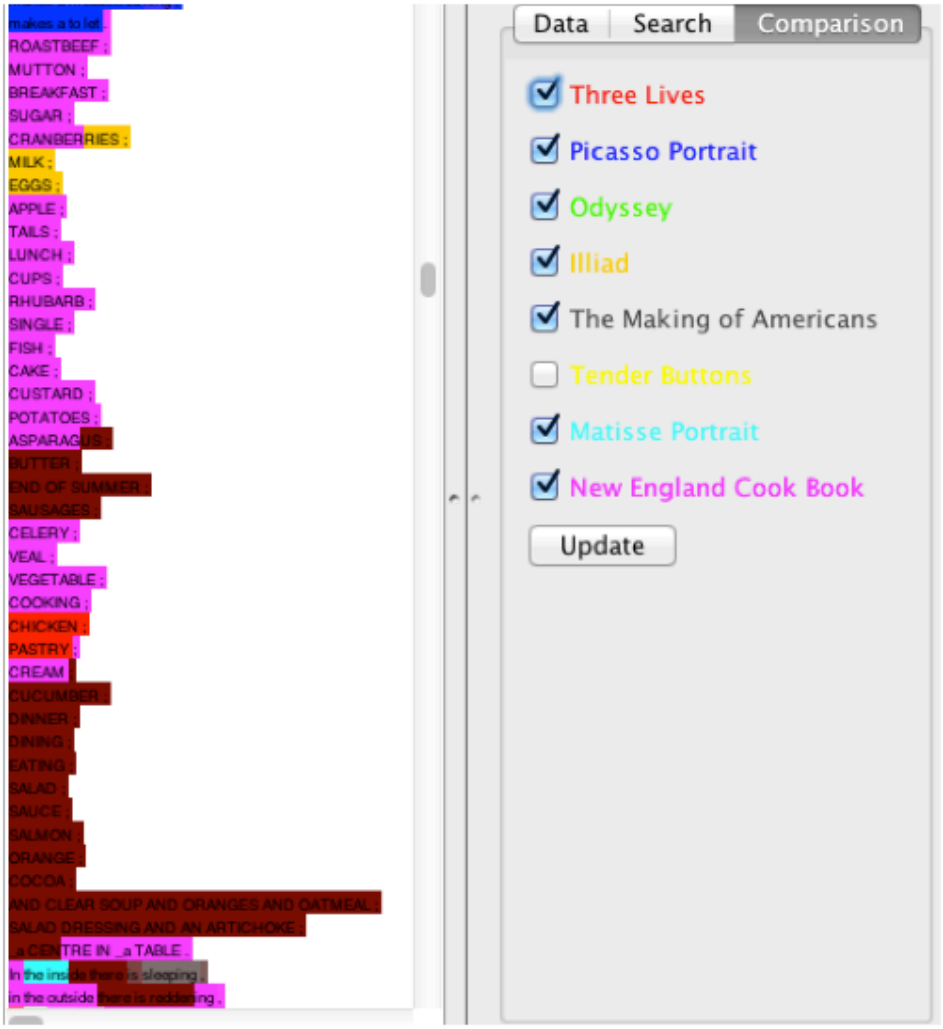

Figure 17.

Second, other questions and hypotheses may be raised concerning how the algorithm and ProseVis work together to generate these visualizations. These latter kinds of questions can be considered in terms of the data sets and the documentation we are providing as well as in respect to articles such as this one. In other words, the goal is not accurate text identification using prosody features, but rather to test hypotheses that consider the sound and prosodic similarities of texts. Part of what we interested in digital humanities are the mistakes we perceive are made by the computer and what these errors reveal about algorithms we are using to gauge the significance of textual features. In other words, one benefit to scholarship represented in this research is determining where the model breaks down and where the ontology must be tweaked. For example, currently, the machine learning system is not being tuned to produce the most accurate classifications: using more context such as a larger window size (i.e., a larger number of phonemes to consider as part of a window) increases the classification accuracy dramatically.[20] As well, if we take parts of speech out of our analysis, our results are less clear. Keeping in mind that we are modeling the possibility of sound *as it could be perceived* opens space for discovery and illumination since what we are not only identifying in this process which text is more like the other (though this is interesting). Rather, by focusing on where the ontology breaks down under the weight of computation, we are learning more about how knowledge representations (our modeling of sound, for instance) are productive for critical inquiry in literary texts.

