The Funny Thing About Incongruity: A Computational Model of Humor in Puns

Justine T. Kao[1] (justinek@stanford.edu), **Roger Levy**[2] (rlevy@ucsd.edu), **Noah D. Goodman**[1] (ngoodman@stanford.edu)
[1]Department of Psychology, Stanford University. [2]Department of Linguistics, UC San Diego.

Abstract

Researchers showed the robot ten puns, hoping that one of them would make it laugh. Unfortunately, no pun in ten did.

What makes something funny? Humor theorists posit that incongruity—perceiving a situation from different viewpoints and finding the resulting interpretations to be incompatible—contributes to sensations of mirth. In this paper, we use a computational model of sentence comprehension to formalize incongruity and test its relationship to humor in puns. By combining a noisy channel model of language comprehension and standard information theoretic measures, we derive two dimensions of incongruity—ambiguity of meaning and distinctiveness of viewpoints—and use them to predict humans' judgments of funniness. Results showed that both ambiguity and distinctiveness are significant predictors of humor. Additionally, our model automatically identifies specific features of a pun that make it amusing. We thus show how a probabilistic model of sentence comprehension can help explain essential features of the complex phenomenon of linguistic humor.

Keywords: Humor; language understanding; probabilistic models

Introduction

Humor plays an essential role in human interactions: it has important positive effects on children's development (Frank & McGhee, 1989), success in the work place (Duncan et al., 1990), coping with illness and traumatic events (Gelkopf & Kreitler, 1996), and marital satisfaction (Ziv & Gadish, 1989). Indeed, in a study on gender differences in desired characteristics of relationship partners, both men and women rated sense of humor as more important than physical attractiveness and earning potential (Stewart et al., 2000). In this paper, we are interested in understanding how this fundamental and ubiquitous phenomenon works from the perspective of cognitive science. What makes something funny? How might defining characteristics of humor shed light on the ways in which the mind processes and evaluates information?

A leading theory of humor posits that incongruity—perceiving a situation from different viewpoints and finding the resulting interpretations to be incompatible—contributes to sensations of mirth (Veale, 2004; Forabosco, 1992; Martin, 2007; Hurley et al., 2011); an idea that dates to Kant's theories about laughter and the sublime (Veatch, 1998). Although there is disagreement about whether incongruity alone is sufficient, most theorists accept that incongruity is necessary for producing humor: as Veale (2004) states, "Of the few sweeping generalizations one can make about humor that are neither controversial or trivially false, one is surely that humor is a phenomenon that relies on incongruity." However, definitions of incongruity are often ambiguous and difficult to operationalize in empirical research. In this paper, we use a computational model of language understanding to formalize a notion of incongruity and test its relationship to humor.

Language understanding in general, and particularly humor, relies on rich commonsense knowledge and discourse understanding. To somewhat limit the scope of our task, we focus on applying formalizations of incongruity to a subset of linguistic humor: puns. Writer and philosopher Henri Bergson defined a pun as "a sentence or utterance in which two ideas are expressed, and we are confronted with only one series of words." This highlights the fact that one sentence must evoke two different interpretations in order to be a pun, which aligns with the concept of incongruity as a requisite of humor.

We develop our model on homophone puns—puns containing words that sound identical to other words in the English language—because the space of possible interpretations of a homophone pun is relatively constrained and well-defined. An example helps to illustrate:

"The magician got so mad he pulled his hare out."

This sentence allows for two interpretations:

(a) The magician got so mad he performed the trick of pulling a rabbit out of his hat.

(b) The magician got so mad he (idiomatically) pulled out the hair on his head.

If the comprehender interprets the word "hare" as itself, he will arrive at interpretation (a); if he interprets the word as its homophone "hair," he will arrive at interpretation (b). The sentence-level differences between interpretations (a) and (b) can thus be approximated by the two interpretations of the observed word "hare." In general, distinct interpretations of a homophone pun hinges on one phonetically ambiguous word, allowing the two lexical forms of the homophone word to stand in for competing interpretations of the entire sentence.

