Pragmatic reasoning in large-scale NLP systems

Christopher Potts

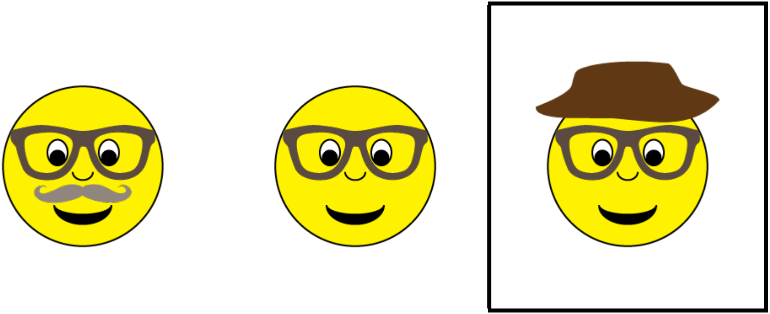
Stanford Linguistics

CU Boulder, February 21, 2020

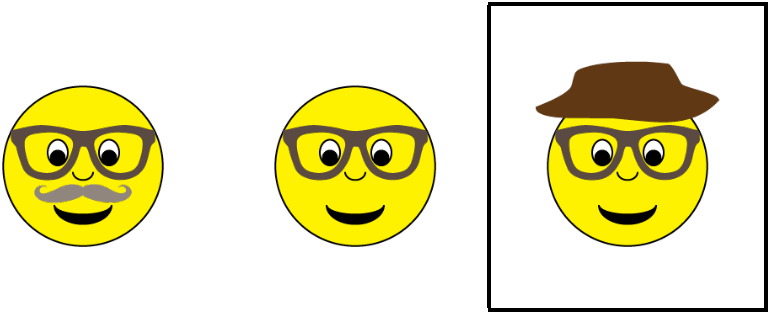
Informativity in context

Generating referring expressions

Generating referring expressions



Generating referring expressions



"The guy with a hat"

Interpreting complex descriptions

Stanford Colors in Context corpus
(Monroe et al. 2017)

Interpreting complex descriptions

Context	Utterance
	blue

Stanford Colors in Context corpus
(Monroe et al. 2017)

Interpreting complex descriptions

Context	Utterance
	blue
	The darker blue one

Stanford Colors in Context corpus
(Monroe et al. 2017)

Interpreting complex descriptions

Context	Utterance
	blue
	The darker blue one
	dull pink not the super bright one

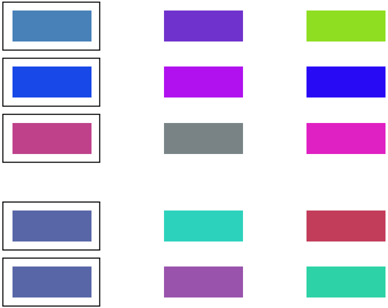
Stanford Colors in Context corpus
(Monroe et al. 2017)

Interpreting complex descriptions

Context	Utterance
	blue
	The darker blue one
	dull pink not the super bright one
	Purple

Stanford Colors in Context corpus
(Monroe et al. 2017)

Interpreting complex descriptions



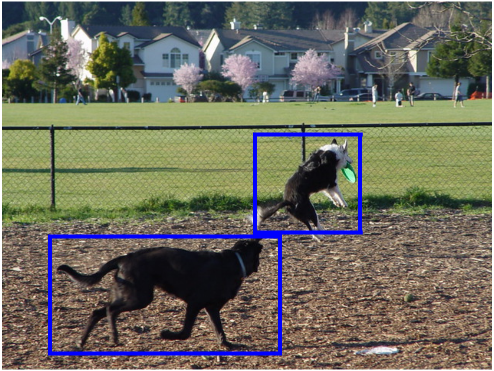
Context	Utterance
	blue
	The darker blue one
	dull pink not the super bright one
	Purple
	blue

Stanford Colors in Context corpus
(Monroe et al. 2017)

Discriminative image labeling

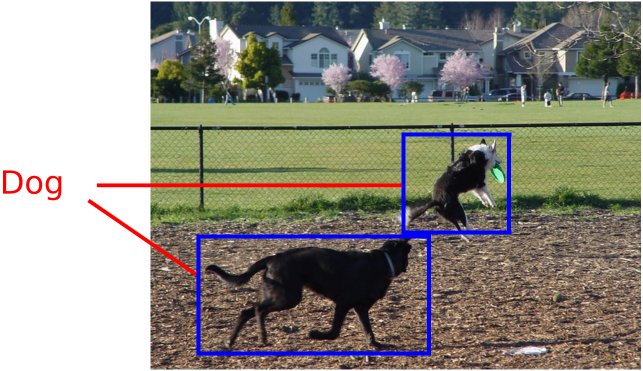
Mao et al. 2016

Discriminative image labeling



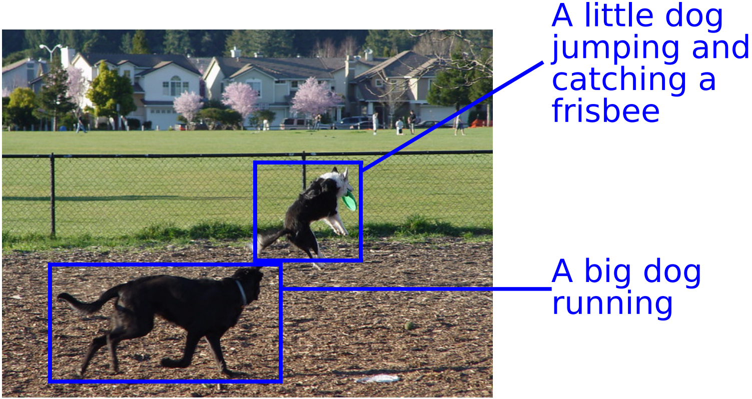
Mao et al. 2016

Discriminative image labeling



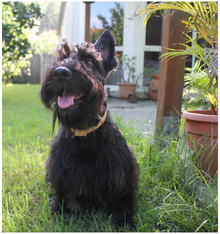
Mao et al. 2016

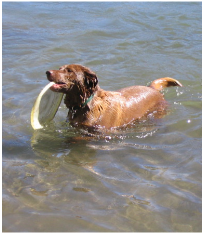
Discriminative image labeling



Mao et al. 2016

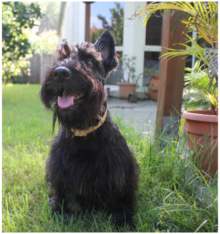
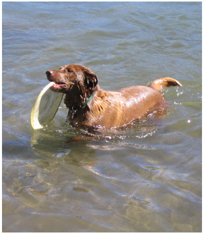
Discriminative image captioning





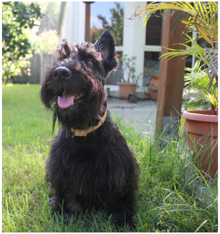
Vedantam et al. 2017; Cohn-Gordon et al. 2018

Discriminative image captioning

Vedantam et al. 2017; Cohn-Gordon et al. 2018

Discriminative image captioning



Vedantam et al. 2017; Cohn-Gordon et al. 2018

Summarization

Tennis champion Serena Williams wobbled into the Third Round of the Australian Open on Thursday.

→

Serena Williams advances to Australian Open Third Round.

Ongoing work with Hanson Lu and Reuben Cohn-Gordon

Summarization

Tennis champion Serena Williams wobbled into the Third Round of the Australian Open on Thursday.

→

Serena Williams advances to Australian Open Third Round.

Sports Champion advances in tournament.

Ongoing work with Hanson Lu and Reuben Cohn-Gordon

Summarization

Tennis champion Serena Williams wobbled into the Third Round of the Australian Open on Thursday.

→

Serena Williams advances to Australian Open Third Round.

Sports Champion advances in tournament.

Williams wobbled on Thursday.

Ongoing work with Hanson Lu and Reuben Cohn-Gordon

Summarization

Tennis champion Serena Williams wobbled into the Third Round of the Australian Open on Thursday.

→

Serena Williams advances to Australian Open Third Round.

Olympic Gold Medalist Venus Williams advanced to the US Open Semi-Finals on Friday.

Sports Champion advances in tournament.

Williams wobbled on Thursday.

Ongoing work with Hanson Lu and Reuben Cohn-Gordon

Summarization

Tennis champion Serena Williams wobbled into the Third Round of the Australian Open on Thursday.

→

Serena Williams advances to Australian Open Third Round.

Olympic Gold Medalist Venus Williams advanced to the US Open Semi-Finals on Friday.

Golfer Lydia Ko eliminated from British Open after finishing 12 over par.

Sports Champion advances in tournament.

Williams wobbled on Thursday.

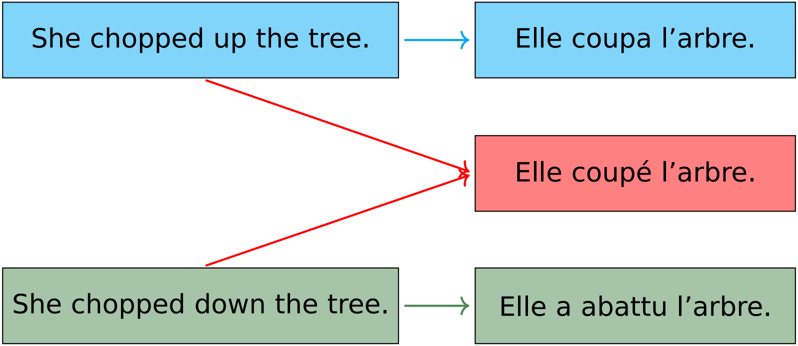
Ongoing work with Hanson Lu and Reuben Cohn-Gordon

Machine translation

She chopped up the tree.	→	Elle coupa l’arbre.
She chopped down the tree.	→	Elle a abattu l’arbre.

