Modeling social meaning in probabilistic semantics/pragmatics

Ciyang Qing
Stanford Linguistics

LINGUIST 130A/230A
March 5, 2020

Joint work with Reuben Cohn-Gordon


Introduction

Truth-conditional analysis
Use-conditional meaning
General discussion

Linguistic variation

Linguistic variation is ubiquitous
- Sounds: /tuh-may-toe/ vs /tuh-mah-toe/
- Syntax: I did nothing vs I didn’t do nothing
- Lexical: queue vs line

Variation carries meaning
- Macrosociological categories
- Micro categories within a local community
- Dynamic construction of style

Eckert, Penelope (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation

Macrosociological categories

/tuh-may-toe/ → the speaker is American
/tuh-mah-toe/ → the speaker is British

First wave (Labov, 1966): correlation between linguistic variables and the macrosociological categories of socioeconomic class, sex, class, ethnicity, age . . .

Microsociological categories

Second wave: use ethnographic methods to explore micro categories within a local community

E.g., social categories in Detroit suburban high schools (Eckert 1989)
  - Jocks (∝ middle-class): standard negation
  - Burnouts (∝ working-class): nonstandard negation
Crucially, such categories are not fully determined by parents’ class.


Our working example: (ING)

Variants of (ING)

(1) a. I am working on my paper -ing (velar) variant
   b. I am workin’ on my paper -in’ (apical) variant

Campbell-Kibler (2009): listener’s perception of -ing
  → Sp is educated, articulated... competent
  → Sp is formal, distant... aloof

Labov (2012): Obama’s use of (ING) in three contexts
  - Barbecue: 72% -in’ (casual)
  - Press conference that follows: 33% -in’ (careful)
  - DNC acceptance speech: 3% -in’ (formal)

Dynamic construction of style

The first two waves focus on static categories
  - Variation simply reflects social category
  - A lack of agency

Third wave: speakers actively use variation as part of their stylistic practice
  - Construction of a certain persona in context

Overview

So, what is social meaning?

Lewis’s advice:
   “In order to say what a meaning *is*, we may first ask what a meaning *does*, and then find something that does that”

A case study for this lecture
  - Obama’s use of (ING) at the barbecue: 72% -in’

How to formally represent the social meanings of the two variants of (ING) and model Obama’s use of them?
  - Truth-conditional vs use-conditional approaches
Reference game

<table>
<thead>
<tr>
<th></th>
<th>hat</th>
<th>glasses</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>k</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Social Meaning Game as a reference game

Burnett (2019): Sp uses (ING) to convey a persona

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>comp. aloof</td>
<td>comp. friendly</td>
<td>incomp. aloof</td>
<td>incomp. friendly</td>
</tr>
<tr>
<td>k</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

- *ING*: competent or aloof
- *-in’*: incompetent or friendly

Obama at the barbecue, prior

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>comp. aloof</td>
<td>comp. friendly</td>
<td>incomp. aloof</td>
<td>incomp. friendly</td>
</tr>
<tr>
<td>k</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Obama at the barbecue, literal listener

Step 1: Point-wise product \([m(\pi)] \cdot \Pr(\pi)\)

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>comp. aloof</td>
<td>comp. friendly</td>
<td>incomp. aloof</td>
<td>incomp. friendly</td>
</tr>
<tr>
<td>k</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
### Obama at the barbecue, literal listener

**Step 2: Normalize each row to get** $P_{\text{lit}}(\pi | m)$

<table>
<thead>
<tr>
<th>$\pi$</th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp. aloof</td>
<td>comp. friendly</td>
<td>incomp. aloof</td>
<td>incomp. friendly</td>
<td></td>
</tr>
<tr>
<td>-ing</td>
<td>0.375</td>
<td>0.25</td>
<td>0.375</td>
<td>0</td>
</tr>
<tr>
<td>-in’</td>
<td>0</td>
<td>0.286</td>
<td>0.428</td>
<td>0.286</td>
</tr>
</tbody>
</table>

