The effect of prosody and speaker voice information on the interpretation of hyperbole

Sunwoo Jeong, Simon Todd, Justine Kao, Noah Goodman, Meghan Sumner

Stanford University
Introduction

• Pragmatically enriched utterances in everyday life: non-literal interpretations
  – Irony: e.g. “That’s just great!”
  – Metaphor: e.g. “You are the light of my life.”
  – **Hyperbole**: e.g. “This place sells the best cheese cake in the universe!”, “This sandwich cost a million dollars!”
Prior studies on hyperbole

• Lexical and syntactic factors that affect the interpretation of hyperbole: McCarthy and Carter (2003), Loewenberg (1982)
  – Lexical items: numeric expressions (*millions of*), intensifying adverbs (*absolutely*), other adverbs (*literally*)
  – Syntactic structures: polysyndetic structure (*loads and loads and loads of*)

• Interactional factors: a shift in footing (Goffman 1979)
Prior studies on hyperbole

• An experimental study by Kao et al. (2014)
  – Subjects are presented with text stimuli (e.g. “It cost 1000 dollars.”) given in an interaction context
  – They are asked to rate the likelihood of 500, 1000, etc. being the actual price of the item.

• Finding: subjects’ responses can be predicted by the price priors and the affect priors
  – A rational speech act model (RSA) that successfully models the experiment data
Our hypotheses

• **Prosodic cues and speaker voice information** given in spoken words will also systematically affect the interpretation of hyperbole
Our hypotheses: the effect of prosody

• Categorical differences in prosodic contours will affect the interpretation of hyperbole
  – declarative contour (vs.) hyperbolic contour

• More incremental adjustments in finer grained prosodic cues will also affect the interpretation of hyperbole
  – higher pitch of the main accented syllable, longer duration of the main accented syllable
Our hypotheses: the effect of speaker voice information

• Macro-level social information about the speaker may influence the interpretation of hyperbole
  – e.g. speaker gender

• More specific impressions about the characteristics of the speaker may also influence the interpretation of hyperbole
  – e.g. sounding impulsive, thrifty, etc.
Methodology

• A perception study using spoken words

• Closely resembles the experiment design in Kao et al. (2014) which used text stimuli

• The stimuli are actual speech that have been carefully manipulated to test the effects of different kinds of prosodies
Methodology

1. Establishing the price priors for the items used in the stimuli

2. Main experiment

3. Post-experiment judgments on different speaker voices
2. Main experiment: stimuli

“I bought X yesterday. It cost Y dollars.”

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfume</td>
<td>a hundred</td>
</tr>
<tr>
<td>China set</td>
<td>a thousand</td>
</tr>
<tr>
<td>Necklace</td>
<td>a million</td>
</tr>
<tr>
<td>Chainsaw</td>
<td></td>
</tr>
<tr>
<td>Dumbell rack</td>
<td></td>
</tr>
<tr>
<td>Lawn mower</td>
<td></td>
</tr>
</tbody>
</table>
2. Main experiment: questions

Questions: How much do you think Y actually cost?

(answer in numeric free response)
2. Main experiment: speakers

- Stimuli recorded by 4 different speakers (2 males, 2 females) in two different prosodies
  - Declarative (L* on the main accented syllable)
  - Hyperbolic (H* on the main accented syllable)

- No manipulations on the declarative prosody recordings
- Further manipulations on the hyperbolic prosody recordings
2. Main experiment: prosodic manipulations

• The main accented syllables were normalized within each gender for the baseline hyperbolic prosodies

• The **pitch** and/or the **duration** of the main accented syllable bearing the pitch accent (**HUNDred, THOUSand, MILLion**) was manipulated, with minimal adjustment to adjacent syllables
2. Main experiment: prosodic manipulations

![Graph showing pitch fluctuations over time with examples of sentences: It cost me a thousand dollars.](image)
2. Main experiment: prosodic manipulations

- Baseline F0 values of the accented syllables multiplied by 1.25 to generate pitch-manipulated hyperbolic prosody tokens

- Baseline durations of the accented syllables multiplied by 1.5 to generate duration-manipulated hyperbolic prosody tokens

