Pragmatic reasoning in large-scale NLP systems

Christopher Potts

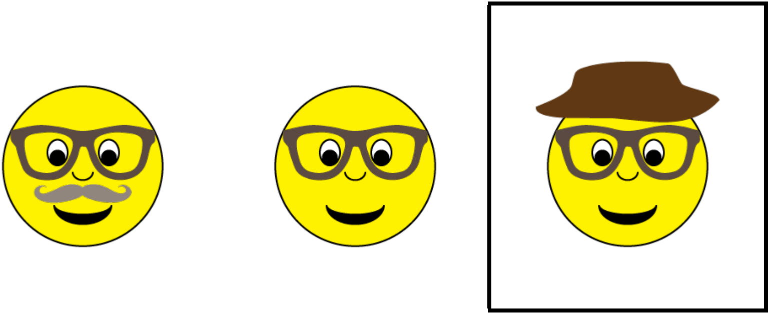
Stanford Linguistics

Stanford NLP Grounding Reading Group, May 13, 2020

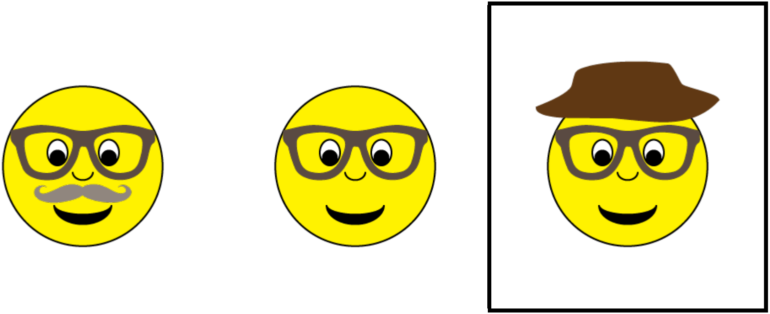
Informativity in context

Generating referring expressions

Generating referring expressions



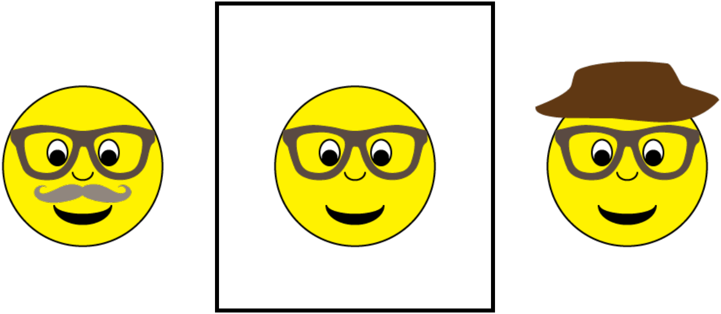
Generating referring expressions



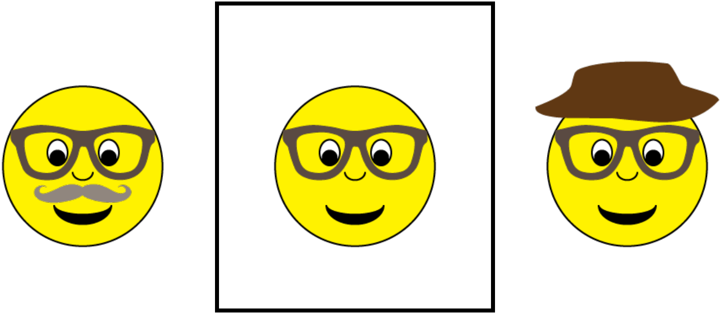
"The guy with a hat"