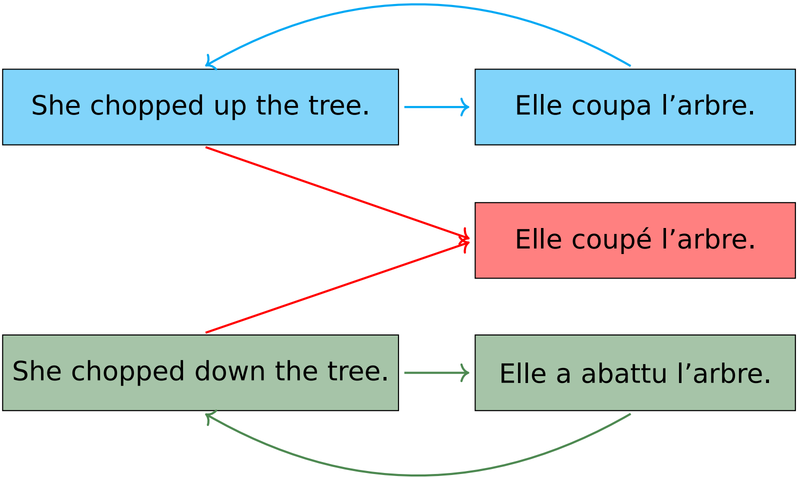
Cohn-Gordon and Goodman 2019

Machine translation



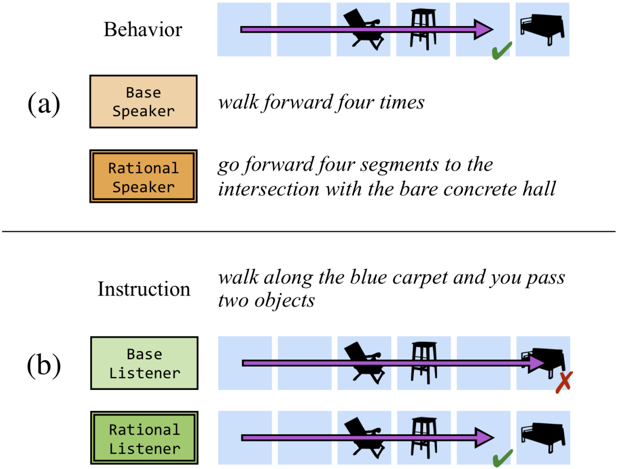
Cohn-Gordon and Goodman 2019

Machine translation



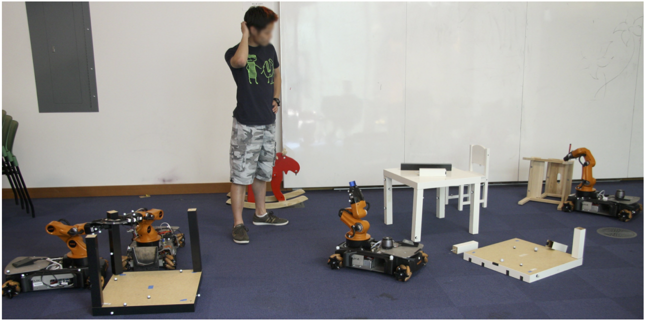
Cohn-Gordon and Goodman 2019

Generating and following instructions



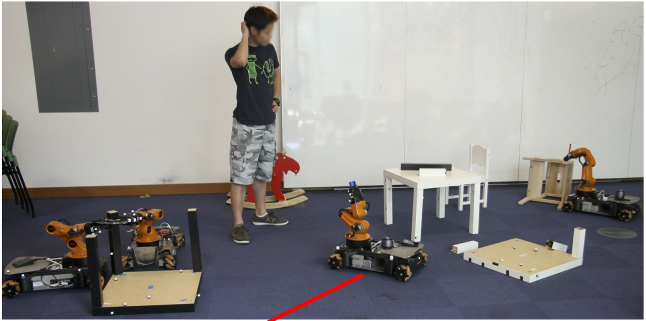
Fried et al. 2018

Collaborative problem solving



Tellex et al. 2014

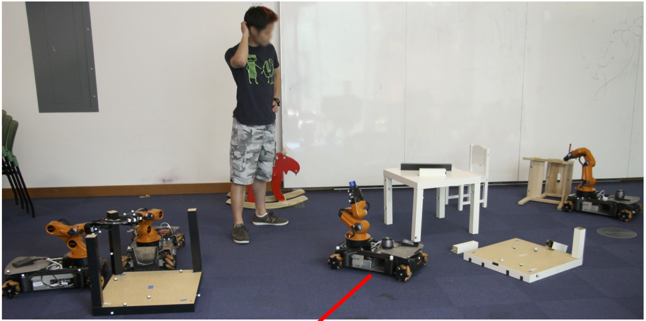
Collaborative problem solving



Help me!

Tellex et al. 2014

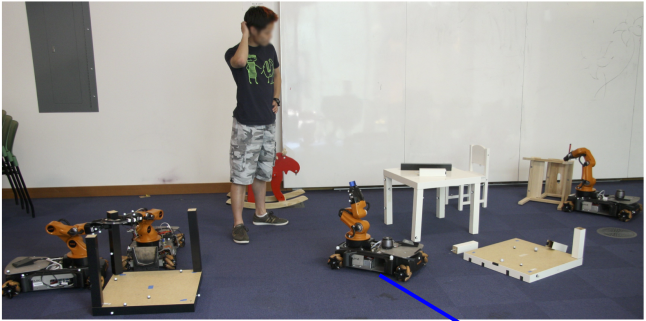
Collaborative problem solving



Hand me
the leg

Tellex et al. 2014

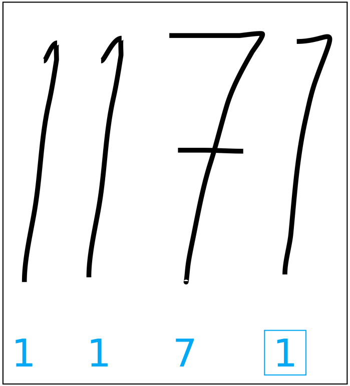
Collaborative problem solving

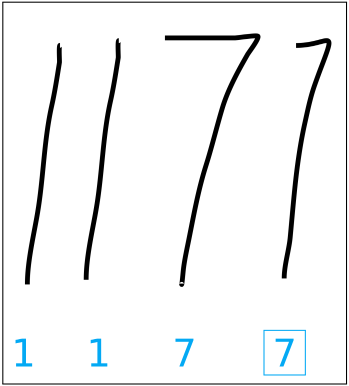


Hand me the white leg on the table

Tellex et al. 2014

Optical character recognition

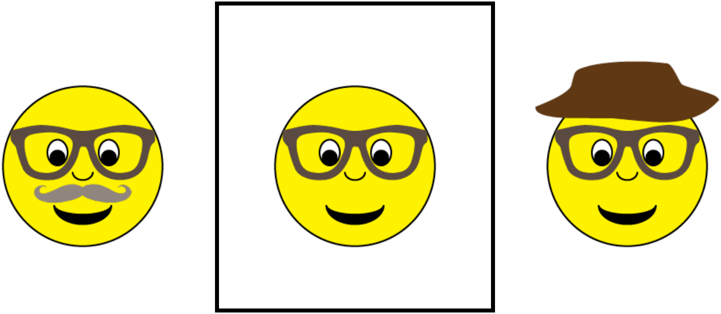




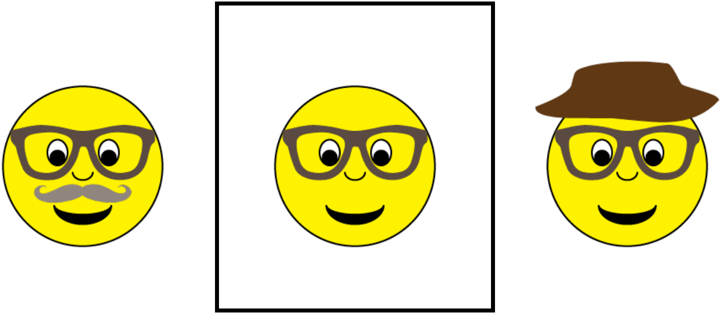
Conversational implicature

A particularized Quantity implicature

A particularized Quantity implicature

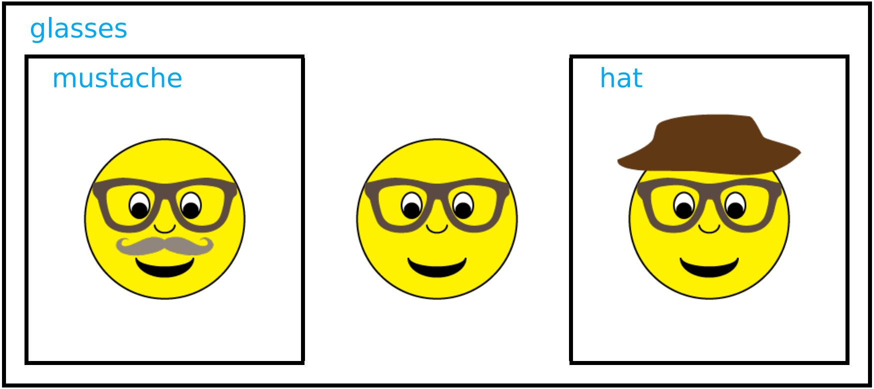


A particularized Quantity implicature



"The guy with glasses"

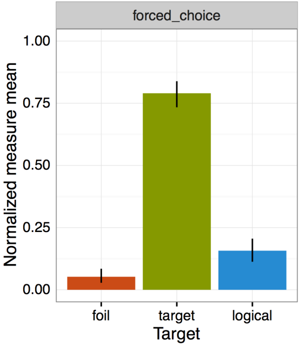
A particularized Quantity implicature



"The guy with glasses"

Reference games

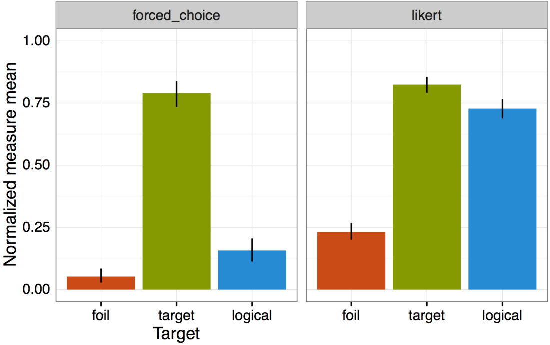
10 experiments, each $N \approx 600$ (4,651 participants). The summary picture:



Frank et al. 2016; `https://github.com/langcog/pragmods`

Reference games

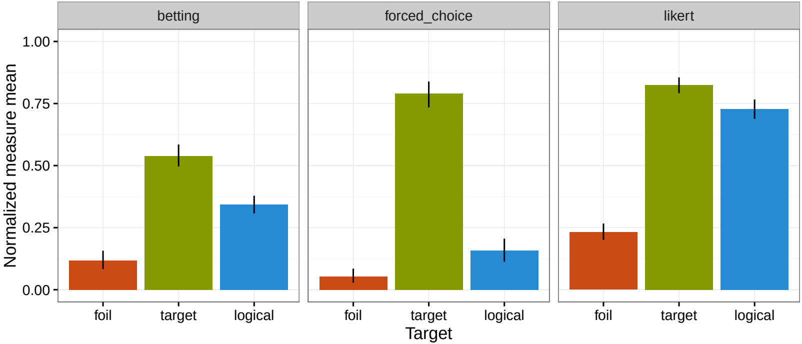
10 experiments, each $N \approx 600$ (4,651 participants). The summary picture:



Frank et al. 2016; `https://github.com/langcog/pragmods`

Reference games

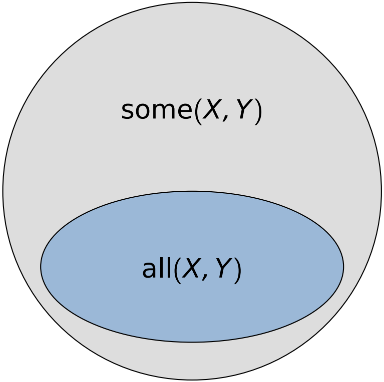
10 experiments, each $N \approx 600$ (4,651 participants). The summary picture:



Frank et al. 2016; `https://github.com/langcog/pragmods`

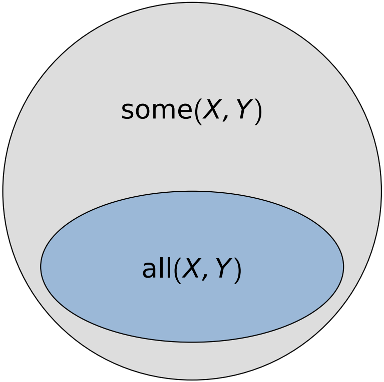
Scalar implicature

John Stuart Mill: *I saw some of your children to-day* invites the inference that I didn't see all of them



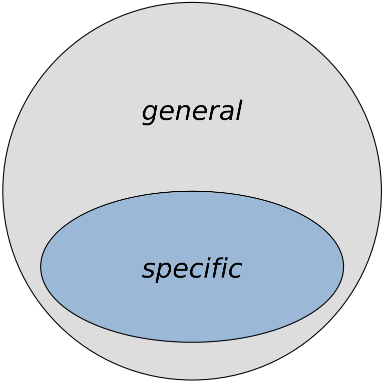
Scalar implicature

John Stuart Mill: *I saw some of your children to-day* invites the inference that I didn't see all of them "not because the words mean it, but because, if I had seen them all, it is most likely that I should have said so."



