A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry

Justine Kao
Psychology Department
Stanford University
Stanford, CA 94305, USA
justinek@stanford.edu

Dan Jurafsky
Linguistics Department
Stanford University
Stanford, CA 94305, USA
jurafsky@stanford.edu

Abstract

What makes a poem beautiful? We use computational methods to compare the stylistic and content features employed by award-winning poets and amateur poets. Building upon existing techniques designed to quantitatively analyze style and affect in texts, we examined elements of poetic craft such as diction, sound devices, emotive language, and imagery. Results showed that the most important indicator of high-quality poetry we could detect was the frequency of references to concrete objects. This result highlights the influence of *Imagism* in contemporary professional poetry, and suggests that concreteness may be one of the most appealing features of poetry to the modern aesthetic. We also report on other features that characterize high-quality poetry and argue that methods from computational linguistics may provide important insights into the analysis of beauty in verbal art.

1 Introduction

Poetry is nerved with ideas, blooded with emotions, held together by the delicate, tough skin of words.

—Paul Engle (1908 -1991)

Many people have experienced the astounding and transformational power of a beautiful poem. However, little empirical research has been done to examine the textual features or mental processes that engender such a sensation. In this paper, we propose a computational framework for analyzing textual features that may be responsible for generating sensations of poetic beauty. We built a poetry corpus consisting of poems by award-winning professional poets and amateur poets, and compared poems in the two categories using various quantitative features. Although there are many reasons why some poems are included in prestigious anthologies and others are never read, such as a poet's fame, we assume that the main reason is that some poems are perceived as more aesthetically pleasing than others and are thus selected by editors. Given this assumption, we believe that the kind of comparison we propose should be the first step towards understanding how certain textual features might evoke aesthetic sensations more than others.

The next sections review previous computational work on poetry and motivate the features we use; we then introduce our corpus, our analyses, and results.

2 Computational aesthetics

Previous research on the computational analysis of poetry focused on quantifying poetic devices such as rhyme and meter (Hayward, 1996; Greene et al., 2010; Genzel et al., 2010), tracking stylistic influence between authors (Forstall et al., 2011), or classifying poems based on the poet and style (Kaplan & Blei, 2007; He et al., 2007; Fang et al., 2009). These studies showed that computational methods can reveal interesting statistical properties in poetic language that allow us to better understand and categorize great works of literature (Fabb, 2006). However, there has been very little work using computational techniques to answer an important question in both poetics and linguistics (Jakobson, 1960): what makes one poem more aesthetically appealing than

another ?

One such attempt is the "aesthetic measure" proposed by mathematician G.D. Birkhoff, who formalized beauty as a ratio between order and complexity (Birkhoff, 1933). Birkhoff found interesting correlations between the measure and people's aesthetic judgments of shapes, sounds, and poems. While the aesthetic measure enjoyed some success in the domain of visual arts (Rigau et al., 2008), it ran into problems of semantics when applied to language. Birkhoff's aesthetic measure judges a poem's beauty based solely on phonemic features, such as alliterations and assonance, rhymes, and musical vowels. The formula does not capture the subtlety of word choice or richness of meaning in poetry. Since Birkhoff's measure only considers phonetic features, it fails to fully quantify the aesthetic value of meaningful poetic texts.

In this paper, we aim to combine computational linguistics with computational aesthetics. We introduce a variety of theoretically-motivated features that target both poetic style and content, and examine whether each feature is a distinguishing characteristic of poetry that is considered to be beautiful by modern experts and critics.

3 Elements of Craft

One demands two things of a poem. Firstly, it must be a well-made verbal object that does honor to the language in which it is written. Secondly, it must say something significant about a reality common to us all, but perceived from a unique perspective

—W. H. Auden (1907 - 1973)

We review several elements of craft that creative writers and critics reference in their analysis and appreciation of poetry. For each feature that we consider in our model, we provide theoretical motivation from creative writing and literary criticism. We then describe how we computed the values of each feature using tools from computational linguistics.

