



Linguistic Features of Humor in Academic Writing

Stephen Skalicky (Corresponding author)

Department of Applied Linguistics & ESL, Georgia State University, United States
E-mail: sskalicky1@gsu.edu

Cynthia M. Berger

Department of Applied Linguistics & ESL, Georgia State University, United States

Scott A. Crossley

Department of Applied Linguistics & ESL, Georgia State University, United States

Danielle S. McNamara

Department of Psychology, Arizona State University, United States

Doi:10.7575/aiac.all.v.7n.3p.248

Received: 13/02/2016

URL: <http://dx.doi.org/10.7575/aiac.all.v.7n.3p.248>

Accepted: 12/04/2016

Abstract

A corpus of 313 freshman college essays was analyzed in order to better understand the forms and functions of humor in academic writing. Human ratings of humor and wordplay were statistically aggregated using Factor Analysis to provide an overall *Humor* component score for each essay in the corpus. In addition, the essays were also scored for overall writing quality by human raters, which correlated ($r = .195$) with the humor component score. Correlations between the humor component scores and linguistic features were examined. To investigate the potential for linguistic features to predict the *Humor* component scores, regression analysis identified four linguistic indices that accounted for approximately 17.5% of the variance in humor scores. These indices were related to text descriptiveness (i.e., more adjective and adverb use), lower cohesion (i.e., less paragraph-to-paragraph similarity), and lexical sophistication (lower word frequency). The findings suggest that humor can be partially predicted by linguistic features in the text. Furthermore, there was a small but significant correlation between the humor and essay quality scores, suggesting a positive relation between humor and writing quality.

Keywords: humor, academic writing, text analysis, essay score, human rating

1. Introduction

Academic writing and humor would seem an unlikely pairing. Especially in contexts of higher education, where students are often ranked and sorted into classes based on diagnostic essays and SAT scores and where academic writing can have serious consequences for students' futures. Traditional advice for academic writing in the United States exhorts writers to compose with clarity and cohesion (e.g., American Psychological Association, 2010) and to respond to the social needs of the audience and surrounding contexts (Palmquist, 2010). Humor, on the other hand, relies on semantic incongruity, linguistic ambiguity, and the violation of pragmatic maxims (Attardo & Raskin, 1991). Thus, traditional advice may compel college writers to avoid humor, because being funny would demonstrate a purposeful lack of clarity and cohesion and disrespect the desires of the audience (i.e., teachers and professors) who tend to expect adherence to academic writing norms.

In contrast to academic writing, everyday language is replete with examples of play and humor. Creativity in language is an important method of communication employed not just by the literary and lyrical, but also by everyday people in everyday speech (Cook, 2000). Indeed, humor has many psychological and social benefits that can work to aid communication between interlocutors (Martin, 2007). Although humor may not serve the immediate rhetorical goals of academic writing, evidence of humor in academic writing would be reflective of this general tendency to be creative and playful when communicating.

However, because no studies have investigated the potential role that humor might play in academic writing, the forms and functions of humor in academic writing remain relatively unknown. As an initial investigation into this topic, our study investigates a corpus of college student academic writing that has been rated for writing quality, creativity, and humor. We take a computational approach to investigate these relations. Specifically, we use correlational and regression analyses to examine relations between linguistic features and humor ratings and the relation between humor and essay quality. Our study addresses the following research questions:

1. Are humor ratings related to ratings of essay quality?
2. Do linguistic features of academic writing (e.g., lexical, rhetorical, cohesive) correlate with ratings of humor in academic writing?
3. What amount of variance in essay humor ratings is accounted for by these linguistic features?

2. Computational detection of humor

The current prevailing linguistic view of humor is a model known as the General Theory of Verbal Humor (GTVH; Attardo & Raskin, 1991), which posits that incongruity between speech scripts or schemas is the primary mechanism underlying humor. The perception and resolution of an apparent incongruity is what results in humor. This theory has some empirical support in both neurobiological (Coulson & Kutas, 2001; Sheridan et al., 2009) and psycholinguistic (Vaid et al., 2003) approaches.

Another method to better understand how linguistic features contribute to humorous incongruity comes from the field of computational linguistics. Specifically, researchers interested in the automatic detection of humor have attempted to distinguish humorous from non-humorous texts. Most work in humor detection has relied on automatic text classification methods. Initial investigations of humorous one-liners (i.e., single sentence jokes) have demonstrated that it is possible to automatically distinguish humorous from non-humorous sentences using a variety of features (Mihalcea & Strapparava, 2005, 2006). Specifically, stylistic features, such as alliteration, antonymy, and adult slang; content-based features of the texts (i.e., words specific to text types); and cohesive features (i.e., semantic overlap) of texts were found to distinguish humorous from non-humorous one-liners using computational text selection methods (Mihalcea & Strapparava, 2005, 2006). Follow up investigations of the content-based features of humorous one-liners found humorous sentences to be more human focused and to contain more negative polarity than non-humorous sentences (Mihalcea & Pulman, 2007). These findings demonstrate that simple linguistic features can be used to automatically detect humor.

Additional studies have expanded this line of investigation to include humorous quotes (Buscaldi & Rosso, 2007), web comments (Reyes et al., 2010), and humorous and ironic tweets (Carvalho et al., 2009; Reyes et al. 2012), all resulting in relatively high classification accuracy rates. However, it is important to note that the humorous texts used in these studies are relatively short, such as one-liner jokes (e.g., Mihalcea & Strapparava, 2005, 2006), quotes (Buscaldi & Rosso, 2007), or tweets (Reyes et al., 2012) and that the features identified as predictive for these short texts may not be universally applicable to all types of humor (Reyes et al., 2010). Indeed, when feature sets from these studies have been applied to more complicated forms of humor, such as user-generated web comments, accuracy levels have dropped (to nearly 50% in the case of Reyes et al., 2010).

Such findings have spurred investigation of humor in longer texts. For instance, Reyes and Rosso (2011) analyzed 3000 ironic review comments from Amazon.com and successfully classified ironic from non-ironic web comments with an accuracy ranging from 70.3 to 78.2%. Burfoot and Baldwin (2009) analyzed a corpus of satirical and non-satirical news texts taken from the Internet, and were able to classify satirical from non-satirical news texts with an accuracy ranging from 78.1 to 79.8%. Finally, Skalicky and Crossley (2015) analyzed a corpus of satirical and non-satirical Amazon.com product reviews using text analysis tools that measured the lexical, semantic, and grammatical properties of the texts. Using discriminant function analysis, their model was able to classify satirical from non-satirical texts with 71.7% accuracy. Together these studies investigated similar types of humor (i.e., satirical irony); however, they all used different linguistic features with differing levels of success (as measured by their classification accuracy).