### Obama at the barbecue, speaker

**Step 1: Transpose**

<table>
<thead>
<tr>
<th>-ing</th>
<th>-in’</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.375</td>
<td>0</td>
</tr>
<tr>
<td>0.25</td>
<td>0.286</td>
</tr>
<tr>
<td>0.375</td>
<td>0.428</td>
</tr>
<tr>
<td>0</td>
<td>0.286</td>
</tr>
</tbody>
</table>

**Step 2: calculate** $\exp(\alpha \cdot \log(P_{\text{lit}}(\pi | m)) + C(m))$

- Burnett (2019) assumes $\alpha = 6$ and costs are 0

<table>
<thead>
<tr>
<th>-ing</th>
<th>-in’</th>
</tr>
</thead>
<tbody>
<tr>
<td>sad</td>
<td>0.0028</td>
</tr>
<tr>
<td>smiles</td>
<td>0.00024</td>
</tr>
<tr>
<td>confused</td>
<td>0.0028</td>
</tr>
<tr>
<td>.UN</td>
<td>0</td>
</tr>
</tbody>
</table>

**Model’s prediction**

$P_S(-\text{in’} | \text{Cool Guy}) = 0.692$

**Labov’s (2012) finding**

- **Barbecue: 72% -\text{in’}**

**Does the model’s prediction capture the empirical data?**

<table>
<thead>
<tr>
<th>-ing</th>
<th>-in’</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.308</td>
<td>0.692</td>
</tr>
<tr>
<td>0.311</td>
<td>0.689</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
What does speaker probability mean?

Model's prediction: $P_S(-\text{'in'} \mid \text{Cool Guy}) = 0.692$

- The probability that the speaker will use -\text{'in'} for the first instance of (ING)
- One-shot probability

Labov’s (2012) finding: 72% -\text{'in'} at the barbecue

- Uses of -\text{'in'} constitute 72% of the instances of (ING)
- Long-term frequency

These two are conceptually different.

- The real question for the model: What happens in the long run? (This is not addressed by Burnett)
- What happens after speaker’s first use of -\text{'in'}?

Obama at the barbecue, literal listener

We assume that after hearing -\text{'in'}, $P_{lit}(\pi \mid -\text{'in'})$ becomes the new prior over personae before the second (ING)

- An agent reasons about the agent 1-level below
- Even if we assume $P_{L}(\pi \mid -\text{'in'})$ becomes the new prior, the same problem will arise (exercise)

Obama at the barbecue, prior II

Now we have a second round of the reference game, which is the same as before except that the prior is $P_{lit}(\pi \mid -\text{'in'})$ from the previous round.

porno.png

Obama at the barbecue, literal listener II

(From now on, all the intermediate steps are skipped)

- $P_{lit}(\pi \mid -\text{'in'}) = Pr(\pi)$ (-\text{'in'} is no longer informative!)
Obama at the barbecue, speaker II

- **-ing**
  - Model’s prediction: $P_S(-in' | \text{Cool Guy}) = .118$
  - Obama is very likely to use **-ing** for the second (ING)

- **-in’**
  - $0.5$
  - $0.882$
  - $0.884$
  - $0$
  - $0.5$
  - $0.118$
  - $0.116$
  - $1$

Obama at the barbecue, literal listener II

- **-ing**
  - $0$
  - $0.4$
  - $0.6$
  - $0$

- **-in’**
  - $0$
  - $0.286$
  - $0.428$
  - $0.286$

$P_{lit}(\pi | -ing)$ becomes the new prior for the third (ING)

Obama at the barbecue, prior III

Round 3 of the game

<table>
<thead>
<tr>
<th>$\pi$</th>
<th>Stern Leader \text{comp. aloof}</th>
<th>Cool Guy \text{comp. friendly}</th>
<th>Asshole \text{incomp. aloof}</th>
<th>Doofus \text{incomp. friendly}</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-ing]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>[-in’]</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$Pr(\cdot)$</td>
<td>0</td>
<td>0.4</td>
<td>0.6</td>
<td>0</td>
</tr>
</tbody>
</table>