A conversational implicature

A conversational implicature

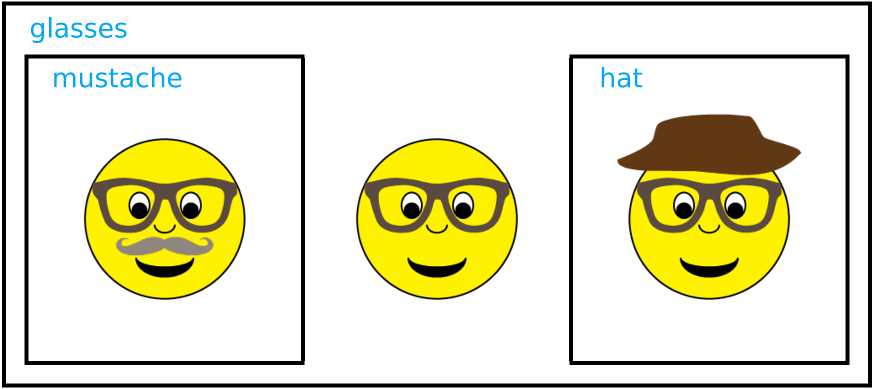


A conversational implicature



“The guy with glasses”

A conversational implicature



“The guy with glasses”

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

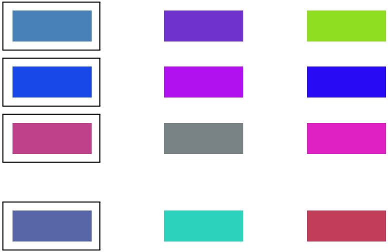
Stanford Colors in Context corpus
(Monroe et al. 2017)

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)

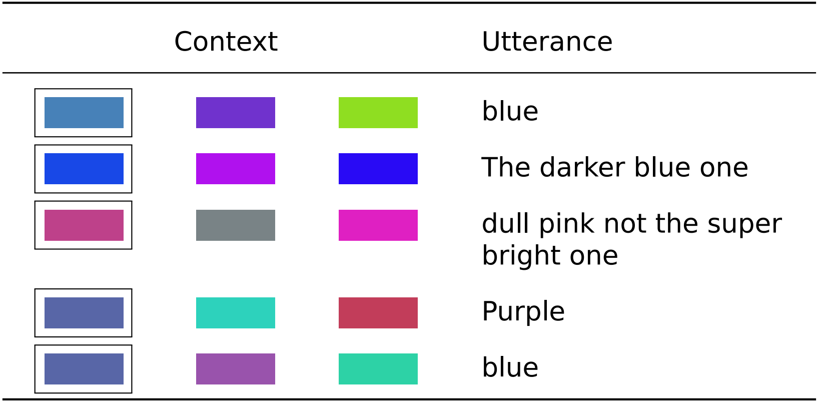
Interpreting complex descriptions



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

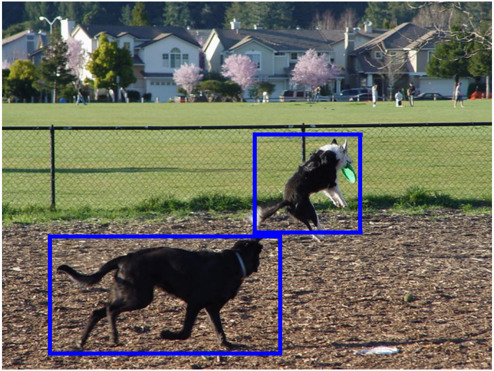


Stanford Colors in Context corpus
(Monroe et al. 2017)

