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 support of different viewpoints—and use them to predict humans' judgments of funniness. Results showed that both ambiguity and support 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

Motivate humor and its importance

(1) Humor is everywhere! It has cognitive, social, and health implications.

Humor theories: incongruity and resolution

(1) Introduce existing theories of humor

(2) Frame incongruity as presence of two equally likely interpretations

(3) Frame resolution as supporting context for each interpretation

(4) Introduce noisy channel and relevance (do this here or after introducing puns?)

Puns as case study

(1) Why language/meaning is hard

(2) Why puns are relatively simpler

(3) Walk through an example pun

Model

(1) Introduce language model (gist + ngram), plot generative model

(2) Introduce formal measures of ambiguity and support

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

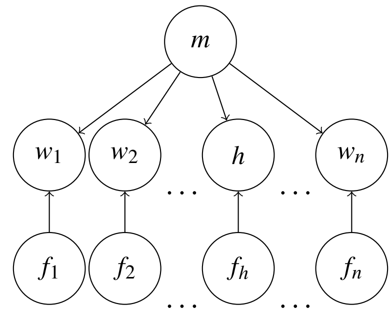


Figure 1: Generative model of a sentence. Each word w_i is generated based on the sentence topic m if the indicator variable f_i puts it in semantic focus; otherwise it is generated as noise (from a trigram distribution).

Measures

Ambiguity

$$\begin{aligned} P(m|\vec{w}) &= \sum_{\vec{f}} P(m,\vec{f}|\vec{w}) & (2)\\ &\propto \sum_{\vec{f}} P(\vec{w}|m,\vec{f})P(m)P(\vec{f}) & (3)\\ &= \sum_{\vec{f}} \left(P(m)P(\vec{f}) \prod_i P(w_i|m,f_i) \right) & (4) \end{aligned}$$

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

Support Under construction...

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

Model of meaning

(1) Describe relatedness ratings as approximation of PMI

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

(2) Describe empirically driven model of meaning

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

Evaluation

Experiment 1: Identical Homophone Puns

(1) Describe identical homophone corpus (should we include de-punned?)

(2) Describe experiment to obtain relatedness ratings

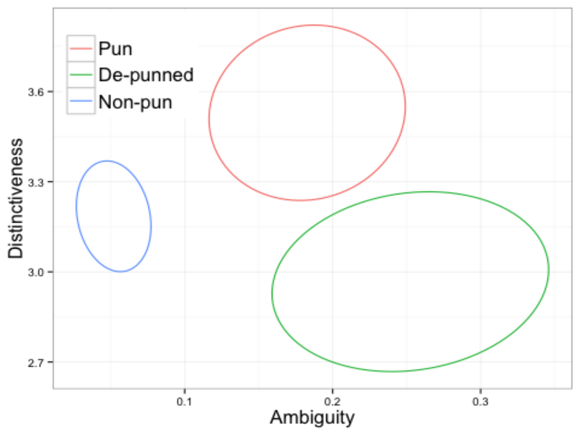


Figure 3: Standard error ellipses of ambiguity and distinctiveness across sentence types. Puns score higher on ambiguity and distinctiveness; de-puns are less supported by distinct focus sets; non-puns have low ambiguity.

(3) Describe experiment to obtain funniness ratings

(4) Show some plots summarizing data

(5) Regression model using ambiguity and support to predict funniness

(6) Plots summarizing results

Experiment 2: Near Homophone Puns

(1) Motivate purpose of examining near homophone puns: more fine-grained. Phonetic distance may introduce more or less noise that affects the delicate balance of ambiguity and support

(2) Describe near homophone corpus

(3) Describe experiment to obtain relatedness ratings

(4) Describe experiment to obtain funniness ratings

(5) Show some plots summarizing data

(6) Regression model using ambiguity and support to predict funniness, without accounting for phonetic distance

(7) Regression model using ambiguity and support to predict funniness, adding phonetic distance (hopefully will get better results)

(8) Table showing focus sets

Discussion

(1) Both experiments and their comparison suggest that funny puns exploit a delicate balance of meanings. By building a model of meanings and formally measuring that balance, we can predict funniness quite well.