3.1 Diction

Aristotle argued that good writing consists of a balance of ordinary words that make the writing comprehensible and strange words that make the writing distinguished (Aristotle, 1998). Several hundred years later, Longinus argued that "noble diction and elevated word arrangement" is one of the primary sources of aesthetic language (Earnshaw, 2007; Longinus, 2001). These early scholars of poetic craft passed down the belief that poetic beauty stems from the level of individual words. In her influential creative writing textbook titled, "Imaginative Writing: The Elements of Craft," Burroway (2007) describes poetry as a high-density form of language. Poetic language is usually intentionally ambiguous and often packs several meanings into a compact passage (Addonizio & Laux, 1997). As a result, each word in a poem carries especially heavy weight and must be carefully selected and digested. Based on these ideas, we decided to examine whether or not good poetry is defined by the use of sophisticated vocabulary.

Diction can be evaluated from two different perspectives: word frequency, a measure of difficulty, and type-token ratio, a measure of diversity.

Word frequency: Psychologists, linguists, and testing agencies often use word frequency to estimate the difficulty and readability of words and sentences (Marks, Carolyn B. et al., 1974; Breland, 1996). Based on these studies, it is reasonable to predict that poems written by professional poets may contain more difficult words and lower average word frequencies than poems written by amateur poets.

We measured average word frequency using a list of top 500,000 most frequent words from the Corpus of Contemporary American English (COCA) (Davies, 2011). An average log word frequency was obtained for each poem by looking up each word in the poem in the word list and summing up the log word frequencies. The total log frequency was then divided by the number of words in the poem to obtain the average.

Type-token ratio: Readability measures and automatic essay grading systems often use the ratio of total word types to total number of words in order to evaluate vocabulary sophistication, with higher type-token ratios indicating more diverse and sophisticated vocabulary (Ben-Simon & Bennett, 2007; Pitler & Nenkova, 2008). We predict that professional poets utilize a larger and more varied vocabulary and avoid using the same word several times throughout a poem. A type-token ratio score was calculated for each poem by counting all the separate instances of words and dividing that num-

ber by the total number of words in the poem.

3.2 Sound Device

Poetry has a rich oral tradition that predates literacy, and traces of this aspect of poetic history can be found in sound devices such as rhyme, repetition, and meter. How a poem sounds is critical to how it is perceived, understood, and remembered. Indeed, most contemporary creative writing handbooks devote sections to defining various sound devices and analyzing notable poetry according to interesting patterns of sound (Burroway, 2007; Addonizio & Laux, 1997).

The sound device features described below were computed using Kaplan's 2006 *PoetryAnalyzer*. *PoetryAnalyzer* utilizes the Carnegie Mellon Pronouncing Dictionary to obtain pronunciations of words in each poem and identify patterns indicative of poetic sound devices.

Perfect and slant end rhyme: Rhyme is one of the most well-known and popular sound devices in poetry. The earliest poets used strict rhyme schemes as a mnemonic device to help them memorize and recite long poems. Research in psychology has confirmed poets' intuitions about the powerful effects of rhyme on perception and learning. For example, an aphorism that contains a rhyme is more likely to be perceived as true than a non-rhyming aphorism with the same meaning (McGlone & Tofighbakhsh, 2000). Exposure to rhymes also enhances phonological awareness in young children and can lead to better reading performances (Bryant et al., 1990).

The *PoetryAnalyzer* identifies end rhymes in poems by examining the phoneme sequences at the end of lines. A window of four line endings is analyzed at a time. If two words in the window have different initial consonants but identical phoneme sequences from the stressed vowel phoneme onward, then an instance of a perfect end rhyme instance is recorded. The final count of perfect end rhymes in a poem is normalized by the total number of words. If two words in the window of four line endings have the same stressed vowel but different phonemes following the stressed vowel, then an instance of a slant end rhyme is recorded. The final count of slant end rhymes in a poem is normalized by the total number of words.

Alliteration and consonance: Alliteration is the repetition of consonant sounds at the beginning of words, and consonance is the repetition of consonant sounds elsewhere. In addition to rhyme, alliteration was used as a powerful mnemonic device in ancient epic poetry (Rubin, 1995). Researchers in psychology and discourse analysis have shown that alliteration reactivates readers' memories for previous information that was phonologically similar to the cue (Lea et al., 2008).

The *PoetryAnalyzer* identifies alliteration and consonance as follows. If the initial phoneme of two consecutive words are identical consonants, the alliteration count is incremented. The total count is then divided by the total number of words to obtain a alliteration score for each poem. If there are at least two identical consonant phonemes in a window of nine syllables, the consonance count is incremented. The count is divided by the total number of words in a poem to obtain a consonance score.