Discriminative image labeling

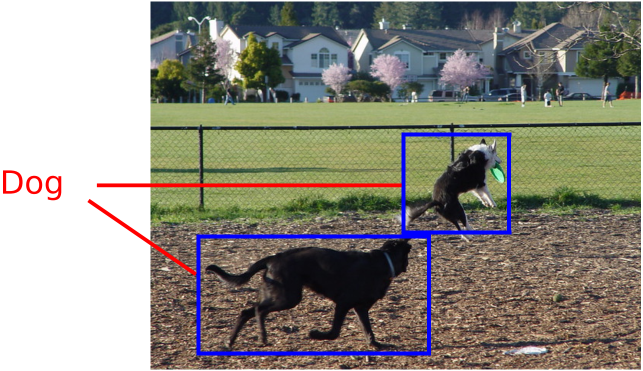
Mao et al. 2016

Discriminative image labeling



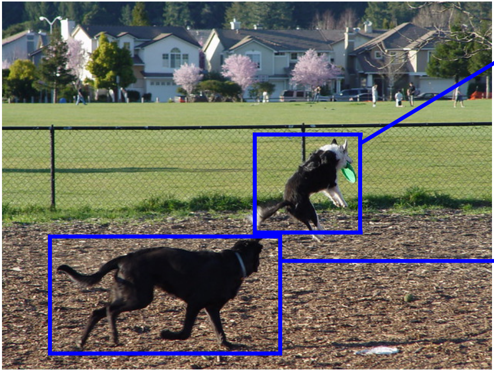
Mao et al. 2016

Discriminative image labeling



Mao et al. 2016

Discriminative image labeling



A little dog jumping and catching a frisbee

A big dog running

Mao et al. 2016

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

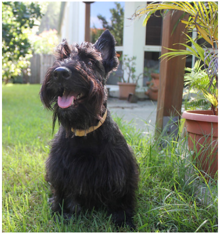
Google Refexp Dataset

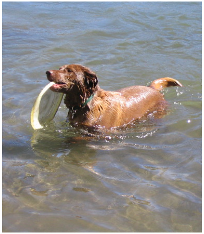
S: “A yellow and black backpack”

Listener

Mao et al. 2016

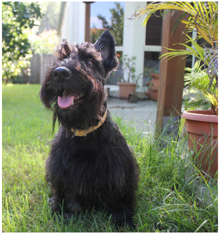
Discriminative image captioning

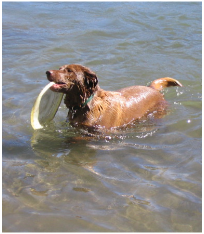




Vedantam et al. 2017; Cohn-Gordon et al. 2018

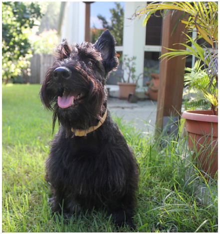
Discriminative image captioning





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

Discriminative image captioning



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

Issue sensitivity

In an article about recent campaign events:

Issue sensitivity

In an article about recent campaign events:

- “Politician Joe Biden speaking on stage”

Issue sensitivity

In an article about recent campaign events:

- “Politician Joe Biden speaking on stage”
- “Elderly, gray-haired politician Joe Biden speaking on stage”

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.

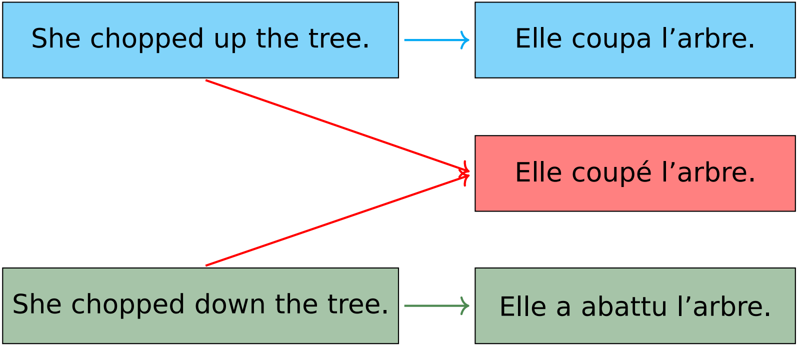
Ongoing work with Hanson Lu and Reuben Cohn-Gordon