Critically, even though the example we gave was a written pun and the reader sees the word "hare" explicitly on the page, the "hair" interpretation is still present and even salient in the context of the sentence. Here we explore the idea that puns such as these arise and are funny when they are due to noisy-channel processing. Noisy channel models of sentence processing posit that language comprehension is a rational process that incorporates uncertainty about surface input to arrive at sentence-level interpretations that are globally coherent (Levy, 2008; Levy et al., 2009). Comprehenders can thus consider multiple word-level interpretations ("viewpoints") to arrive at more than one interpretation of a sentence, each coherent but potentially incongruous with each other. The notion of incongruity thus fits naturally into a noisy channel model of sentence comprehension.

Our purposes for developing a formal model of linguistic humor are two-fold. First, we wish to formalize the concept

of incongruity and test assumptions adopted by leading theories in humor research. Secondly, we aim to show that a noisy channel of language processing allows for flexible context selection and sentence comprehension that gives rise to sophisticated linguistic and social meaning such as humor.

Model

Incongruity is a property of the interpretations derived from a sentence. In order to formalize incongruity, we first describe a probabilistic model of sentence comprehension. Our model aims to infer the topic of a sentence (a coarse representation of its meaning) from the observed words. Unlike previous such models, however, we take a noisy channel approach, assuming that the comprehender maintains uncertainty over which words reflect the sentence topic and which are noise. From this model we derive two quantities that may contribute to humor: *ambiguity* and *distinctiveness*. Intuitively, if the resulting interpretation is unambiguous, then no incongruity exists and the sentence is unlikely to be funny. However, since many ambiguous sentences are not funny (e.g. "I went to the bank"), ambiguity alone is insufficient. This is because the interpretations of such sentences are not supported by distinct topical subsets of the sentence (or "viewpoints"). In other words, there must be a set of words in the sentence that support one interpretation and a set that supports the other, and these two sets must be different or "distinct" from each other in order to evoke a sense of incongruity.

Assume our sentence is composed of a vector of content words $\vec{w} = \{w_1, \ldots, w_i, h, w_{i+1}, \ldots, w_n\}$, including a phonetically ambiguous word h. We will use a simple generative model for $\vec{w}$ (see Figure 1): given the latent sentence topic m, each word is generated independently by first deciding if it reflects the topic (the indicator variable f_i). If so it is sampled based on semantic relevance to m; if not it is sampled from a fixed unigram prior over words. We thus view the sentence as a mixture of topical and non-topical words. Similar approaches have been used in generative models of language to account for words that provide non-semantic information, such as topic models that incorporate syntax (Griffiths et al., 2005). Our model is motivated by the important role that semantic priming plays in lexical disambiguation during sentence processing (Seidenberg et al., 1982; Burke & Yee, 1984); while ignoring the other non-semantic factors of interpretation (which may also be important).

We make the simplifying assumption that the plausible candidate topics m of the sentence correspond to the potential interpretations of the homophone word h, which are constrained by phonetic similarity to two alternatives, m_1 and m_2. For example, in the magician pun described above, h is the phonetically ambiguous target word "hare," and m_1 and m_2 are the candidate interpretations *hare* and *hair*. The two potential topics of the sentence can be identified by the two interpretations *hare* and *hair*. This assumption reduces the ill-defined space of sentence meanings to the simple proxy of alternate spellings for phonetically ambiguous words.

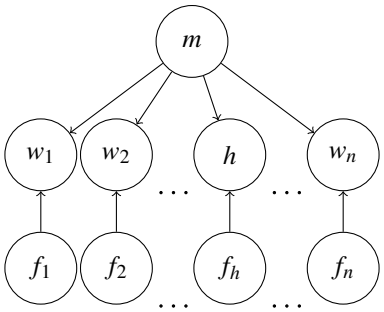


Figure 1: Generative model of a sentence. Each word w_i is generated based on the sentence topic m if the indicator variable f_i puts it in semantic focus; otherwise it is generated as noise (from a unigram distribution).