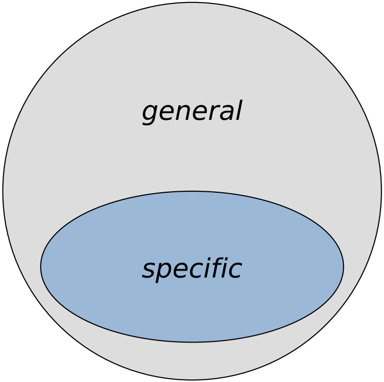
Scalar implicature

Generalization: Using a general term invites the inference that its more specific, salient alternatives are inappropriate.



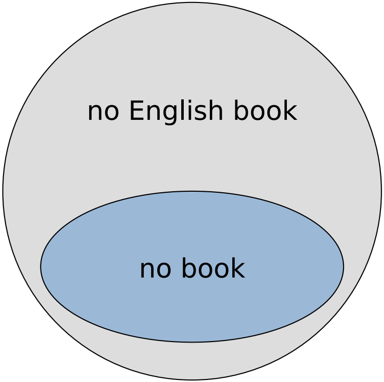
Scalar implicature

Generalization: Using a general term invites the inference that its more specific, salient alternatives are inappropriate.



Scalar implicature

Strawson and Grice: *there are no English books here* invites the inference that there are books here.

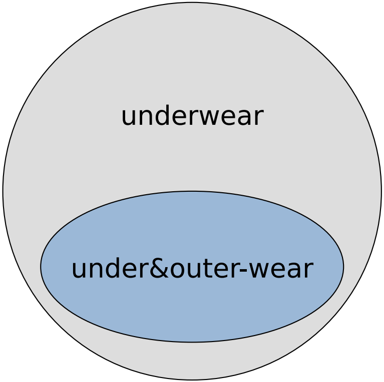


Scalar implicature

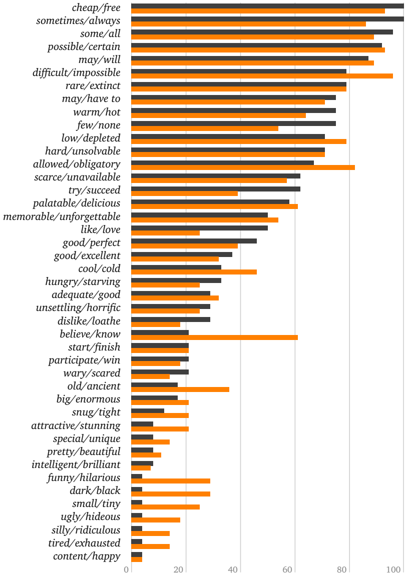
Chris Potts: Watching TV in your underwear – that's a scalar implicature!

Scalar implicature

Chris Potts: Watching TV in your underwear – that's a scalar implicature!

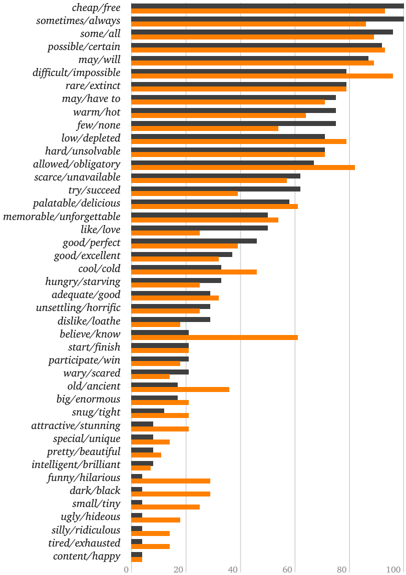


Scalar diversity



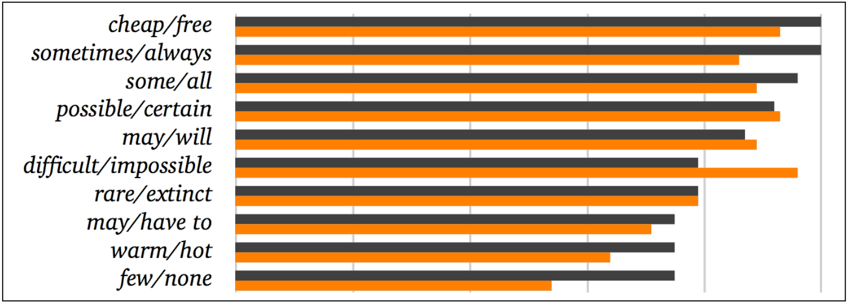
van Tiel et al. 2016;
see also Degen 2013

Scalar diversity



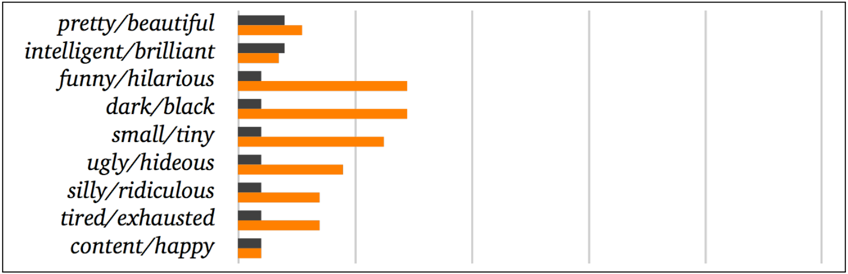
van Tiel et al. 2016;
see also Degen 2013

Scalar diversity



van Tiel et al. 2016;
see also Degen 2013

Scalar diversity



van Tiel et al. 2016;
see also Degen 2013

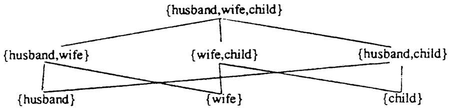
Partial-order implicature

Hirschberg 1985

Partial-order implicature

A: Do you speak German?

B: My husband does.

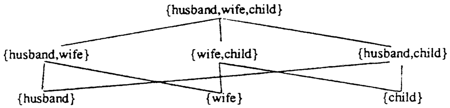


Hirschberg 1985

Partial-order implicature

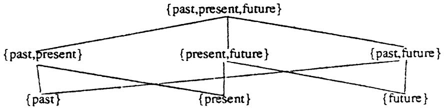
A: Do you speak German?

B: My husband does.



A: Are you on your honeymoon?

B: Well, I was.



Hirschberg 1985

I-implicature

Levinson (2000): “what is simply described is stereotypically exemplified”

I-implicature

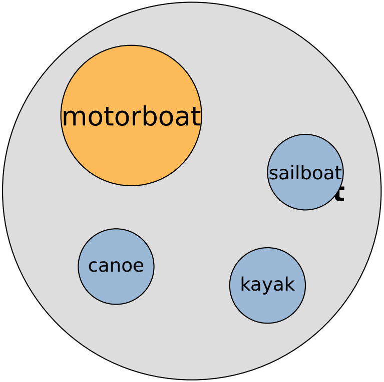
Levinson (2000): “what is simply described is stereotypically exemplified”

1. At a busy marina in water-skiing country:
   “boat” interpreted as *motorboat*

I-implicature

Levinson (2000): "what is simply described is stereotypically exemplified"

1. At a busy marina in water-skiing country:
   "boat" interpreted as *motorboat*



I-implicature

Levinson (2000): “what is simply described is stereotypically exemplified”

1. At a busy marina in water-skiing country:
   “boat” interpreted as *motorboat*
2. Kim is in France. (in Paris)

I-implicature

Levinson (2000): “what is simply described is stereotypically exemplified”

1. At a busy marina in water-skiing country:
   “boat” interpreted as *motorboat*
2. Kim is in France. (in Paris)
3. I hit the button and it started. (causation)

I-implicature

Levinson (2000): “what is simply described is stereotypically exemplified”

1. At a busy marina in water-skiing country:
   “boat” interpreted as *motorboat*
2. Kim is in France. (in Paris)
3. I hit the button and it started. (causation)
4. Sandy finished the book. (reading)

The Rational Speech Acts model (RSA)

Origin story

Origin story

- Rosenberg and Cohen (1964): early Bayesian model of production and comprehension

Origin story

- Rosenberg and Cohen (1964): early Bayesian model of production and comprehension
- Lewis (1969): signaling systems

Origin story

- Rosenberg and Cohen (1964): early Bayesian model of production and comprehension
- Lewis (1969): signaling systems
- Rabin (1990): recursive strategic signaling

Origin story

- Rosenberg and Cohen (1964): early Bayesian model of production and comprehension
- Lewis (1969): signaling systems
- Rabin (1990): recursive strategic signaling
- Camerer et al. (2004): cognitive hierarchy models for games of conflict and coordination

Origin story

- Rosenberg and Cohen (1964): early Bayesian model of production and comprehension
- Lewis (1969): signaling systems
- Rabin (1990): recursive strategic signaling
- Camerer et al. (2004): cognitive hierarchy models for games of conflict and coordination
- Franke (2009) and Jäger (2007): iterated best response

Origin story

- Rosenberg and Cohen (1964): early Bayesian model of production and comprehension
- Lewis (1969): signaling systems
- Rabin (1990): recursive strategic signaling
- Camerer et al. (2004): cognitive hierarchy models for games of conflict and coordination
- Franke (2009) and Jäger (2007): iterated best response
- Golland et al. (2010): pragmatic listeners and probabilistic compositionality

Origin story

- Rosenberg and Cohen (1964): early Bayesian model of production and comprehension
- Lewis (1969): signaling systems
- Rabin (1990): recursive strategic signaling
- Camerer et al. (2004): cognitive hierarchy models for games of conflict and coordination
- Franke (2009) and Jäger (2007): iterated best response
- Golland et al. (2010): pragmatic listeners and probabilistic compositionality
- Frank and Goodman (2012): *very* sophisticated pragmatic agents and a new Bayesian foundation

Pragmatic listeners

Pragmatic listeners

Literal listener

$$L_{\text{lit}}(state \mid msg) = \frac{[\![msg, state]\!] P(state)}{\sum_{state'} [\![msg, state']\!] P(state')}$$

Pragmatic listeners

Pragmatic speaker

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg) - C(msg)\right)\right)}{\sum_{msg'} \exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg') - C(msg')\right)\right)}$$

Literal listener

$$L_{\text{lit}}(state \mid msg) = \frac{[\![msg, state]\!] P(state)}{\sum_{state'} [\![msg, state']\!] P(state')}$$

Pragmatic listeners

Pragmatic listener

$$L_{\text{prag}}(state \mid msg) = \frac{S_{\text{prag}}(msg \mid state) P(state)}{\sum_{state'} S_{\text{prag}}(msg \mid state') P(state')}$$

Pragmatic speaker

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg) - C(msg)\right)\right)}{\sum_{msg'} \exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg') - C(msg')\right)\right)}$$

Literal listener

$$L_{\text{lit}}(state \mid msg) = \frac{[\![msg, state]\!] P(state)}{\sum_{state'} [\![msg, state']\!] P(state')}$$