- Both durations and F0 values manipulated
## 2. Main experiment: design

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Item</th>
<th>Said Price</th>
<th>Manipulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male 1</td>
<td>Perfume Necklace</td>
<td>100</td>
<td>Declarative Hyperbolic</td>
</tr>
<tr>
<td>Male 2</td>
<td>China set</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>Female 1</td>
<td>Lawn mower</td>
<td>1000000</td>
<td></td>
</tr>
<tr>
<td>Female 2</td>
<td>Dumbbell rack</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chainsaw</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(P/D)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Condition 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(P)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Condition 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(D)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Condition 3</td>
</tr>
</tbody>
</table>
1. Establishing baselines

- Price prior experiment

\[ D \text{ bought } E. \text{ How much do you think } E \text{ cost?} \]

\( D \): feminine (Nicole), masculine (Jerry), or gender-neutral (Skyler) names

\( E \): stereotypically feminine items (perfume, necklace, china set), stereotypically masculine items (lawnmower, dumbbell rack, chainsaw); 6 items in total
1. Price priors: results
3. Subjective judgments on speaker voices

- A sentence with neutral propositional content spoken by the 4 speakers
  
  “I think he walked around the pond.”

- Questions related to impressions about the speaker voice
3. Subjective judgments on speaker voices

- **Like-me rating**: How much do you think the speaker is like you?
  - on a scale from 1-100

- **Impulsiveness**: Is the speaker likely to spend money without planning to?
  - yes/no

- **Thriftiness**: Is the speaker likely to save more than 20% of his/her income annually?
  - yes/no
3. Subjective judgments on speaker voices

• **Income**: What do you think is the most likely annual income of the speaker?
  – less than 50,000
  – between 50,000-100,000
  – more than 100,000)

• **Trendiness**: Is the speaker likely to be up to date on recent style trends? (yes/no)
Procedures

• Price prior experiment
  – 40 subjects
  – Less than 5 minutes

• Main experiment / post-experiment judgments
  – 150 subjects (50 for each of the 3 conditions)
  – Allowed multiple breaks, 30-45 minutes

• All recruited from Amazon Mechanical Turk
Interpreting the results

• Roughly, less frequent literal interpretations are interpreted as more hyperbolic readings

• Roughly, lower estimated actual prices are interpreted as stronger, more hyperbolic reading

• Some caveats (the complex pragmatics of hyperbole)
The semantics and pragmatics of hyperbole

• The Gricean view and the cooperative principle (Grice 1975)
  – The speakers are cooperative agents whose aim is to make relevant and helpful contributions to the conversation

• More and more exaggeration not always equivalent to stronger hyperbole

• Must be mindful of the pragmatics of hyperbole when interpreting the results
The semantics and pragmatics of hyperbole

“This perfume cost a million dollars!”

- Rule out the literal interpretation based on prior knowledge
- Reason that there must be a specific goal as to why the speaker decided to convey the non-literal meaning
- Derive affective subtext (usually, negative affect – e.g. Gibbs et al. 1991)
Results: the case of said price 100

“It cost a hundred dollars.”

Answers from the subjects (estimated actual price)
Results: the case of said price Million

"It cost a million dollars."

Answers from the subjects (estimated actual price)
Results: the case of said price 1000

“It cost a thousand dollars.”

Answers from the subjects (estimated actual price)
Results: the effect of prosody

• 3 different mixed effects models for different said price bins (hundred / thousand / million)

  – Main dependent variable: estimated actual price (responses from the subjects)
  – Independent variables: prosody, price prior means, price prior standard deviations, item sex
  – Random effects: speakers, subjects
Results: the effect of prosody

• At least three-way distinction between the effects of different prosodies:
  – Declarative
  – Hyperbolic (base)
  – Hyperbolic + Pitch + Duration
Results: the effect of prosody

• Baseline hyperbolic prosodies elicit significantly more hyperbolic responses (i.e. lower estimated actual price) than declarative prosodies (p <0.1)

• Hyperbolic prosodies with both pitch and duration manipulations elicit significantly more hyperbolic responses (i.e. lower estimated actual price) than declarative prosodies (p <0.001)
Results: the effect of prosody
Results: The effect of speaker voice information