Using the above generative model, we can infer the joint probability distribution $P(m, \vec{f}|\vec{w})$ of the sentence topic m and the indicator variables $\vec{f}$ that determine whether each word is in semantic focus. This distribution can be factorized into:

$$P(m, \vec{f}|\vec{w}) = P(m|\vec{w})P(\vec{f}|m, \vec{w}) \quad (1)$$

The two terms on the right-hand side are the basis for our derivations of measures for ambiguity and distinctiveness respectively. Ambiguity means the presence of two similarly likely interpretations and can be quantified as a summary of the distribution $P(m|\vec{w})$. Distinctiveness measures the degree to which two interpretations are supported by "distinct" viewpoints of the sentence, which we represent as the divergence between sets of words that are in semantic focus given the two values of m; it can be quantified as a summary of the distribution $P(\vec{f}|m, \vec{w})$. Together, these two measures constitute our formalization of incongruity.

Ambiguity Let M denote the distribution $P(m|\vec{w})$, a binomial distribution over the two meaning values m_1 and m_2 given the observed words. If the entropy of this distribution is low, this means that the probability mass is concentrated on only one meaning, and the alternative meaning is unlikely given the observed words. If entropy is high, this means that the probability mass is more evenly distributed among m_1 and m_1, and the two interpretations are similarly likely given the contexts. The entropy of $P(m|\vec{w})$ is thus a natural measure of the degree of ambiguity present in a sentence. We compute $P(m|\vec{w})$ as follows:

$$P(m|\vec{w}) = \sum_{\vec{f}} P(m, \vec{f}|\vec{w}) \quad (2)$$

$$\propto \sum_{\vec{f}} P(\vec{w}|m, \vec{f})P(m)P(\vec{f}) \quad (3)$$

$$= \sum_{\vec{f}} \left(P(m)P(\vec{f}) \prod_i P(w_i|m, f_i) \right) \quad (4)$$

We approximate $P(m)$ as the unigram frequency of the words that represent m. For example, $P(m = \textit{hare})$ is approximated

as $P(m = \text{“hare”})$. We also assume a uniform probability that each words is in focus—hence $P(\vec{f})$ is a constant. As for $P(m|\vec{w})$, note that it is driven in part by the semantic relationship between m and $\vec{w}$ and in part by the prior probability of m, which we approximate using the unigram probability of the words m_1 and m_2. From the generative model,

$$P(w_i|m,f_i) = \begin{cases} P(w_i), & \text{if } f = 0 \\ P(w_i|m), & \text{if } f = 1 \end{cases}$$

Once we derive $P(m|\vec{w})$, we then compute its entropy as a measure of ambiguity.

Distinctiveness We next turn to the distribution over focus sets, given sentence topic. This may be computed as follows:

$$P(\vec{f}|m,\vec{w}) \propto P(\vec{w}|m,\vec{f})P(\vec{f}|m) \quad (5)$$

Since $\vec{f}$ and m are independent, $P(\vec{f}|m) = P(\vec{f})$.

Let F_1 denote the distribution $P(f|m_1,\vec{w})$ and F_2 denote the distribution $P(f|m_2,\vec{w})$. F_1 and F_2 represent the distributions over semantic focus sets assuming the sentence topic m_1 and m_2, respectively. We use a symmetrized Kullback-Leibler divergence score $D_{KL}(F_1||F_2) + D_{KL}(F_2||F_1)$ to measure the distance between F_1 and F_2. This score measures how “distinct” the semantic focus sets are given m_1 and m_2. A low KL score would indicate that meanings m_1 and m_2 are supported by similar subsets of the sentence; a high KL score would indicate that m_1 and m_2 are supported by distinct subsets of the sentence.