Assonance: Assonance is the repetition of vowel sounds. Similar to consonants, different vowel sounds also have their own characteristics and effects. Long vowels take longer to utter and draw out the rhythm and pacing of the line, while short vowels feel brief and urgent (Burroway, 2007).

We calculated an assonance score for each poem in the same fashion as we did for the consonance score, except the target phonemes are vowels instead of consonants.

3.3 Affect

Studies have shown that poetry allows mental health patients to explore and reinterpret their emotions in useful ways. Through reading and writing poetry, patients are able to freely express their thoughts without the constraints of form and logic (Harrower, 1972). On the other hand, critics of poetry therapy have suggested that writing poetry may be harmful to psychological health, because it allows the poet to immerse herself in an inexplicable emotion without having to make sense or order out of it (Stirman & Pennebaker, 2001). For example, Silverman & Will (1986) claimed that Sylvia Plath's poetry may have undermined her control mechanisms and contributed to her death. If reading good poetry is found to be cathartic and therapeutic, do skilled poets make more references to psychological states and explore the emotional world with more depth and intensity?

We examine this question using several existing sentiment lexicons available for sentiment analysis research. One is the Harvard General Inquirer, which consists of 182 word categories, including basic sentiment categories, categories for concrete objects, and categories for abstract concepts (Stone et al., 1966). Another sentiment lexicon is the Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001). While the General Inquirer was designed for content analysis, LIWC was designed to facilitate the understanding of individuals' cognitive and emotional states through text analysis. As a result, most of the categories in LIWC involve mental activity, with over $4,500$ words related to affective, social, and cognitive processes. Six categories from the Harvard General Inquirer and two categories from LIWC were selected because they are most suitable for our purpose of analyzing elements of poetic craft. These features are summarized in Table 1.

3.4 Imagery

One of the most important and oft-repeated piece of advice for writers is the following: "Show, don't tell." Burroway (2007) interprets this as meaning: "Use concrete, significant details that address the senses." Effective imagery allows readers to bring in their own associations to understand and truly experience a new emotion, and skilled poets and writers are able to pick out specific sensory details that evoke deeper abstractions and generalizations.

The appeal of concrete imagery may have roots in processes that facilitate learning and memory. Previous research has shown that concrete noun pairs are easier to memorize than abstract noun pairs, which suggests that imagery can enhance the learning of word pairings (Paivio et al., 1966). Other studies have shown that mental imagery facilitates relational association between concepts (Bower, 1970). Furthermore, Jessen et al. (2000) found neural correlates that suggest that concrete nouns are processed differently in the brain than abstract nouns. One of the reasons why we find poetic imagery striking may be due to the psychological power of imagery to evoke rich associations formed by culture and personal experience.

Another reason why imagery is an essential element of poetic craft is that it allows writers to avoid

Feature	Examples
Word frequency	–
Type-token ratio	–
Perfect end rhyme	*floor / store*
Slant end rhyme	*bred / end*
Alliteration	*frozen field*
Consonance	*brown skin hung*
Assonance	*shallower and yellowed*
Positive outlook	*able; friend*
Negative outlook	*abandon; enemy*
Positive emotion	*happiness; love*
Negative emotion	*fury; sorrow*
Phys. wellbeing	*alive; eat*
Psych. wellbeing	*calm; adjust*
Object	*boat; leaf*
Abstract	*day; love*
Generalization	*none; all*

Table 1: Summary of features

falling into cliche, which is the bane of the creative writer's existence. Burroway (2007) writes, "flat writing is... full of abstractions, generalizations, and judgments. When these are replaced with nouns that call up a sense image and with verbs that represent actions we can visualize, the writing comes alive." Many abstract and common concepts can be embodied or evoked by surprising imagery. In our analysis, we predict that skilled poets are more likely to describe concrete objects and less likely to reference abstract concepts. We measure the degree to which a poem contains concrete details rather than abstractions and generalizations using categories from the Harvard General Inquirer (see Table 1).

