Informativeness in language production and comprehension

Judith Degen
02/07/2016
Stanford Linguistics Colloquium
Joint work with

Caroline Graf
Robert Hawkins
Noah Goodman

Mike Tanenhaus
What can I help you with?
Sorry, I missed that.
linguistic signal
world knowledge

linguistic signal
world knowledge

context

linguistic signal
world
knowledge
reasoning
context
linguistic
signal
PRAGMATIC
Outline

I. **Overinformativeness** in production
   — why do we do it?

II. **Underinformativeness** in production
    — how do we deal with it in comprehension?
Outline

I. **Overinformativeness** in production
   — why do we do it?

II. **Underinformativeness** in production
    — how do we deal with it in comprehension?

- models
- corpora
- experiments
Part I

Production of referring expressions

Degen, Graf, Hawkins, & Goodman, in prep.
CONTENT SELECTION

Which features to include in a referring expression?
The Cooperative Principle
Grice 1975

“Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.”

**Quantity-1:** Make your contribution as informative as required.  
**Quantity-2:** Don’t make your contribution more informative than necessary.  
**Manner:** Be brief and orderly; avoid ambiguity and obscurity.
Overinformative referring expressions — color/size asymmetry

**size sufficient**

*the big lightbulb*

75-80%

**color sufficient**

*the green lightbulb*

8-10%

Deutsch 1976; Pechmann 1989; Sedivy 2003; Gatt et al. 2011; many others
Overinformative referring expressions — color/size asymmetry

- Size sufficient:
  - The big lightbulb
  - 75-80%

- Color sufficient:
  - The green lightbulb
  - 8-10%

1. Speakers produce overinformative referring expressions
2. More overinformative color than size mentions

Deutsch 1976; Pechmann 1989; Sedivy 2003; Gatt et al. 2011; many others
Overinformative referring expressions — color/size asymmetry

- size sufficient
- color sufficient

- the big lightbulb
- the green lightbulb

75-80% the big green lightbulb 8-10%

1. speakers produce overinformative referring expressions
2. more overinformative color than size mentions

Deutsch 1976; Pechmann 1989; Sedivy 2003; Gatt et al. 2011; many others
Computational models of REs

- Greedy Algorithm
  Dale 1989

- Incremental Algorithm
  Dale & Reiter 1995

- PRO
  Gatt et al 2013
Computational models of REs

- Greedy Algorithm
  Dale 1989
- Incremental Algorithm
  Dale & Reiter 1995
- PRO
  Gatt et al 2013

Informativeness
Computational models of REs

- Greedy Algorithm
  - Dale 1989
  - Informativeness

- Incremental Algorithm
  - Dale & Reiter 1995
  - Preferences

- PRO
  - Gatt et al 2013
Computational models of REs

- **Greedy Algorithm**
  - Dale 1989
  - Informativeness

- **Incremental Algorithm**
  - Dale & Reiter 1995
  - Preferences

- **PRO**
  - Gatt et al 2013
  - Probabilities
Computational models of REs

- **Greedy Algorithm**
  - Dale 1989

- **Incremental Algorithm**
  - Dale & Reiter 1995

- **PRO**
  - Gatt et al 2013

- **Rational Speech Act (RSA)**
  - Frank & Goodman 2012
Computational models of REs

- **Greedy Algorithm**
  Dale 1989

- **Incremental Algorithm**
  Dale & Reiter 1995

- **PRO**
  Gatt et al 2013

- **Rational Speech Act (RSA)**
  Frank & Goodman 2012
Probabilistic pragmatics

RSA models

Reference
Frank & Goodman, 2012; Qing & Franke, 2015; Degen & Franke, 2012; Stiller et al., 2011; Franke & Degen, 2015

Cost-based Quantity implicatures
Degen et al., 2013; Rohde et al., 2012

Scalar implicatures
Goodman & Stuhlmüller, 2013; Degen et al., 2015

Embedded implicatures
Potts et al., in press; Bergen et al., in press

M-implicatures
Bergen et al., 2012

Figurative meaning
Kao et al., 2013; 2014; 2015; Kao & Goodman, to appear

Gradable adjectives
Lassiter & Goodman, 2013; 2015; Qing & Franke, 2014

In the works:
- collective predication: Scontras & Goodman
- I-implicatures: Poppels & Levy
- overinformativeness: Degen & Goodman
- generics: Tessler & Goodman
- modals: Herbstritt & Franke
- vague quantifiers: Schöller & Franke

Probabilistic pragmatics

RSA
models
The RSA framework

Frank & Goodman 2012

\[ O = \{ \text{on}, \text{green}, \text{off} \} \]
\[ U = \{ \text{big}, \text{small}, \text{green}, \text{black} \} \]
The RSA framework

Frank & Goodman 2012

\[ O = \{ \text{ } , \text{ } , \text{ } , \text{ } \} \]
\[ U = \{ \text{big, small, green, black} \} \]