Pragmatic listeners

Pragmatic listener

$L_{\text{prag}}(\textit{state} \mid \textit{msg}) =$ **pragmatic speaker** × state prior

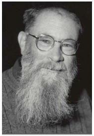
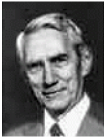
Pragmatic speaker

$S_{\text{prag}}(\textit{msg} \mid \textit{state}) =$ **literal listener** − message costs

Literal listener

$L_{\text{lit}}(\textit{state} \mid \textit{msg}) =$ **lexicon** × state prior

A simple example

			
beard	1	0	0
glasses	1	1	0
tie	0	1	1

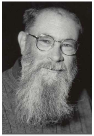
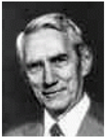
L_{prag}

S_{prag}

L_{lit}

$[\![\cdot]\!]$

A simple example

			
beard	**1**	0	0
glasses	.5	.5	0
tie	0	.5	.5

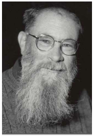
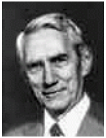
L_{prag}

S_{prag}

L_{lit}

$[\![\cdot]\!]$

A simple example

	beard	*glasses*	*tie*
	**.67**	.33	0
	0	**1**	0
	0	0	**1**

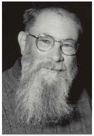
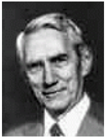
L_{prag}

S_{prag}

L_{lit}

$[\![\cdot]\!]$

A simple example

			
beard	**1**	0	0
glasses	.25	**.75**	0
tie	0	0	**1**

L_{prag}

S_{prag}

L_{lit}

$[\![\cdot]\!]$

Pragmatic speakers

Pragmatic speakers

Literal speaker

$$S_{\text{lit}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log[\![msg, state]\!] - C(msg)\right)\right)}{\sum_{msg'} \exp\left(\alpha\left(\log[\![msg', state]\!] - C(msg')\right)\right)}$$

Pragmatic speakers

Pragmatic listener

$$L_{\text{prag}}(state \mid msg) = \frac{S_{\text{lit}}(msg \mid state) P(state)}{\sum_{state'} S_{\text{lit}}(msg \mid state') P(state')}$$

Literal speaker

$$S_{\text{lit}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log[\![msg, state]\!] - C(msg)\right)\right)}{\sum_{msg'} \exp\left(\alpha\left(\log[\![msg', state]\!] - C(msg')\right)\right)}$$

Pragmatic speakers

Pragmatic speaker

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{prag}}(state \mid msg) - C(msg)\right)\right)}{\sum_{msg'} \exp\left(\alpha\left(\log L_{\text{prag}}(state \mid msg') - C(msg')\right)\right)}$$

Pragmatic listener

$$L_{\text{prag}}(state \mid msg) = \frac{S_{\text{lit}}(msg \mid state) P(state)}{\sum_{state'} S_{\text{lit}}(msg \mid state') P(state')}$$

Literal speaker

$$S_{\text{lit}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log[\![msg, state]\!] - C(msg)\right)\right)}{\sum_{msg'} \exp\left(\alpha\left(\log[\![msg', state]\!] - C(msg')\right)\right)}$$

Pragmatic speakers

Pragmatic speaker

$$S_{\text{prag}}(msg \mid state) = \textbf{pragmatic listener} - \text{message costs}$$

Pragmatic listener

$$L_{\text{prag}}(state \mid msg) = \textbf{literal speaker} \times \text{state prior}$$

Literal speaker

$$S_{\text{lit}}(msg \mid state) = \textbf{lexicon} - \text{message costs}$$

Major achievements

Major achievements

- M-implicatures (Bergen et al. 2016)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)
- Implicatures and compositionality (Potts et al. 2016)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)
- Implicatures and compositionality (Potts et al. 2016)
- Hyperbole (Kao et al. 2014b)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)
- Implicatures and compositionality (Potts et al. 2016)
- Hyperbole (Kao et al. 2014b)
- Metaphor (Kao et al. 2014a)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)
- Implicatures and compositionality (Potts et al. 2016)
- Hyperbole (Kao et al. 2014b)
- Metaphor (Kao et al. 2014a)
- Irony (Cohn-Gordon and Bergen 2019)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)
- Implicatures and compositionality (Potts et al. 2016)
- Hyperbole (Kao et al. 2014b)
- Metaphor (Kao et al. 2014a)
- Irony (Cohn-Gordon and Bergen 2019)
- Politeness (Yoon et al. 2016, 2017)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)
- Implicatures and compositionality (Potts et al. 2016)
- Hyperbole (Kao et al. 2014b)
- Metaphor (Kao et al. 2014a)
- Irony (Cohn-Gordon and Bergen 2019)
- Politeness (Yoon et al. 2016, 2017)
- Social meaning (Burnett 2017; Qing and Cohn-Gordon 2019)

Major achievements

- M-implicatures (Bergen et al. 2016)
- Implicature blocking (Potts and Levy 2015)
- Implicatures and compositionality (Potts et al. 2016)
- Hyperbole (Kao et al. 2014b)
- Metaphor (Kao et al. 2014a)
- Irony (Cohn-Gordon and Bergen 2019)
- Politeness (Yoon et al. 2016, 2017)
- Social meaning (Burnett 2017; Qing and Cohn-Gordon 2019)
- Adaptation (Schuster and Degen 2019)

Joint inference

$$L_{\text{prag}}(state, Context \mid msg)$$

$$S_{\text{prag}}(msg \mid state, Context)$$

RSA/ML hybrids

Motivation

Motivation

- Discriminative image labeling

Motivation

- Discriminative image labeling
- Image captioning

Motivation

- Discriminative image labeling
- Image captioning
- Summarization

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition

- Scalability

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition
- Scalability
- Sensitivity to variation

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition
- Scalability
- Sensitivity to variation
- Bounded rationality

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition

- Scalability
- Sensitivity to variation
- Bounded rationality
- New kinds of model assessment

Motivation

- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition

- Scalability
- Sensitivity to variation
- Bounded rationality
- New kinds of model assessment
- Impact

A formative challenge

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg) - C(msg)\right)\right)}{\boxed{\sum_{msg'}} \exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg') - C(msg')\right)\right)}$$

Notational conventions

Notational conventions

Logical semantic agents

$$S_{\text{prag}}(msg \mid state; Context, [\![\cdot]\!])$$

Notational conventions

Logical semantic agents

$$S_{\text{prag}}(msg \mid state; Context, [\![\cdot]\!])$$

$$S^{[\![\cdot]\!]}_{\text{prag}}(msg \mid state)$$

Notational conventions

Logical semantic agents

$$S_{\text{prag}}(msg \mid state; Context, [\![\cdot]\!])$$

$$S^{[\![\cdot]\!]}_{\text{prag}}(msg \mid state)$$

Learned agents

$$\mathbf{S}_{\text{prag}}(msg \mid state; Context, \theta)$$

Notational conventions

Logical semantic agents

$$S_{\text{prag}}(msg \mid state; Context, [\![\cdot]\!])$$

$$S^{[\![\cdot]\!]}_{\text{prag}}(msg \mid state)$$

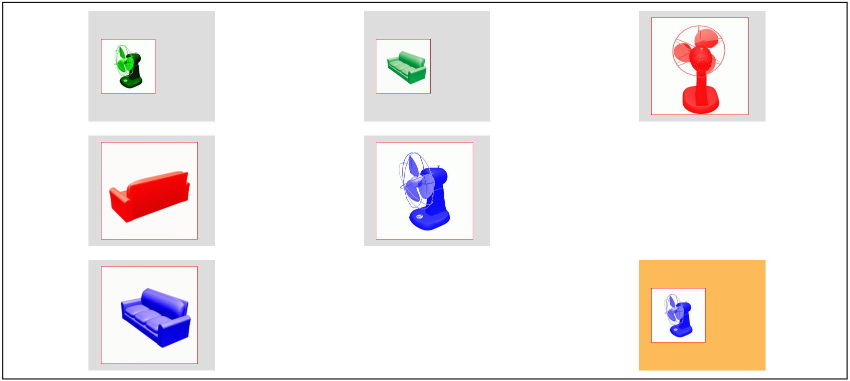
Learned agents

$$\mathbf{S}_{\text{prag}}(msg \mid state; Context, \theta)$$

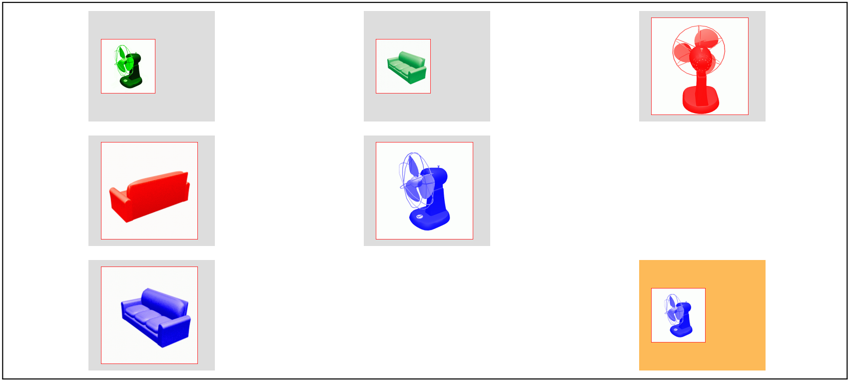
$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state)$$

RSA learning objectives

TUNA furniture example

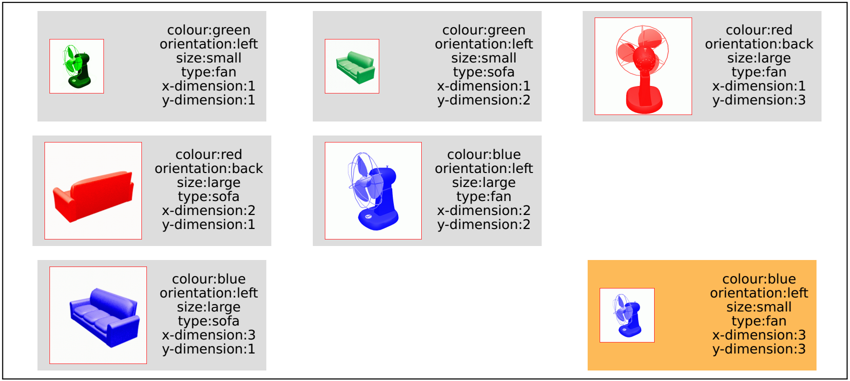


TUNA furniture example



Utterance: “blue fan small”

TUNA furniture example



Utterance: “blue fan small”
Utterance attributes: [*colour:blue*]; [*size:small*]; [*type:fan*]

TUNA people example

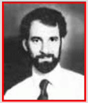
TUNA people example

Utterance: “The bald man with a beard”

TUNA people example

age:old
hairColour:light
hasBeard:1
hasGlasses:0
hasHair:0
hasShirt:1
hasSuit:0
hasTie:0
type:person

age:young
hairColour:dark
hasBeard:0
hasGlasses:0
hasHair:1
hasShirt:1
hasSuit:0
hasTie:0
type:person



age:young
hairColour:dark
hasBeard:1
hasGlasses:0
hasHair:1
hasShirt:1
hasSuit:0
hasTie:1
type:person

age:young
hairColour:dark
hasBeard:1
hasGlasses:0
hasHair:1
hasShirt:0
hasSuit:1
hasTie:1
type:person

