Play on Words: Predicting Punniness with Statistics and Semantics

Justine Kao
Psychology Department
Stanford University
justnek@stanford.edu

Jireh Tan
Statistics Deparmtment
Stanford University
jirehtan@stanford.edu

Abstract

Linguistic humor is a puzzling psychological phenomenon that can be better understood through formal models of incongruity. Through behavioral data and computational modeling, we show that sentences containing phonetically ambiguous words are perceived as funny when different types of contextual information support incongruous interpretations of the ambiguous item. By modeling the tension between contexts using n-gram language models, part-of-speech tagging, and semantic analysis, we propose a computational model for pun understanding that may have implications on language processing, humor, and cognitive mechanisms behind other creative uses of language.

1 Introduction

Humor is ubiquitous in everyday social interaction, serving the purpose of establishing common ground, forming friendships and alliances, and the simple act of provoking laughter. Despite the many social benefits of humor, there is not yet a clear understanding of linguistic and cognitive mechanisms that give rise to the sensation that we all call "funny." In this paper, we propose and test a framework for pun identification and funniness prediction motivated by techniques in Natural Language Processing. These techniques include language modeling, part-of-speech tagging, and computational semantics. Through this framework, we hope to uncover important linguistic factors that allow multiple levels of contextual information to work together to create sophisticated social meaning in language.

We focus on an important type of linguistic humor in our analysis: homophone puns. Homophone puns are sentences that contain one or more words that are phonetically ambiguous and that can be interpreted in multiple ways given the surrounding context. Consider the following sentence:

(1) Being able to fit size fourteen shoes is quite a *[feet]*.

The phoneme sequence *[feet]* has two lexical forms: *feat* and *feet*. The interpretation *feat* is more likely given the preceding words *quite a*, while the interpretation *feet* is more likely given the semantically related word *shoes*.

There are two reasons why we focus on homophone puns in particular. The first is because the competing interpretations of a homophone pun are relatively shallow and easily recovered using the surface encoding. Given the properties of homophone words, we do not need to incorporate sophisticated methods to identify the area of ambiguity in a homophone pun or its possible interpretations. We simply need to locate the phonetically ambiguous word and look up its homophones in order to uncover the candidate meanings.

The second reason for choosing homophone puns is because it nicely incorporates three aspects of linguistic information—phonetic, because homophone words are phonetically ambiguous; syntactic, because puns often contain word sequences that are syntactically incoherent (e.g. "quite a *feet*"); and semantic, because the reason why sentence (1) is a pun and not simply a grammatically anomalous sentence is because the word *feet* is semantically related to

the word *shoes*. Given that NLP research has made significant progress in uncovering statistical properties underlying these three aspects of language, we believe that a computational analysis of homophone puns is the ideal case study to shed light on how different aspects of language interact to generate incongruity and humor.

Our paper is composed of two tasks: classification and prediction. In the first, we seek to design and test features that classify homophone pun sentences from non-pun sentences that contain homophone words. In the second, we seek to design and test features that predict the varying degrees of funniness within homophone puns. Together, these two tasks reveal increasingly fine-grained levels of features that contribute to ambiguity, incongruity, and punniness.

2 Background

Existing theories on humor agree that ambiguity is an essential component of most if not all jokes (Vaid & Ramachandran, 2001; Bekinschtein, Davis, Rodd, & Owen, 2011; Vaid, Hull, Heredia, Gerkens, & Martinez, 2003). (Koestler, 1964) argued that people engage in bisociative thinking when processing humorous input, meaning we simultaneously juggle multiple incongruous interpretations of a single situation. While context in normal discourse usually produces an unequivocal interpretation of ambiguous input, humorous utterances tend to contain a certain degree of unresolvable incongruity. This can be illustrated by comparing puns to other ambiguous sentences such as garden paths. Garden path sentences offer a single "correct" interpretation of an ambiguous item, which may be difficult to discover at first but eventually emerges after reanalysis (Ferreira & Henderson, 1991; Altmann, Garnham, & Dennis, 1992). On the other hand, multiple interpretations of an ambiguous item are simultaneously "correct" in puns, and ambiguity resolution is often incomplete.