**Literal listener**

\[ P_{L_0}(o|u) = U(o|\{u \text{ is true of } o\}) \]
\[ [[u]] : O \rightarrow \{ \text{true, false} \} \]
The RSA framework

Frank & Goodman 2012

\[ O = \{ \text{ Emitting bulb, Green light, Black light } \} \]
\[ U = \{ \text{ big, small, green, black } \} \]

Literal listener

\[ P_{L_0}(o|u) = \mathcal{U}(o|\{u \text{ is true of } o\}) \]
\[ [[u]] : O \rightarrow \{ \text{true, false} \} \]
The RSA framework
Frank & Goodman 2012

\[ O = \{ \text{big, small, green, black} \} \]

\[ U = \{ \text{big, small, green, black} \} \]

**Literal listener**

\[ P_{L_0}(o|u) = \mathcal{U}(o|\{u \text{ is true of } o\}) \]

\[ [[[u]] : O \rightarrow \{\text{true, false}\} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]
The RSA framework
Frank & Goodman 2012

\[ O = \{ \text{object} \} \]
\[ U = \{ \text{big}, \text{small}, \text{green}, \text{black} \} \]

**Literal listener**

\[ P_{L_0}(o|u) = U(o|\{u \text{ is true of } o\}) \]
\[ [[u]] : O \rightarrow \{ \text{true, false} \} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]
The RSA framework
Frank & Goodman 2012

\[ O = \{ \text{light}, \text{green}, \text{black} \} \]
\[ U = \{ \text{big, small, green, black} \} \]

**Literal listener**

\[ P_{L_0}(o|u) = \mathcal{U}(o|\{u \text{ is true of } o\}) \]
\[ [[u]] : O \rightarrow \{ \text{true, false} \} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]
The RSA framework

Frank & Goodman 2012

\[ O = \{ \text{on}, \text{big}, \text{green}, \text{black} \} \]

\[ U = \{ \text{big, small, green, black} \} \]

**Literal listener**

\[ P_{L_0}(o|u) = \mathcal{U}(o\{u \text{ is true of } o\}) \]

[[u]] : O \rightarrow \{ \text{true, false} \}

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

---

\( \lambda = 1 \)

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>big</td>
<td>0.6</td>
</tr>
<tr>
<td>black</td>
<td>0.4</td>
</tr>
<tr>
<td>green</td>
<td>0.6</td>
</tr>
<tr>
<td>small</td>
<td>0.2</td>
</tr>
</tbody>
</table>
The RSA framework
Frank & Goodman 2012

\[ O = \{\text{on}, \text{off}, \text{on} \} \]
\[ U = \{\text{big, small, green, black}\} \]

**Literal listener**

\[ P_{L_0}(o|u) = \mathcal{U}(o|\{u \text{ is true of } o\}) \]
\[ [[u]] : O \to \{\text{true, false}\} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

### Manner

- **big**
- **black**
- **green**
- **small**

### Utterance Probability

\[ \lambda = 1 \]

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>big</td>
<td>0.6</td>
</tr>
<tr>
<td>black</td>
<td>0.2</td>
</tr>
<tr>
<td>green</td>
<td></td>
</tr>
<tr>
<td>small</td>
<td></td>
</tr>
</tbody>
</table>

**Diagram**

- **Utterance**
  - big
  - black
  - green
  - small

- **Probability**
  - 0.0
  - 0.2
  - 0.4
  - 0.6
The RSA framework

Frank & Goodman 2012

\[ O = \{ \text{light bulb}, \} \]
\[ U = \{ \text{big, small, green, black} \} \]

**Literal listener**

\[ P_{L_0}(o|u) = \mathcal{U}(o|\{u \text{ is true of } o\}) \]
\[ [[u]] : O \rightarrow \{ \text{true, false} \} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

obvious problem: no complex utterances

\[ \lambda = 1 \]

\[
\begin{array}{cc}
\text{big} & 0.6 \\
\text{black} & 0.4 \\
\text{green} & 0.2 \\
\text{small} & 0.0 \\
\end{array}
\]
The RSA framework

Frank & Goodman 2012

\[ O = \{ \text{big, small, green, black, big green, small green, small black} \} \]

\[ U = \{ \text{big, small, green, black} \} \]

**Literal listener**

\[ P_{L_0}(o|u) = U(o|\{ u \text{ is true of } o \}) \]

\[ [[u]] : O \rightarrow \{ \text{true, false} \} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

obvious problem: no complex utterances
Utterance semantics & cost

**Intersective semantics**

\[[u]\] = \[[u_1]\] \land \[[u_2]\]

\[[\text{big green}]\] = \[[\text{big}]\] \land \[[\text{green}]\]

**Cost**

\[C(u) = C(u_1) + C(u_2)\]
Utterance semantics & cost

**Intersective semantics**

\[
[u] = [u_1] \land [u_2]
\]

\[
[big \ green] = [big] \land [green]
\]