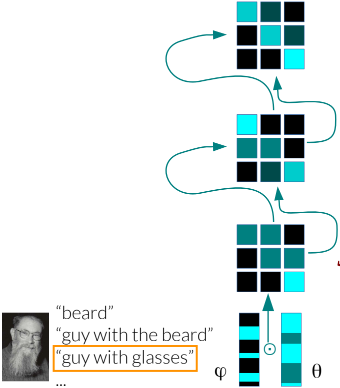
age:young
hairColour:dark
hasBeard:0
hasGlasses:0
hasHair:1
hasShirt:0
hasSuit:1
hasTie:1
type:person

age:young
hairColour:dark
hasBeard:1
hasGlasses:0
hasHair:1
hasShirt:1
hasSuit:0
hasTie:0
type:person

age:young
hairColour:dark
hasBeard:0
hasGlasses:0
hasHair:1
hasShirt:0
hasSuit:1
hasTie:1
type:person

Utterance: “The bald man with a beard”
Utterance attributes: [*hasBeard:1*]; [*hasHair:0*]; [*type:person*]

Forward inference



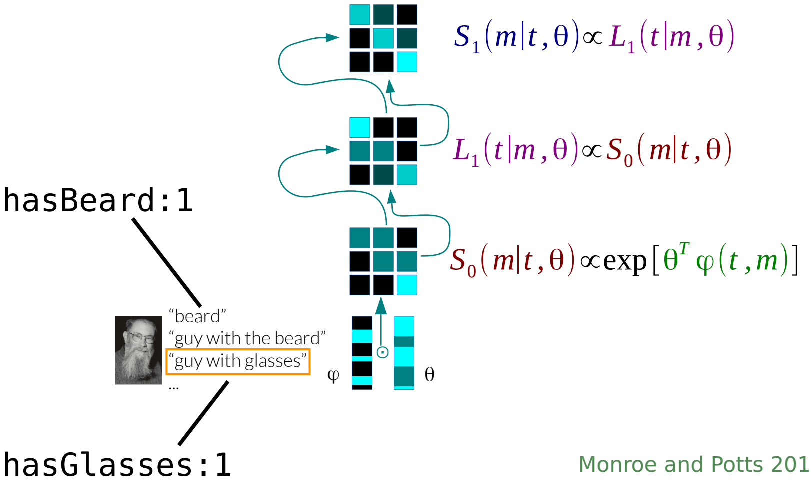
$$S_1(m|t,\theta) \propto L_1(t|m,\theta)$$