Previous research on automatic humor generation and recognition has uncovered interesting properties of linguistic humor. (Binsted, 1996) used framing and phonetic relationships to automatically generate puns; (Mihalcea & Strapparava, 2006) and semantic2010 identified stylistic and semantic features that helped distinguish one-liner jokes from non-humorous sentences. However, most of the work in computational humor has focused either on utilizing joke-specific templates and schemata, or identifying linguistic features that strongly predict humorous intent (e.g. slang and alliteration). The former type of studies is restricted to identifying jokes with a very specific format and structure, while the latter type falls short of testing or building upon deeper and more general theories of humor involving the management of ambiguity. We hope that our work will be able to move beyond these two types of approaches and directly utilize humor theories on linguistic incongruence to identify humorous texts. Given that humor theorists view ambiguity and incongruent interpretations as an essential component of jokes (Martin, 2007; Hurley et al., 2011), we aim to quantify the rather fuzzy concepts of ambiguity and incongruity and explore whether human ratings of funniness can be modeled by these formalizations.

3 Classification of Puns and Non-Puns

In our first task, we explore and test features to classify homophone pun sentences and non-pun sentences that also contain homophone words. Most of the work was in the feature selection process: we had to build features that could give us a lot of information about the pun sentence. We note that the typical bag-of-words model typically used in other text classification tasks (such as spam filtering) is practically useless here. The fact that a word is a pun means that the word can appear across many different contexts with many different meanings; the joy of a pun arises in the moment of disambiguation. In this case, we had to build features that actively looked at contextual clues: in particular syntactic coherence with neighboring words and semantic coherence with content words. The idea was that we can build a good classification model by using features that the human brain takes into account whenever it encounters a pun.

3.1 Corpus

We constructed a corpus consisting of 40 homophone puns and 80 non-pun sentences that contain the same homophone items as the pun sentences. Ta-

Table 1: Example sentence in each category.

Pun	Thieves have muscles of *steal*.
Non-pun	The boy wanted to *steal* the candy.
Non-pun	The ship is made of wood and *steel*.

ble 1 shows an example sentence in each category.

The 40 pun sentences were taken in their original forms from a website called "Pun of the Day," which contains a over a thousand original puns. We selected 40 puns in which the ambiguous item is a single phonetically ambiguous word. We also filtered through the puns such that no two puns in the collection have the same ambiguous item.

The non-pun sentences were constructed to match each pun sentence on its ambiguous word. The sentences were either taken from Wikipedia or dictionary entries of the ambiguous lexical item (in this example, *"steel"* and *"steal"*) or hand-written when the entries were too technical to sound natural. This design ensured that puns and non-pun sentences contained the same phonetically ambiguous words, a control that was not used in previous work on automatic humor recognition (Mihalcea & Strapparava, 2006; Reyes, Buscaldi, & Rosso, 2010).

3.2 Classification features

Our features were motivated by the idea that the existence of multiple incongruent interpretations of a sentence is key to its funniness. We hypothesized that sentences containing phonetically ambiguous words are perceived as funny when different types of contextual information support conflicting interpretations of the ambiguous item to a similar degree, thus preventing the reader from using context to unequivocally disambiguate among the candidate interpretations. To this end, we identified several kinds of contexts that contribute to the interpretation of an ambiguous word. By deriving various measures based on the degree to which different contexts support different interpretations, we hoped to come up with formalizations of "ambiguity" and "incongruity" that can successfully classify humorous puns.