**Cost**

\[
C(u) = C(u_1) + C(u_2)
\]
Utterance semantics & cost

**Intersective semantics**

\[ [[u]] = [[u_1]] \land [[u_2]] \]

\[ [[\text{big green}]] = [[\text{big}]] \land [[\text{green}]] \]

**Cost**

\[ C(u) = C(u_1) + C(u_2) \]

![Graph showing the probability of objects being big, big green, and green, with the probability of big being less than the probability of big green.](image)
Utterance semantics & cost

**Intersective semantics**
\[
[[u]] = [[u_1]] \land [[u_2]]
\]
\[
[[\text{big green}]] = [[\text{big}]] \land [[\text{green}]]
\]

**Cost**
\[
C(u) = C(u_1) + C(u_2)
\]

RSA will not produce overinformative REs...

Gatt et al 2013; Westerbeek et al 2015
Utterance semantics & cost

Intersective semantics

$$[[u]] = [[u_1]] \land [[u_2]]$$

$$[[\text{big green}]] = [[\text{big}]] \land [[\text{green}]]$$

Cost

$$C(u) = C(u_1) + C(u_2)$$

RSA will not produce overinformative REs…

Gatt et al 2013; Westerbeek et al 2015

…with deterministic semantics
Motivation for non-deterministic semantics?

Modifiers differ:

- size adjectives are vague and context-dependent in a way that color adjectives are not
  
  *Kennedy & McNally 2005*

- color is intrinsically salient in a way that size is not
  
  *Arts et al 2011; Gatt et al 2013*

- size adjectives are judged to be more subjective than color adjectives
  
  *Scontras, Degen, & Goodman in press*
Non-deterministic semantics

**Literal listener**

\[ P_{L_0}(o \mid u) \propto \begin{cases} 
1 - \varepsilon & \text{if } [[u]](o) = \text{true} \\
\varepsilon & \text{otherwise}
\end{cases} \]
Non-deterministic semantics

Literal listener

\[ P_{L_0}(o|u) \propto \begin{cases} 
1 - \epsilon & \text{if } [[u]](o) = \text{true} \\
\epsilon & \text{otherwise}
\end{cases} \]
Non-deterministic semantics

Literal listener

\[ P_{L_0}(o\mid u) \propto \begin{cases} 1 - \epsilon & \text{if } [u](o) = \text{true} \\ \epsilon & \text{otherwise} \end{cases} \]

![Probability distribution chart](chart.png)
Non-deterministic semantics

Literal listener

\[ P_{L_0}(o|u) \propto \begin{cases} 
1 - \epsilon & \text{if } [[u]](o) = \text{true} \\
\epsilon & \text{otherwise}
\end{cases} \]

Pragmatic speaker

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]
Non-deterministic semantics

**Literal listener**

\[ P_{L_0}(o|u) \propto \begin{cases} 1 - \epsilon & \text{if } [[u]](o) = \text{true} \\ \epsilon & \text{otherwise} \end{cases} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

Two free fidelity parameters:

\[ \text{fid(size)} \quad \text{fid(color)} \]
Non-deterministic semantics

**Literal listener**

\[ P_{L_0}(o|u) \propto 1 - \epsilon \quad \text{if } [[u]](o) = \text{true} \]
\[ \epsilon \quad \text{otherwise} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

Two free fidelity parameters:
- \( \text{fid}(\text{size}) \)
- \( \text{fid}(\text{color}) \)

Two free cost parameters:
- \( C(\text{size}) \)
- \( C(\text{color}) \)
Non-deterministic semantics

**Literal listener**

\[ P_{L_0}(o|u) \propto \begin{cases} 1 - \epsilon & \text{if } [[u]](o) = \text{true} \\ \epsilon & \text{otherwise} \end{cases} \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

Two free fidelity parameters:

- \( \text{fid}(\text{size}) \)
- \( \text{fid}(\text{color}) \)

Two free cost parameters:

- \( C(\text{size}) \)
- \( C(\text{color}) \)

![Diagram](chart.png)
Non-deterministic semantics

Literal listener

\[ P_{L_0}(o|u) \propto \begin{cases} 1 - \epsilon & [[u]](o) = \text{true} \\ \epsilon & \text{otherwise} \end{cases} \]

Pragmatic speaker

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

Two free fidelity parameters:

\[ \text{fid}(\text{size}) \quad \text{fid}(\text{color}) \]

Two free cost parameters:

\[ C(\text{size}) \quad C(\text{color}) \]

\[ \text{color-size asymmetry!} \]

\[ \begin{array}{ll}
\text{fid}(\text{size}) &= 0.8 \\
\text{fid}(\text{color}) &= 0.999 \\
\lambda &= 15 \\
C(\text{size}) &= 0.1 \\
C(\text{color}) &= 0.1 \\
\end{array} \]
Non-deterministic semantics