The current study follows in a similar manner to these studies to the degree that we investigate texts that have previously been identified as relatively more or less humorous and use automatic text classification methods. However, unlike previous studies, we do not examine texts that are traditionally considered to have an *a priori* humorous purpose. Indeed, because academic writing is not typically associated with humor, instances of humor in these texts may even be working against the genre within which the authors are operating. Importantly, our study differs from existing text classification studies because we are attempting to predict human ratings of humor in essays, and, in turn, to better understand what human raters attend to when evaluating humor (or the lack thereof). Thus, our approach is similar to that used by researchers attempted to predict human ratings of quality using linguistic indices (e.g., Crossley & McNamara, 2010; Deane, 2014; McNamara, Crossley, & McCarthy, 2010; Pitler & Nenkova, 2008). These studies generally identify which linguistic features in a text strongly associate with analytic ratings of writing quality, and whether essay scores are attributable to assumed features of quality (such as cohesion and lexical sophistication). Like these studies, the current study is based on the notion that the same phenomenon occurs when raters are asked to judge the presence or absence of humor in academic writing. In other words, regardless of what elements of writing the raters believe they are attending to when rating essays as more or less humorous, there may be subtle linguistic features that associate more or less strongly with these ratings. Identification of such linguistic features can provide a better understanding of the linguistic features of humor itself, as it is manifested in academic writing.

3. Methods

This study investigates the linguistic features related to humor in undergraduate student writing and examines the relation between humor and writing quality. To do so, we first examine human judgments of humor using a number of linguistic indices taken from three text analysis tools: The Tool for the Automatic Assessment of Lexical Sophistication (TAALES; Kyle & Crossley, 2015), The Tool for the Automatic Analysis of Cohesion (TAACO; Crossley, Kyle, & McNamara, 2015), and The Writing Assessment Tool (WAT; McNamara, Crossley, & Roscoe, 2013). We then use these indices to predict human ratings of humor in the essays in order to better understand the text features that are associated with humor in student writing.

3.1 Corpus

The corpus for this study comprised 313 timed essays written by undergraduate freshman composition students at Mississippi State University (MSU). Students were given 25 minutes to respond to one of two randomly assigned SAT prompts. No referencing to outside sources was allowed. All student writers were native speakers of English. The essays in the corpus had been previously rated for overall essay quality using a standardized holistic grading scale (1-6)

commonly used when assessing SAT essays. Collection and background of this corpus is further described in Crossley and McNamara (2011).

3.2 Human ratings

Two separate pairs of trained raters scored each essay using a rubric designed to assess either essay quality (holistically) or essay creativity (analytically). The holistic quality rubric was designed using a standardized rubric associated with the essay portion of the SAT test. The analytic creativity rubric contained seven subscales related to idea generation and style. Four subscales were related to idea generation (fluency, flexibility, originality, and elaboration) whereas three subscales were related to style (humor, metaphor and simile, and word play. Each subscale was rated on a scale of 1 – 6, with raters informed that the distance between each value on the scale was equal. The two rubrics are included in Appendix C. Raters possessed either Masters or Doctoral degrees in English, and all had at least two years of experience teaching writing at the university level.

Each pair of raters first trained with a rubric using a practice set of 20 essays (not included in the MSU corpus) until they reached an inter-rater reliability of at least $r = .60$ for the analytic scores and $r = .70$ for the holistic scores (holistic scores generally reach a higher consensus and thus have a higher threshold). The raters then scored the remainder of the 313 essays independently. After the scoring was completed, differences between the raters' scores were calculated. If the difference was greater than two points for any sub-scale, the two raters adjudicated their scores, and average score between the two raters was computed for each subscale. For the creativity rubric, this process brought most adjudicated scores down to a difference of two or less, but some scores remained at a difference of two or more. Correlations and Kappas for the raters' scores after adjudication are reported below.

Table 1. Inter-rater reliability for essay scores.

Scale	Correlations	Kappa
Holistic quality	0.789	0.745
Fluency	0.801	0.763
Flexibility	0.647	0.642
Originality	0.573	0.533
Elaboration	0.707	0.703
Humor	0.718	0.715
Metaphor and Simile	0.686	0.683
Word play	0.492	0.488

3.3 Linguistic variables

The indices we extracted from TAALES, TAACO, and WAT were pre-selected based on perceived and known links between humor and linguistic features. TAALES is a text analysis tool designed to measure the overall lexical sophistication of a text and includes over 150 different lexical measurements related to lexical frequency, lexical range, psycholinguistic word information, and academic language. TAACO measures the cohesion properties of a text by incorporating over 150 indices related to word overlap, type-token ratios, and use of connectives, as well as local (sentence-to-sentence) and global (paragraph-to-paragraph) measures of cohesion within a text. WAT is a text analysis tool designed to assess overall writing quality and includes a variety of writing-specific lexical, rhetorical, and cohesion indices. Specifically, WAT reports the incidence of certain lexical categories indicative of rhetorical style. These include exemplification, hedges, amplifiers, downtoners, copular verbs, and private and public verbs. WAT also uses latent semantic analysis (LSA; Landauer et al., 2007) to measure cohesion by calculating the semantic overlap (i.e., conceptually related words and phrases across a text) between sentences and paragraphs. In addition, WAT reports on a variety of indices related to lexical sophistication, key word use, and n-grams. The indices selected from these three tools are discussed below.

3.3.1 Basic text properties

We selected basic properties of the text, such as number of words per text, number of total lemmas per text, number of total word types per text, and average sentences per text because the length of the essay may be related to a greater probability for humor to be expressed. Basic text descriptive indices were calculated using WAT.

3.3.2 Grammatical and semantic word properties

We included the WAT word part of speech (POS) type indices related to incidences of pronoun types, verb types, adverbs, adjectives, and nouns because previous humor studies have shown that humor exhibits unique semantic features, such as human-centric language (Mihalcea et al., 2010) and descriptiveness (Reyes et al., 2012). Additionally, we included word indices from WAT designed to measure the overall incidences of negative or positive words in each text based on a number of investigations that have identified negative semantic meanings or polarity as indicative of humor (e.g., Campbell & Katz, 2012; Reyes et al., 2012).

