Towards a unifying computational account of reference production and comprehension

Judith Degen
Feb 24, 2021
LingLangLunch, Brown University
Reference comprehension and production

Comprehension:
How do listeners predict the referent of (possibly ambiguous) referring expressions in real time?

Production:
How do speakers decide which features of an object to include in a referring expression?

**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.

Grice 1975
“Click on the yellow…”

Comprehension: contrastive inferences

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
<td>competitor</td>
</tr>
<tr>
<td>distractor</td>
<td>distractor</td>
</tr>
</tbody>
</table>

**no contrast**

Listener expect speakers not to be over-informative.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
<td>competitor</td>
</tr>
<tr>
<td>distractor</td>
<td>contrast</td>
</tr>
</tbody>
</table>

**contrast**

**Quantity-2**

Sedivy et al 1999; Sedivy 2003; Grodner & Sedivy 2011; Heller et al 2008; Ryskin et al 2019; Rubio-Fernández & Jara-Ettinger 2018; Aparicio et al 2018; Alsop et al 2018
Production: redundant referring expressions

size sufficient  
the small pin  
75-80%

the small blue pin  
8-10%

color sufficient

Quantity-2
the blue pin

Speakers produce seemingly overinformative referring expressions.

Reference comprehension and production

Comprehension:
Listeners draw inferences based on the expectation that speakers not be over-informative.

Production:
Speakers produce seemingly overinformative referring expressions.

HOW TO RESOLVE?
world
knowledge
reasoning
context

PRAGMATICS

linguistic
signal
Probabilistic pragmatics

Franke & Jäger, 2016; Goodman & Frank, 2016; Scontras, Tessler, & Franke 2018

Reference
Frank & Goodman, 2012; Qing & Franke, 2015; Degen & Franke, 2012; Stiller et al., 2015; Franke & Degen, 2015; Degen et al, 2020

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., 2016; Bergen et al., 2016

M-implicatures
Bergen et al., 2012

Figurative meaning
Kao et al., 2013; 2014; 2015; Cohn-Gordon & Bergen, under review

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

Adjective ordering
Hahn et al 2018; Scontras et al 2019

Other
plural predication Scontras & Goodman 2017
I-implicatures Poppels & Levy, 2016
generics Tessler & Goodman, 2019
modals Herbstritt & Franke, 2017
vague quantifiers Schöller & Franke, 2017
convention formation Hawkins et al 2018; 2019
questions Hawkins et al 2015
pragmatic adaptation Schuster & Degen, 2020
exhaustivity inferences Javangula & Degen in prep
atypicality inferences Kratvchenko & Demberg
social meaning Burnett 2017; 2019

Probabilistic pragmatics

Franke & Jäger, 2016; Goodman & Frank, 2016; Scontras, Tessler, & Franke 2018

Reference
Frank & Goodman, 2012; Qing & Franke, 2015; Degen & Franke, 2012; Stiller et al., 2015; Franke & Degen, 2015; Degen et al, 2020

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., 2016; Bergen et al., 2016

M-implicatures
Bergen et al., 2012

Figurative meaning
Kao et al., 2013; 2014; 2015; Cohn-Gordon & Bergen, under review

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

Adjective ordering
Hahn et al 2018; Scontras et al 2019

Other
plural predication Scontras & Goodman 2017
I-implicatures Poppels & Levy, 2016
generics Tessler & Goodman, 2019
modals Herbstritt & Franke, 2017
vague quantifiers Schöller & Franke, 2017
convention formation Hawkins et al 2018; 2019
questions Hawkins et al 2015
pragmatic adaptation Schuster & Degen, 2020
exhaustivity inferences Javangula & Degen in prep
atypicality inferences Kratvchenko & Demberg
social meaning Burnett 2017; 2019
The Rational Speech Act framework (RSA)

Comprehension: contrastive inferences

"Click on the yellow…"

$P_{L_1}(r \mid u) \propto P_{S_1}(u \mid r) \cdot P(r)$

Production: redundant referring expressions

the small blue pin
la tachuela pequeña

$P_{S_1}(u \mid r) = e^{\alpha(\ln P_{L_0}(r \mid u) - C(u))}$

$P_{L_0}(r \mid u) \propto [[u]](r) \cdot P(r)$
PART I

Comprehension of referring expressions: contrastive inferences

Kreiss & Degen 2020, in prep
Contrastive Inference (CI)

...is elicited by the contrastive function of adjectives.

“Click on the (adj) ... !”

Contrastive Inference (CI)
... is elicited by relative, but not (?) absolute adjectives.

“Click on the (adj) ... !”
Contrastive Inference (CI) … is elicited by all relative adjectives and absolute adjectives with a low production probability.

Speech production

Proportion of mentioning color

Proportion of mentioning color

Sedivy 2003; Westerbeek et al. 2015; Rubio-Fernandez 2016; Mitchell et al. 2013
Rational Speech Act Model

\[ P_{L1}(r | u) \propto P_{S1}(u | r) \times P_{S1}(r) \]

“Click on the yellow…”

![Diagram with images and labels](image)
Rational Speech Act Model

\[ P_{L_1}(r|u) \propto P_{S_1}(u|r) \times P_{S_1}(r) \]

“Click on the yellow … !”