4 Methods

4.1 Materials

In order to test the defining features of beautiful poetry described in the section above, we constructed a corpus containing poems that vary in quality and "beauty" by some established standard. One way to do this would be to randomly sample poems from various sources and ask experts to rate them for quality and beauty. However, such a method can be expensive and time-consuming. A more efficient way of achieving a similar effect is to sample poems from pre-existing categories, such as poems written by

skilled professional poets versus poems written by amateur poets. We assume that award-winning poets produce poems that experts would consider "better" and more beautiful than poetry written by amateurs. Although there might be exceptions, since for example experts may consider some poems written by amateur poets to be very beautiful and sophisticated, these pre-existing categories for the most part should be a good approximation of expert opinions.

One hundred poems were selected from sixty-seven professional poets whose work was published in a collection of Contemporary American Poetry (Poulin & Waters, 2006). The poets produced most of their work towards the middle and end of the 20th century and are considered some of the best contemporary poets in America (e.g., Louise Gluck, Mary Oliver, Mark Strand, etc.). All of the poets are listed in the website of the Academy of American Poets and many have won prestigious awards. This serves as confirmation that the poets in this collection are widely acclaimed and that their craft is acknowledged and celebrated by poetry experts and literary critics.

We randomly selected one to three poems from each poet, proportionate to the number of poems each poet had in the collection. When an excessively long poem (over 500 words) was selected, we removed it and replaced it with a different poem from the same poet. This served as a rough control for the length of the poems in the corpus. The final selection of one hundred professional poems ranged from 33 to 371 words in length with an average length of 175 words. We believe that these poems are a good representation of work produced by the best and most celebrated poets of our time.

In addition, one hundred poems were selected from amateur poets who submitted their work anonymously to a free and open-to-all website, aptly called "Amateur Writing" (www.amateur-writing.com). At the time of selection, the website had over 2500 amateur poem submissions by registered users. The website contains a diverse set of poems submitted by amateur writers with a wide range of experience and skill levels. We randomly selected one hundred poems from the website and corrected for misspellings and obvious grammatical errors in the poems to control for the effect of basic language skills. The final selection of amateur poems ranged from 21 to 348 words in length with an average length of 136 words.

4.2 Procedures

We implemented the 16 features described in section 3, each of which target one of three separate domains: style, sentiment, and imagery. The sound device scores were computed using *PoetryAnalyzer* (Kaplan & Blei, 2007). For each category taken from the General Inquirer, scores were calculated using the General Inquirer system available on a server (Inquirer, 2011). A score for a certain category is the number of words in a poem that appear in the category normalized by the length of the poem. For the two categories taken from LIWC, scores were calculated by counting the number of words in each poem that match a word stem in the LIWC dictionary and dividing it by the total number of words. A score for each of the features was derived for every poem in the poetry corpus. All scores were then standardized to have zero mean and unit variance across poems.

5 Results and Analysis

To measure the effect of each variable on the likelihood of a poem being written by a professional or an amateur poet, we constructed a logistic regression model in R (R: A Language and Environment for Statistical Computing). For model selection, we used the step-wise backward elimination method. This method begins by building a model using all 16 feature variables. It then recursively eliminates variables that do not significantly contribute to explaining the variance in the data according to the Akaike information criterion (AIC), which measures the amount of information lost when using a certain model. The selection method stops when further eliminating a variable would result in significant loss of information and model fit. The final logistic regression model for the predictors of professional versus amateur poetry is summarized in the formula above (Table 2). Note that the variables included in the final model might not all be statistically significant.

Results show that poem type (professional or amateur) is significantly predicted by eight different variables ($p < 0.05$): type token ratio, perfect

Probability(poem type = professional $|X$), where

$$X\beta \quad -0.6071 =$$

-0.5039 *	average log word frequency	+
0.6646 *	type token ratio	+
0.4602 *	slant end rhyme frequency	+
-2.1 *	perfect end rhyme frequency	+
-0.6326 *	alliteration frequency	+
-1.0701 *	positive outlook words	+
-0.7861 *	negative emotional words	+
-0.5227 *	psychological words	+
1.3124 *	concrete object words	+
-1.2633 *	abstract concept words	+
-0.836 *	generalization words	

Table 2: Model formula

Feature variable	Odds	p-value
type token ratio	1.94	0.0308
perfect end rhyme frequency	0.12	$5.06e^{-7}$
alliteration frequency	0.53	0.0188
positive outlook words	0.34	0.0130
negative emotional words	0.46	0.0244
concrete object words	3.72	0.0002
abstract concept words	0.28	0.0027
generalization words	0.43	0.0035