$$L_1(t|m,\theta) \propto S_0(m|t,\theta)$$

$$S_0(m|t,\theta) \propto \exp\left[\theta^T \varphi(t,m)\right]$$

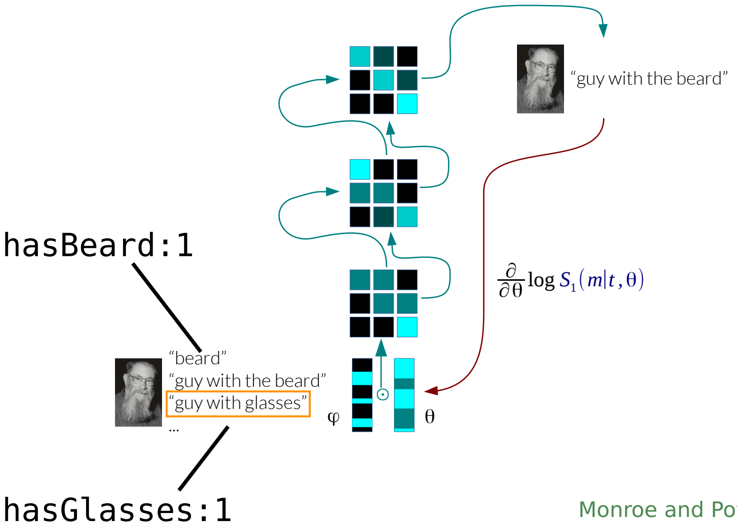
Monroe and Potts 2015

Forward inference



Monroe and Potts 2015

Optimization

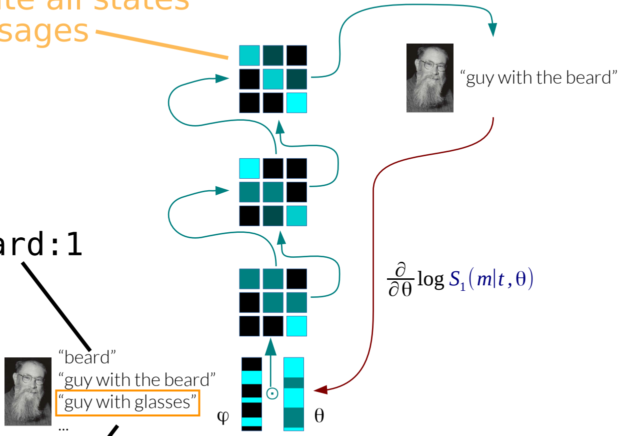


Monroe and Potts 2015

Optimization

Enumerate all states
and messages

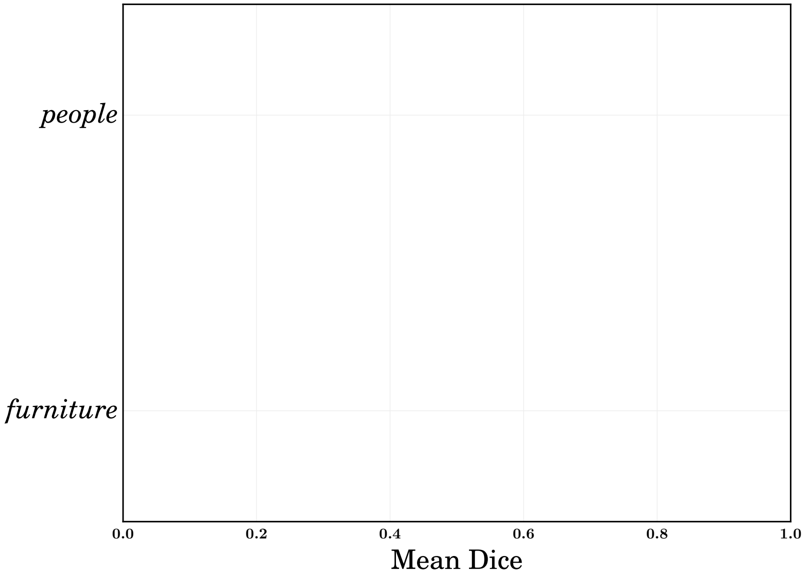
hasBeard:1



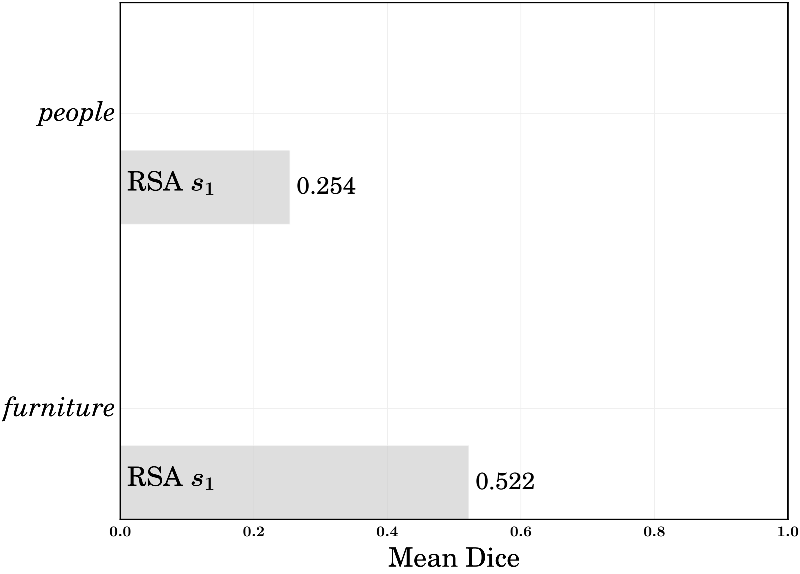
hasGlasses:1

Monroe and Potts 2015

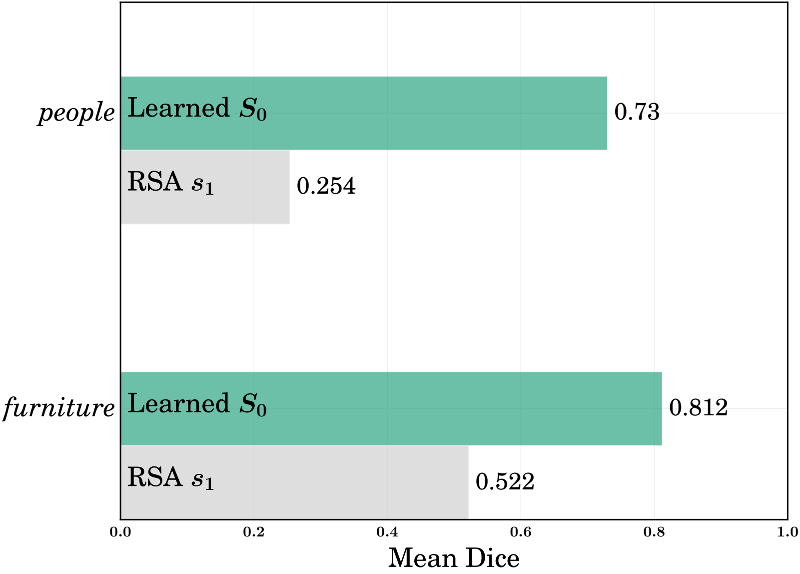
TUNA Results



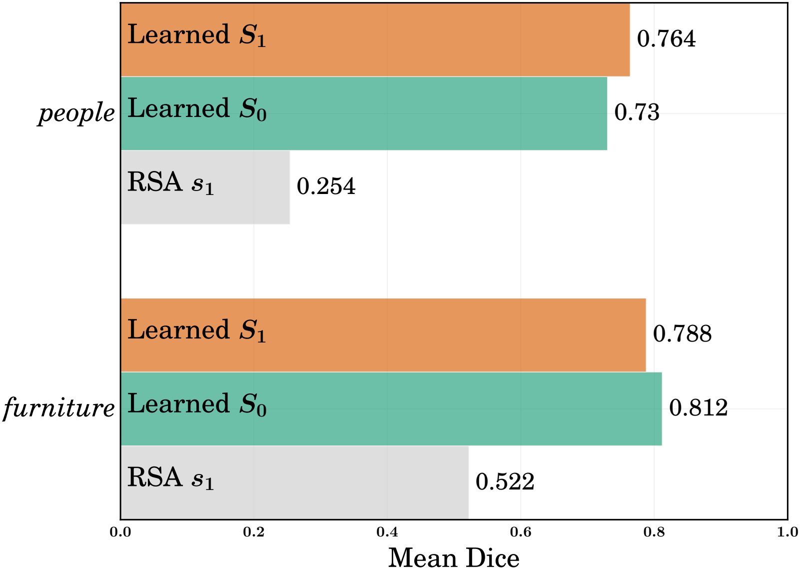
TUNA Results



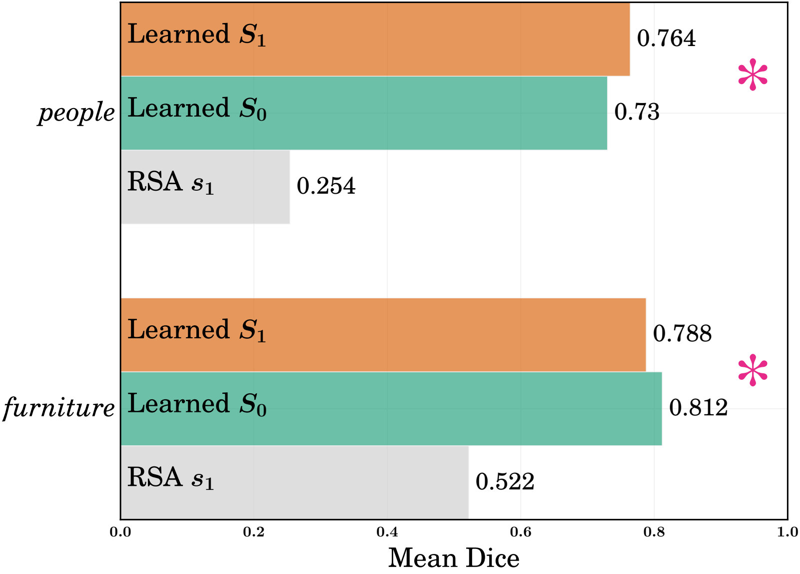
TUNA Results



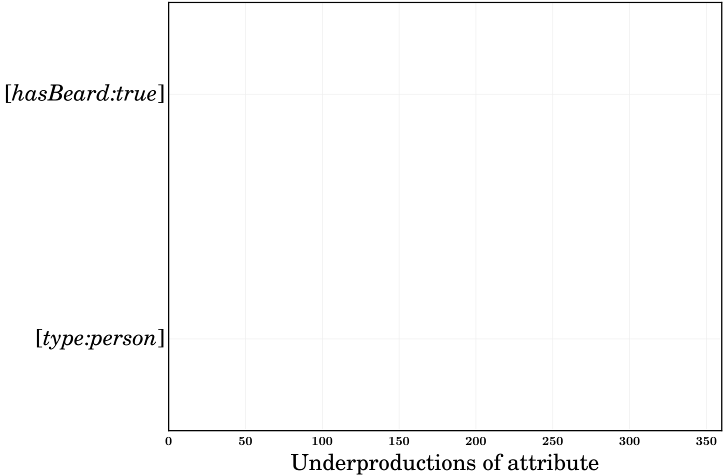
TUNA Results



TUNA Results

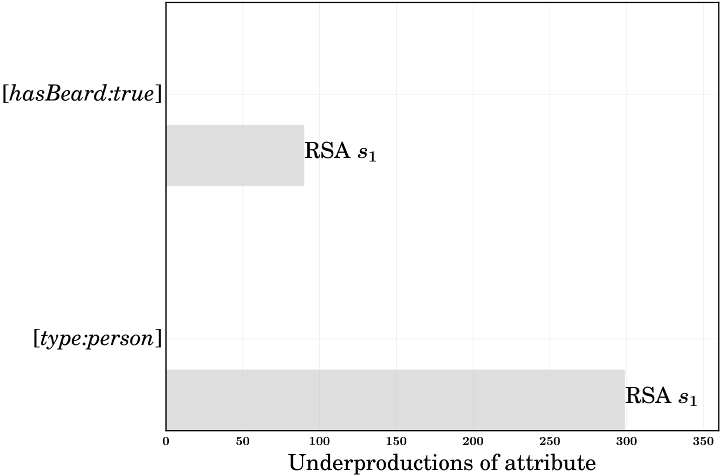


Error analysis



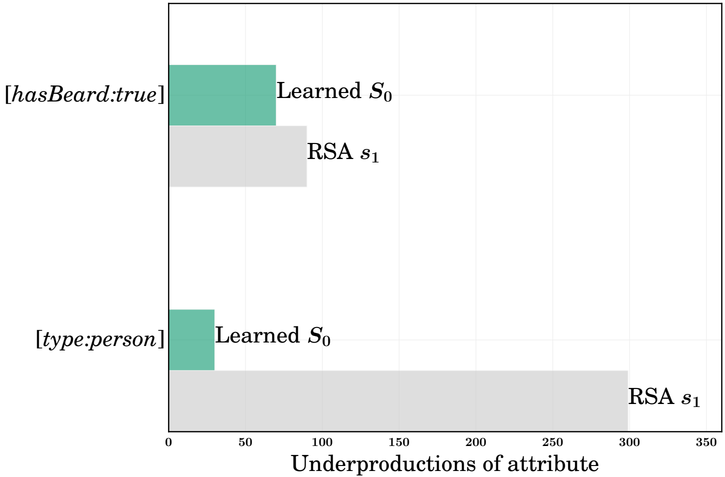
(Lower is better!)

Error analysis



(Lower is better!)

Error analysis



(Lower is better!)

Error analysis

(Lower is better!)

Limitations

- Scalability 😟
- Sensitivity to variation 😃
- Bounded rationality 😃
- New kinds of model assessment 😃
- Impact 😟

Introspective speakers from Google

Generation and Comprehension of Unambiguous Object Descriptions

Junhua Mao[2*] Jonathan Huang[1] Alexander Toshev[1] Oana Camburu[3] Alan Yuille[2,4] Kevin Murphy[1]
[1]Google Inc. [2]University of California, Los Angeles [3]University of Oxford [4]Johns Hopkins University
{mjhustc@,yuille@stat.}ucla.edu, oana-maria.camburu@cs.ox.ac.uk
{jonathanhuang,toshev,kpmurphy}@google.com

Context-aware Captions from Context-agnostic Supervision

Ramakrishna Vedantam[1] Samy Bengio[2] Kevin Murphy[2] Devi Parikh[3] Gal Chechik[2]
[1]Virginia Tech [3]Georgia Institute of Technology [2]Google
[1]vrama91@vt.edu [3]parikh@gatech.edu [2]{bengio,kpmurphy,gal}@google.com

Mao et al. 2016; Vedantam et al. 2017

Google Refexp Dataset

Mao et al. 2016

Google Refexp Dataset

Mao et al. 2016

Google Refexp Dataset

S: “A backpack”

Mao et al. 2016

Google Refexp Dataset

S: “A backpack”

Listener

Mao et al. 2016

Google Refexp Dataset

~~S: “A backpack”~~

Listener

Mao et al. 2016

Google Refexp Dataset

Mao et al. 2016

Google Refexp Dataset

S: “A yellow and black backpack”

Mao et al. 2016

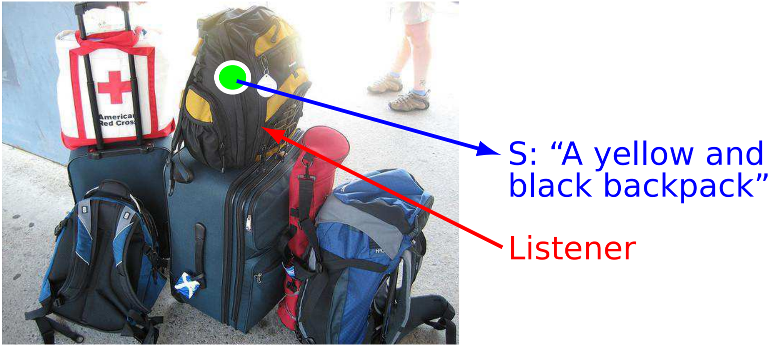
Google Refexp Dataset

S: “A yellow and black backpack”

Listener

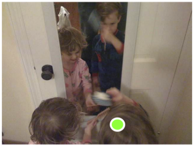
Mao et al. 2016

Google Refexp Dataset



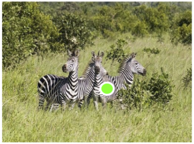
Mao et al. 2016

Google Refexp Dataset examples



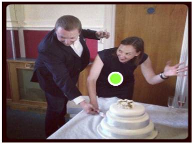
A boy brushing his hair while looking at his reflection.

A young male child in pajamas shaking around a hairbrush in the mirror.



Zebra looking towards the camera.

A zebra third from the left.



The woman in black dress.

A lady in a black dress cuts a wedding cake with her new husband.

A bird that is close to the baby in a pink shirt.

A bird standing on the shoulder of a person with its tail touching her face.

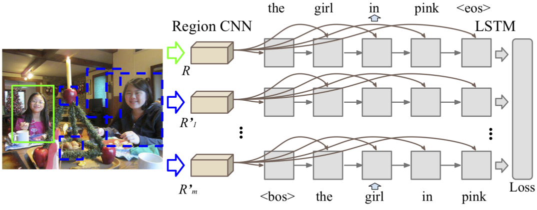
Mao et al. 2016

Maximum Mutual Information Training

Neural listener objective

Where an example is a message, a set of entities I, and a entity $ent \in I$:

$$J'(\theta) = -\sum_{n=1}^{N} \log \frac{\mathbf{S}^{\theta}_{\text{lit}}(msg_n \mid ent_n; I_n)}{\sum_{ent' \in I_n} \mathbf{S}^{\theta}_{\text{lit}}(msg_n \mid ent_n; I_n)}$$



Mao et al. 2016

Maximum Mutual Information Training

Neural listener objective

Where an example is a message, a set of entities I, and a entity $ent \in I$:

$$J'(\theta) = -\sum_{n=1}^{N} \log \frac{\mathbf{S}^{\theta}_{\text{lit}}(msg_n \mid ent_n; I_n)}{\sum_{ent' \in I_n} \mathbf{S}^{\theta}_{\text{lit}}(msg_n \mid ent_n; I_n)}$$

Max margin objective

To speed up training and make it more stable, they approximate the abovean max-margin objective that compares each target with a single randomly chosen distractor.

Mao et al. 2016

Introspective image captioners

Target Image:

Distractor Image:

Speaker:
An airplane is flying in the sky.

Introspective Speaker:
A large passenger jet flying through a blue sky.

Vedantam et al. 2017

Introspective speaker training

$$\Delta(I, state, state') =$$

$$\text{argmax}_{msg} \Bigg[\lambda \log \mathbf{S}^{\theta}_{\text{lit}}(msg \mid state; I_n) + (1 - \lambda) \log \frac{\mathbf{S}^{\theta}_{\text{lit}}(msg \mid state; I_n)}{\mathbf{S}^{\theta}_{\text{lit}}(msg \mid state'; I_n)} \Bigg]$$

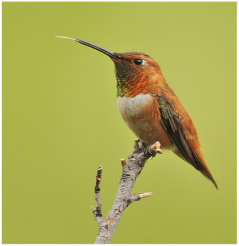
Proportional to a standard RSA $\mathcal{L}_1$.

Vedantam et al. 2017

Diagnosing the role of introspection

Target image and class

Rufous Hummingbird



Justifications vary with λ

fully discriminative

λ **= 0.00** tarsals orange white brown wings wings orange tail dark an primaries

λ **= 0.30** This is a brown bird with a brown wing and a long pointy beak.

λ **= 0.50** This bird is **brown with red on its neck** and has a long , pointy beak.

λ **= 0.70** This is a bird with a **white belly , brown wing and a red throat**.