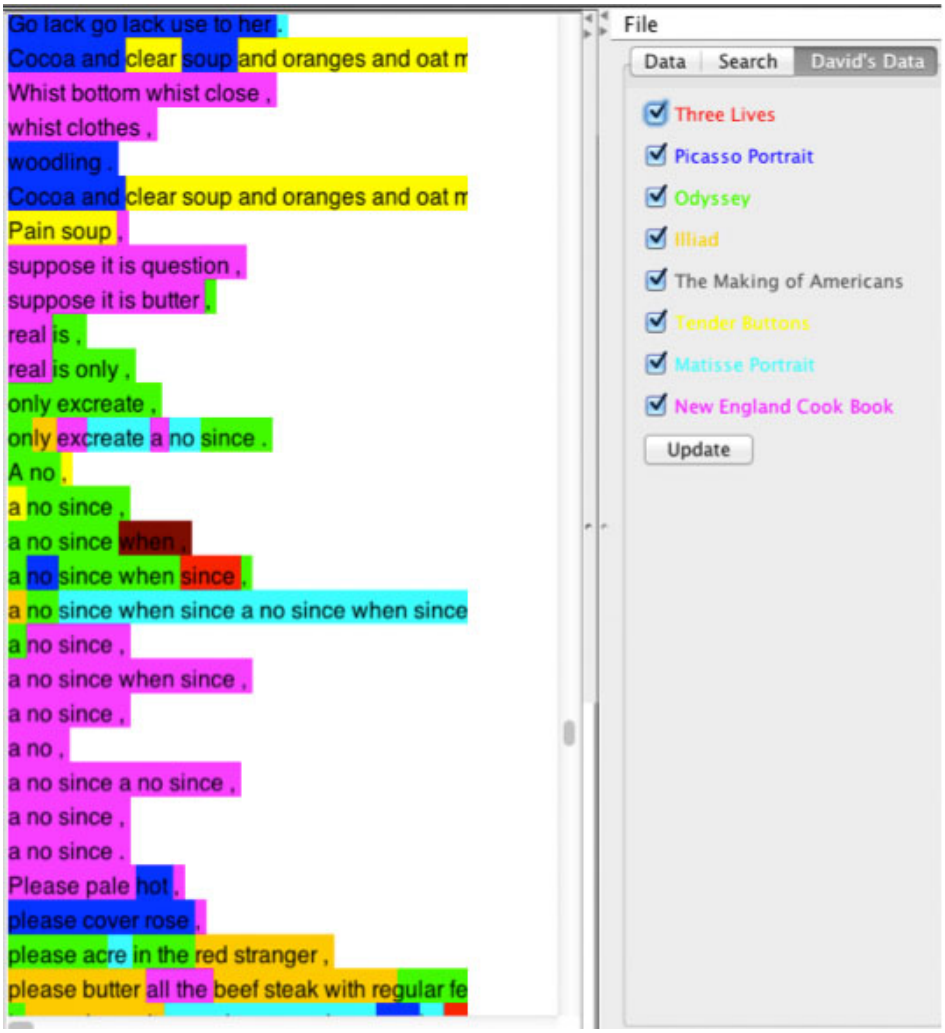

Figure 18.

Conclusion

Previously, digital humanities scholars have also used phonetic symbolic research to create tools that mark and analyze sound in poetry. For instance, Marc Plamondon created AnalysePoems to analyze the Representative Poetry Online (RPO) website (http://rpo.library.utoronto.ca). Plamondon's goal with AnalysePoems was to "automate the identification of the dominant metrical pattern of a poem and to describe some basic elements of the structure of the poem such as the number of syllables per line and whether all lines are of the same syllabic length or whether there are variations in the syllabic length of the lines in an identifiable pattern" [Plamondon 2006, 128]. Like our project, AnalysePoems is not a tool that attempts to represent the "reality" of a spoken poem, a feat that is impossible because of the ephemeral elements of performance of which a poetry reading comprises. Instead, AnalysePoems is "built on the prosodic philosophy that a full scansion of a poe[Plamondon 2006, 129]. Plamondon's work has been important for the development of the processes described here as it creates a precedent and model for analyzing sound from a perspective that also values pre-speech (aurality) and phonetic symbolism. 51