**Literal listener**

\[ P_{L_0}(o|u) \propto \left\{ \begin{array}{ll} 1 - \epsilon & \text{if } [[u]](o) = \text{true} \\ \epsilon & \text{otherwise} \end{array} \right. \]

**Pragmatic speaker**

\[ P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))} \]

Two free fidelity parameters:

- \( \text{fid}(\text{size}) \)
- \( \text{fid}(\text{color}) \)

Two free cost parameters:

- \( C(\text{size}) \)
- \( C(\text{color}) \)

If modifiers don’t “work perfectly”:

adding modifiers adds information
Independent empirical evidence for RSA with non-deterministic semantics?
Scene variation

Koolen et al 2013, Davies & Katsos 2013
Scene variation

low variation

high variation

more redundant color use in high-variation scenes

Koolen et al 2013, Davies & Katsos 2013
Scene variation

more redundant color use in high-variation scenes

Koolen et al 2013, Davies & Katsos 2013

non-deterministic RSA predicts this result
Independent quantitative evidence for non-deterministic RSA?
Scene variation

scene variation increases probability of redundancy
Scene variation increases probability of redundancy

Proportion of total distractors that don’t share target value on insufficient dimension

\[
\frac{n_{\text{diff}}}{n_{\text{total}}}
\]
Scene variation increases probability of redundancy.

\[
\frac{n_{\text{diff}}}{n_{\text{total}}}
\]
proportion of total distractors that don't share target value on insufficient dimension.

- **sufficient dimension:** size
- **insufficient dimension:** color

\[
\frac{n_{\text{red}}}{n_{\text{total}}} = \frac{2}{4} = .5
\]
Scene variation increases probability of redundancy

\[
\frac{n_{\text{diff}}}{n_{\text{total}}} \quad \text{proportion of total distractors that don't share target value on insufficient dimension}
\]

sufficient dimension: size
insufficient dimension: color

greater proportion = more variation
Model predictions

Prediction: increase in redundant adjective use with increasing scene variation for color but not size
Interactive reference game experiment

- 58 pairs of participants on Mechanical Turk
- random assignment to speaker/listener role
- 72 trials (half targets, half fillers)
- 36 object types
- on all target trials, one of size or color was sufficient
- **scene variation manipulation:**
  - total number of distractors (2, 3, 4)
  - number of distractors that shared the insufficient feature value with target
Speaker’s perspective

You: the stapler
listener: which one??
You: big purple
Listener’s perspective

speaker: the stapler
You: which one??
speaker: big purple

Round 1 of 72
Results

1. more redundant adjective use with greater scene variation
2. greater effect of scene variation for color than size
Results

1. more redundant adjective use with greater scene variation
2. greater effect of scene variation for color than size

Bayesian Data Analysis
Results

1. more redundant adjective use with greater scene variation
2. greater effect of scene variation for color than size

Bayesian Data Analysis
Posterior predictive

$R^2 = .73$
Posterior over parameters

Fidelity:
inferred size fidelity lower than inferred color fidelity

Cost:
inferred size and color costs similar (with tendency towards costlier size)
Interim summary

if modifiers are noisy, adding modifiers adds utility

RSA with noisy truth functions captures this:

overinformative referring expressions
Interim summary

if modifiers are noisy, adding modifiers adds utility

RSA with noisy truth functions captures this:

everinformative referring expressions
Interim summary

if modifiers are noisy, adding modifiers adds utility

RSA with noisy truth functions captures this:

- overinformative referring expressions
- rational redundant referring expressions
Interim summary

if modifiers are noisy, adding modifiers adds utility

RSA with noisy truth functions captures this:

- overinformative referring expressions
- rational redundant referring expressions

level of reference


Kreiss, Degen, et al in prep

Kreiss, Degen, et al in prep
Part II
——
Comprehension of scalar expressions
Scalar implicature

(1) John: Why is Ann happy? Mary: She found some of her marbles. **Inference**: Ann found some, but not all her marbles.
Scalar implicature

(1) John: Why is Ann happy?  
Mary: She found some of her marbles.  
**Inference**: Ann found some, but not all her marbles.

(2) John: Who came to the party?  
Mary: Ann or Greg.  
**Inference**: Either Ann or Greg came, but not both.
Scalar implicature

(1) John: Why is Ann happy?  
Mary: She found some of her marbles.  
**Inference**: Ann found some, but not all her marbles.

(2) John: Who came to the party?  
Mary: Ann or Greg.  
**Inference**: Either Ann or Greg came, but not both.

(3) John: How was your date?  
Mary: It was OK.  
**Inference**: The date was OK, but not great.
Scalar implicature

(1) John: Why is Ann happy?
Mary: She found some of her marbles.
**Inference**: Ann found *some*, but not *all* her marbles.