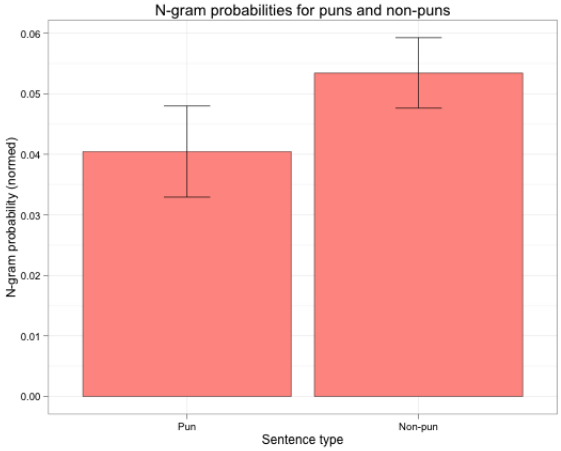

Figure 1: N-gram probabilities for puns and non-pun sentences

3.2.1 Context as neighboring words

The first set of features were developed based on n-gram probabilities. We observed that pun sentences often contain word sequences that are syntactically anomalous (e.g. "My computer is so slow it *hertz*"). Based on this intuition, we computed n-gram probabilities for all 120 sentences using Google n-grams as an approximate measure of syntactic coherence. We hypothesized that puns would be less syntactically coherent (have lower n-gram probabilities) than non-pun sentences and tested this intuition.

Surprisingly, results did not support our predictions. While Figure 1 shows that pun sentences on average have lower n-gram probabilities than non-pun sentences, the difference was not significant ($p = 0.17$). We thus decided that raw n-gram probabilities would not be helpful features for our classification task.

Instead, we looked into the n-gram probabilities of the sentences in which the observed homophone word—the word that originally appeared in the sentence—was replaced by its homophone. For example, for the pun sentence in Table 1, we computed the n-gram probabilities for both "Thieves have muscles of *steal*" and "Thieves have muscles of *steel*". We call the first sentence the *observed* sentence, and the second sentence the *unobserved* sentence. Figure 2 shows an interaction between sentence type and observed and unobserved n-gram

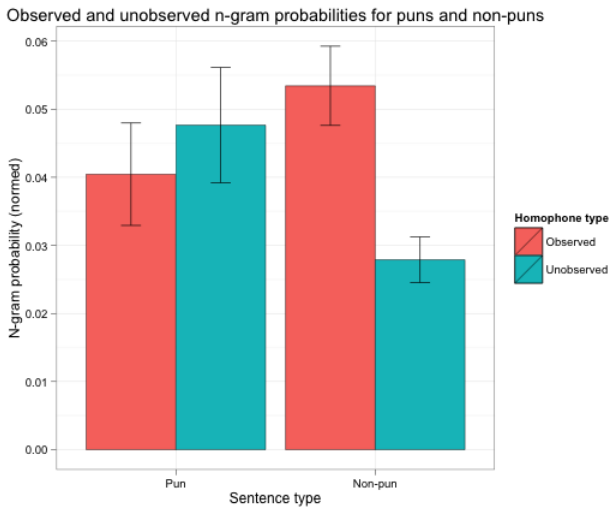

Figure 2: Observed and unobserved n-gram probabilities for puns and non-pun sentences

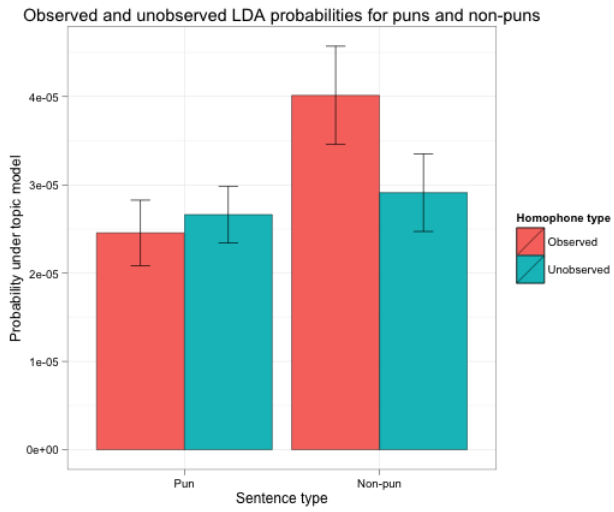

Figure 3: Observed and unobserved n-gram probabilities for puns and non-pun sentences

probability.

It appears that while the observed version of non-pun sentences are significantly more coherent than their unobserved counterparts, the observed and unobserved versions of pun sentences are similarly coherent as measured by n-gram probabilities. As a result, we decided to use the ratio between observed and unobserved n-gram probabilities as a feature for classification. This feature essentially measures the degree to which the observed interpretation is more likely given the immediately neighboring context than the unobserved interpretation.