Evaluation

By generating a large corpus of sentences involving the same words and measuring subjective funniness of each sentence we can evaluate the contribution of each of our quantitative measures, ambiguity and distinctiveness, to humor. We evaluate our model and measures on a set of 235 sentences, consisting of 65 puns, 40 “de-punned” control sentences that are matched with a subset of the puns, and 130 non-pun control sentences that match the puns in containing the same phonetically ambiguous words.

Materials

We selected 40 pun sentences from a large collection of puns on a website called Pun of the Day, which contains over one thousand puns. Puns were selected such that the ambiguous item is a single phonetically ambiguous word, and no two puns in the collection have the same ambiguous item. To obtain more homophone pun items, a research assistant generated an additional 25 pun sentences based on a separate list of homophone words.

We constructed 40 sentences to be minimally different from the pun sentences that we collected from “Pun of the Day,” which we will call de-punned sentences. A second research assistant who was blind to the hypothesis was asked to replace one word in each of the pun sentences (without changing the homophone word itself) so that the sentence is still grammatical but is no longer a pun. This resulted in sentences that differed from the pun sentences by one word each.

The 130 non-pun sentences were chosen to match each pun sentence on its ambiguous word as well as the alternative homophone. The sentences were taken from an online version of Heinle’s Newbury House Dictionary of American English (`http://nhd.heinle.com/`). We selected sample sentences included in the definition of the homophone word. This design ensured that puns, de-punned, and non-pun sentences all contain the same set of phonetically ambiguous words. Table 1 shows example sentences from each category.

Type	Example
Pun	The magician got so mad he pulled his hare out.
De-pun	The professor got so mad he pulled his hare out.
Non-pun	The hare ran rapidly across the field.
Non-pun	Some people have lots of hair on their heads.

Table 1: Example sentences from each category

Human ratings of semantic relatedness

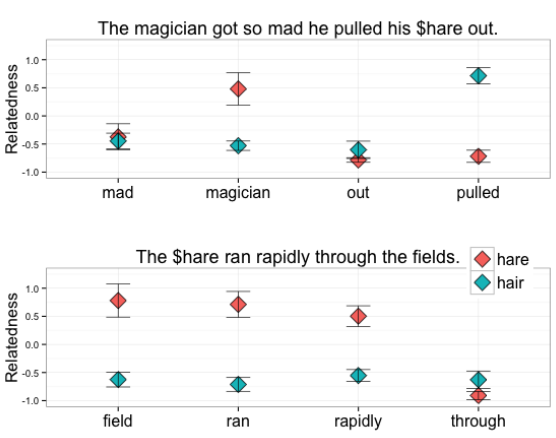
As described in the model section, computing our measures requires the prior probabilities of meanings $P(m)$ (approximated as the unigram probabilities of the words that denote the meanings), the prior probabilities of words $P(w)$, and the conditional probabilities of each word in the sentence given a meaning $P(w|m)$. While we computed $P(w)$ and $P(m)$ directly from the Google Web unigram corpus, $P(w|m)$ is difficult to obtain through traditional topic models trained on corpora due to data sparsity. However, since each meaning we consider has a single word as proxy, we may approximate $P(w|m)$ using an empirical measure of the semantic relatedness between w and m, denoted $R(c,m)$. We use $R(c,m)$ as a proxy for point wise mutual information between c and m, defined as follows:

$$R(w,m) = \log \frac{P(w,m)}{P(w)P(m)} = \log \frac{P(w|m)}{P(w)} = \log P(w|m) - \log P(w)$$