**Generalization** Grice 1975; Horn 1972, 2004
By uttering the *weaker* alternative from a scale of a *weaker* and a *stronger* alternative, the speaker implicates the negation of the *stronger* alternative.
Scalar implicature

(1) John: Why is Ann happy?  
Mary: She found some of her marbles.  
**Inference**: Ann found some, but not all her marbles.

**Generalization** Grice 1975; Horn 1972, 2004  
By uttering the weaker alternative from a scale of a weaker and a stronger alternative, the speaker implicates the negation of the stronger alternative.

**Pragmatic interpretation**  
…some, but not all…

**Literal interpretation**  
…some, and possibly all…
Scalar implicature

(1) John: Why is Ann happy?
Mary: She found some of her marbles.
**Inference**: Ann found **some**, but not **all** her marbles.

**Generalization** Grice 1975; Horn 1972, 2004
By uttering the **weaker** alternative from a scale of a **weaker** and a **stronger** alternative, the speaker implicates the negation of the **stronger** alternative.

**Pragmatic interpretation**
...some, but not all...

**Literal interpretation**
...some, and possibly all...
RSA for scalar implicature

\[ M = \{ m \_\exists, m \_\exists \_\forall, m \_\forall \} \]
\[ U = \{ u \_\text{none}, u \_\text{some}, u \_\text{all} \} \]
\[ [[u \_\text{none}]] = \{ m \_\exists \} \]
\[ [[u \_\text{some}]] = \{ m \_\exists \_\forall, m \_\forall \} \]
\[ [[u \_\text{all}]] = \{ m \_\forall \} \]

Ann found some of her marbles
RSA for scalar implicature

**Literal listener**

$$P_{L_0}(m|u) = \mathcal{U}(m|\{u \text{ is true of } m\})$$

$$M = \{m_{\neg \exists}, m_{\exists \neg \forall}, m_{\forall}\}$$

$$U = \{u_{\text{none}}, u_{\text{some}}, u_{\text{all}}\}$$

$$[[u_{\text{none}}]] = \{m_{\neg \exists}\}$$

$$[[u_{\text{some}}]] = \{m_{\exists \neg \forall}, m_{\forall}\}$$

$$[[u_{\text{all}}]] = \{m_{\forall}\}$$

"some"

Ann found some of her marbles
RSA for scalar implicature

Literal listener

\[ P_{L_0}(m|u) = \mathcal{U}(m|\{u \text{ is true of } m\}) \]

\[ M = \{m_{\neg \exists}, m_{\exists \neg A}, m_A\} \]
\[ U = \{u_{\text{none}}, u_{\text{some}}, u_{\text{all}}\} \]
\[ [[u_{\text{none}}]] = \{m_{\neg \exists}\} \]
\[ [[u_{\text{some}}]] = \{m_{\exists \neg A}, m_A\} \]
\[ [[u_{\text{all}}]] = \{m_A\} \]

Ann found some of her marbles
RSA for scalar implicature

**Literal listener**

\[ P_{L0}(m|u) = U(m|\{u \text{ is true of } m\}) \]

**Pragmatic speaker**

\[ P_{S1}(u|m) \propto e^{\lambda \cdot (\ln P_{L0}(m|u))} \]

\[ M = \{ m_{\neg \exists}, m_{\exists \neg \forall}, m_{\forall} \} \]
\[ U = \{ u_{\text{none}}, u_{\text{some}}, u_{\text{all}} \} \]
\[ [[u_{\text{none}}]] = \{ m_{\neg \exists} \} \]
\[ [[u_{\text{some}}]] = \{ m_{\exists \neg \forall}, m_{\forall} \} \]
\[ [[u_{\text{all}}]] = \{ m_{\forall} \} \]

Ann found some of her marbles
RSA for scalar implicature

**Literal listener**

\[ P_{L_0}(m|u) = \mathcal{U}(m|\{u \text{ is true of } m\}) \]

**Pragmatic speaker**

\[ P_{S_1}(u|m) \propto e^{\lambda \cdot (\ln P_{L_0}(m|u))} \]

\[ M = \{m_{\exists}, m_{\exists-\forall}, m_{\forall}\} \]

\[ U = \{u_{\text{none}}, u_{\text{some}}, u_{\text{all}}\} \]

\[ [[u_{\text{none}}]] = \{m_{\exists}\} \]

\[ [[u_{\text{some}}]] = \{m_{\exists-\forall}, m_{\forall}\} \]

\[ [[u_{\text{all}}]] = \{m_{\forall}\} \]

---

*Ann found some of her marbles*
RSA for scalar implicature

**Literal listener**

\[ P_{L_0}(m|u) = \mathcal{U}(m|\{u \text{ is true of } m\}) \]

**Pragmatic speaker**

\[ P_{S_1}(u|m) \propto e^{\lambda \cdot (\ln P_{L_0}(m|u))} \]