3.2.2 Context as semantically related words

Previous research has shown that semantic relatedness and thematic fit guide the disambiguation of ambiguous sentences (Binder, Duffy, & Rayner, 2001; Tanenhaus & Lucas, 1987). Based on these findings, we posit that semantically related words beyond those immediately neighboring the ambiguous homophone also contribute to its interpretation. We thus sought to quantify semantic relatedness between the different interpretations of the homophone word and the rest of the sentence.

We first approached this problem using topic modeling, or specifically Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003). We trained an LDA model on the American National Corpus with 300 topics and computed the likelihood of each sentence under the topic model to obtain a "topical coherence score." A sentence with a high likelihood under the topic model is more topically coherent.

Figure 3 shows that the observed version of non-pun sentences are more topically coherent (had higher probabilities under the LDA model) than the pun sentences. However, after closer inspection, we realized that several word pairs that are considered semantically related based on intuition (e.g. *hare* and *magician*) received low topical coherence under the topic due to the fact that those particular words and words related to them were never observed in the same document. Many "topics" that puns exploit rarely appear in the kinds of corpora that topic models tend to be trained on. As a result, we decided to incorporate more direct and empirically derived measures of semantic relatedness.

We constructed a total of 474 word pairs for the observed version of the sentences, one for each content word in the sentence paired with the observed homophone. We constructed a separate 474 word pairs for the unobserved version of the sentences, one for each content word in the sentence paired with the unobserved homophone. We then asked 80 subjects on Amazon Mechanical Turk to rate the semantic relatedness of those word pairs. Each subject rated roughly 120 word pairs on a scale of 1 to

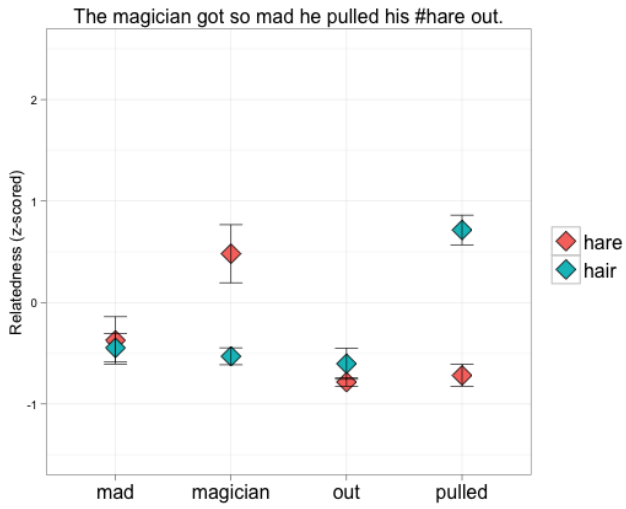

Figure 4: Relatedness ratings with observed and unobserved homophones in pun sentence

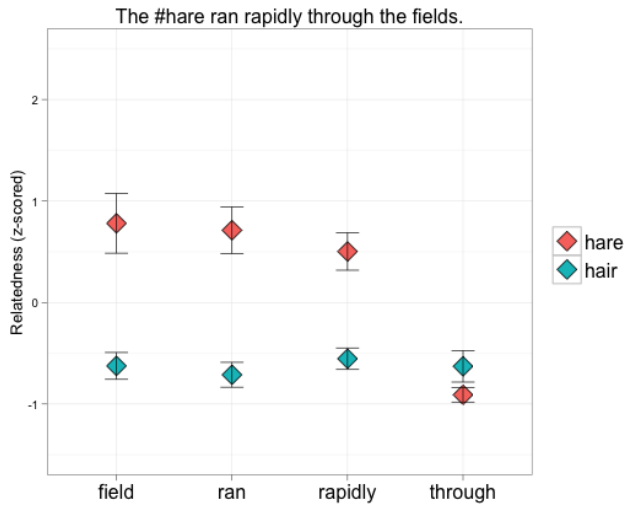

Figure 5: Relatedness ratings with observed and unobserved homophones in non-pun sentence

7. Ratings were standardized (z-scored) within each subject.

Using empirically measured relatedness, we were able to get at much more fine-grained and commonsense based measures of semantic relatedness than the ones provided by a topic model. Figure 4 shows the relatedness ratings of each content word in the pun sentence and the two homophone interpretations. We see that it matches our intuition and commonsense judgment of the word *magician* being more related to the *hare* interpretation than the *hair* interpretation. Figure 5 shows the relatedness ratings of each content word in the non-pun sentence and the two homophone interpretations.