3.3.3 Textual cohesion

We used indices related to semantic overlap, lexical diversity, and givenness reported by TAACO and WAT to capture textual cohesion in student essays based on previous results showing greater semantic distance of shared topics and themes in humorous texts (Mihalcea & Strapparava, 2006; Mihalcea et al., 2010). Because incongruity is widely recognized as an element of humor (Martin, 2007), we hypothesize that greater semantic distance between words, higher lexical diversity, and relatively less givenness within a text may be more predictive of humor.

3.3.4 Rhetorical devices

While all of these texts were written under the purview of an academic genre, we presume that student essays containing humor will contain fewer overt markers of academic writing. One way to measure this is through the frequency of rhetorical devices commonly associated with academic writing. Thus, we included indices that calculate the use of classic rhetorical phrases used to conclude an essay (e.g., “In closing”) or to state a concluding opinion, such as “I think...” or “I believe...”. These indices were calculated using WAT.

3.3.5 Word frequency

Measurements of word frequency indicate how often a particular word is used in a given corpus. Word frequency is typically provided for single words. In addition, frequency can also be calculated for n-grams (i.e., two or more words that frequently pattern together). While few studies have explicitly used word frequency measures in automated assessments of humor (cf. Reyes & Rosso, 2012, who included n-gram frequency in a computational model to predict ironic texts), success in the computational *generation* of humor has relied on the exploitation of simple, unambiguous lexical items in order to generate riddles, puns, and one-liner jokes (Ritchie, 2004). Therefore, we predict that humor in student essays will involve relatively frequent words. Word frequency measures were obtained with TAALES and WAT.

3.3.6 Psycholinguistic properties of words

Several indices indicative of the psycholinguistic properties of words were included. These included word familiarity, imagability, concreteness, meaningfulness, and age of acquisition. To our knowledge, only one previous study of humor has considered this range of word properties: Skalicky and Crossley (2015) found that satire included more concrete words than did non-satire. Further evidence of the potential importance of these indices comes from studies in ironic and figurative language processing, which demonstrate that word salience (i.e., concreteness, familiarity) is crucial for ironic interpretations (Cronk & Schweigert, 1992). Because humor and irony are closely related (Simpson, 2003), we examine whether humorous texts include more familiar, imageable, concrete, and meaningful words. All measures of essays’ psycholinguistic properties were calculated using TAALES and WAT.

4. Statistical Analysis

An exploratory factor analysis was conducted to examine relations between the seven analytic creativity subscales obtained by human raters and to develop weighted component scores based on co-occurrence factors in the ratings found in the creativity rubric (see Results section below). Results from that factor analysis revealed two factors: a *Creativity* component score and a *Humor* component score. Because the current study is primarily concerned with how humor is manifested linguistically in academic writing, only the *Humor* component score was analyzed further in this study. This *Humor* score was used as a dependent variable in a regression analysis to examine the potential for linguistic variables to explain humor in academic writing.

For the selected variables described above, we first removed non-normally distributed indices. We then conducted correlations between the *Humor* component score and the remaining indices to assess which indices reported a meaningful and significant relation ($p < .05$, indicating at least a small effect size; $r \geq .10$) with the *Humor* component. Correlations amongst the indices that demonstrated a meaningful and significant relation were then checked for instances of multicollinearity. If any two indices were highly collinear ($r > .90$), only the index with the strongest relation to the *Humor* component score was retained. Finally, we discarded any of the remaining indices that we were unable to justify theoretically for inclusion. The remaining indices ($n = 24$) were entered as predictor variables into a stepwise multiple regression in order to explain the variance in the *Humor* component scores.

Before carrying out the regression analysis, we divided the student essays into training and test sets using a 67/33 split (67% training, 33% test; Witten et al. 2011), which allowed for cross-validation of the regression model. If a model derived from a training set predicts the outcome variable in the test set at a similar accuracy rate as the training set, the regression model can be considered stable. We first obtained a model from the essays comprising the training set. We then applied that model to the test set to assess its predictive power and overall generalizability.

5. Results

5.1 Scoring subscales

An exploratory factor analysis was conducted using the human scores on the creativity rubric to investigate potential subscales for the ratings. A Bartlett’s test of sphericity was statistically significant ($p < .001$), and the Kaiser-Meyer-Olkin measure of sampling adequacy reported .693, indicating underlying structures. The scree plot suggested the extraction of two factors, which was also supported by the percent of variance explained by the initial Eigenvalues between the second and third factors. The principal axis factoring using a varimax rotation also identified two factors. The items that loaded onto the first factor, which we labeled *Creativity*, were *fluency*, *flexibility*, *elaboration*, *originality*, and *metaphor*. The items that loaded onto the second factor, which we labeled *Humor*, were *humor* and *word play*. All items loaded onto their respective factors with eigenvalues $> .500$ (see Table 2). The *Creativity* and *Humor* subscales were both calculated by weighting the items based on their Eigen weights in the factors and averaging these weighted scores across the items for each factor. For this study, we only focus on the *Humor* subscale, which was used in a subsequent regression analysis, along with the previously discussed linguistics variables, in order to examine the potential for language features to predict the presence of humor in the essays.

Table 2. Factor analysis: Eigen loadings for components

Item	Creativity component	Humor component
Fluency	0.890	
Flexibility	0.832	
Elaboration	0.809	
Originality	0.535	
Metaphor	0.509	
Humor		0.824
Word play		0.615

5.2 Humor component scores and essay quality

The average humor component score for the essays was $M = 1.96$ ($SD = 0.59$). The average essay quality was $M = 3.29$ ($SD = 0.98$). The correlation between the *Humor* component scores and the holistic essay quality scores was $r(313) = .195$, $p < .001$, indicating a small (yet significant) relation (Cohen, 1992).

5.3 Correlations between humor component scores and linguistic indices

As an initial step to identify indices that best predict essays' *Humor* scores, we discarded indices that were non-normally distributed, were not theoretically related to humor, or did not demonstrate a significant correlation with the humor component score ($r \geq .10$, $p > .05$). The 34 remaining indices were then checked for multicollinearity. If any two indices were highly collinear ($r > .90$), the index with the weakest relation to the *Humor* component score was removed. This resulted in the removal of 10 additional indices, and a total of 24 linguistic indices. Correlations between these 24 indices and the *Humor* component score are displayed in Table 3.