**Pragmatic listener**

\[ P_{L_1}(m|u) \propto P_{S_1}(u|m) \cdot P(m) \]

\[ M = \{m_{\neg \exists}, m_{\exists \neg \forall}, m_{\forall}\} \]

\[ U = \{u_{\text{none}}, u_{\text{some}}, u_{\text{all}}\} \]

\[ [[u_{\text{none}}]] = \{m_{\neg \exists}\} \]

\[ [[u_{\text{some}}]] = \{m_{\exists \neg \forall}, m_{\forall}\} \]

\[ [[u_{\text{all}}]] = \{m_{\forall}\} \]

Ann found some of her marbles
RSA for scalar implicature

**Literal listener**

\[ P_{L_0}(m|u) = \mathcal{U}(m|\{u \text{ is true of } m\}) \]

**Pragmatic speaker**

\[ P_{S_1}(u|m) \propto e^{\lambda \cdot (\ln P_{L_0}(m|u))} \]

**Pragmatic listener**

\[ P_{L_1}(m|u) \propto P_{S_1}(u|m) \cdot P(m) \]

- \( M = \{m_{\neg \exists}, m_{\exists \neg \forall}, m_{\forall}\} \)
- \( U = \{u_{\text{none}}, u_{\text{some}}, u_{\text{all}}\} \)
- \([u_{\text{none}}] = \{m_{\neg \exists}\}\)
- \([u_{\text{some}}] = \{m_{\exists \neg \forall}, m_{\forall}\}\)
- \([u_{\text{all}}] = \{m_{\forall}\}\)

"some"

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\neg \exists)</td>
<td>0.00</td>
</tr>
<tr>
<td>(\exists \neg \forall)</td>
<td>0.25</td>
</tr>
<tr>
<td>(\forall)</td>
<td>0.50</td>
</tr>
<tr>
<td>All</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Ann found some of her marbles
RSA for scalar implicature

**Literal listener**

\[ P_{L_0}(m|u) = \mathcal{U}(m|\{u \text{ is true of } m\}) \]

**Pragmatic speaker**

\[ P_{S_1}(u|m) \propto e^{\lambda \cdot (\ln P_{L_0}(m|u))} \]

**Pragmatic listener**

\[ P_{L_1}(m|u) \propto P_{S_1}(u|m) \cdot P(m) \]

Let's consider the sentence: "Ann found some of her marbles."
Context in scalar implicature


• **utterance alternatives** that the speaker could have used, but didn’t Katzir 2007; Fox & Katzir 2011

• conversational goal or Question Under Discussion (QUD) Roberts 1996; 2012
Alternatives and the QUD

**Question Under Discussion** (explicit or implicit)
(1) What does she look like?
(2) What are some features of Sally?

Harry: She has a good personality.

**Alternatives**

Jess: So which one is she?

Harry: Attractive.

Jess’s inference: Sally is **attractive**, but not **beautiful**.
Number alternatives in processing “some”  
Degen & Tanenhaus 2015  

Naturalness ratings  
- Do listeners have **expectations** about use of *some*? $P(\text{some} \mid M)$  
- Do expectations depend on the contextual availability of **number alternatives**? $P(\text{some} \mid M, C)$
Number alternatives in processing “some”

Degen & Tanenhaus 2015

Naturalness ratings

- Do listeners have *expectations* about use of *some*? $P(\text{some} \mid M)$
- Do expectations depend on the contextual availability of *number alternatives*? $P(\text{some} \mid M, C)$
The gumball paradigm

You got some of the gumballs

360 participants on MTurk

Independent variables:

- set size in lower chamber: 0 - 13
- quantifier: some, all, none, (one, two, …)
- presence of number terms
Expectations of use for *some*

*some* is a dispreferred alternative for small sets ($p < .0001$)
Expectations of use for *some*

*some* is a dispreferred alternative for small sets (*p* < .0001) especially when numbers are available alternatives (*p* < .01).
You got some of the orange gumballs
Processing alternatives online

Degen & Tanenhaus 2016

absent: some/all

present: some/all/two/three/four/five
Processing alternatives online

absent: some/all

present: some/all/two/three/four/five

You got some of the orange gumballs.
You got some of the orange gumballs.

implicatures are slower to process when number alternatives are contextually available.

absent: *some/all*  
present: *some/all/two/three/four/five*
Context in RSA: alternatives

Absent: \( U = \{\text{none, some, all}\} \)
Present: \( U = \{\text{none, some, all, one, two, three, four}\} \)
Context in RSA: alternatives

Absent: $U = \{\text{none, some, all}\}$
Present: $U = \{\text{none, some, all, one, two, three, four}\}$

Pragmatic speaker
state = 2

![Diagram showing probability distribution for utterances with and without context.](image)
Context in RSA: alternatives

Absent: $U = \{\text{none, some, all}\}$
Present: $U = \{\text{none, some, all, one, two, three, four}\}$

Pragmatic speaker
state = 4
Context in RSA: alternatives

Absent: $U = \{\text{none, some, all}\}$
Present: $U = \{\text{none, some, all, one, two, three, four}\}$

both “some” and “all” less expected when numbers are present
QUD effects on scalar implicature

Degen & Goodman 2014

Does the QUD modulate scalar implicature strength?