**Contrastive Inference!**

<table>
<thead>
<tr>
<th></th>
<th>Target</th>
<th>Competitor</th>
<th>Distractor</th>
<th>No Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Probability</strong></td>
<td>0%</td>
<td>50%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Contrastive inference!
Prediction
Prediction

TARGET COMPETITOR CONTRAST DISTRACTOR

TARGET COMPETITOR

DISTRACTOR CONTRAST

=≠
1. Production experiment to obtain model predictions
2. Model predictions using production data
3. Comprehension experiment to assess model predictions

2 x 2 x 2 design: contrast presence, target typicality, competitor typicality
Production experiment

- Interactive reference game
- MTurk, n=141 participant pairs (112)
- 60 trials (32 critical; 28 fillers)

Production results

contrast absent

contrast present

Object to be communicated

\[ P_{L_1}(r \mid u) \propto P_{S_1}(u \mid r) \]

\[ P(r) \quad \text{uniform} \]
Comprehension experiment

Click on the yellow banana!

- incremental decision task rather than eye-tracking study
- competitor typicality manipulated between-subjects
- MTurk, n=79 (73)
  40 per competitor typicality condition
- 55 trials (20 critical; 35 fillers)

Qing et al. 2018, Degen et al under review
Comprehension results

contrast absent

contrast present

Proportion of target selections

human  RSA
Comparison of predicted and empirical target selections

Simple pragmatic RSA listener captures variability in contrastive inference strength by reasoning about (variability in) modifier production probabilities.

$r = .92$ at condition level
$r = .60$ at item level
Interim summary I

A simple pragmatic RSA listener captures variability in contrastive inference strength by reasoning about (variability in) modifier production probabilities.

Some consequences:

- effects of adjective function or semantics only operative via their effect on listeners’ production expectations
- no need for “default descriptions”
PART II

Production of referring expressions: redundant modification

Degen et al 2020; Waldon & Degen 2021; Kursat & Degen to appear
1. Redundant color and size modification
   Degen et al 2020

2. Model extension: color typicality
   Degen et al 2020

3. Model extension: cross-linguistic variability
   Waldon & Degen 2021
Production: redundant referring expressions

**size sufficient**
- the small pin
- 75-80%

**color sufficient**
- the small blue pin
- 8-10%

Speakers produce seemingly overinformative referring expressions.

The RSA framework
Frank & Goodman 2012

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

**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

\( \lambda = 1 \)

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Manner</th>
</tr>
</thead>
<tbody>
<tr>
<td>big</td>
<td>black</td>
</tr>
<tr>
<td>green</td>
<td>small</td>
</tr>
</tbody>
</table>

Probability

- big: 0.60
- black: 0.40
- green: 0.20
- small: 0.10
Utterance semantics & cost

**Intersective semantics**

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

\[
[b\text{ig green}] = [b\text{ig}] \land [g\text{reen}]
\]

**Cost**

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

RSA does not produce overinformative REs...

Gatt et al 2013; Westerbeek et al 2015

...with deterministic Boolean semantics
Motivation for relaxed semantics?

Modifiers differ:

- Size adjectives are more vague and context-dependent than color adjectives.
- Color is more salient than size.

Scontras, Degen, & Goodman 2017

Arts et al 2011; Gatt et al 2013

- Size adjectives are judged to be more subjective than color adjectives.

Scontras, Degen, & Goodman 2017
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**

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

If modifiers don’t “work perfectly”, adding modifiers adds information.

- Color-size asymmetry
  - \( \text{fid}(\text{size}) = 0.8 \)
  - \( \text{fid}(\text{color}) = 0.999 \)
  - \( \lambda = 15 \)
  - \( C(\text{size}) = 0.1 \)
  - \( C(\text{color}) = 0.1 \)
Interactive reference game results

![Graph showing utterance probability against scene variation for color and size redundant items. The graph compares empirical data with model predictions, with error bars indicating variability.](image)
Model fit

$R^2 = .73$
Interim summary II

A continuous-semantics RSA speaker captures variability in overmodification by reasoning about a continuous semantics literal listener: overmodification is more likely, the less inferable the intended referent is from the simply modified description.

A consequence:

“overinformative” —> “usefully redundant”
Extending the model to within-color variability
Extending the continuous semantics

**Literal listener**

\[ P_{L_0}(o|u) \propto [[u]](o) \]
\[ [[u]](o) = \text{typicality}(u, o) \]

**Pragmatic speaker**

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

How typical is \( o \) for \( u \)?

- “banana”
- “yellow banana”
- “brown banana”
- “brown”
- …
RSA predictions with continuous semantics

typicality("banana", \( \text{blue banana} \)) = .98

typicality("blue banana", \( \text{blue banana} \)) = .98

typicality("banana", \( \text{red apple} \)) = .01

typicality("banana", \( \text{yellow banana} \)) = .98
Redundancy more likely when probability of confusion is high
Independent empirical evidence for RSA with continuous semantics?