Table 3: Odds ratios and p values of significant predictors of professional poetry

Professional		Amateur	
Word	**Count**	**Word**	**Count**
tree	29	thing	40
room	20	wall	12
thing	18	bed	11
grass	17	clock	7
wall	14	room	7
flower	13	tree	6
glass	13	leave	6
floor	13	gift	5
car	12	mirror	4
dirt	11	flower	4
[. . .]	538	[. . .]	103
Proportion	4.1%	Proportion	1.5%
Type count	250	Type count	85

Table 4: Concrete words

end rhyme frequency, alliteration frequency, positive outlook words, negative emotional words, concrete object words, abstract concept words, and generalization words. The other nine variables: average log word frequency, slant end rhyme frequency, assonance, consonance, negative outlook words, positive emotional words, physical well-being words, and psychological words did not have significant predictive value. While positive outlook and positive emotion were highly correlated ($r = 0.54$), as were negative outlook and negative emotion ($r = 0.53$), there was no collinearity among the variables in the final logistic regression model selected by the backward elimination method.

The model predicts the likelihood of the poem type (professional or amateur) using the formula described in Table 2. The influence of each feature is represented by the coefficient β for each variable. A positive value for a coefficient increases the likelihood of a poem being written by a professional. For example, type token ratio and concrete object words have positive coefficient values; thus higher type token ratios and more concrete object words increase the likelihood of a poem being a professional poem. A negative value for a coefficient decreases the likelihood of a poem being written by a professional. For example, perfect end rhyme frequency has a negative coefficient value, and thus higher perfect end rhyme frequencies decrease the likelihood of a poem being written by a professional poet. The relative odds and p-values of each significant predictor variable are presented in Table 3.

In summary, professional poems have significantly higher type-token ratios, contain fewer perfect end rhymes, fewer instances of alliteration, fewer positive outlook words, fewer negative emotional words, more references to concrete objects, less references to abstract concepts, and fewer generalizations. From the odds ratios, we can see that the most significant predictors of professional poetry are fewer perfect end rhymes and more references to concrete objects.

6 Discussion

What are skilled poets doing differently from amateurs when they write beautiful poetry? Based on results from our regression model, it appears that Aristotle may have been wrong about diction, at least for modern poetry. The words in professional po-

Professional		Amateur	
Word	**Count**	**Word**	**Count**
day	40	day	54
night	31	time	33
year	25	beauty	25
time	20	soul	16
death	11	night	15
new	9	new	14
morning	8	moment	13
childhood	7	christmas	12
hour	7	think	11
afternoon	7	future	9
[. . .]	139	[. . .]	143
Proportion	1.8%	Proportion	2.6%
Type count	82	Type count	75

Table 5: Abstract words

Professional		Amateur	
Word	**Count**	**Word**	**Count**
all	63	all	82
nothing	26	never	46
never	19	always	43
always	14	nothing	21
every	11	every	15
any	10	forever	14
anything	5	anything	7
nobody	5	any	6
everything	5	everything	5
forever	3	everyone	4
Proportion	$< 1\%$	Proportion	1.8%

Table 6: Generalization words

etry are not significantly more unusual or difficult than words used by amateur writers. This suggests that contemporary poets are not interested in flowery diction or obscure words, but are focused on using ordinary words to create extraordinary effects.

However, professional poets do use more distinct word types. The 100 poems written by professional poets contain a total of $18,304$ words and $4,315$ distinct word types (23.57%). The 100 poems written by amateur poets contain a total of $14,046$ words and $2,367$ distinct word types (16.85%), a much smaller portion. In aggregate, professional poets have a larger and more varied vocabulary than amateur poets. Moreover, professional poets use a significantly larger number of word types within each poem. Although professional poets do not use more difficult and unusual words, higher type-token ratio is a significant predictor of professional poetry, suggesting that professional poems may be distinguished by a richer set of words.

The results on sound devices provide interesting insight into the current stylistic trends of contemporary professional poetry. While sound devices have a long history in poetry and are considered a feature of poetic beauty, contemporary professional poets now use these devices much less often than amateur poets. Sound devices that were traditionally important in poetry for mnemonic purposes, such as rhyme and alliteration, are more prevalent in amateur poems. Even subtle and sophisticated sound devices like slant rhyme, consonance, and assonance are not significant indicators of professional poetry. These results suggest that repetition of sound is becoming a less aesthetically significant poetic device among contemporary masters of poetry.