Obama at the barbecue, literal listener III

- **-ing**
  - $0$
  - $0.4$
  - $0.6$
  - $0$

- **-in’**
  - $0$
  - $0.4$
  - $0.6$
  - $0$

$P_{lit}(\pi | -in’) = P_{lit}(\pi | -ing) = Pr(\pi)$

- Neither variant is informative!
Obama at the barbecue, speaker III

- **ing** - **in’**

- **Model’s prediction**
  \[ P_S(-\text{in’} \mid \text{Cool Guy}) = 0.5 \]

  ▶ Obama is equally likely to use **-ing** or **-in’** for the third (ING)

- **What happens next and in the long run?**

Obama at the barbecue, literal listener III

<table>
<thead>
<tr>
<th>♦</th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp. aloof</td>
<td>comp. friendly</td>
<td>incomp. aloof</td>
<td>incomp. friendly</td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{array}{cccc}
\pi & \text{Stern Leader} & \text{Cool Guy} & \text{Asshole} & \text{Doofus} \\
\text{comp. aloof} & 0 & 0.4 & 0.6 & 0 \\
\text{comp. friendly} & 0.4 & 0.6 & 0 & 0 \\
\text{incomp. aloof} & 0.6 & 0 & 0 & 0 \\
\text{incomp. friendly} & 0 & 0 & 0 & 0 \\
\end{array}
\]

\[ P_{\text{lit}}(\pi \mid -\text{in’}) = P_{\text{lit}}(\pi \mid -\text{ing}) = \text{Pr}(\pi) \]

▶ The prior will not change for future uses of (ING)

▶ \[ P_S(-\text{in’} \mid \text{Cool Guy}) = 0.5 \text{ in the long run} \]

Summary of Burnett’s (2019) model

Key assumptions

▶ A reference game with personae

▶ Social meaning is truth-conditional

Prediction

▶ Chance-level use of either variant in the long run :-(

Problem

▶ Repeated use of the same variant is not informative

▶ A consequence of the truth-conditional analysis

Our proposal

Social meaning as use-conditional meaning

Use-conditional meaning (Kaplan 1999): to know the meaning of **oops** is to know when it is felicitously used, i.e., iff Sp observed a minor mishap

Our probabilistic formulation: to know the meaning of **oops** is to know how likely a normal competent speaker would use it under various circumstances
Our proposal

Instead of

\[ P_{\text{lit}}(\pi | m) \propto \Pr(\pi) \cdot [m](\pi) \]

Truth-conditional

Assume a hypothetical stereotypical speaker \( S_0 \)

\[ P_{\text{lit}}(\pi | m) \propto \Pr(\pi) \cdot S_0(m | \pi) \]

Use-conditional

\[ \begin{array}{cccc}
\pi & \text{Stern Leader} & \text{Cool Guy} & \text{Asshole} & \text{Doofus} \\
\hline
S_0(\text{-ing} | \pi) & 0.7 & 0.3 & 0.1 & 0.01 \\
S_0(\text{-in'} | \pi) & 0.3 & 0.7 & 0.9 & 0.99 \\
\end{array} \]

Obama at the barbecue, prior

Step 1: Calculate \( \Pr(\pi) \cdot S_0(m | \pi) \)

\[ \begin{array}{cccc}
\pi & \text{Stern Leader} & \text{Cool Guy} & \text{Asshole} & \text{Doofus} \\
\hline
\text{-ing} & 0.21 & 0.06 & 0.03 & 0.002 \\
\text{-in'} & 0.09 & 0.14 & 0.27 & 0.198 \\
\Pr(\cdot) & 0.3 & 0.2 & 0.3 & 0.2 \\
\end{array} \]

Step 2: Normalize each row

\[ \begin{array}{cccc}
\pi & \text{Stern Leader} & \text{Cool Guy} & \text{Asshole} & \text{Doofus} \\
\hline
\text{-ing} & 0.695 & 0.199 & 0.099 & 0.007 \\
\text{-in'} & 0.129 & 0.201 & 0.387 & 0.284 \\
\end{array} \]