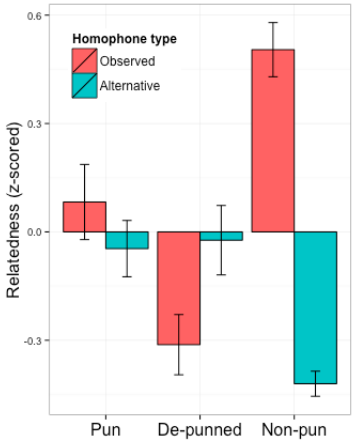
We assume that human ratings of relatedness between two words $R'(w,m)$ approximate true relatedness up to an additive constant z. With the proper substitutions and transformations,

$$P(w|m) = e^{R'(w,m)+z}P(w) \quad (6)$$

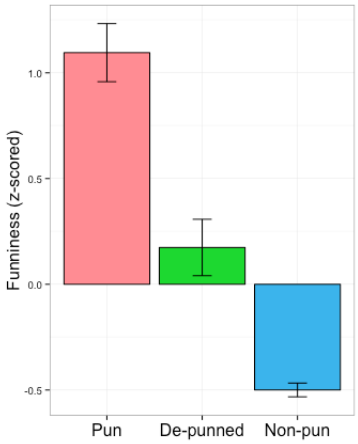
To obtain $R'(w,m)$ for each of the words w in the stimuli sentences, we recruited 200 subjects on Amazon’s Mechanical Turk to rate distinct word pairs on their semantic relatedness. Since it is difficult to obtain the relatedness rating of a word with itself, we used a free parameter r and fit it to data. Function words were removed from each of the sentences in our dataset, and the remaining words were paired with each of the interpretations of the homophone sequence (e.g., for the pun in Figure 1, “magician” and “hare” is a legitimate word pair, as well as “magician” and “hair”). This resulted in 1460



(a) Relatedness of each word with candidate meanings



(b) Average relatedness



(c) Average funniness ratings

Figure 2: (a) In the example pun (top), two candidate meanings of *h* are each more related to a subset of the content words. In the non-pun, only one candidate meaning is more related. (b) Content words are similarly related to both candidate meanings in puns; more related to alternative meanings in de-puns; more related to observed meanings in non-pun. (c) Funniness varies across the sentence types in a pattern that reflects the balance of relatedness to candidate meanings shown in (b).

distinct word pairs. Each subject saw 146 pairs of words in random order and were asked to rate how related each word pair is using a scale from 1 to 10. The average split-half correlation of the relatedness ratings was 0.916, indicating that semantic relatedness was a reliable measure.

Figure 2(a) shows the relatedness of each content word with the two homophone interpretations for two example sentences. In the top sentence, which is a pun, the word "magician" is rated as significantly more related to "hare" than it is to "hair", while the word "pulled" is rated as significantly more related to "hair" than it is to "hare." In the bottom sentence, which is a non-pun, all words except the neutral word "through" are more related to the word "hare" than to "hair."

Figure 2(b) shows the average relatedness ratings of words and the two homophone interpretations across the three types of sentences. In pun sentences, the average relatedness of words to the two homophone interpretations are not significantly different. In the de-punned sentences, the average relatedness of words to the alternative meaning is significantly higher than to the observed meaning. In the non-pun sentences, the average relatedness of words to the observed meaning is significantly higher than to the alternative meaning. These analyses suggest that relatedness ratings for the two candidate meanings capture the presence or absence of multiple interpretations in a sentence. It further supports our model's prediction that ambiguity of meaning and the distinctiveness of supporting context words can help distinguish among the three types of sentences.

Human Ratings of Funniness

We obtained funniness ratings of the 235 sentences from 100 subjects on Amazon's Mechanical Turk. Each subject read roughly 60 sentences in random order, counterbalanced for the sentence types, and rated each sentence on funniness and correctness. The average split-half correlation of funniness ratings was 0.83. Figure 2(c) shows the average funniness ratings of puns, de-punned, and non-pun sentences. Pun sentences are rated as significantly funnier than de-punned sentences, and de-punned sentences are rated as significantly funnier than non-pun sentences ($F(2,232) = 415.3, p < 0.0001$). Figure 2 (b) and Figure 2 (c) together suggest that the balance of relatedness between the two interpretations is a predictor of funniness.