Machine translation

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

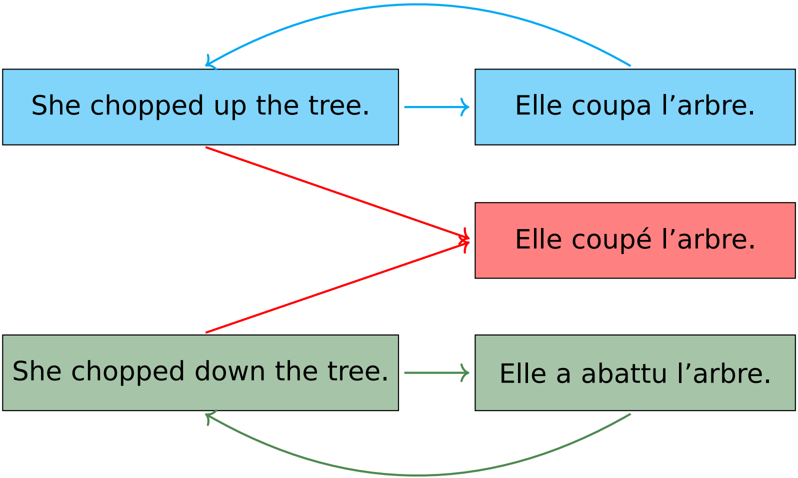
Cohn-Gordon and Goodman 2019

Machine translation



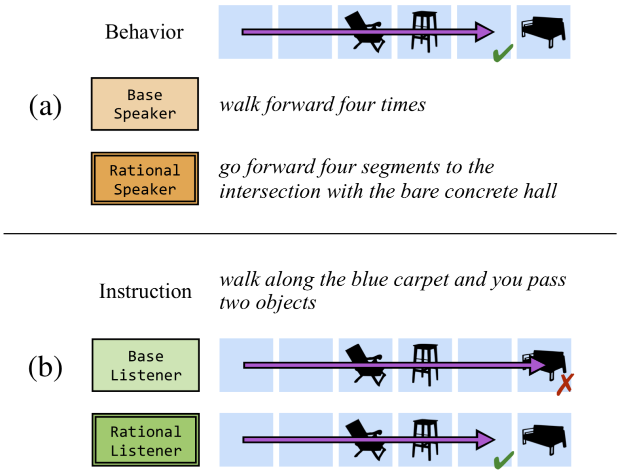
Cohn-Gordon and Goodman 2019

Machine translation



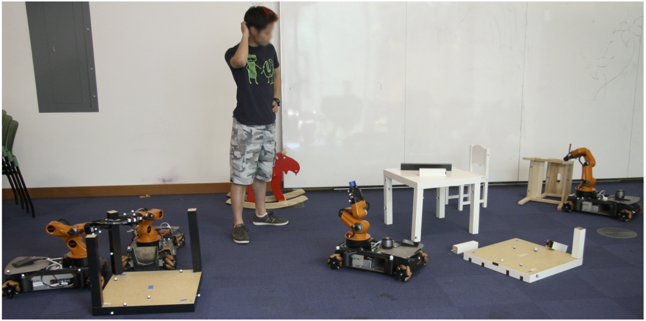
Cohn-Gordon and Goodman 2019

Generating and following instructions



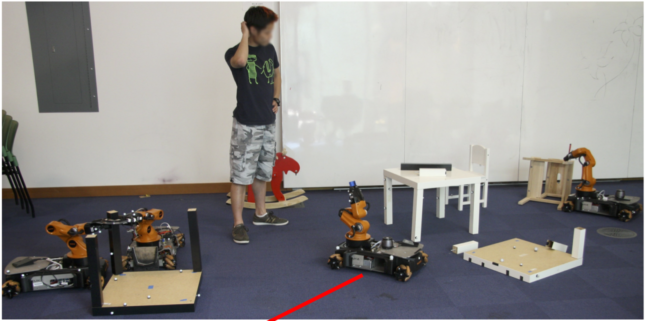
Fried et al. 2018

Collaborative problem solving



Tellex et al. 2014

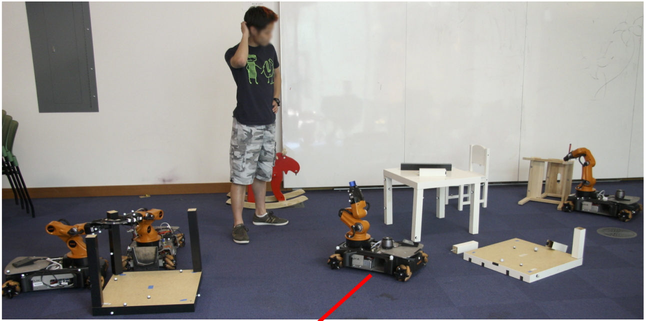
Collaborative problem solving



Help me!