Obama at the barbecue, literal listener
Obama at the barbecue, speaker

(Still assuming $\alpha = 6$ and 0 costs)

-\text{ing} -\text{in'}

$\approx 1$ $\approx 0$

Model's prediction

$P_S(-\text{in'} \mid \text{Cool Guy}) = 0.515$

$\approx 0$

$\approx 1$

Obama is basically indifferent between -\text{in'} and -\text{ing}$

Suppose he uses -\text{in'}; what happens next?

Obama at the barbecue, literal listener

\begin{align*}
\pi &\quad \text{Stern Leader} & \text{Cool Guy} & \text{Asshole} & \text{Doofus} \\
-\text{ing} & 0.695 & 0.199 & 0.099 & 0.007 \\
-\text{in'} & 0.129 & 0.201 & 0.387 & 0.284
\end{align*}

$P_{\text{in'}}(\pi \mid -\text{in'})$ becomes the new prior for the second (ING)

Obama at the barbecue, prior II

\begin{align*}
\pi &\quad \text{Stern Leader} & \text{Cool Guy} & \text{Asshole} & \text{Doofus} \\
S_0(-\text{ing} \mid \pi) & 0.7 & 0.3 & 0.1 & 0.01 \\
S_0(-\text{in'} \mid \pi) & 0.3 & 0.7 & 0.9 & 0.99 \\
Pr(\cdot) & 0.129 & 0.201 & 0.387 & 0.284
\end{align*}

Obama at the barbecue, literal listener II

\begin{align*}
\pi &\quad \text{Stern Leader} & \text{Cool Guy} & \text{Asshole} & \text{Doofus} \\
-\text{ing} & 0.47 & 0.314 & 0.201 & 0.015 \\
-\text{in'} & 0.048 & 0.174 & 0.431 & 0.348 \\
Pr(\cdot) & 0.129 & 0.201 & 0.387 & 0.284
\end{align*}
Obama at the barbecue, speaker II

(Still assuming $\alpha = 6$ and 0 costs)

-ing  -in’

$\approx 1 \approx 0$

Model’s prediction

$P_S(-\text{in’} \mid \text{Cool Guy}) = 0.028$

- Obama is very likely to use -ing
- What happens next if he does?

Obama at the barbecue, literal listener II

$\pi$

<table>
<thead>
<tr>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ing</td>
<td>0.47</td>
<td>0.314</td>
<td>0.201</td>
</tr>
<tr>
<td>-in’</td>
<td>0.048</td>
<td>0.174</td>
<td>0.431</td>
</tr>
</tbody>
</table>

$P_{\text{lit}}(\pi \mid -\text{ing})$ will become the new prior for the third instance of (ING)

Obama at the barbecue, prior III

$\pi$

<table>
<thead>
<tr>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ing</td>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>-in’</td>
<td>0.3</td>
<td>0.7</td>
<td>0.9</td>
</tr>
</tbody>
</table>

$P_{\text{lit}}(\pi \mid -\text{ing})$ will become the new prior for the third instance of (ING)

Obama at the barbecue, literal listener III

$\pi$

<table>
<thead>
<tr>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ing</td>
<td>0.742</td>
<td>0.212</td>
<td>0.045</td>
</tr>
<tr>
<td>-in’</td>
<td>0.253</td>
<td>0.395</td>
<td>0.325</td>
</tr>
</tbody>
</table>

$P_{\text{lit}}(\pi \mid -\text{ing})$ will become the new prior for the third instance of (ING)

- The -in’ variant is still informative!
Obama at the barbecue, speaker III

(Still assuming $\alpha = 6$ and 0 costs)

- ing  -in'

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.998</td>
<td>0.022</td>
<td>0.023</td>
<td>0.977</td>
<td></td>
</tr>
</tbody>
</table>