	Estimate	Std. Error	p value
Intercept	-0.699	0.180	< 0.0001
Ambiguity	1.338	0.245	< 0.0001
Distinctiveness	0.183	0.053	< 0.0001

Table 2: Regression coefficients using ambiguity and distinctiveness to predict funniness ratings

Results

Following the derivations described in the model section and using the relatedness measures described above, we computed an ambiguity and distinctiveness value for each of the 235 sentences. Our model has two free parameters—the additive constant z in equation (6) and the constant r that indicates the relatedness of a word with itself —which we optimized using R^2 in the linear regression summarized in Table 2. As predicted, ambiguity differs significantly across sentence types ($F(2,232) = 25.42, p < 0.0001$) and correlates significantly with human ratings of funniness across the 235 sentences ($r = 0.33, p < 0.0001$). Furthermore, distinctiveness scores differ significantly across sentence types as well ($F(2,232) = 5.76, p < 0.005$) and correlates signifi-

$m1$	$m2$	Type	Sentence and Semantic Focus Sets	Amb.	Disj.	Funniness
hare	hair	Pun	The **magician** got so mad he **pulled** his **hare** out.	0.570	3.405	1.714
		De-pun	The professor got so mad he **pulled** his **hare** out.	0.575	2.698	0.328
		Non-pun	The **hare ran rapidly** through the **fields**.	0.055	2.791	−0.400
		Non-pun	Most **people** have **lots** of **hair** on their **heads**.	$2.76E^{-5}$	3.920	−0.343
tiers	tears	Pun	It was an **emotional wedding**. Even the **cake** was in **tiers.**	0.333	3.424	1.541
		De-pun	It was an **emotional wedding**. Even the **mother-in-law** was in **tiers**.	0.693	2.916	0.057
		Non-pun	**Boxes** are **stacked** in **tiers** in the warehouse.	0.018	3.203	−0.560
		Non-pun	**Tears** ran down her **cheeks** as she watched a **sad movie**.	$1.73E^{-5}$	4.397	−0.569

Table 3: Semantic focus sets, ambiguity/disjointedness scores, and funniness ratings for two groups of sentences. Words in red are in semantic focus with m_1; green with m_2; blue with both. Semantic focus sets for all sentences can be found at `http://www.stanford.edu/~justinek/Pun/focusSets.html`

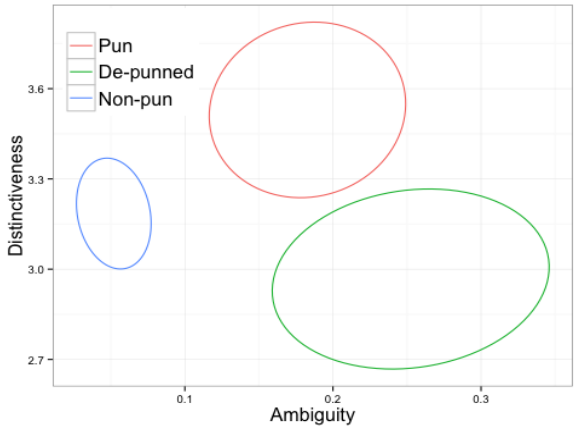


Figure 3: Standard error ellipses of ambiguity and distinctiveness across sentence types. Puns score higher on ambiguity and distinctiveness; de-puns are less supported by distinct focus sets; non-puns have low ambiguity.