In terms of affect, our results suggest that poems by professional poets are not more negatively emotional—at least not explicitly. On the contrary, amateur poets are significantly more likely to reference negative emotions than professional poets. Our results reveal an interesting distinction between words with positive and negative outlooks and connotations versus words that reference positive and negative emotions. While the two pairs of categories are strongly correlated, they capture different aspects of a text's emotional content. The positive and negative outlook categories contain many words that are not emotions but may evoke certain emo-

tional attitudes, such as *clean* and *death*. The fact that professional poets are significantly less likely to use explicitly negative emotion words than amateur poets, but not significantly less likely to use negatively connotative words, suggests that professional poets may evoke more negative sentiment through connotation rather than explicit descriptions.

As predicted, poems written by professional poets contain significantly more words that reference objects and significantly less words about abstract concepts and generalizations. This result suggests that professional poets follow the sacred rule of "show, don't tell" and let images instead of words convey emotions, concepts, and experiences that stick to readers' minds. Professional poets not only use more object words than amateur poets (698 counts versus 205), but they also use a larger and more diverse set of object words (250 types versus 85), as shown in Table 4. Professional poets reference natural objects very often, such as *tree, grass,* and *flower*. On the other hand, the most frequent concrete object word in amateur poems is the extremely vague word *thing*. This suggests that even when amateur poets reference concrete objects, they do not use words that provide specific sensory details.

Our analysis supports the idea that Imagism has strongly influenced the ways in which modern poets and literary critics think about literary writing. Literary critic I.A. Richards argued that image clusters and patterns of imagery are keys to deeper meaning in literary works, and that critics should pay close attention to these patterns in order to understand "the language of art" beneath the surface ordinary language (Richards, 1893). Not only are concrete images able to render the world in spectacular detail, they also provide windows into particular experiences on which readers can project their own perceptions and interpretations.

Consistent with our predictions and with the aesthetic ideals of Imagism, professional poets also make significantly fewer direct references to abstract and intangible concepts (Table 5). If the deeper meaning of a poem is conveyed through imagery, abstract words are no longer needed to reference concepts and experiences explicitly. Moreover, amateur poets use significantly more words concerned with generalizations, as shown in Table 6. While amateur poets embrace the human impulse to generalize, the

skilled poet must learn to extract and report unique details that single out each experience from the rest.

Overall, our results suggest that professional poets are more likely to show, while amateur poets have a tendency to tell. This difference marks the most significant distinction between contemporary professional and amateur poetry in our analysis and may be an essential aspect of craft and poetic beauty.

7 Future directions

Categorizing poetry as professional or amateur is a rather course measure of quality. In order to identify defining features of more fine-grained levels of poetic skill, future work could compare award-winning poetry with poems written by less prestigious but also professionally trained poets. Experimenting with different databases and lexicons for affect and imagery could also be helpful, such as word-emotion associations (Mohammad & Turney, 2011) and imageability ratings (Coltheart, 1981). More sophisticated methods of dealing with sense ambiguities and meaning compositionality in the words in affective lexicons (Socher et al., 2011) should be applied to help enhance and improve upon our current analyses.

While our approach reveals interesting patterns that shed light on elements of poetic sophistication, conclusions from the analysis need to be tested using controlled experiments. For example, does modifying a professional poem to include less concrete words make people perceive it as less beautiful?

In summary, our framework provides a novel way to discover potential features of poetic beauty that can then be experimentally tested and confirmed. By applying both stylistic and content analyses to the quantitative assessment of contemporary poetry, we were able to examine poetic craft on a representative set of poems and reveal potential elements of skill and sophistication in modern poetry.

Acknowledgments

References

Addonizio, K., & Laux, D. (1997). *The Poet's Companion: A guide to the pleasures of writing poetry*. W. W. Norton and Company.

Aristotle (1998). Poetics. *The Critical Tradition: Classical Texts and Contemporary Trends*.

Ben-Simon, A., & Bennett, R. E. (2007). Toward more substantively meaningful automated essay scoring. *The journal of Technology, Learning, and Assessment*.

Birkhoff, G. (1933). *Aesthetic Measure*. Kessinger Publishing.

Bower, G. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 529–533.