Tellex et al. 2014

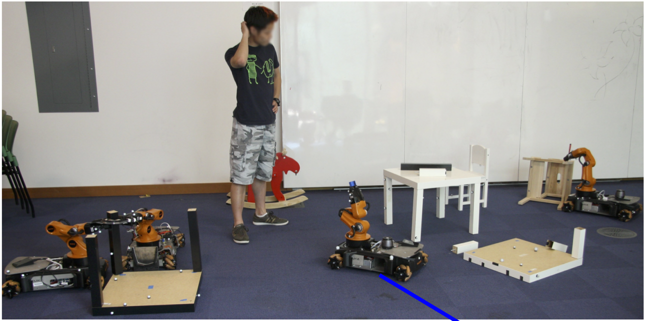
Collaborative problem solving



Hand me
the leg

Tellex et al. 2014

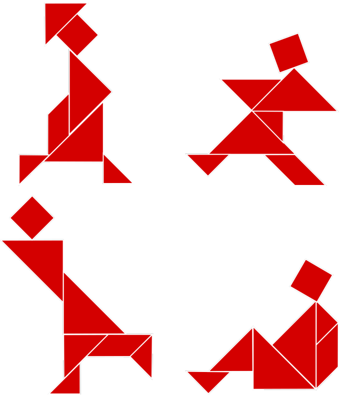
Collaborative problem solving



Hand me the white leg on the table

Tellex et al. 2014

Continual adaptation



Round 1: All right, the next one looks like a person who's ice skating, except they're sticking their arms out in front.

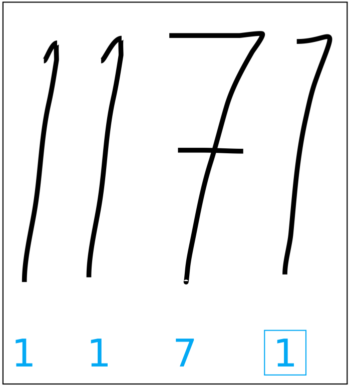
Round 2: Um, the next one's the person ice skating that has arms out.

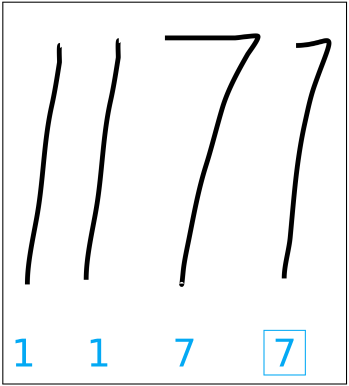
[. . .]

Round 6: The ice skater.

Clark and Wilkes-Gibbs 1986; DeVault et al. 2005; Paetzel et al. 2014; Hawkins et al. 2017

Optical character recognition





The Rational Speech Acts model (RSA)

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

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

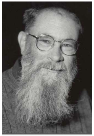
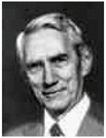
Pragmatic speaker

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

Literal listener

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

A simple example

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

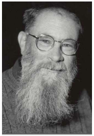
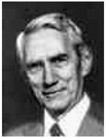
L_{prag}

S_{prag}

L_{lit}

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

A simple example

			
beard	**1**	0	0
glasses	.50	.50	0
tie	0	.50	.50

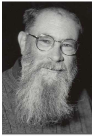
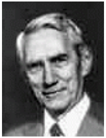
L_{prag}

S_{prag}

L_{lit}

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

A simple example

	beard	*glasses*	*tie*
	**.67**	.33	0
	0	.50	.50
	0	0	**1**

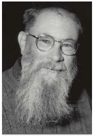
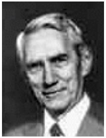
L_{prag}