The correlations between the *Humor* component score and the linguistic indices are generally weak. Collectively, however, they tell a coherent story. They indicate that the essays scored as more humorous are longer, more descriptive (i.e., more adverbs, more adjectives, more infinitives, greater negativity, more verbs, fewer nouns, greater concreteness), use more distinctive, sophisticated language (i.e., more unique bigrams, less frequent content words), and less cohesive (i.e., lower semantic similarity, greater lexical diversity, less overlap, fewer connectives, lower givenness, fewer conclusion words). Hence, on their own, the correlations provide some insight into the linguistic nature of the humor scores.

5.4 Regression analysis to predict humor component scores

Correlations do not address the question regarding which of those features in the *Humor* component scores influence judgments made by human raters. To address this question, a step-wise regression was conducted to assess which of the 24 indices collectively explained the variance in the *Humor* component score. The regression model, $F(4, 193) = 10.650$, $p < .001$, $r = .425$, $R^2 = .181$, demonstrated that four predictor variables explained 18% of the variance for the 198 essays in the training set (see Table 4). When the model was applied to the test set, the model yielded, $r = .419$, $R^2 = .175$, indicating that the four predictor variables explained 17.5% of the variance in the *Humor* component score for the 115 essays in the test set, and that the model can therefore be considered stable.

Of the four significant predictor variables, three were reported by WAT (*Incidence of adverbs*, *Incidence of adjective predicates*, *Semantic similarity: paragraph-to-paragraph*) and one was reported by TAALES (*Word frequency content words: Kucera-Francis*). The first two of these variables were positive predictors of the *Humor* component score, meaning that as they increased, so did the *Humor* component score. The final two were negative predictors, meaning that as their scores decreased, the *Humor* component score increased. In other words, more adverbs and adjective predicates resulted in higher *Humor* component scores, whereas lower semantic similarity between paragraphs and lower word frequency resulted in higher *Humor* component scores. As such, the regression tells a similar story as the correlations, the essays with more humor were more descriptive (i.e., more adverbs, more adjectives), use more distinctive, sophisticated language (i.e., less frequent content words), and less cohesive (i.e., lower semantic similarity).

Table 3. Correlations between humor component score and computational indices

Index	<i>M</i>	<i>SD</i>	<i>r</i>	Construct	Tool
Incidence of adverbs	26.55	11.42	0.298***	Rhetorical	WAT
Total number of sentences in text	20.06	6.93	0.256***	Rhetorical	WAT
Number of unique bigrams	495.15	182.12	0.247***	Lexical	TAALES
Semantic similarity: Sentence to sentence	0.22	0.06	-0.244***	Cohesion	WAT
Lexical diversity D	73.42	24.17	0.242***	Cohesion	WAT
Overlap of word stems	0.13	0.04	-0.230***	Cohesion	WAT
Adjacent overlap content words: Essay	0.10	0.03	-0.222***	Cohesion	TAACO
Semantic similarity: Paragraph to paragraph	0.20	0.06	-0.221***	Cohesion	WAT
Incidence of adjective predicates	9.83	5.99	0.213***	Rhetorical	WAT

Incidence of infinitives	7.81	4.41	0.203***	Rhetorical	WAT
Density of logical connectives	162.50	76.75	-0.203***	Cohesion	WAT
Incidence of not	4.59	3.39	0.188**	Rhetorical	WAT
Lexical diversity MTLT	77.22	19.34	0.178**	Cohesion	WAT
Givenness	0.33	0.04	-0.170**	Cohesion	WAT
Adjacent overlap nouns: Essay	0.03	0.01	-0.168**	Cohesion	TAACO
Incidence of motion verbs	106.48	60.98	0.154**	Rhetorical	WAT
Word frequency content words: KF	1418.75	245.45	-0.149**	Lexical	TAALES
Word concreteness (Pavio)	83.70	21.72	0.141*	Lexical	WAT
Word frequency content words: Brown	2.04	0.15	-0.132*	Lexical	TAALES
Incidence of conclusion words	7.21	2.50	-0.124*	Rhetorical	WAT
Incidence of concluding statements	4.11	1.54	-0.120*	Rhetorical	WAT
Incidence of noun phrases	266.37	24.50	-0.119*	Rhetorical	WAT
Average frequency of content words: KF	225.96	25.48	-0.118*	Lexical	TAALES
Incidence of plural nouns	83.71	29.41	-0.112*	Rhetorical	WAT

For correlations, * indicates $p < .05$, ** indicates $p < .10$, and *** indicates $p < .001$

Table 4. Stepwise regression analysis and significance values for linguistic indices predicting humor component scores

Entry	Index added	r	R^2	R^2 change	B	B	S.E.	T
1	Incidence of adverbs	0.295	0.087	0.087	0.010	0.193	0.004	2.432
2	Semantic similarity: Paragraph to paragraph	0.362	0.131	0.044	-1.805	-0.208	0.577	-3.128
3	Word frequency content words: KF	0.403	0.163	0.032	-0.001	-0.220	0.000	-3.205
4	Incidence of adjective predicates	0.425	0.181	0.018	0.016	0.169	0.007	2.074

B = unstandardized β ; B = standardized; S.E. = standard error. Estimated constant term is 2.720; all t significant at $< .05$

6. Discussion

This study analyzed a corpus of undergraduate essays in order to better understand the linguistic forms and features of humor in student academic writing. In addition, we also examined the relations between judgments of humor and essay quality. Because humor and creativity serve important roles in communication (Cook, 2000; Martin, 2007), it is important to understand the manner in which humor functions in academic writing, and whether or not humor and essay quality are linked. In general, our results indicate that four linguistic features are predictive of humor in academic writing. We also found a small but positive link between humor and essay quality. Our final model selected four linguistic indices which successfully accounted for 17.5% of the variance in *Humor* scores, suggesting that higher incidences of adverbs and adjective predicates and lower paragraph-to-paragraph semantic similarity and word frequency account for approximately one fifth of the variance in the *Humor* score component. In the remainder of this section, we will discuss these indices in detail and provide examples from essays that loaded the highest into the *Humor* component score.