Another tool that was built to examine how the "phonetic/phonological structure of a poem may contribute to its meaning and emotional power" is Pattern-Finder [Smolinsky and Solokoff 2006, 339]. Smolinksky and Sokoloff's hypothesis — "that feature-patterning is the driving force in the 'music' of the poetry" [Smolinsky and Solokoff 2006, 340] is also important for the creation of our visualization tool, ProseVis. With this tool, we are also interested in allowing readers to identify patterns in the analyzed texts by facilitating their ability to highlight different aspects or "features" of data such as parts of speech, syllables, stress, and groups of vowel and consonant sounds. Like the creators of Pattern-Finder, we are also interested in allowing readers to make groupings of consonant sounds that include plosives, frications, and 52

affricates and groupings of vowels that include those formed in the front or the back of the mouth. Phonetic symbolic research and the creation of these tools demonstrate a precedent for facilitating readings that use these features to analyze text for meaning.

Practically speaking, our system for predictive modeling and the ProseVis tool are in the early stages of development but we are encouraged with the results so far. We predicted that *Tender Buttons* and *The New England Cook Book* would be most similar and that *The Odyssey* and *The Iliad* should be most similar and these predictions were confirmed on our first attempt, but our work in determining whether or not analyzing sound or interpreting with sound in these ways is critically productive requires much more research, development, and experimentation. Future development plans include collecting more use cases from multiple experts and doing cross validation studies before we can have high confidence that we have a useful system that compares well to expert predictions. Similarly, ProseVis has only been used by a few scholars. Use case studies are needed to establish if and how examining sound in this way can be useful to or change the nature of scholarship in areas of text and sound. 53

At the same time, while usability studies are still a future goal in the project, developing the algorithm and the ProseVis interface has already been productive in terms of interrogating the efficacy of our underlying theories of aurality. Our work is a new and promising approach to comparing texts based on prosody, but what is equally promising is that we are ultimately basing our ontology for creating a machine learning algorithm on an underlying logic of potential and inexact sounds as they are anticipated in text. Further, the "success" of the comparison of sounds between texts is based on the extent to which the computer is "confused" about these possibilities. The fact that this is a "best guess" methodology, which stems from theories in artificial intelligence and knowledge representation and is based on potentials and probabilities, suggests that the algorithm and the tool incorporate and function within a space that invites hypothesis generation and discoveries in the sound of text. 54

Notes

[1] Both projects seek to analyze prosodic elements in audio files and in corresponding text files. In any case, they are studying everyday speech rather than the more crafted language that typically informs literary texts.

[2] Other collaborators include humanities professors from Stanford, George Mason University, and University of Illinois at Urbana-Champaign. All of the use cases within the project include research on how humanities scholars can use textual analytics and predictive modeling with visualizations to help scholars interpret large digital collections. In particular, we are demonstrating how humanities scholars can use the open-source SEASR environment developed by the informatics experts at the University of Illinois for research.

[3] Meandre is explained more fully later in the discussion. In brief, it is an environment that allows for assembling and executing data flows. A data flow is a software application consisting of software components that process data. Processing can include, for example, an application that accesses a data store, one that transforms the data from that store and analyzes it with textual analysis, and one that visualizes the transformed results.

[4] A wonderful example of the different ways that people can say the phrase "than I did" can be found as part of the "Prosody Datasets" (https://confluence.cornell.edu/display/prosody/Prosody+Datasets) from the "Harvesting Speech Datasets for Linguistic Research on the Web" project.

[5] See Shrum and Lowrey 2007, 40-47 for an extensive discussion concerning this history.

[6] See http://mary.dfki.de/.

[7] More information on these elements may be needed: in "accent" the "L" indicates a low pitch and "H" indicates a high pitch; a character followed by * gives the pitch of the stressed syllable; pitches preceding and following the stressed pitch are separated from the stressed pitch by "+" and "!" represents a downstep onto the following pitch; "g2p_method" indicates how the phonetics were found; "breakindex" indicates the type of boundary: "2" = a potential boundary location; "3" = an intermediate phrase break; "4" = an intra-sentential phrase break; "5"= a sentence-final boundary; "6" = a paragraph-final boundary; "tone" indicates the tone ending the previous phrase

[8] For an example, see [Clement 2008].