**Literal listener**

\[ P_{L_0}(o|u) \propto [[u]](o) \]

\[ [[u]](o) = \text{typicality}(u, o) \]

**Pragmatic speaker**

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

1. Typicality norming
2. Production study
3. Model evaluation
Production study: interactive reference game experiment
Production study

- 60 pairs of participants on Mechanical Turk
- random assignment to speaker/listener role
- 42 trials
- varied contextual informativeness of utterances:

<table>
<thead>
<tr>
<th>Presence of Same Type</th>
<th>Presence of Color Competitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a) informative</td>
<td>(1b) informative-cc</td>
</tr>
<tr>
<td>(1c) overinformative</td>
<td>(1d) overinformative-cc</td>
</tr>
</tbody>
</table>

presence of same type x presence of color competitor
Model evaluation

A continuous-semantics RSA speaker captures variability in color modification by reasoning about a continuous semantics literal listener: the less inferable the intended referent is from the unmodified noun, the more likely modification is.

<table>
<thead>
<tr>
<th>Semantics</th>
<th>empirical</th>
<th>fixed plus empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>empirical</td>
<td>-1474.6 (4)</td>
<td>-1354.4 (7)</td>
</tr>
<tr>
<td>fixed</td>
<td>-1434.8 (4)</td>
<td>-1321.9 (7)</td>
</tr>
<tr>
<td>none</td>
<td>-1372.9 (2)</td>
<td><strong>-1209.8 (5)</strong></td>
</tr>
</tbody>
</table>

$r(250) = .94$
Final extension: cross-linguistic variability

Color overmodification less likely in Spanish (with post-nominal adjectives) than in English. Rubio-Fernández 2016

Strong argument for the role of incrementality.

<table>
<thead>
<tr>
<th>Utterances</th>
<th>Size-sufficient (SS) scene</th>
<th>Color-sufficient (CS) scene</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td>( O_{\text{big_blue}} ), ( O_{\text{big_red}} ), ( O_{\text{small_blue}} )</td>
<td>( O_{\text{small_red}} ), ( O_{\text{big_red}} ), ( O_{\text{small_blue}} )</td>
</tr>
</tbody>
</table>
Incremental RSA
Cohn-Gordon, Goodman, & Potts 2018

\[ L_{0}^{\text{INCR}}(r|c,i) \propto \mathcal{X}^{D}(c,i,r) \cdot P(r) \]

\[ \mathcal{X}^{D}(c,i,r) = \frac{|u|[u]^{D}(r)=1 \wedge u \text{ is a continuation of } c+i}{|u|u \text{ is a continuation of } c+i} \]

\[ S_{1}^{\text{INCR}}(i|c,r) \propto e^{\alpha(L_{0}^{\text{INCR}}(r|c,i)-C(i))} \]

\[ S_{1}(u|r) = \prod_{j=1}^{n} S_{1}^{\text{INCR}}(i_{j}|c = [i_{1}...i_{j-1}], r) \]

Continuous-Incremental RSA
Waldon & Degen 2021

\[ \mathcal{X}^{C}(c,i,r) = \frac{\sum[[u]]^{C}(r):u \text{ is a continuation of } c+i}{|u|u \text{ is a continuation of } c+i} \]

proportion of applicable continuations

sum of semantic values over number of continuations
Model predictions
Continuous-Incremental RSA

English size-sufficient

Analogous contexts (symmetric in Incremental RSA)

Spanish color-sufficient
Continuous-Incremental RSA

Combining incremental and continuous RSA

• provides some support for Rubio-Fernández’s claim that modification is generally less useful post-nominally

• makes interesting novel prediction for flipped color/size overmodification asymmetry in post-nominal adjective languages

Much more empirical work needed!
Reference comprehension and production

**Comprehension:**
Listeners draw inferences based on the expectation that speakers not be over-informative.

**Production:**
Speakers produce seemingly overinformative referring expressions.

**Quantity-2**  **Quantity-2**

**HOW TO RESOLVE?**
Reference comprehension and production

Comprehension:
Listeners draw inferences based on the expectation that speakers be sufficiently contextually informative.

Production:
Speakers produce usefully redundant referring expressions.

RSA: useful modeling framework to formalize the link between production and comprehension via probabilistic inference.
The literal listener rule can be written as follows:

```javascript
var objects = [{color: 'blue', shape: 'square', string: 'blue square'},
               {color: 'blue', shape: 'circle', string: 'blue circle'},
               {color: 'green', shape: 'square', string: 'green square'}]

var utterances = ['blue', 'green', 'square', 'circle']

var objectPrior = function() {
    var obj = uniformDraw(objects)
    return obj.string
}

var meaning = function(utterance, obj) {
    _.includes(obj, utterance)
}

var literalListener = function(utterance) {
    Infer({model: function() {
        var obj = objectPrior();
        var uttTruthVal = meaning(utterance, obj);
        condition(uttTruthVal == true)
        return obj
    }})
}

viz.table(literalListener('blue'))
```

### Exercises:

1. In the model above, objectPrior() returns a sample from a uniformDraw over the possible objects of reference. What happens when the listener’s beliefs are not uniform over the
Thank you!