(2) A relatively simple model of language and meaning can shed light on complex phenomena

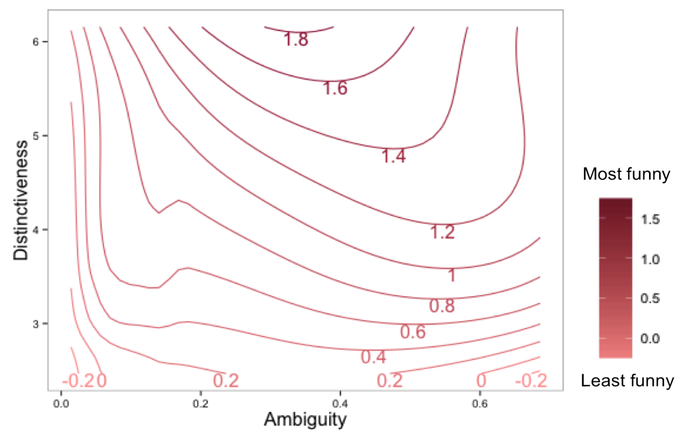
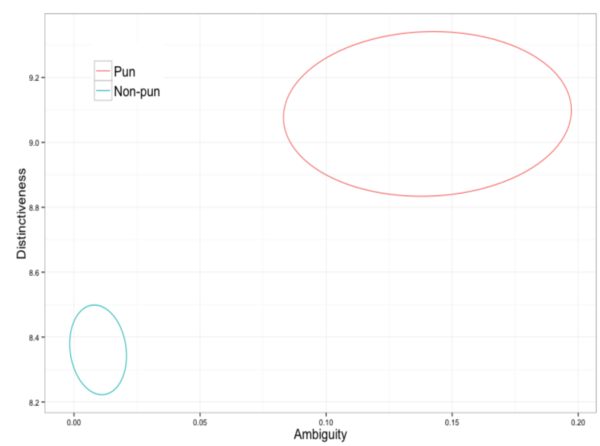
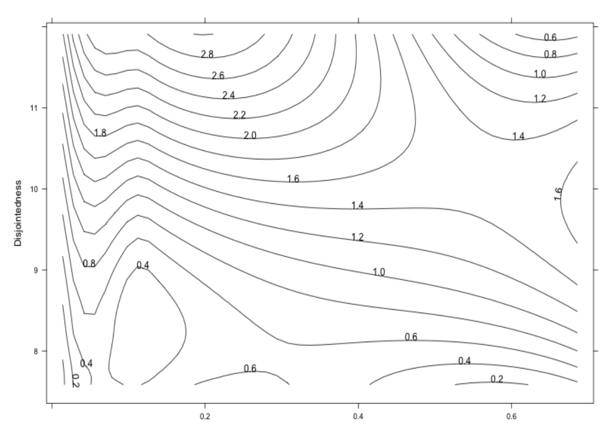
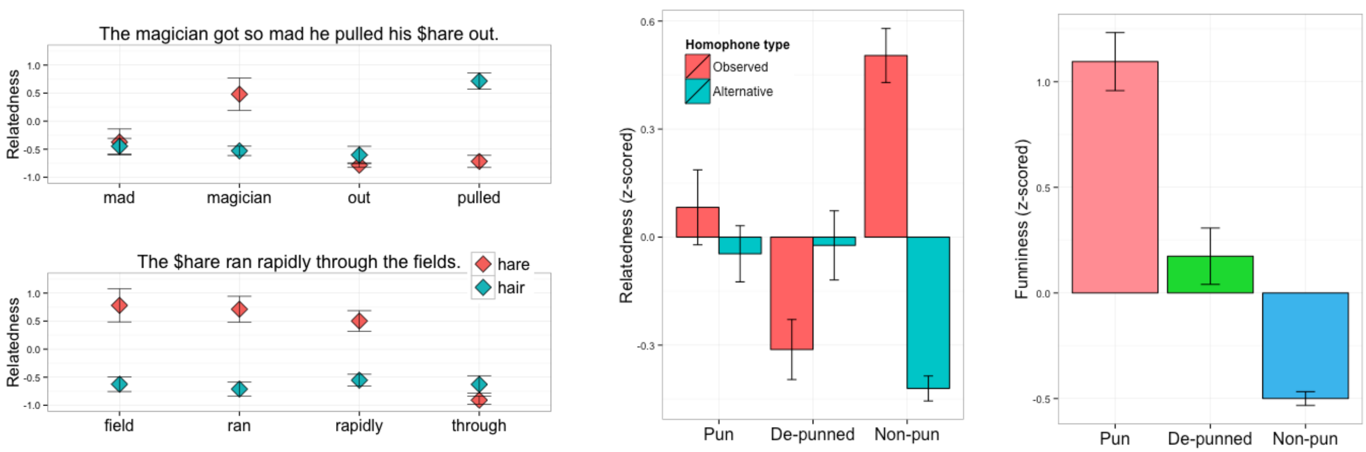


Figure 4: Funniness contours smoothed using a 2-D Loess regression with ambiguity and disjointedness measures as predictors. Sentences become funnier as they move to high ambiguity and distinctiveness space.







(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).