λ **= 1.00** A small sized bird that has a very long and pointed bill.

context blind

Distractor class

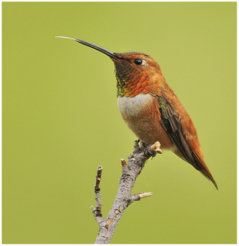
Ruby throated Hummingbird

Vedantam et al. 2017

Diagnosing the role of introspection

Target image and class

Rufous Hummingbird

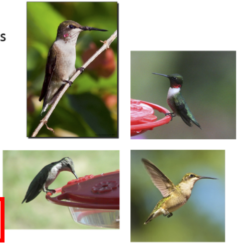


Justifications vary with λ

fully discriminative	
λ **= 0.00**	tarsals orange white brown wings wings orange tail dark an primaries
λ **= 0.30**	This is a brown bird with a brown wing and a long pointy beak.
λ **= 0.50**	This bird is **brown with red on its neck** and has a long , pointy beak.
λ **= 0.70**	This is a bird with a **white belly , brown wing and a red throat**.
λ **= 1.00** context blind	A small sized bird that has a very long and pointed bill.

Distractor class

Ruby throated Hummingbird

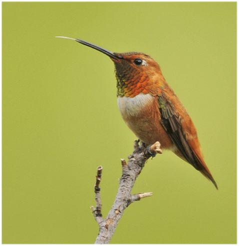


Vedantam et al. 2017

Diagnosing the role of introspection

Target image and class

Rufous Hummingbird

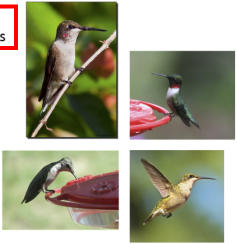


Justifications vary with λ

fully discriminative	
λ = 0.00	tarsals orange white brown wings wings orange tail dark an primaries
λ = 0.30	This is a brown bird with a brown wing and a long pointy beak.
λ = 0.50	This bird is **brown with red on its neck** and has a long , pointy beak.
λ = 0.70	This is a bird with a **white belly , brown wing and a red throat**.
λ = 1.00 context blind	A small sized bird that has a very long and pointed bill.

Distractor class

Ruby throated Hummingbird

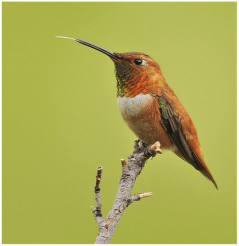


Vedantam et al. 2017

Diagnosing the role of introspection

Target image and class

Rufous Hummingbird



Justifications vary with λ

fully discriminative

λ = 0.00 tarsals orange white brown wings wings orange tail dark an primaries

λ = 0.30 This is a brown bird with a brown wing and a long pointy beak.

λ = 0.50 This bird is **brown with red on its neck** and has a long , pointy beak.

λ = 0.70 This is a bird with a **white belly , brown wing and a red throat**.

λ = 1.00 A small sized bird that has a very long and pointed bill.

context blind

Distractor class

Ruby throated Hummingbird

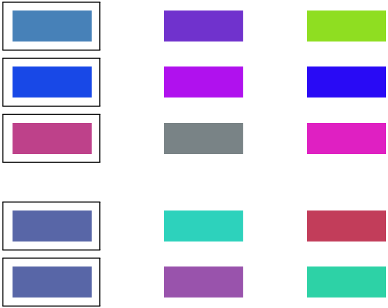
Vedantam et al. 2017

Modular neural RSA

Papers employing these general techniques

- Andreas and Klein. 2016. Reasoning about pragmatics with neural listeners and speakers. *EMNLP*.
- Fried, Andreas, Klein. 2018. Unified pragmatic models for generating and following instructions. *NAACL*.
- Monroe, Hawkins, Goodman, Potts. Colors in context: A pragmatic neural model for grounded language understanding. *TACL*.
- Monroe, Hu, Jong, Potts. 2018. Generating bilingual pragmatic color references. *NAACL*.

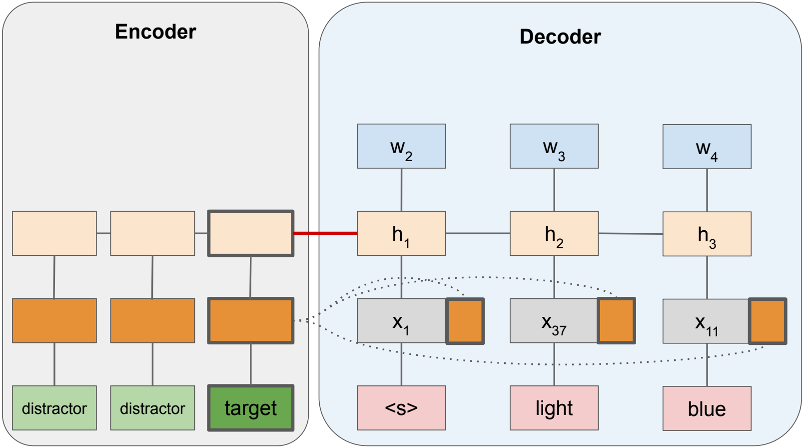
Stanford English Colors in Context corpus



Context	Utterance
	blue
	The darker blue one
	dull pink not the super bright one
	Purple
	blue

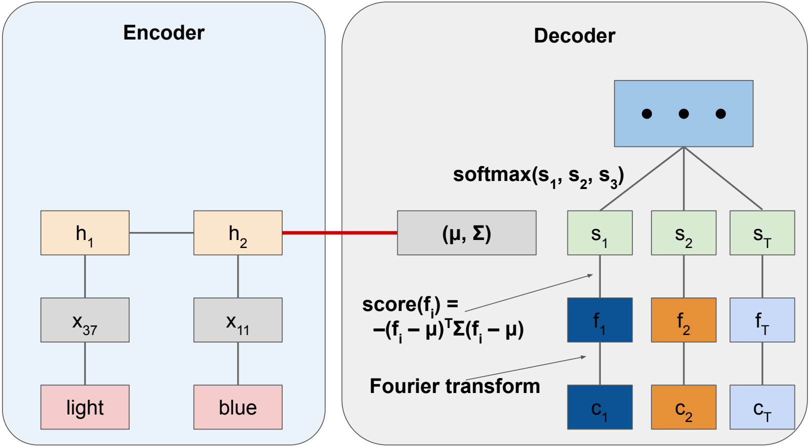
Stanford Colors in Context corpus (Monroe et al. 2017).
Chinese version available as well (Monroe et al. 2018)!

Literal neural speaker S_{lit}^{θ}



Monroe et al. 2017

Neural literal listener $\mathcal{L}_0$



Monroe et al. 2017

The formative challenge again

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg) - C(msg)\right)\right)}{\boxed{\sum_{msg'}} \exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg') - C(msg')\right)\right)}$$

Neural pragmatic agents

Neural pragmatic agents

Neural pragmatic speaker (Andreas and Klein 2016)

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathcal{L}_0(state \mid msg)}{\sum_{msg' \in X} \mathcal{L}_0(state \mid msg')}$$

with X a sample from $\mathbf{S}^{\theta}_{\text{lit}}(msg \mid state)$ such that $msg \in X$.

Neural pragmatic agents

Neural pragmatic speaker (Andreas and Klein 2016)

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathcal{L}_0(state \mid msg)}{\sum_{msg' \in X} \mathcal{L}_0(state \mid msg')}$$

with X a sample from $\mathbf{S}^{\theta}_{\text{lit}}(msg \mid state)$ such that $msg \in X$.

Neural pragmatic listener

$$\mathcal{L}_1(state \mid msg) \propto \mathbf{S}^{\theta}_{\text{prag}}(msg \mid state)$$

Neural pragmatic agents

Neural pragmatic speaker (Andreas and Klein 2016)

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathcal{L}_0(state \mid msg)}{\sum_{msg' \in X} \mathcal{L}_0(state \mid msg')}$$

with X a sample from $\mathbf{S}^{\theta}_{\text{lit}}(msg \mid state)$ such that $msg \in X$.

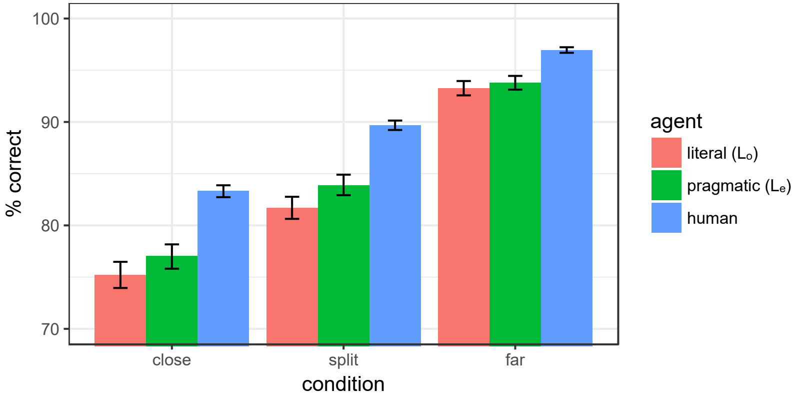
Neural pragmatic listener

$$\mathcal{L}_1(state \mid msg) \propto \mathbf{S}^{\theta}_{\text{prag}}(msg \mid state)$$

Blended neural pragmatic listener

Weighted combination of $\mathcal{L}_0$ and $\mathcal{L}_1$.

English Colors in Context results



Monroe et al. 2017

Incremental pragmatics

The speaker normalization challenge again

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg) - C(msg)\right)\right)}{\boxed{\sum_{msg'}} \exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg') - C(msg')\right)\right)}$$

The speaker normalization challenge again

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg) - C(msg)\right)\right)}{\boxed{\sum_{msg'}} \exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg') - C(msg')\right)\right)}$$

Sampling

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathcal{L}_0(state \mid msg)}{\sum_{msg' \in X} \mathcal{L}_0(state \mid msg')}$$

with X a sample from $\mathbf{S}^{\theta}_{\text{lit}}(msg \mid state)$ such that $msg \in X$.

The speaker normalization challenge again

$$S_{\text{prag}}(msg \mid state) = \frac{\exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg) - C(msg)\right)\right)}{\boxed{\sum_{msg'}} \exp\left(\alpha\left(\log L_{\text{lit}}(state \mid msg') - C(msg')\right)\right)}$$

Sampling

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathcal{L}_0(state \mid msg)}{\sum_{msg' \in X} \mathcal{L}_0(state \mid msg')}$$

with X a sample from $\mathbf{S}^{\theta}_{\text{lit}}(msg \mid state)$ such that $msg \in X$.

Cohn-Gordon et al. (2018)

Not good enough!

The incremental RSA model

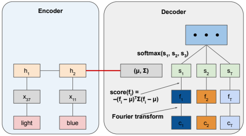
$$L_0^{\mathrm{WORD}}(state \mid [word_1 \ldots word_n], word)$$

Cohn-Gordon et al. 2018, 2019

The incremental RSA model

$$[\![\cdot]\!]$$

$$L_0^{\mathrm{WORD}}(\mathit{state} \mid [\mathit{word}_1 \ldots \mathit{word}_n], \mathit{word})$$

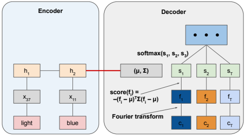


Cohn-Gordon et al. 2018, 2019

The incremental RSA model

$$[\![\cdot]\!]$$

$$L_0^{\text{WORD}}(\textit{state} \mid [\textit{word}_1 \ldots \textit{word}_n], \textit{word})$$



$$S_1^{\text{WORD}}(\textit{word} \mid [\textit{word}_1 \ldots \textit{word}_n], \textit{state}) = \frac{L_0^{\text{WORD}}(\textit{state} \mid [\textit{word}_1 \ldots \textit{word}_n], \textit{word})}{\sum_{\textit{word}' \in \mathbf{V}} L_0^{\text{WORD}}(\textit{state} \mid [\textit{word}_1 \ldots \textit{word}_n], \textit{word}')}$$

Cohn-Gordon et al. 2018, 2019

Evaluation

Using a separately trained S_0, apply Bayes Rule to create a simple classifier (see also Newman et al. 2020). Each context contains the target and 9 distractors.