Implicit QUD
QUD effects on scalar implicature

Degen & Goodman 2014

Does the QUD modulate scalar implicature strength?

Implicit QUD

all? Did the speaker find all of the marbles?
QUD effects on scalar implicature

Degen & Goodman 2014

Does the QUD modulate scalar implicature strength?

Implicit QUD

all? Did the speaker find all of the marbles?
QUD effects on scalar implicature

Degen & Goodman 2014

Does the QUD modulate scalar implicature strength?

Implicit QUD

all? Did the speaker find all of the marbles?
I found all / some of the marbles.
QUD effects on scalar implicature

Does the QUD modulate scalar implicature strength?

Implicit QUD

all? Did the speaker find all of the marbles?
I found all / some of the marbles.
QUD effects on scalar implicature
Degen & Goodman 2014

Does the QUD modulate scalar implicature strength?

Implicit QUD

all? Did the speaker find all of the marbles?
   I found all / some of the marbles.

any? Did the speaker find any of the marbles?
QUD effects on scalar implicature
Degen & Goodman 2014

Does the QUD modulate scalar implicature strength?

Implicit QUD

all? Did the speaker find all of the marbles?
I found all / some of the marbles.

any? Did the speaker find any of the marbles?
QUD effects on scalar implicature
Degen & Goodman 2014

Does the QUD modulate scalar implicature strength?

Implicit QUD

**all?** Did the speaker find all of the marbles?
I found all / some of the marbles.

**any?** Did the speaker find any of the marbles?
I found all / some of the marbles.
QUD effects on scalar implicature

Does the QUD modulate scalar implicature strength?

Implicit QUD  —> manipulated via cover stories

**all?** Did the speaker find all of the marbles?
I found all / some of the marbles.

**any?** Did the speaker find any of the marbles?
I found all / some of the marbles.
Task and results

48 participants on Mechanical Turk
Task and results

48 participants on Mechanical Turk
Task and results

48 participants on Mechanical Turk
Task and results

48 participants on Mechanical Turk

Ann found this box:

She called out to her husband: 'I found some of the marbles!'

Is Ann's statement true?

literal pragmatic

Yes No

Proportion of 'No' responses

<table>
<thead>
<tr>
<th>QUD</th>
<th>Proportion of 'No' responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>all?</td>
<td>0.62 ± 0.04</td>
</tr>
<tr>
<td>any?</td>
<td>0.27 ± 0.04</td>
</tr>
</tbody>
</table>

* indicates a statistically significant difference.
Task and results

48 participants on Mechanical Turk

The QUD modulates scalar inference strength
Task and results

Ann found this box:

She called out to her husband: 'I found some of the marbles!'

Is Ann's statement true?

- literal
- pragmatic

- Yes
- No

Proportion of 'No' responses

The QUD modulates scalar inference strength

48 participants on Mechanical Turk

see also Degen 2013
Context in RSA: QUD

Degen & Goodman 2014

\[ Q = \{q_{\text{all}{}?}, q_{\text{any}{}?}\} \]

\[ P_{L_1}(m|u, q) \propto P_{S_1}(u|m, q) \cdot P(m) \]

(assuming uniform prior on QUDs and independence of QUD and actual state of the world)
Context in RSA: QUD

Degen & Goodman 2014

\[ Q = \{ q_{\text{all?}}, q_{\text{any?}} \} \]

\[ P_{L_1}(m|u, q) \propto P_{S_1}(u|m, q) \cdot P(m) \]

(assuming uniform prior on QUDs and independence of QUD and actual state of the world)

**Pragmatic speaker**
Context in RSA: QUD

\[ Q = \{ q_{\text{all?}}, q_{\text{any?}} \} \]

\[ P_{L_1}(m|u, q) \propto P_{S_1}(u|m, q) \cdot P(m) \]

(assuming uniform prior on QUDs and independence of QUD and actual state of the world)

**Pragmatic speaker**

**Pragmatic listener**
Context in RSA: QUD

\[ Q = \{q_{\text{all}?>}, q_{\text{any}?}\} \]
\[ P_{L_1}(m|u, q) \propto P_{S_1}(u|m, q) \cdot P(m) \]

(assuming uniform prior on QUDs and independence of QUD and actual state of the world)

**Pragmatic speaker**

**Pragmatic listener**

RSA captures QUD effects in scalar implicature
What is the **natural distribution** of scalar implicatures?
Scalar implicatures in the wild

1. I like some country music.

2. It would certainly help them to appreciate some of the things we have here.

3. You sound like you have some small ones in the background.
Combining corpora & the web

Degen 2015

1. extracted all 1390 utterances containing *some* from the Switchboard corpus of spoken American English

2. collected implicature strength ratings for each item on MTurk
Speaker A: i mean, they just have beautiful, beautiful homes and they have everything. the kids only wear name brand things to school and it's one of these things,

Speaker B: oh me. well that makes it hard for you, doesn't it.