[9] Please see [Plamondon 2006]; [Sapir 1929]; [Saussure 1916]; [Shrum and Lowrey 2007]; [Smolinsky and Solokoff 2006] on phonetic

symbolism; [Bolinger 1986] and [Cole 2011] on prosody and intonation; and [Tsur 1992] on speech perception.

[10] This flow and ProseVis are available for trial at http://tclement.ischool.utexas.edu/ProseVis/.

[11] A simple approach works well if all books were of the same size resulting in the same number of phrase windows, but they are not. If we do not modify the instance-based algorithm, it will tend to predict the class of the largest book. To address this problem, we normalize the predicted probabilities under the assumption that all classes are equally likely.

[12] Our distance function is simply a count of the number of mismatches between the windows, which ranges from 0 to 14. The function that computes example weights on the basis of distance is: weight = 1 / (distance p) where p is the “distance weighting power”. When p is set to a maximum, only the single nearest neighbor is used to make predictions and when p is set to a minimum, all memorized examples contributed equally. For any given problem there is a “sweet spot” where the highest accuracy is achieved. This optimal parameter setting is different for each variation of the problem and is also affected by the number of training examples.

[13] Doing the trillions of comparisons per bias point is too much work to be quickly solved using a single CPU. Fortunately, we now have access to highly parallel computing system with GPUs (Graphical Processing Units). The fastest GPU used in this project is a GTX 580, which has 512 cores (tiny CPUs), all of which are simultaneously used to solve the problem. Using GPUs has made it possible for us to approach a problem of this scale. To make these comparisons efficiently, we implemented the instance-based algorithm on a GPU using NVIDIA's CUDA framework. Using the GTX 580 with 512 GPU cores simultaneously, the analysis only took 47 minutes.

[14] ProseVis has been developed in a two-stage process, first as VerseVis: Visualizing Spoken Language Features in Text by graduate students Christine Lu, Leslie Milton, and Austin Myers as part of a graduate course in visualization with Ben Shneiderman at the University of Maryland, College Park. Megan Monroe further developed the prototype as ProseVis under the auspices of this grant.

[15] Another display issue we discovered concerns that fact that while each syllable takes approximately the same amount of time to speak, variations in letter count and letter width can render one sound as significantly longer than another in the display. To address the final issue of variable syllable width, we built a display option that normalizes every syllable to a uniform width. Using this option, syllables are displayed without text, using a series of equally sized blocks. This gives the user a more normalized view that does not confuse syllable size in terms of actual character-size on the page with syllable size in terms of timing and cadence.

[16] A list of the consonant and vowel sounds is available at http://mary.opendfki.de/wiki/USEnglishSAMPA. However, over the course of testing ProseVis, we uncovered two additional vowel components, @U and EI, which are now included in the implementation. Nearly every syllable in the data contains a vowel component. Across the sample Stein files (“Matisse”, “Picasso”, and “Miss Furr and Miss Skeene”) that we analyzed during testing, we only found four words that contained a syllable that did not include a vowel component. The words were “Struggle,” “Skeene,” “Stay,” “Stayed.” In all four of these words (which accounted for 67 instances across all files), the leading “S” is separated into its own sound. The reason for this breakdown is that sounds in the OpenMary data are broken into parts to correspond to how words are spoken, not necessarily according to syllables. Some sounds lack one or more of these components, and the absence of any such component will be assigned a color as well.

[17] In this example, the “Picasso” and “Matisse” portraits are deselected. Otherwise, their “colors” would override the others. This is explained in further detail below.

[18] The texts by Stein were works that were written during the same time period as *Tender Buttons*. The works by Homer and Joyce were chosen based on preliminary work Clement has done comparing the repetition patterns in *The Making of Americans* to these texts [Clement 2008], [Clement 2012]. The translations of the *Iliad* and the *Odyssey* represent those editions that scholars have identified in both Joyce [Schork 1998, 122] and Stein’s [Watson 2005] libraries.