The index that contributed the most to the regression model was incidence of adverbs (8.7%), which loaded positively into the model, meaning that essays with higher *Humor* scores tended to contain higher numbers of adverbs. The following excerpt comes from an essay that received a *Humor* component score of 5.45 and essay quality score of 2.5 (both scales ranged from 1-6). The author was responding to a prompt on the nature of heroes and celebrities. Adverbs have been italicized for ease of identification:

“Anyway, heroes are cool because they don't even care what you think. They will just wake up and silently think to themselves, “Yep, it's time to be awesome today.” They don't even exclaim that in their heads because that would be so unnecessary and foolish. Alternatively, celebrities wake up all scared and unsure of themselves hoping that the world will approve of them because, “I'm not sure I'll be awesome today...hope everything goes smoothly today and I don't crash my car into a fire hydrant while sneaking away at 3 a.m. to cheat on my wife! ...Cause man that would stink.”

This example demonstrates the author's frequent use of adverbs to modify verbs (“goes smoothly”), adjectives (“all scared”), and entire clauses (“Alternatively, ...”). Semantically, adverbs are typically employed to express degree, convey attitudes, or modify actions (Biber et al., 2002). In this particular excerpt (and in other essays in the corpus), the adverbs function to qualify elements of the narrative characterization of heroes and celebrities in a manner that intensified the actions described. The effect of such purposefully exaggerated narration is both comical and vernacular in tone. In this regard, the narration in the above excerpt is more descriptive, and mirrors spoken language, rather than academic registers.

The second-strongest index in our model was paragraph-to-paragraph semantic similarity, which added 4.4% more to our model's R^2 value. This index is a measure of cohesion that uses latent semantic analysis to calculate the semantic similarity among paragraphs within an essay. This index loaded negatively into our model, meaning essays with higher *Humor* scores tended to have lower paragraph-to-paragraph semantic similarity. In other words, funnier essays were more likely to contain paragraphs whose topics were semantically inconsistent relative to surrounding paragraphs' topics. Of the four indices in the model, semantic similarity has a direct relation to incongruity models of humor (Martin, 2007), as disruptions in the semantic cohesion of an essay may signal to a reader that a section of the essay should *not* be interpreted as academic writing but instead as a humorous aside.

As an example, the same essay quoted above demonstrated a lack of paragraph-to-paragraph similarity in paragraphs three and four of the essay (see Appendix A for the full essay).

“Heroes set out to decide whether or not they approve of the world. If not, they change it by any means necessary without resorting to celebrity-like tactics because that would so totally defeat the purpose of their heroic deeds. If everyone looked up to heroes then the world would have many fewer celebrities in the future. Everyone would become all modest, smart, strong, self-reliant and wise. That's a nice idea but if you think for a minute, fewer celebrities means fewer fools to laugh at which means fewer examples of what not to do. Normal people learn from their mistakes while wise people learn from the mistakes of fools.

Heroes may not always be popular with the law. Batman, for example, was constantly hunted by the police for being a vigilante and for littering. It is a little known fact that batman does not pick up his soda cans. This just goes to show that even though heroes have the best interest of the world in mind, they may not always be perfect themselves.”

The discussion of Batman (a fictional comic book superhero) aligns well with the thesis and topic of this essay, but the author's decision to include this example of Batman's fictional misdemeanors to support the topic sentence of the paragraph is in stark contrast to the previous paragraph, which argued that heroes are distinct from celebrities using very straightforward and academic vocabulary. As a result, the above excerpt contains relatively anomalous lexical choices (e.g., “soda cans” and “littering”) compared to the previous paragraph. In general, this essay is marked by the author's shifts between content topics and writing styles from paragraph to paragraph. Our model suggests that the humor in this essay may have thus been signaled in part by a lack of semantic cohesion between paragraphs.

Word frequency of content words was the third-strongest contributor in our model, explaining 3.2% of the total variance in *Humor* component scores. This index loaded negatively, suggesting that essays with higher *Humor* scores tended to have lower content word frequency. *Content word* is used here to refer to a noun, lexical verb, adjective, or adverbs (as opposed to a *function word*, which typically expresses a grammatical relation, e.g., prepositions). Content words with relatively low word frequency in the essay quoted above were *tactics*, *royal*, *transgression*, *hydrant*, and *vigilante*, among others. Recall that the frequency of words is a measure of their relative use in language, meaning that less-frequent words are less-commonly encountered, and also more distinctive. Of course, infrequently encountered words are not inherently humorous. Rather, we would argue that authors who tend to use humor are using more distinctive language, and as such, are more likely to exhibit rich vocabularies, or lexical sophistication, for which the use of low-frequency language is a strong indicator.

Adjective predicates were the fourth and final significant contributor to the *Humor* component score in our model, explaining 1.8% of the variance in the overall model. This index loaded positively, suggesting that essays with higher *Humor* component scores contained a higher number of adjective predicates. Adjective predicates are single- or multiple-word adjective phrases that modify the subject of a sentence. As opposed to attributive adjectives (which almost always precede a noun phrase in English), adjective predicates are part of the main verb phrase in a clause and are typically preceded by a copular verb (e.g., *be*, *seems*, *appears*). Thus, in the sentence “The dog is brown,” *brown* is an adjective predicate. The following sentence illustrates the use of adjective predicates from the essay quoted above: “Everyone would become all *modest*, *smart*, *strong*, *self-reliant* and *wise*.” Here we see a string of adjective predicates (italicized above) following the copular verb *become*. Adjective predicates are further unique from attributive adjectives in that their occurrence after the main verb makes them more likely to express new information about the subject of a sentence than previously given information (Chafe, 1976). In this regard, adjective predicates are syntactically poised to redefine a topic, rather than to merely modify it. One interpretation of the ability of adjective predicates to predict humor is that humorous academic texts are more likely to redefine their topic matter in a manner that is comical, surprising, or deprecating. Evidence of this tendency can be found in the example essay and throughout humorous essays in our corpus as a whole.

These findings have several implications. First, the linguistic features that emerged as significant in this study differ from those seen in other computational studies of humor (e.g., Carvalho et al., 2009; Mihalcea & Strapparava, 2005, 2006; Reyes et al. 2012; Skalicky & Crossley, 2015). This is not surprising, given that the humor analyzed here was markedly different from previously studied humor, such as one-liners, humorous quotes, or ironic tweets, and agrees with observations made claiming feature sets from one descriptive study of humor may not match others (Reyes et al., 2010).