We observe from these two examples that pun sentences tend to have content words that prefer different interpretations, while content words in non-pun sentences tend to unanimously prefer the observed interpretation. Based on this observation, we computed a measure of how much, on average, the observed interpretation is "preferred" by the semantic context, where preference is measured by the difference between the semantic relatedness ratings for the observed homophone and the unobserved homophone.

Figure 6 confirms the intuition that content words in puns seem to prefer both homophones equally on average, while content words in non-puns strongly prefer the observed homophone. This serves as our formalization of "semantic ambiguity"—the weaker the semantic preference for either homophone, the more ambiguous the interpretation of the homophone in context. Furthermore, we predict that sentences where the homophone word is ambiguous in context are more likely to be puns.

3.3 Building a Classification Model

For this task, we chose to use a support vector machine to classify the puns. We decided that the SVM would be the most appropriate classifier for this task, as SVMs are able to capture nonlinearities in the data through the kernel trick, and in general are more resistant to the problem of high-dimensionality than logistic regression. We used a radial kernel and proceeded to tune the hyperparameters, C and γ, by using a grid search. Over the grid of values for C and γ, we selected the pair that led to the best 10-fold cross-validation error. We found that we could get a good cross-validation error of 0.1143368 by using $C = 10, \gamma = 0.001$.

Given that SVMs are resistant to the addition of variables, we also wanted an idea of which features were most useful in contributing to the separation of the data, in order to test theories of humor. To do this, we examined each measure in turn. We found that merely including the semantic ambiguity

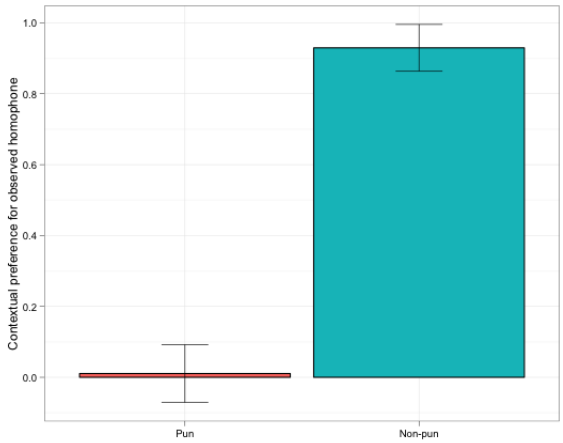

Figure 6: Average semantic preference for observed homophone in pun and non-pun sentences

measure described in Section 3.2.2 led to a cross-validation error of 0.1197091, implying that most of the separation in the data was successfully captured by semantic ambiguity. This implies that puns are most successfully distinguished from non-puns by taking into account not only its local coherence with neighboring words, but rather its global coherence with the semantic theme of the rest of the sentence.

4 Predicting Funniness

After designing and testing features that successfully distinguish between pun and non-pun sentences, we now move on to more carefully examine pun sentences with the goal of predicting human ratings of funniness.

4.1 Corpus

We constructed a corpus consisting of 80 homophone pun sentences. 40 of the pun sentences, which we call *original puns*, were taken in their original forms from a website called "Pun of the Day," which contains a over a thousand original puns. We selected 40 puns in which the ambiguous item is a single phonetically ambiguous word. We also filtered through the puns such that no two puns in the collection have the same ambiguous item.

The other 40 pun sentences, which we call *modified puns*, were constructed by changing the lexical form of the phonetically ambiguous item in the original puns. We replaced the ambiguous word in the original pun with its homophone. "Thieves have muscles of *steel*" is an example of a modified pun, where we replaced *"steal"* with the lexical form of the alternate interpretation, *"steel"*. This allows us to test whether or not boosting the likelihood of a particular interpretation by spelling it out directly changes the perceived funniness of the pun.