Model's prediction $P_{S\,\text{-in'}}(\text{Cool Guy}) = 0.977$

- Obama is very likely to use -in'
- What happens next if he does?
- $\text{in'}, \text{ing}, \text{in'}, \text{?}$

Obama at the barbecue, literal listener III

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.742</td>
<td>0.212</td>
<td>0.045</td>
<td>0.0003</td>
<td></td>
</tr>
</tbody>
</table>

$P_{\text{lit}}(\pi | \text{-in'})$ will become the new prior for the fourth instance of (ING)

Obama at the barbecue, prior IV

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.539</td>
<td>0.361</td>
<td>0.099</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>

$S_0(\text{-ing} | \pi)$

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

$S_0(\text{-in'} | \pi)$

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.7</td>
<td>0.9</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>

$\text{Pr}(\cdot)$

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.253</td>
<td>0.395</td>
<td>0.325</td>
<td>0.027</td>
<td></td>
</tr>
</tbody>
</table>

Obama at the barbecue, literal listener IV

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.113</td>
<td>0.412</td>
<td>0.436</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

$\text{Pr}(\cdot)$

<table>
<thead>
<tr>
<th></th>
<th>Stern Leader</th>
<th>Cool Guy</th>
<th>Asshole</th>
<th>Doofus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.253</td>
<td>0.395</td>
<td>0.325</td>
<td>0.027</td>
<td></td>
</tr>
</tbody>
</table>

- The -in' variant is still informative!
- This is so even after -in', -ing, -in'
Obama at the barbecue, speaker IV

(Still assuming $\alpha = 6$ and 0 costs)

- $\text{ing}$ $\approx 1$ $\approx 0$
- $\text{in'}$ $\approx 0$ $\approx 1$
- $\approx 0$ $\approx 1$
- $\approx 0$ $\approx 1$

Model's prediction

$P_S(\text{in'} | \text{Cool Guy}) = 0.688$

$\pi$ stem leader cool guy asshole doofus

$S_0(\text{in'} | \pi)$ 0.7 0.3 0.1 0.01
$S_0(\text{in'} | \pi)$ 0.3 0.7 0.9 0.99

The best strategy to convey the Cool Guy persona is to produce $\text{-in'}$ at the rate of $S_0$, i.e., one needs to perform the persona

Properties of expressive content

Potts (2007): Expressives such as $\text{damn}$ and $\text{bastard}$

(2) That bastard Kresge is famous

$\rightarrow$ Sp feels negatively about Kresge

$\rightarrow$ largely independent of the descriptive meanings

Independence

$\rightarrow$ always about the utterance situation itself

Non-displaceability

$\rightarrow$ not propositional and can be hard to pin down

Descriptive ineffability

$\rightarrow$ performative in that the very act of utterance conveys the meaning

Immediacy

$\rightarrow$ strengthened when repeated without redundancy

Repeatability

Properties of social meaning

(3) I am working on my paper $\rightarrow$ Sp is competent $\rightarrow$ Sp is aloof

Social meaning formally represented as $S_0(\text{ing | } \pi)$

$\rightarrow$ largely independent of the descriptive meanings

Independence

$\rightarrow$ always about the utterance situation itself

Non-displaceability

$\rightarrow$ not propositional and can be hard to pin down

Descriptive ineffability

$\rightarrow$ performative in that the very act of utterance conveys the meaning

Immediacy

$\rightarrow$ strengthened when repeated without redundancy

Repeatability
Conclusion

Linguistic variation carries meaning
▶ Dynamic construction of style
▶ Can be modeled as a reference game

Social meaning as use-conditional meaning
▶ Parallel to expressive content

Use-conditional meaning in RSA models
▶ Assume a hypothetical stereotypical speaker $S_0$
▶ $P_{\text{lit}}(\pi \mid m) \propto \Pr(\pi) \cdot S_0(m \mid \pi)$ Use-conditional
▶ $P_{\text{lit}}(\pi \mid m) \propto \Pr(\pi) \cdot [m](\pi)$ Truth-conditional