Secondly, it may be that authors who employ humor in academic writing do so cautiously, aware of the exhortations to write concisely, directly, and to remain on point (e.g., American Psychological Association, 2010; Palmquist, 2010). As a result, linguistic features typical of academic writing remain dominant, even in more humorous essays. For example, the essay quoted above, despite receiving the highest *Humor* component score, still contains both the typical rhetorical organization of an academic essay, including opening, body, and concluding paragraphs, and a paragraph structure that

includes both topic and concluding sentences. Furthermore, the primary function of humor in this essay was to provide humorous examples that served to support the author's overall thesis. Therefore, the essay demonstrates that it is possible to use humor to support the larger rhetorical demands of academic writing, although the low essay quality score for our exemplar essay demonstrates that it will not always be successful.

Moreover, despite wordplay's connection to the manipulation of linguistic forms and semantic meanings of words (Cook, 2000), measurements that might have captured linguistic features such as repetition, alliteration, and ambiguity did not account for a large percentage of the *Humor* component score. This suggests that for both wordplay and humor, raters may attend to other features of the texts not measureable by the text tools employed in this study. These may be larger, rhetorical or pragma-linguistic devices, such as genre conventions or the voice of the author (Devitt, Reiff, & Bawarshi, 2004). Furthermore, actual incidences of explicit humor were relatively rare. No essays attempted humor through canned jokes or puns. Instead, humor was typically signaled through sarcasm, derisive comments about the subject matter, or fantastical descriptions of fictional characters. In other words, having a high *Humor* component score did not necessarily mean that the essay included jokes or attempted to be explicitly funny, but rather, that the raters perceived some elements of wordplay or humor in the essay that created a tone more accurately described as playful, whimsical, or wry.

Importantly, though, our results found a small positive correlation between essay quality and humor ratings ($r = .195$). This suggests that humor may be a contributing factor to holistic ratings of essay quality. In order to illustrate the positive correlation between humor and academic writing, we briefly discuss another essay from the corpus, which had an essay quality score of 6 and a humor score of 4 (see Appendix B for full essay). The prompt for this essay asked students to discuss the inherent tension between a desire to be unique and the reality that it is difficult to make truly unique contributions to the world. In this essay, the student employed irony, wordplay, and negative sarcastic evaluation. The student began the essay by stating that unoriginality is inevitable, and pointed out the irony inherent in constant recycling of styles in the fashion industry:

“However, no matter how much effort the designers for Versace put into a gown, it is almost guaranteed that Chanel produced nearly the same dress twenty years ago.”

However, when the student turned to focus on the context of a local university and town, a number of negative evaluations through sarcasm (which may result in humor depending on the reader) were apparent:

“More immediate examples of this principle can be seen on campus at [name of university]. One cannot turn a corner without seeing girls in Nike running shorts. These particular shorts were designed for exercising, not for sitting in class. It is a trend that was spawned by a sorority, probably as a joke, and unfortunately caught on to the point where it is the norm for girls here in [name of town] to walk around in gym shorts all day long. It would be understandable if they intended to work out after class, but from the looks of most of them they do not do much in the way of exercise. The fraternity trend is Ralph Lauren Polo shirts. Fraternity boys have a polo shirt in every color: long-sleeved, short-sleeved, no-sleeved. These shirts cost well over eighty dollars, so their parents are probably not happy that these shirts are the only acceptable form of clothing for fraternities.”

In this example, the student opens with a jab targeted at other students who wear exercise clothes for purposes other than exercising, before implying that these same people are in need of exercise. The author then turns their ire towards fraternity styles, using a parallel play on the hyphenated adjectival “-sleeved” to joke that some polo shirts have *no* sleeves. The author ends the paragraph with the observation that parents must be upset over the high cost of this style. What is interesting about this paragraph is that it serves two functions: to add support for the overall argument using examples, while at the same time mocking members of the author's local community.

Both of the exemplar essays use humor as a means to support their claims. The difference between this essay and the previous essay is primarily in the humorous example that is used. In the first essay, the fictional superhero Batman is discussed, whereas in this essay, the author targets real members of the local community. It may be that the function of humor in this second essay worked to build rapport between the essay rater and author (a recognized function of humor; Martin, 2007), especially if the essay rater shared similar feelings towards members of fraternities or those who wear exercise clothing outside of a gymnasium. However, the humor in this essay was also more congruent with the rest of the writing, unlike the first example, and the author was better able to cloak the humor behind the typical diction of academic writing. It may be, then, that humor does have a place in academic writing, but only if students employ it carefully and subtly.

7. Conclusion

In this study, we have demonstrated the ability to predict a portion of the variance in raters' perceptions of humor and wordplay in academic writing. This task is challenging because academic writing is not a genre in which humor would be expected to occur. Nonetheless, we have offered initial evidence suggesting that humor or wordplay in academic writing may be signaled via descriptive language, such as adverbs and adjective predicates, along with a lack of semantic cohesion between paragraphs and the use of more sophisticated words. We have also demonstrated a small yet significant relation between the use of humor in academic essays and human perceptions of essay quality, one that warrants further investigation.

To our knowledge, no student is expressly instructed to be funny in academic writing. Yet, as this analysis demonstrates, student attempts at humor in academic writing do occur. While we have identified some of the linguistic forms and functions of humor in student essays, further research is needed to investigate the attested relation between

essay quality and humor. The features identified here can also be used in future studies examining a wider range of contexts and writing proficiency in order to contribute to a better understanding of how humor functions in academic writing.