Table 2 shows an example sentence in each category.

Table 2: Example sentence in each pun category.

Original pun	Thieves have muscles of *steal*.
Modified pun	Thieves have muscles of *steel*.

4.2 Human Ratings of Funniness

We recruited subjects on Amazon Mechanical Turk to obtain human funniness ratings of the sentences. Each subject read 40 sentences, 10 from each type of sentence, and rated the funniness of each sentence on a scale from 1 to 7. Subjects were also randomly quizzed on the content of 5 sentences in order to ensure that they read each sentence carefully. After filtering out non-native English speakers and removing responses from subjects who answered 2 or more content questions incorrectly, 64 subjects' responses were collected and analyzed. Ratings were standardized for each subject, and the analysis was conducted on mean ratings for each sentence over subjects.

Consistent with our predictions, results show that people rated pun sentences as significantly funnier than non-pun sentences. People also gave significantly higher ratings for original pun sentences than for the manipulated pun sentences, with a significant interaction between sentence type and manipulation.

We performed a spit-half correlation test on the funniness ratings to get a sense of reliability. Correlation was 0.53 ($p < 0.0001$).

4.3 Funniness Features

After exploring formalizations of ambiguity to classify puns and non-pun sentences in the previous section, we now further explore formalizations of "incongruity." What makes the interpretations of a homophone word more incongruous in one sentence

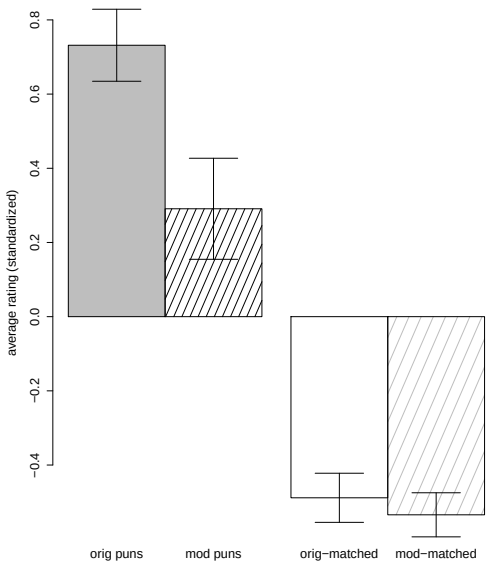

Figure 7: Mean pun ratings by category.

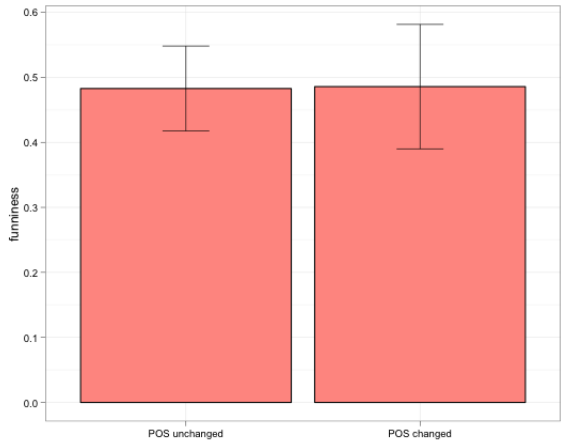

Figure 8: Mean pun ratings by category.

than in another? Does the level of incongruity between interpretations predict humans' subjective ratings of funniness?

4.3.1 Incongruity of homophone part of speech

We proposed that two interpretations of a homophone word are most incongruous when they are the most different. One way of measuring how "different" two interpretations are is through part-of-speech tags. If the two interpretations of a homophone word belong to different parts-of-speech, the two interpretations of the entire sentence given the different homophones must be more different since the syntactic structure of the sentence needs to be reconstructed. We thus predicted that homophone puns in which the two interpretations of the homophone belong to different parts-of-speech are funnier than homophone puns in which the two interpretations of the homophone belong to the same part-of-speech.