Breland, H. M. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science*, *7*(2), pp. 96–99.

Bryant, P., Maclean, M., Bradley, L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology*, *26*(3).

Burroway, J. (2007). *Imaginative Writing: The Elements of Craft*. Pearson, 2 ed.

Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*(4), 497–505.

Davies, M. (2011). Word frequency data from the Corpus of Contemporary American English (COCA). Downloaded from http://www.wordfrequency.info on May 10, 2011.

Earnshaw, S. (Ed.) (2007). *The Handbook of Creative Writing*. Edinburgh University Press.

Fabb, N. (2006). Generated metrical form and implied metrical form. *Formal approaches to poetry*, (pp. 77–91).

Fang, A. C., Lo, F., & Chinn, C. K. (2009). Adapting nlp and corpus analysis techniques to structured imagery analysis in classical chinese poetry. In *Proceedings of the Workshop on Adaptation of Language Resources and Technology to New Domains*, AdaptLRTtoND '09, (pp. 27–34).

Forstall, C., Jacobson, S., & Scheirer, W. (2011). Evidence of intertextuality: investigating paul the deacon's angustae vitae. *Literary and Linguistic Computing*, *26*(3), 285–296.

Genzel, D., Uszkoreit, J., & Och, F. (2010). Poetic statistical machine translation: rhyme and meter. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, (pp. 158–166). Association for Computational Linguistics.

Greene, E., Bodrumlu, T., & Knight, K. (2010). Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, (pp. 524–533).

Harrower, M. (1972). *The therapy of poetry*. Oryx, London.

Hayward, M. (1996). Analysis of a corpus of poetry by a connectionist model of poetic meter. *Poetics*, *24*(1), 1–11.

He, Z., Liang, W., Li, L., & Tian, Y. (2007). Svm-based classification method for poetry style. In *Machine Learning and Cybernetics, 2007 International Conference on*, vol. 5, (pp. 2936–2940). IEEE.

Inquirer, H. G. (2011). How the general inquirer is used and a comparison of general inquirer with other text-analysis procedures.

Jakobson, R. (1960). Closing statement: Linguistics and poetics. *Style in language*, *350*, 377.

Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000). The concreteness effect: Evidence for dual coding and context availability. *Brain and Language*, *74*(1), 103 – 112.

Kaplan, D. (2006). Computational analysis and visualized comparison of style in american poetry. Unpublished undergraduate thesis.

Kaplan, D., & Blei, D. (2007). A computational approach to style in american poetry. In *IEEE Conference on Data Mining*.

Lea, R., Rapp, D., Elfenbein, A., Mitchel, A., & Romine, R. (2008). Sweet silent thought: Alliteration and resonance in poetry comprehension. *Psychological Science*, *19*(709).

Longinus (2001). On sublimity. *The Norton Anthology of Theory and Criticism*.

Marks, Carolyn B., Doctorow, Marleen J., & Wittrock, M. C. (1974). Word frequency and reading comprehension. *The Journal of Educational Research*, *67*(6), 259–262.

McGlone, M., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*, 424–428.

Mohammad, S., & Turney, P. (2011). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, *59*(000), 1–24.

Paivio, A., Yuille, J., & Smythe, P. (1966). Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-asscoiate learning. *Canadian Journal of Psychology*, *20*(4).

Pennebaker, J., Francis, M., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Erlbaum.

Pitler, E., & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. In *Empirical Methods in Natural Language Processing*, (pp. 186–195).

Poulin, A., & Waters, M. (2006). *Contemporary American Poetry*. Houghton Mifflin Company, eighth ed.

Richards, I. (1893). *Practical criticism: a study of literary judgment*. Transaction Publishers.

Rigau, J., Feixas, M., & Sbert, M. (2008). Informational aesthetics measures. In *IEEE Computer Graphics and Applications*.

Rubin, D. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes*. New York: Oxford University Press.

Silverman, M., & Will, N. (1986). Sylvia plath and the failure of emotional self-repair through poetry. *Psychoanal Q*, *55*, 99–129.

Socher, R., Pennington, J., Huang, E., Ng, A., & Manning, C. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 151–161). Association for Computational Linguistics.

Stirman, S. W., & Pennebaker, J. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*(4), 517–22.

Stone, P., Dunphry, D., Smith, M., & Ogilvie, D. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press.