References

- American Psychological Association. (2010). *Publication manual of the American psychological association*. Washington DC: American Psychological Association.
- Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: Joke similarity and joke representation model. *Humor: International Journal of Humor Research* 3(4), 293-347.
- Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English*. Essex, GB: Longman.
- Burfoot, C., & Baldwin, T. (2009). Automatic satire detection: Are you having a laugh? *Association for Computational Linguistics International Joint Conference on Natural Language Processing 2009 Conference Short Papers*, 161-164.
- Buscaldi, D., & Rosso, P. (2007). Some experiments in humour recognition using the Italian wikiquote collection. In F. Masulli, S. Mita, & G. Pasi (eds.), *Applications of Fuzzy Sets Theory*, 464-468. Berlin, DE: Springer Berlin Heidelberg.
- Campbell, J. D. & Katz, A. N. (2012). Are there necessary conditions for inducing a sense of sarcastic irony? *Discourse Processes* 49(6), 459-480.
- Carvalho, P., Sarmiento, L., Silva, M., & de Oliveira, E. (2009). Clues for detecting irony in user-generated contents: Oh ...!! It's "so easy" ; -). *TSA '09 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*, 53-56.
- Chafe, W. R. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view in subject and topic. In Charles. N. Li (ed.), *Subject and topic*, 25-56. New York, NY: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cook, G. (2000). *Language play, language learning*. New York, NY: Oxford University Press.
- Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, 316(2), 71-74.
- Cronk, B. C. & Schweigert, W. A. (1992). The comprehension of idioms: The effects of familiarity, literalness, and usage. *Applied Psycholinguistics*, 13(2), 131-146.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 1-11.
- Crossley, S.A. & McNamara, D.S. (2010) Cohesion, coherence, and expert evaluations of writing proficiency. In Catrambone, R. and Ohlsson, S. (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pp. 984-989. Cognitive Science Society, Austin, TX.
- Crossley, S. A., & McNamara, D. S. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. (1236-1241). Austin, TX: Cognitive Science Society.
- Deane, P. (2014). Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks. *ETS Research Report Series, 2014(1)*, 1-23.
- Devitt, A. J., Reiff, M. J., & Bawarshi, A. S. (2004). *Scenes of writing: Strategies for composing with genres*. Pearson/Longman.
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757-786.
- Landauer, T., McNamara, D.S., Dennis, S., & Kintsch, W. (2007). *LSA: A road to meaning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Martin, R. A. (2007). *The psychology of humor: An integrative approach*. San Diego, CA: Elsevier.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86.
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural Language Processing in an Intelligent Writing Strategy Tutoring System. *Behavior Research Methods*, 45(2), 499-515.
- Mihalcea, R. & Pulman, S. (2007). Characterizing humour: An exploration of features in humorous texts. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*, 337-347. New York: Springer Berlin Heidelberg.
- Mihalcea, R. & Strapparava, C. (2005). Making computers laugh: Investigations in automatic humor recognition. *Association for Computational Linguistics (ACL) Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 531-538.
- Mihalcea, R. & Strapparava, C. (2006). Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, 22(2), 126-142.
- Mihalcea, R., Strapparava, C., & Pulman, S. (2010). Computational models for incongruity detection in humour. In Alexander Gelbukh (Ed.), *Computational linguistics and intelligent text processing*, 364-374. Berlin, DE: Springer Berlin Heidelberg.
- Palmquist, M. (2010). *Joining the conversation: Writing in college and beyond*. New York, NY: Bedford/St. Martin's.
- Pitler, E. & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. *Association for Computational Linguistics (ACL) Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 186-195.

- Reyes, A., Pothast, M., Rosso, P., & Stein, B. (2010). Evaluating humour features on web comments. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, 1138-1141.
- Reyes, A. & Rosso, P. (2011). Mining subjective knowledge from customer reviews: A specific case of irony detection. *Association for Computational Linguistics (ACL) Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, 118-124.
- Reyes, A., Rosso, P., & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74, 1-12.
- Reyes, A., Rosso, P., & Veale, T. (2013). A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, 47(1), 239-268.
- Ritchie, G. D. (2004). *The linguistic analysis of jokes*. New York, NY: Routledge.
- Simpson, P. (2003). *On the discourse of satire: Towards a stylistic model of satirical humor*. Philadelphia, PA: John Benjamin's Publishing Company.
- Sheridan, H., Reingold, E., & Daneman, M. (2009). Using puns to study contextual influences on lexical ambiguity resolution: Evidence from eye movements. *Psychonomic Bulletin & Review*, 16(5), 875-881.
- Skalicky, S., & Crossley, S. (2015). A statistical analysis of satirical Amazon. com product reviews. *The European Journal of Humour Research*, 2(3), 66-85.
- Vaid, J., Hull, R., Heredia, R., Gerkens, D., & Martinez, F. (2003). Getting a joke: The time course of meaning activation in verbal humor. *Journal of Pragmatics*, 35(9), 1431-1449.
- Witten, I. H., Frank, E., & Hall, M. (2011). *Data mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann Publishers.

Appendix A: Example Essay 1

Throughout the world and even more so in America today, celebrities are given an almost royal rank in society. They can do no wrong and all that is associated with them is success. A simple public apology and small donation to charity is more than enough to satisfy some form of repentance for any exposed transgression. Heroes do not seek out the fame and admiration of a nation. They do what is right and necessary for the simple idea of justice. Heroes deserve the title of "role model".

Celebrities may fall into the category of hero but it is very rare for them to avoid the corrupting temptations that separates them from that title. I can't think of what to write in such a short amount of time. Anyway, heroes are cool because they don't even care what you think. They will just wake up and silently think to themselves, "Yep, it's time to be awesome today." They don't even exclaim that in their heads because that would be so unnecessary and foolish. Alternatively, celebrities wake up all scared and unsure of themselves hoping that the world will approve of them because, "I'm not sure I'll be awesome today...hope everything goes smoothly today and I don't crash my car into a fire hydrant while sneaking away at 3 a.m. to cheat on my wife! ...Cause man that would stink [sic]."

Heroes set out to decide whether or not they approve of the world. If not, they change it by any means necessary without resorting to celebrity-like tactics because that would so totally defeat the purpose of their heroic deeds. If everyone looked up to heroes then the world would have many fewer celebrities in the future. Everyone would become all modest, smart, strong, self-reliant and wise. That's a nice idea but if you think for a minute, fewer celebrities means fewer fools to laugh at which means fewer examples of what not to do. Normal people learn from their mistakes while wise people learn from the mistakes of fools.

Heroes may not always be popular with the law. Batman, for example, was constantly hunted by the police for being a vigilante and for littering. It is a little known fact that batman does not pick up his soda cans. This just goes to show that even though heroes have the best interest of the world in mind, they may not always be perfect themselves.

This is not to say that heroes or celebrities should or should not be looked up to. This is to say that people should do what truly makes them happy. To make any distinction more specific than that would go against the will of the individual.