[19] This figure shows ProseVis opened twice because if *Tender Buttons* and *NECB* were both open in the same view, comparing them would be difficult. For example, by selecting a book like the cook book in which most of the sounds correlate to that book, the presentation is overwhelmed by the colors representing that book. Deselecting *NECB* would mean that it would also be deselected in the *Tender Buttons* view. We plan to address this in future development but in this case it is most efficient to open the tool twice.

[20] We ran experiments with the window sizes of 1, 2, 4, 8, 16, 32, and 64. Figure 18 shows the accuracy of each of these experiments. The accuracy is the ability of each window to predict the book it came from for all books. As an example from Figure 13, with a window size of 14, the accuracy of predicting each book varies from 76% for “Picasso” to 29% for *Ulysses*.

Works Cited

Barthes 1978 Barthes, Roland. *Image-Music-Text*. New York: Hill and Wang, 1978.

Becker et al 2006 Becker, S., M. Shröder and W. Barry. *Rule-based Prosody Prediction for German Text-to-Speech Synthesis*. 2006. http://www.dfki.de/~schroed/articles/becker_etal2006.pdf.

Bernstein 1998 Bernstein, Charles. *Close Listening: Poetry and the Performed Word*. Oxford and New York: Oxford University Press, 1998.

Bolinger 1986 Bolinger, D. *Intonation and Its Parts: Melody in Spoken English*. Stanford: Stanford University Press, 1986.

Chow and Steintrager 2012 Chow, R., and J. Steintrager. "In Pursuit of the Object of Sound: An Introduction". *differences* 22: 2-3 (2011), pp. 1-9. http://differences.dukejournals.org/content/22/2-3/1.full.pdf.

Clement 2008 Clement, Tanya. "'A thing not beginning or ending': Using Digital Tools to Distant-Read Gertrude Stein's The Making of Americans". *Literary and Linguistic Computing* 23: 3 (2008), pp. 361-382.

Clement 2012 Clement, T. "The story of one: the rhetoric of narrative and composition in The Making of Americans by Gertrude Stein". *Texas Studies in Literatureand Language* 54: 3 (2012), pp. 426-448.

Cole 2011 Cole, J. "Respondent to Rooth, M. and Wagner, M. on Harvesting Speech Datasets for Linguistic Research on the Web". Presented at *Digging Into Data Conference, Washington DC*, sponsored by National Endowment for the Humanities (June 2011).

Davis et al 1993 Davis, R., R.H. Schrobe and P. Szolovits. "What is a Knowledge Representation?". *AI Magazine* 14: 1 (1993), pp. 17-33. http://www.medg.lcs.mit.edu/ftp/psz/k-rep.html.

Derrida 1991 Derrida, Jacques. "Signature, Event, Context". In Peggy Kamuf, ed., *A Derrida Reader: Between the Blinds*. Hemel Hempstead: Harvester Wheatsheaf, 1991.

Drucker 2011 Drucker, Johanna. "Humanities Approaches to Graphical Display". *Digital Humanities Quarterly* 5: 1 (2011). http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Flanders 2009 Flanders, Julia. "The Productive Unease of 21st-century Digital Scholarship". *Digital Humanities Quarterly* 3: 3 (2009).

Hrushovski 1968 Hrushovski, B. *Poetry*. New Haven: Yale University Press, 1968.

Jolas 1929 Jolas, E. *Introduction*. http://www.davidson.edu/academic/english/Little_Magazines/transition/manifesto.html.

MARY TTS MARY. *Adding support for a new language to MARY TTS*. *MARY Text To Speech*. September 8 2011. http://mary.opendfki.de/wiki/NewLanguageSupport.

McGann 2005 McGann, Jerome. "Culture and Technology: The Way We Live Now, What Is to Be Done?". *New Literary History* 36: 1 (2005), pp. 71-82.

Meyer 2001 Meyer, Steven. *Irresistible Dictation: Gertrude Stein and the Correlations of Writing and Science*. Stanford: Stanford University Press, 2001.

Monk 1998 Monk, Craig. "Sound Over Sight: James Joyce and Gertrude Stein in Transition". In John Brannigan Geoff Ward and Julian Wolfreys, eds., *Re: Joyce: Text, Culture, Politics*. Basingstoke: Macmillan, 1998. pp. 17-59.