We used the Stanford Part-of-Speech tagger to identify pun sentences in which the two homophone interpretations belong to different grammatical categories. We then compared the funniness ratings of pun sentences in which the parts-of-speech were the same to pun sentences in which the parts-of-speech switched between interpretations. However, results did not support our hypothesis. Instead, Figure 8 suggests that whether parts-of-speech switches between interpretations does not predict the funniness of the pun at all. We thus neglected this feature as a predictor of funniness, since it was perhaps too coarse-grained of a measure for incongruity and distance between interpretations.

4.3.2 Incongruity of homophone meaning distributions given semantic context

Our next measure focused solely on using semantic relatedness to quantify incongruity between interpretations. We did this using a probabilistic model of punniness that formalizes the incongruity theory of humor. Interpretations of ambiguous items are made probabilistically given surrounding context, and the contexts in focus are selected probabilistically as well.

Suppose sentence S is composed of words $w = \{w_1, w_2, \ldots, w_l\}$ and ambiguous phoneme sequence h. Suppose h can be interpreted in two ways: $\{h_o, h_u\}$, where h_o is the *observed* interpretation and h_u is the *unobserved* interpretation.

We assume that at a given point in time, only a subset of w_i in w contributes to the interpretation of h. This is a "noisy" model in which listeners assume that speakers have a probability of uttering spurious or incorrect words that do not contribute to the interpretation of the ambiguous word. In other words, only the words in focus at a point in time are considered context words and contribute to the interpreta-

tion of the homophone word.

Next we need to compute the probability of a meaning given a context (without having observed the homophone sequence itself). Using a generative model, we assume that words $\{w_1, w_2, \ldots, w_n\}$ in the context are generated independently given an interpretation h. Thus

$$P(h|w_1, \ldots, w_n) = \frac{P(w_1, \ldots, w_n|h)P(h)}{P(w_1, \ldots, w_n)}$$

$$= P(h) \prod_i^n \frac{P(w_i|h)}{P(w_i)}$$

Since $P(w_i|h)$ is difficult to obtain empirically, we approximate it through word pair relatedness ratings of (w_i, h).

$$P(w|h) = \frac{P(w, h)}{P(h)}$$

Thus

$$P(h|w_1, \ldots, w_n) = P(h) \prod_i^n \frac{P(w_i|h)}{P(w_i)}$$

$$= P(h) \prod_i^n \frac{\frac{P(w_i,h)}{P(h)}}{P(w_i)} = P(h)^{1-n} \prod_i^n \frac{P(w_i, h)}{P(w_i)}$$

In log space,

$$\log P(h|w_1, \ldots, w_n) =$$

$$(1-n) \log P(h) + \sum_i^n \log P(w_i, h) - \sum_i^n \log P(w_i)$$

We obtained relatedness ratings $R(w_i, h)$ for word pairs (w_i, h), as described in section 3.2.2. We construe of relatedness measures $R(w_i, h)$ as point wise mutual information (PMI) between w_i and h.

$$R(w_i, h) = \log \frac{P(w, h)}{P(w)P(h)}$$

Thus

$$\log P(h|w_1, \ldots, w_n) = \log P(h) + \sum_i^n R(w_i, h)$$

and we can directly use empirical relatedness ratings for the measure.

Since we propose that puns are funniest when the interpretations of the homophone word are the most incongruous, we predict that funniness ratings are positively correlated with the distance between probability distributions of meanings h given different contexts. Thus, we wish to select two contexts c_1 and c_2 within a sentence for which the distributions $P(m|c_1)$ and $P(m|c_2)$ are maximally different. We measure the difference between the two distributions using symmetrized KL divergence.

Since the possible number meanings m are constrained by the ambiguous phoneme sequence h, we are only interested cases where $m = h_o$ and $m = h_u$. As a result, we approximate the difference between the two distributions using a partial KL divergence measure that considers only the probability of the meaning when $m = h_o$ or $m = h_u$. Our measure of funniness is thus simply

$$\max_{c_1,c2} \left[D_{KL}(P(h|c_1)||P(h|c_2)) + D_{KL}(P(h|c_2)||P(h|c_1)) \right]$$

To recap, this measure approximates the degree to which interpretations of the homophone sequence are different given different contexts. Interestingly,

$$\arg\max_{c_1,c2} \left[D_{KL}(P(h|c_1)||P(h|c_2)) + D_{KL}(P(h|c_2)||P(h|c_1)) \right]$$

gives us the contexts that most supports each context. As a sanity check, we were very happy to find that the measure selects sensible contexts that intuitively felt the most different in their support for either homophone. Table 3 lists some examples.

