Pun documentation

September 18, 2012

1 Sentence sources

Puns $\times 40 \leftarrow$ punoftheday.com

Nonpuns $\times 80 \leftarrow$ Newbury House Dictionary: http://nhd.heinle.com

2 Terminology

"Original" and "modified" describe whether a word or sentence is presented as it naturally occurs in the wild (web); "observed" and "unobserved" describe whether the word is actually spelled out to the reader

Original pun = pun sentence in its original form as discovered on punoftheday.com, e.g. "Thieves have muscles of steal."

Original homophone = target homophone as spelled in the original sentence, e.g. "steal"

Modified pun = pun sentence in which we replaced the original homophone with its alternative spelling, e.g. "Thieves have muscles of steel."

Modified homophone = the alternative spelling of the homophone that was originally in the sentence

Original non pun = non pun sentence in its original form as discovered in the dictionary example sentences, e.g "The hare ran rapidly through the fields"

Modified non pun = non pun sentence in which we replaced the original homophone with its alternative spelling, e.g. "The hair ran rapidly through the fields."

Observed homophone = the spelling in which the target homophone appears in a sentence. e.g., in the original pun: "Thieves have muscles of *steal*," the original homophone "steal" is the observed homophone; while in the modified pun: "Thieves have muscles of *steel*," the modified homophone "steel" is the observed homophone.

Unobserved homophone = the spelling in which the target homophone does not appear in a sentence–namely an alternative phonetically identical word that is different from how the homophone is actually spelled out. e.g., in the original pun: "Thieves have muscles of *steal*," "steel" is the unobserved homophone.

3 Funniness measures

Measure1: Observed relatedness

A big value indicates that the content words strongly support the observed homophone.

measure1.ave = average relatedness of content words to the observed homophone

measure1.max = maximal relatedness of content words to the observed homophone

Measure 2: Relatedness difference between observed and unobserved

A small value indicates that the observed and unobserved homophones receive balanced support

measure2.ave = average difference between the relatedness of a content word to the observed and unobserved homophones

measure2.max = maximal difference between the relatedness of a content word to the observed and unobserved homophones

Measure 3: Absolute relatedness difference between observed and unobserved

A large value indicates that the observed and unobserved homophones are strongly supported by content words (but unclear which homophone receives stronger support)

measure3.ave = average absolute difference between observed and unobserved relatedness

Measure 4: Degree to which homophones have strong and biased supporters

Like Measure 3, a large value indicates that the observed and unobserved homophones are strongly supported by content words (but unclear which homophone receives stronger support)

measure4.max = $max(R(w, h_{Obs}) - R(w, h_{Unobs})) + max(R(w, h_{Uobs}) - R(w, h_{Obs}))$

an observed supporter is a word w such that $R(w, h_{Obs}) > R(w, h_{Unobs})$

measure4.sum = $sum(R(observedSupporter, h_{Obs})) + sum(R(unobservedSupporter, h_{Unobs}))$

Measure 5: Difference between strength of supporters for observed and unobserved homophone

Like Measure 2, a small value indicates that the observed and unobserved homophones receive balanced support from content words.

measure5.max = $max(R(w, h_{obs}) - R(w, h_{unobs})) - max(R(w, h_{unobs}) - R(w, h_{obs}))$

measure5.sum = $sum(R(observedSupporter, h_{Obs})) - sum(R(unobservedSupporter, h_{Unobs}))$