Appendix B: Sample Essay Two

The question of whether or not people can ever truly be original is one that is posed often in western society. Particularly, Americans try to be unique when creating outfits to wear to class, work, or out for a night on the time. However, these efforts are wasted because western society looks towards those in the public eye for ideas on who to be. Women look to their favorite celebrities for tips on how they should wear their hair, clothes, and accessories. Men seek assistance from their favorite athletes and musicians for inspiration for a new hair cut or a new pair of shoes. Not only do Americans look to those in the public eye for hints on how to better fit into society, but they also look to each other for inspiration. This example can be seen from the styles on red carpets in Hollywood, to the runways of New York City, and even on the campus of Mississippi State University.

The trends that flow through Hollywood are infectious. They stem from runways and designer boutiques and reach as far as small, rural towns in the south much like Starkville, Mississippi. During award show season celebrities turn to their most trusted stylists to dress them for red carpets and after parties. The stylists they go to follow a regime of the fit, color, and design pattern that is in that season. Thus, some celebs end up wearing very similar gowns and tuxes or, even worse, the same exact outfit. The "who wore it better" war ensues, and said actor is forever humiliated when their opponent wins. As much as people try to be unique in Hollywood, there will always be someone who has worn a similar outfit whether it was last season or last decade.

Another case where imitation is avoided but never truly avoidable is on the runways of New York City during fashion week. Runway shows are elaborate, confusing, and sometimes ghastly experiences. Designers do their best to make clothes so ridiculously flashy in order to stand out. However, no matter how much effort the designers for Versace put into a gown, it is almost guaranteed that Chanel produced nearly the same dress twenty years ago. Designers go to great lengths in an effort to make them stand out from their competitors. Unfortunately to stay on the radar, they must conform to a certain style or else their prized creations may forever be lost in a Macy's one day sale bin.

More immediate examples of this principle can be seen on campus at Mississippi State University. One cannot turn a corner without seeing girls in Nike running shorts. These particular shorts were designed for exercising, not for sitting in class. It is a trend that was spawned by a sorority, probably as a joke, and unfortunately caught on to the point where it is the norm for girls here in Starkville to walk around in gym shorts all day long. It would be understandable if they intended to work out after class, but from the looks of most of them they do not do much in the way of exercise. The fraternity trend is Ralph Lauren Polo shirts. Fraternity boys have a polo shirt in every color: long-sleeved, short-sleeved, no-sleeved. These shirts cost well over eighty dollars, so their parents are probably not happy that these shirts are the only acceptable form of clothing for fraternities.

Whether it is on the red carpet or on campus, people relentlessly and shamelessly follow trends. There is no hope for individuality, not just because people imitate one another, but because it really has all been done. Maybe someday a brilliant designer will find a way to make shirts out of styrofoam and originality will live again.

Appendix C: Rubrics Used for Human Ratings

Analytical rating form

Read each essay carefully and then assign a score on each of the points below. For the following evaluations, you will need to use a grading scale between 1 (minimum) and 6 (maximum).

We present here a description of the score as a guide using the example of *does not meet the set criterion in any way* versus *meets the set criterion in every way*. For example, a grade of 1 would relate to not meeting the criterion in any way, and a grade of 4 would relate to somewhat meeting the criterion. The distance between each grade (e.g., 1-2, 3-4, 4-5) should be considered equal. Thus, a grade of 5 (*meets the criterion*) is as far above a grade of 4 (*somewhat meets the criterion*) as a grade of 2 (*does not meet the criterion*) is above a grade of 1 (*does not meet the criterion in any way*).

Score	Definition
1	Does not meet the criterion in any way
2	Does not meet the criterion
3	Almost meets the criterion but not quite
4	Meets the criterion but only just
5	Meets the criterion
6	Meets the criterion in every way

Part	Score
1. Ideas	
<i>1.1 Fluency</i> The essay contains many unique ideas within the essay.	1 2 3 4 5 6
<i>1.2 Flexibility</i> The essay contains a variety of different ideas (e.g., many different categories of ideas).	1 2 3 4 5 6
<i>1.3 Originality</i> The essay contains ideas that are unique across essays.	1 2 3 4 5 6
<i>1.4 Elaboration</i> The essay includes information that expands on the main idea(s) contained in the essay.	1 2 3 4 5 6

2. Style	
<i>2.1 Humor</i> The essay attempts to provoke laughter or amusement.	1 2 3 4 5 6
<i>2.2 Metaphor & Simile (cognitive style)</i> The essay involves original comparisons that construe entities outside of their content domain(s).	1 2 3 4 5 6
<i>2.3 Word Play (linguistic style)</i> The essay includes the use of sounds, meanings, or forms of words that are unexpected or original.	1 2 3 4 5 6

SAT Scoring Rubric

SCORE OF 6: An essay in this category demonstrates clear and consistent mastery, although it may have a few minor errors. A typical essay effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas exhibits skillful use of language, using a varied, accurate, and apt vocabulary demonstrates meaningful variety in sentence structure is free of most errors in grammar, usage, and mechanics.

SCORE OF 5: An essay in this category demonstrates reasonably consistent mastery, although it will have occasional errors or lapses in quality. A typical essay effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position is well organized and focused, demonstrating coherence and progression of ideas exhibits facility in the use of language, using appropriate vocabulary demonstrates variety in sentence structure is generally free of most errors in grammar, usage, and mechanics.

SCORE OF 4: An essay in this category demonstrates adequate mastery, although it will have lapses in quality. A typical essay develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position is generally organized and focused, demonstrating some coherence and progression of ideas exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary demonstrates some variety in sentence structure has some errors in grammar, usage, and mechanics

SCORE OF 3: An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses: develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice lacks variety or demonstrates problems in sentence structure contains an accumulation of errors in grammar, usage, and mechanics.

SCORE OF 2: An essay in this category demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses: develops a point of view on the issue that is vague or seriously limited, and demonstrates weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas displays very little facility in the use of language, using very limited vocabulary or incorrect word choice demonstrates frequent problems in sentence structure contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured.

SCORE OF 1: An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses: develops no viable point of view on the issue, or provides little or no evidence to support its position is disorganized or unfocused, resulting in a disjointed or incoherent essay displays fundamental errors in vocabulary demonstrates severe flaws in sentence structure contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.