Table 3: Example contexts selected by KL

The magician was so mad he pulled his *hare* out.	
c_1 : magician	c_2: mad, out, pulled
Thieves have muscles of *steal*.	
c_1 : thief	c_2: muscle
Authors in jail have their *prose* and cons.	
c_1 : authors	c_2: jail, cons

4.4 Model Predictions of Funniness

We constructed a linear regression model to predict human funniness ratings using the measure of meaning incongruity given semantic context described in

the previous section. Since we observed that the KL measure seemed to be log-distributed, we transformed it into log space prior to constructing a linear regression model. Results are described below.

	Estimate	t value	Pr(>|t|)
(Intercept)	0.2679	2.61	0.0111
log(KL)	0.0569	2.61	0.0111

The correlation between log(KL) and funniness ratings for each pun sentence was 0.3 ($p = 0.011$). Although this was not a very strong correlation, we would like to note that the funniness ratings themselves carry a lot of noise since they were subjective and had only a split-half correlation of 0.5. Given these factors and the inherent difficulty of deriving measures of humor, we were quite pleased with the result we obtained from measuring the KL distance between meaning distributions given different contexts.

5 Discussion

The linguistic and psychological principles governing successful verbal humor are complex and often seemingly intangible. In this paper, our goal was to begin quantifying some of the abstract principles and theories that humor theorists have championed for decades.

First we set out on the daunting task of identifying humorous puns from a set of phonetically ambiguous sentences. The features we developed and our classification results suggested that quantifying both syntactic and semantic ambiguity allowed us to perform fairly well at classifying homophone puns. In particular, semantic ambiguity—as measured by semantic relatedness of the rest of the sentence to the homophone word—alone obtained promising results. This suggests that the kind of incongruity that generates humor arises at the level of semantic coherence instead of at the level of local or syntactic coherence. This possibility was further confirmed by our findings in the funniness prediction task, where incongruity of part-of-speech came nowhere close to predicting human ratings of funniness compared to the incongruity in meaning distribution given semantic context.

Overall, the features we propose and the results we obtained shed some light on the direction towards which research on computational humor could turn. Formalizing ambiguity and incongruity involves understanding the various levels of information that contribute to language interpretation as well as the sometimes conflicting roles of syntactic and semantic context. By separately examining the effects that different kinds of context have on incongruity and humor, we may begin to gain a deeper understanding of how language gives rise to social meaning.

References

Altmann, G., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, *31*(5), 685–712.

Bekinschtein, T., Davis, M., Rodd, J., & Owen, A. (2011). Why clowns taste funny: the relationship between humor and semantic ambiguity. *The Journal of Neuroscience*, *31*(26), 9665.

Binder, K., Duffy, S., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language*, *44*(2), 297–324.

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

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Boyd-Graber, J., Blei, D., & Zhu, X. (2007). A topic model for word sense disambiguation. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (emnlp-conll)* (pp. 1024–1033).

Cai, J., Lee, W., & Teh, Y. (2007). Improving word sense disambiguation using topic features. In *Proc. of emnlp.*

Ferreira, F., & Henderson, J. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745.

Griffiths, T., & Steyvers, M. (2002). A probabilistic approach to semantic representation. In *Proceedings of the 24th annual conference of the cognitive science society* (pp. 381–386).

Koestler, A. (1964). The act of creation. *New York*.

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.

Tanenhaus, M., & Lucas, M. (1987). Context effects in lexical processing. *Cognition*, *25*(1), 213–234.

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.

Vaid, J., & Ramachandran, V. S. (2001). Laughter and humor. In C. Blakemore, S. Jennett, & S. Jennett (Eds.), *The oxford companion to the body* (p. 426-427). Oxford University Press.