Speaker A: well it does, you know. it really does because i'm a single mom and i have a thirteen year old now and uh, you know, it does.

Speaker B: oh, me.

Speaker A: i mean, we do it to a point but uh, not to where she feels different ,

Speaker B: yeah.

Speaker A: but some of them are very rich

but some, but not all of them are very rich

How similar is the statement with 'some, but not all' (green) to the statement with 'some' (red)?

Very different meaning 1 2 3 4 5 6 7 Same meaning

Continue
Default prediction
Variation in inference strength

large amount of variation in inference strength
Just noise?

No. Systematically stronger inferences with…
Just noise?

No. Systematically stronger inferences with…

Partitive

...partitive some-NPs.

Some of them are very rich
Just noise?

No. Systematically stronger inferences with...

- **Partitive** some-NPs.
  - Some *of them* are very rich

- **Previously mentioned** embedded NP referents.
  - Some *of them* are very rich
Just noise?

No. Systematically stronger inferences with...

- **Partitive** some-NPs. Some *of them* are very rich
  - new
  - inferable
  - mentioned

- **Previous mention** embedded NP referents. Some *of them* are very rich
  - new
  - inferable
  - mentioned

- **Grammatical function** some-NPs in **subject** position. Some *of them* are very rich
  - other
  - subject
Model fit

by-participant intercepts only

$r = .16$
Model fit

after adding fixed effects of context

\[ r = .66 \]
after adding by-item random effects

Model fit

Predicted rating vs. Empirical rating

$r = .99$
Summary

I. **Overinformativeness** in production
   redundant referring expressions are rational when modifiers are noisy

II. **Underinformativeness** in production
   listeners make efficient use of context in drawing scalar inferences

models  corpora  experiments
Summary

I. **Overinformativeness** in production
   redundant referring expressions are rational when
   modifiers are noisy

II. **Underinformativeness** in production
   listeners make efficient use of context in drawing scalar
   inferences

**NO CONFLICT**

models  corpora  experiments
world
knowledge
reasoning
context

linguistic
signal
...as informative as required
...as informative as required
...no more informative than necessary

linguistic signal

world
knowledge
reasoning
context
Communicatively efficient system

...as informative as required
...no more informative than necessary

world
knowledge
reasoning
context

linguistic
signal
Future directions I

Extension of probabilistic models to **online processing**

Feb 17/18  CSLI workshop: bridging computational and psycholinguistic approaches to meaning
Future directions I

Extension of probabilistic models to online processing

Some of the people at the dinner drank beer.

Feb 17/18 CSLI workshop: bridging computational and psycholinguistic approaches to meaning
Future directions I

Extension of probabilistic models to **online processing**

Some of the people at the dinner drank beer.

Feb 17/18  **CSLI workshop: bridging computational and psycholinguistic approaches to meaning**
Future directions I

Extension of probabilistic models to **online processing**

Some of the people at the dinner drank beer.

Feb 17/18  **CSLI workshop: bridging computational and psycholinguistic approaches to meaning**
Future directions I

Extension of probabilistic models to online processing

Some of the people at the dinner drank beer.

Feb 17/18  CSLI workshop: bridging computational and psycholinguistic approaches to meaning
Future directions

Extension of probabilistic models to **online processing**

Some of the people at the dinner drank beer.

Feb 17/18  **CSLI workshop: bridging computational and psycholinguistic approaches to meaning**
Future directions I

Extension of probabilistic models to **online processing**

Some of the people at the dinner drank beer.

Feb 17/18  **CSLI workshop: bridging computational and psycholinguistic approaches to meaning**
Future directions II

**Development** of inferencing across the lifespan — acquisition — adaptation — aging
Yildirim, Degen, Tanenhaus, & Jaeger 2016

**Resource limitations** and **individual differences**
working memory — cognitive control
Franke & Degen 2016

**Cross-linguistic** pragmatics
Degen & Franke in prep

**Social** meaning
Burnett & Degen in prep
Thank you

**Collaborators**
- Michael Franke
- Noah Goodman
- Caroline Graf
- Robert Hawkins
- Mike Tanenhaus

**Research assistants**
- Chelsea Marsh
- Laurel Raymond
- Dana Subik
- Andrew Watts

**Funding**
- Swiss National Science Foundation Early Postdoc.Mobility grant to JD
- [XPrag.de](http://www.xprag.de) DFG priority program grant to JD
- James S. McDonnell grant to NG
- ONR grant N00014-13-1-0788 to NG
- NIH grant HD27206 to MT