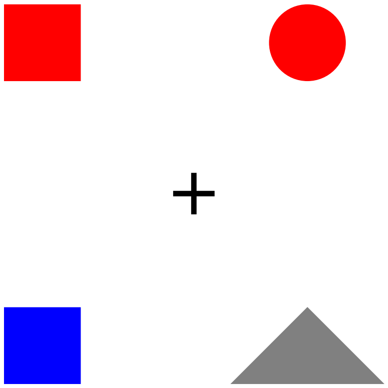
	Visual Genome	
Model	**Common objects**	**Confusable objects**
Char S_0	48.9	47.5
Char S_1	**68.0**	**65.9**
Word S_0	57.6	53.4
Word S_1	60.6	57.6

Cohn-Gordon et al. 2018

Papers employing these general techniques

- Cohn-Gordon and Goodman. 2019. Lost in machine translation: a method to reduce meaning loss. *NAACL*.
- Cohn-Gordon, Goodman, Potts. 2018. Pragmatically informative image captioning with character-level inference. *SCiL*.
- Cohn-Gordon, Goodman, Potts. 2019. An incremental iterated response model of pragmatics. *SCiL*.

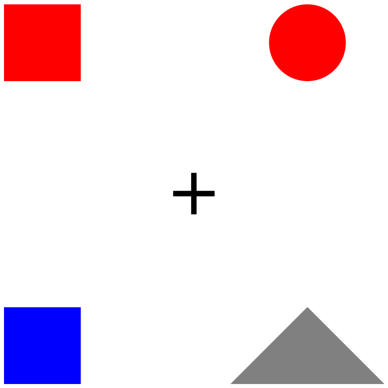
Linguistic evidence for incremental pragmatics



Sedivy et al. 1999; Sedivy 2007; Grodner and Sedivy 2008

Linguistic evidence for incremental pragmatics

Touch

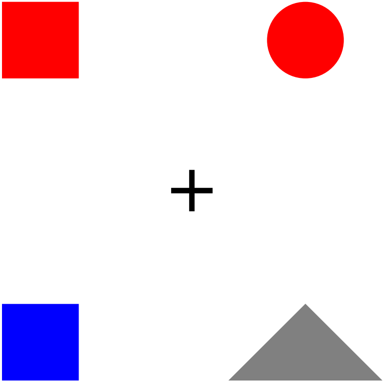


Sedivy et al. 1999; Sedivy 2007; Grodner and Sedivy 2008

Linguistic evidence for incremental pragmatics

Touch

the



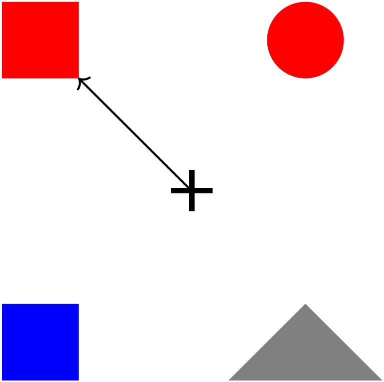
Sedivy et al. 1999; Sedivy 2007; Grodner and Sedivy 2008

Linguistic evidence for incremental pragmatics

Touch

the

red



Sedivy et al. 1999; Sedivy 2007; Grodner and Sedivy 2008

Conclusions

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.
3. Informativity in context motivates refinements to the NLP tasks themselves.

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.
3. Informativity in context motivates refinements to the NLP tasks themselves.
4. Next frontiers:

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.
3. Informativity in context motivates refinements to the NLP tasks themselves.
4. Next frontiers:
   - More complex open domains

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.
3. Informativity in context motivates refinements to the NLP tasks themselves.
4. Next frontiers:
   - More complex open domains
   - Richer contextual representation

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.
3. Informativity in context motivates refinements to the NLP tasks themselves.
4. Next frontiers:
   - More complex open domains
   - Richer contextual representation
   - Individualization and adaptation

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.
3. Informativity in context motivates refinements to the NLP tasks themselves.
4. Next frontiers:
   - More complex open domains
   - Richer contextual representation
   - Individualization and adaptation
   - ...

1. Iterated response models of pragmatics successfully capture a wide range of linguistic phenomena.
2. Many NLP tasks benefit from such models as well.
3. Informativity in context motivates refinements to the NLP tasks themselves.
4. Next frontiers:
   - More complex open domains
   - Richer contextual representation
   - Individualization and adaptation
   - ...

Thanks!

References I

Jacob Andreas and Dan Klein. 2016. Reasoning about pragmatics with neural listeners and speakers. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1173–1182. Association for Computational Linguistics.

Leon Bergen, Roger Levy, and Noah D. Goodman. 2016. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9(20).

Heather Burnett. 2017. Indexicality, utility and conceptual spaces variation and change in grammatical gender in the *Assemblée Nationale*. Talk at Stanford University.

Colin F. Camerer, Teck-Hua Ho, and Juin-Kuan Chong. 2004. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898.

Reuben Cohn-Gordon and Leon Bergen. 2019. Verbal irony, pretense, and the common ground. Ms. Stanford University and UC San Diego.

Reuben Cohn-Gordon and Noah Goodman. 2019. Lost in machine translation: A method to reduce meaning loss. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 437–441, Minneapolis, Minnesota. Association for Computational Linguistics.

Reuben Cohn-Gordon, Noah D. Goodman, and Christopher Potts. 2018. Pragmatically informative image captioning with character-level inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 439–443, Stroudsburg, PA. Association for Computational Linguistics.

Reuben Cohn-Gordon, Noah D. Goodman, and Christopher Potts. 2019. An incremental iterated response model of pragmatics. In *Proceedings of the Society for Computation in Linguistics*, pages 81–90, Washington, D.C. Linguistic Society of America.

Judith Degen. 2013. A corpus-based study of *Some* (but not *All*) implicatures. Ms., University of Rochester.

Michael C. Frank, Andrés Gómez Emilsson, Benjamin Peloquin, Noah D. Goodman, and Christopher Potts. 2016. Rational speech act models of pragmatic reasoning in reference games. Ms., Stanford University.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998.

Michael Franke. 2009. *Signal to Act: Game Theory in Pragmatics*. ILLC Dissertation Series. Institute for Logic, Language and Computation, University of Amsterdam.

Daniel Fried, Jacob Andreas, and Dan Klein. 2018. Unified pragmatic models for generating and following instructions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1951–1963, New Orleans, Louisiana. Association for Computational Linguistics.

References II

Albert Gatt, Anja Belz, and Eric Kow. 2009. The TUNA-REG Challenge 2009: Overview and evaluation results. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, pages 174–182, Athens, Greece. Association for Computational Linguistics.

Dave Golland, Percy Liang, and Dan Klein. 2010. A game-theoretic approach to generating spatial descriptions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 410–419, Stroudsburg, PA. ACL.

Daniel J. Grodner and Julie Sedivy. 2008. The effects of speaker-specific information on pragmatic inferences. In *The Processing and Acquisition of Reference*, pages 239–272, Cambridge, MA. MIT Press.

Julia Hirschberg. 1985. *A Theory of Scalar Implicature*. Ph.D. thesis, University of Pennsylvania.

Gerhard Jäger. 2007. Game dynamics connects semantics and pragmatics. In Ahti-Veikko Pietarinen, editor, *Game Theory and Linguistic Meaning*, pages 89–102. Elsevier, Amsterdam.

Justine T. Kao, Leon Bergen, and Noah D. Goodman. 2014a. Formalizing the pragmatics of metaphor understanding. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 719–724, Wheat Ridge, CO. Cognitive Science Society.

Justine T. Kao, Jean Y. Wu, Leon Bergen, and Noah D. Goodman. 2014b. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33):12002–12007.

Stephen C. Levinson. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press, Cambridge, MA.

David Lewis. 1969. *Convention*. Harvard University Press, Cambridge, MA. Reprinted 2002 by Blackwell.

Junhua Mao, Jonathan Huang, Alexander Toshev, Oana Camburu, Alan L. Yuille, and Kevin Murphy. 2016. Generation and comprehension of unambiguous object descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 11–20. IEEE.

Will Monroe, Robert X. D. Hawkins, Noah D. Goodman, and Christopher Potts. 2017. Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, 5:325–338.

Will Monroe, Jennifer Hu, Andrew Jong, and Christopher Potts. 2018. Generating bilingual pragmatic color references. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2155–2165, Stroudsburg, PA. Association for Computational Linguistics.

Will Monroe and Christopher Potts. 2015. Learning in the Rational Speech Acts model. In *Proceedings of 20th Amsterdam Colloquium*, Amsterdam. ILLC.

Benjamin Newman, Reuben Cohn-Gordon, and Christopher Potts. 2020. Communication-based evaluation for natural language generation. In *Proceedings of the Society for Computation in Linguistics*, Washington, D.C. Linguistic Society of America.

Christopher Potts, Daniel Lassiter, Roger Levy, and Michael C. Frank. 2016. Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, 33(4):755–802.

References III

Christopher Potts and Roger Levy. 2015. Negotiating lexical uncertainty and speaker expertise with disjunction. In *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society*, pages 417–445, Berkeley, CA. Berkeley Linguistics Society.

Ciyang Qing and Reuben Cohn-Gordon. 2019. Use-conditional meaning in Rational Speech-Act models. In *Proceedings of Sinn und Bedeutung 23*, volume 2, pages 253–265.

Matthew Rabin. 1990. Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170.

Seymour Rosenberg and Bertram D. Cohen. 1964. Speakers' and listeners' processes in a word communication task. *Science*, 145(3637):1201–1203.

Sebastian Schuster and Judith Degen. 2019. Speaker-specific adaptation to variable use of uncertainty expressions. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pages 2769–2775, Montreal. Cognitive Science Society.

Julie Sedivy. 2007. Implicatures in real-time conversation: A view from language processing research. *Philosophy Compass*, 2/3:475–496.

Julie C. Sedivy, Michael K. Tanenhaus, Craig G. Chambers, and Gregory N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2):109–147.

Stefanie Tellex, Ross A. Knepper, Adrian Li, Thomas M. Howard, Daniela Rus, and Nicholas Roy. 2014. Asking for help using inverse semantics. In *Proceedings of Robotics: Science and Systems*.

Bob van Tiel, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts. 2016. Scalar diversity. *Journal of Semantics*, 33(1):137–175.

Ramakrishna Vedantam, Samy Bengio, Kevin Murphy, Devi Parikh, and Gal Chechik. 2017. Context-aware captions from context-agnostic supervision. *arXiv:1701.02870*.

Erica J. Yoon, Michael Henry Tessler, Noah D. Goodman, and Michael C. Frank. 2016. Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pages 2771–2776. Cognitive Science Society.

Erica J Yoon, Michael Henry Tessler, Noah D Goodman, and Michael C Frank. 2017. "i won't lie, it wasn't amazing": Modeling polite indirect speech. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.