- 3 different linear regression models for different said price bins (hundred / thousand / million)
  - Main dependent variable: estimated actual price (responses from the subjects)
  - Independent variables: prosody, price prior means, price prior standard deviations, item sex, impulsiveness, thriftiness, estimated speaker income, trendiness
  - Interactions between prosody and 4 kinds of speaker characteristics
Results: the effect of speaker voice information

• There was a weak speaker gender effect ($p > 0.05$) – female speakers elicited significantly lower estimated actual price (more hyperbole)

• However, this effect disappeared when more fine-grained speaker judgments were included as predictors (e.g. impulsiveness)

• Surface gender effects most likely mediated other more specific characteristics (cf) Ochs 1992, Podesva 2007)
Results: the effect of speaker voice information

• Other things being equal, in general, if the speaker is perceived as being:
  – **Impulsive** ➔ significantly **more hyperbolic interpretations**
  – **Thrifty** ➔ significantly **less hyperbolic interpretations**
  – **higher income** ➔ significantly **more hyperbolic interpretations**
By item: the effect of impulsiveness

- Higher peaks in the 1000 area (literal interpretation) for the non-impulsive cases
- More responses in 500 – 800 region for the impulsive cases
By item: the effect of perceived speaker income

- Perceived low income speakers elicit considerably higher literal interpretations
  - esp. in the case of lawn mower (combined with higher pitch)
Interactions between prosody and speaker voice information

• Complex, multi-faceted interactions between prosody and speaker voice information
  – If the speaker is perceived to be upper-class, higher pitch elicits significantly more hyperbolic interpretations
  – If the speaker is perceived to be lower-class, higher pitch elicits significantly less hyperbolic interpretations (more literal interpretations)
Estimated speaker income: low

Declarative Hyperbolic Hyp+Duration Hyp+Pitch Hyp+Pitch+Duration

Higher Pitch: Literal interpretation

Longer duration: Hyperbolic interpretation

estimated actual price

factor (manipulation)
Estimated speaker income: mid

- Higher Pitch: Hyperbolic interpretation
- Longer duration: Literal interpretation

- Declarative
- Hyperbolic
- Hyp+Duration
- Hyp+Pitch
- Hyp+Pitch+Duration

estimated actual price
Estimated speaker income: high

Declarative

Hyperbolic

Hyp+Duration

Hyp+Pitch

Hyp+Pitch+Duration

Longer duration: Literal interpretation

Higher Pitch: Hyperbolic interpretation

estimated actual price

factor (manipulation)
Interactions between prosody and speaker voice information

• Second example
  – If the speaker is perceived to be **impulsive**, the effect of higher pitch and longer duration become more pronounced.
  – If the speaker is perceived to be **not impulsive**, the effect of higher pitch and longer duration is considerably mitigated.
Speaker impulsiveness: yes
Speaker impulsiveness: no

![Box plot showing estimated and actual prices for different factors (manipulation).]

- Declarative
- Hyperbolic
- Hyp + Duration
- Hyp + Pitch
- Hyp + Pitch + Duration

The box plot compares the estimated and actual prices across different factors, with the x-axis representing the factor of manipulation and the y-axis showing the price range.
Interactions between prosody and speaker voice information

• A complex interaction between prosodic cues and speaker voice information

• A single prosodic cue such as higher pitch does not always categorically index a single kind of affective or pragmatic meaning

• Different emphasis compared to previous works (e.g. Paeschke 2004, Ko et al. 2014)
Conclusion

- Categorical differences in prosodic contours systematically affect the interpretation of hyperbole.

- Finer-grained differences in prosodic cues (e.g. higher pitch, longer duration) also systematically affect the interpretation of hyperbole.
Conclusion

• Speaker voice information such as perceived impulsiveness of the speaker also significantly influence the interpretation of hyperbole

• The study suggests important three-way interactions between prosody, pragmatic meaning, and social meaning
Selected bibliography


We thank Chris Potts, Penny Eckert, and Rob Podesva for their insightful comments.

We also thank the four speakers who helped us with the stimuli creation, as well as numerous other participants who helped us during our recording sessions.

This project was supported in part by funding from the Clayman Institute for Gender Research.