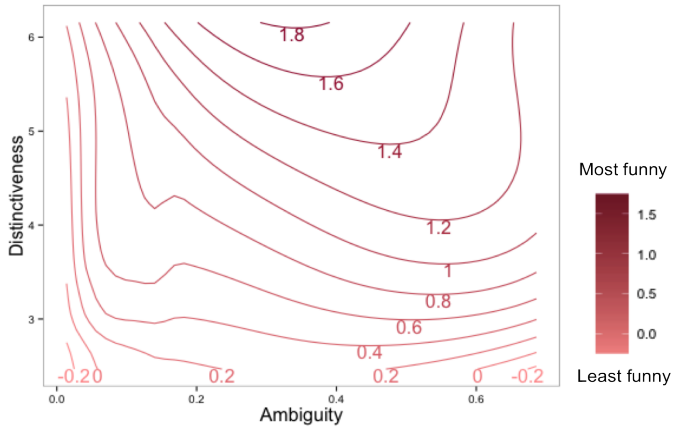


Figure 4: Funniness contours smoothed using a 2-D Loess regression with ambiguity and disjointedness measures as predictors. Sentences become funnier as they move to high ambiguity and distinctiveness space.

cantly with human ratings of funniness ($r = 0.21, p < 0.005$).

A linear regression showed that both ambiguity and distinctiveness are significant predictors of funniness. Together, the two predictors capture a modest but significant amount of the variance in funniness ratings ($F(2,232) = 20.86, R^2 = 0.145, p < 0.001$; see Table 2). Using both ambiguity and distinctiveness as dimensions that formalize incongruity, we can distinguish among puns, non-puns, and de-punned sentences, as shown in Figure 3. Figure 3 shows the standard error ellipses for each of the three sentence types in the two-dimensional space of ambiguity and distinctiveness. Although there is a fair amount of noise in the predictors (likely due to simplifying assumptions, the need to use empirical measures of relatedness, and the inherent complexity of humor) we see that pun sentences tend to cluster at a space with higher ambiguity and distinctiveness. While de-punned sentences are also high on ambiguity (e.g. it is ambiguous whether the word "hare" in "The professor got so mad he pulled his hare out" should be interpreted as *hair*), they tend to have lower distinctiveness measures. Non-puns score the lowest on ambiguity with moderate distinctiveness measures.

Figure 4 shows the funniness contours in the two-dimensional ambiguity-distinctiveness space smoothed using a 2-D Loess regression. Not only do the three types of sentences differ along the two dimensions, but sentences become funnier as they increase in ambiguity and distinctiveness. These results suggest that our measures of incongruity capture an important aspect of humor in pun sentences.

Beyond predicting the funniness of a sentence, our model can also tell us which particular features of a pun make it amusing. By finding the most likely semantic focus sets $\vec{f}$ given each latent meaning variable m and the observed words, we can identify words in a funny sentence that are critical to producing incongruity and humor. Table 3 shows the most likely semantic focus sets given each meaning for two groups of sentences. Sentences in each group contain the same pair of candidate meanings for the target word h. However, they differ in measures of ambiguity, distinctiveness, and funniness. Words in the most likely focus sets given m_1 are in red; words in the most likely focus sets given m_2 are in green; and

words in the most likely focus sets of both meanings are in dark blue. We observe that visually, the two pun sentences (which are significantly funnier) have more distinctive and balanced sets of focus words for each meaning than other sentences in their groups. De-punned sentences tend to have fewer words in support of m_1, and non-pun sentences tend to have no words in support of the interpretation that was not observed. Moreover, imagine if you were asked to explain why the two pun sentences are funny. The colorful words in each pun sentence—for example, the fact that magicians tend to perform magic tricks with hares, and people tend to be described as pulling out their hair when angry—are what one might intuitively use to explain why the sentence is a pun. Our model thus provides a natural way of not only formalizing incongruity and using it to predict when a sentence is a pun, but also to explain what aspects of a pun make it funny.

Discussion

Researchers in artificial intelligence have argued that given the importance of humor in human communication, computers need to generate and detect humor in order to interact with humans more effectively (Mihalcea & Strapparava, 2006). However, most work in computational humor has focused either on joke-specific templates and schemata (Binsted, 1996; Kiddon & Brun, 2011) or surface linguistic features that predict humorous intent (Mihalcea & Strapparava, 2006; Reyes et al., 2010). Our work moves beyond these approaches and directly utilizes a model of sentence comprehension to derive theory-driven measures of humor.