Murphy 1991 Murphy, M.S. "Familiar Strangers: The Household Words of Gertrude Stein's Tender Buttons". *Contemporary Literature* 32: 3 (1991), pp. 383-402.

Ong 2002 Ong, Walter. *Orality and Literacy: The Technologizing of the Word*. London and New York: Routledge, 2002.

Peterson 1996 Peterson, C.L. "The Remaking of Americans: Gertrude Stein's 'Melanctha' and African-American Musical Traditions". In H.B. Wonham, ed., *Criticism and the Color Line: Desegregating American Literary Studies*. New Brunswick: Rutgers UP, 1996. pp. 140-157.

Plamondon 2006 Plamondon, M.R. "Virtual Verse Analysis: Analysing Patterns in Poetry". *Literary and Linguistic Computing* 21 (March 2006), pp. 127-141.

Pound 2007 Pound, Scott. "The Difference Sound Makes: Gertrude Stein and the Poetics of Intonation". *ESC: English Studies in Canada* 33: 4 (2007), pp. 25-35.

Salkind 2011 Salkind, W. *Two Portraits: Matisse and Picasso*. *Gertrude Stein Aloud*. January 2011. http://www.gertrudesteinaloud.com/matissepicasso.php.

Sapir 1929 Sapir, E. “A study in phonetic symbolism”. *Journal of Experimental Psychology* 12 (1929), pp. 225-239.

Saussure 1916 Saussure, Ferdinand de. *Course in General Linguistics*. Translated by W. Baskin. New York: McGraw-Hill, 1916.

Schork 1998 Schork, R.J. *Greek and Hellenic culture in Joyce*. Florida: University Press of Florida, 1998.

Shrum and Lowrey 2007 Shrum, L.J., and T.J. Lowrey. “Sounds Convey Meaning: The Implications of Phonetic Symbolism for Brand Name Construction”. In Tina M. Lowrey, ed., *Psycholinguistic Phenomena in Marketing Communications*. Mahwah: Lawrence Erlbaum, 2007. pp. 39-58.

Smolinsky and Solokoff 2006 Smolinsky, S., and C. Sokoloff. “Introducing the Pattern-Finder”. Presented at *Digital Humanities Conference, Paris* (2006).

Soderstrom et al 2003 Soderstrom, M. Seidl, D.G.K. Nelson and P.W. Jusczyk. “The prosodic bootstrapping of phrases: Evidence from prelinguistic infants”. *Journal of Memory and Language* 49: 2 (2003), pp. 249-267.

Sowa 2000 Sowa, J.F. *Knowledge representation*. Pacific Grove, CA: Brooks Cole Publishing Co, 2000.

Stein 1988a Stein, Gertrude. “Matisse”. In *Writings 1903-1932*. New York: Library of America, 1988. pp. 278-281.

Stein 1988b Stein, Gertrude. “Picasso”. In *Writings 1903-1932*. New York: Library of America, 1988. pp. 282-282.

Stein 1990 Stein, Gertrude. “What are Masterpieces and Why Are There So Few of Them?”. In B.K. Scott and M.L. Broe, eds., *The Gender of Modernism*. Bloomington: Indiana University Press, 1990. pp. 495-501.

Stein1988c Stein, Gertrude. “Portraits and Repetition”. In *Lectures in America*. London: Virago, 1988. pp. 165-206.

Tsur 1992 Tsur, Reuven. *What Makes Sound Patterns Expressive?: the Poetic Mode of Speech Perception*. Durham: Duke University Press, 1992.

Turner 1905 Turner, A.M. *The New England Cook Book: The Latest and the Best Methods for Economy and Luxury at Home*. Boston: Chas. E. Brown, 1905.

Unsworth 2002 Unsworth, John. “What is Humanities Computing, and What is Not?”. *Jahrbuch für Computerphilologie* 4 (2002). http://computerphilologie.uni-muenchen.de/jg02/unsworth.html.

Watson 2005 Watson, Dana Cairns. *Gertrude Stein and the Essence of What Happens*. Nashville: Vanderbilt University Press, 2005.