While the measures we developed account for a significant amount of variance in funniness ratings, there are several ways to improve our model of language in order to more accurately capture the subtleties of linguistic humor. By making the simplifying assumption that semantic association drives sentence comprehension, we disregarded the sequential structure of language that is often important for understanding a pun. For example, "The actors had one great movie after another. They were on a role." scores high on funniness but low on our measures because it leverages the idiomatic expression "on a roll" to boost the interpretation *roll*. Since our bag-of-words model does not account for word sequences, the measures we derive fail to fully capture the incongruity of many pun sentences that contain idiomatic expressions. In future work, we aim to incorporate information about the sequential structure of a sentence to further improve our language model and measures of incongruity.

In this paper, we showed how a basic model of sentence comprehension can illuminate incongruous sentence interpretations with rich social and linguistic meaning. Although our task in this paper is limited in scope, we believe that it represents a step towards developing models of language that can explain complex phenomena such as humor. From the perspective of language understanding, such phenomena can serve as probes for developing models of language that account for the subtleties of linguistic behavior. We hope that our work contributes to research in humor theory, computational humor, and language understanding, with the aim to one day understand what makes us laugh and build robots that appreciate the wonders of word play.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship to JTK and a John S. McDonnell Foundation Scholar Award to NDG. We thank Stu Melton and Mia Polansky for helping prepare part of the materials used in this paper.

References

Binsted, K. (1996). Machine humour: An implemented model of puns.

Burke, D., & Yee, P. (1984). Semantic priming during sentence processing by young and older adults. *Developmental Psychology*, *20*(5), 903.

Duncan, W., Smeltzer, L., & Leap, T. (1990). Humor and work: Applications of joking behavior to management. *Journal of Management*, *16*(2), 255–278.

Forabosco, G. (1992). Cognitive aspects of the humor process: The concept of incongruity. *Humor: International Journal of Humor Research; Humor: International Journal of Humor Research*.

Frank, M., & McGhee, P. (1989). *Humor and children's development: A guide to practical applications*. Routledge.

Gelkopf, M., & Kreitler, S. (1996). Is humor only fun, an alternative cure or magic? the cognitive therapeutic potential of humor. *Journal of Cognitive Psychotherapy*, *10*(4), 235–254.

Griffiths, T., Steyvers, M., Blei, D., & Tenenbaum, J. (2005). Integrating topics and syntax. *Advances in neural information processing systems*, *17*, 537–544.

Hurley, M., Dennett, D., & Adams, R. (2011). *Inside jokes: Using humor to reverse-engineer the mind*. Mit Pr.

Kiddon, C., & Brun, Y. (2011). Thats what she said: double entendre identification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 89–94).

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 234–243).

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090.

Martin, R. (2007). *The psychology of humor: An integrative approach*. Academic Press.

Mihalcea, R., & Strapparava, C. (2006). Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, *22*(2), 126–142.

Reyes, A., Buscaldi, D., & Rosso, P. (2010). The impact of semantic and morphosyntactic ambiguity on automatic humour recognition. In H. Horacek, E. Mtais, R. Muoz, & M. Wolska (Eds.), *Natural language processing and information systems* (Vol. 5723, p. 130-141). Springer Berlin / Heidelberg.

Seidenberg, M., Tanenhaus, M., Leiman, J., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive psychology*, *14*(4), 489–537.

Stewart, S., Stinnett, H., & Rosenfeld, L. (2000). Sex differences in desired characteristics of short-term and long-term relationship partners. *Journal of Social and Personal Relationships*, *17*(6), 843–853.

Veale, T. (2004). Incongruity in humor: Root cause or epiphenomenon? *Humor-International Journal of Humor Research*, *17*(4), 419–428.

Veatch, T. (1998). A theory of humor. *Humor*, *11*, 161–215.

Ziv, A., & Gadish, O. (1989). Humor and marital satisfaction. *The journal of social psychology*, *129*(6), 759–768.