S_{prag}

L_{lit}

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

A simple example

			
beard	**1**	0	0
glasses	.40	**.60**	0
tie	0	0.33	**.67**

L_{prag}

S_{prag}

L_{lit}

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

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 in linguistics

- 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

- 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

Logical semantic agents

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

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

Learned agents

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

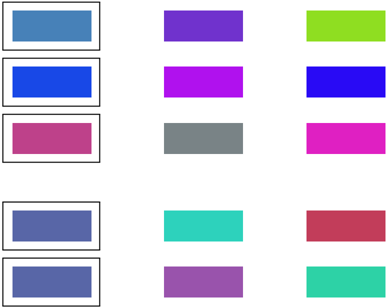
$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state)$$

Modular neural RSA

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*.
- 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*.

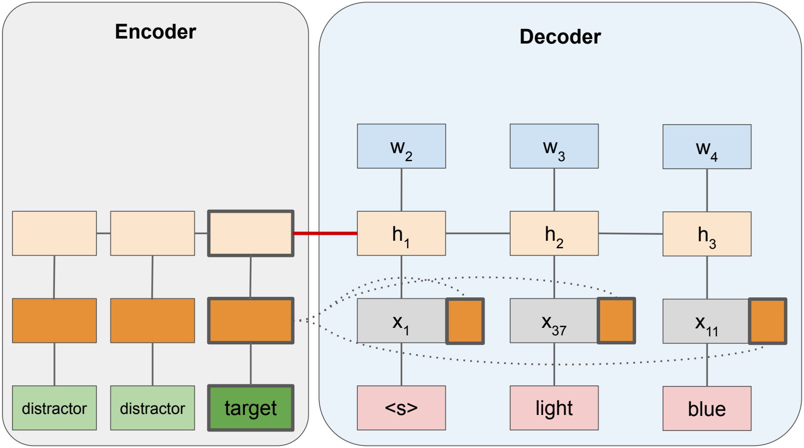
Stanford English Colors in Context corpus



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

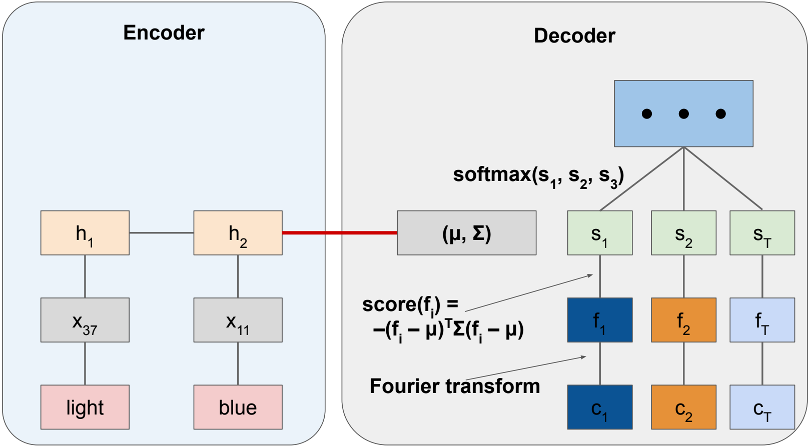
Stanford Colors in Context corpus (Monroe et al. 2017).
Chinese version available as well (Monroe et al. 2018)!

Literal neural speaker S_{lit}^{θ}



Monroe et al. 2017

Neural literal listener $\mathbf{L}_0^\theta$



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

Speaker (Andreas and Klein 2016)

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathbf{L}^{\theta}_{0}(state \mid msg)}{\sum_{msg' \in X} \mathbf{L}^{\theta}_{0}(state \mid msg')}$$

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

Neural pragmatic agents

Speaker (Andreas and Klein 2016)

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathbf{L}^{\theta}_{0}(state \mid msg)}{\sum_{msg' \in X} \mathbf{L}^{\theta}_{0}(state \mid msg')}$$

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

Listener

$$\mathbf{L}^{\theta}_{1}(state \mid msg) \propto \mathbf{S}^{\theta}_{\text{prag}}(msg \mid state)$$

Neural pragmatic agents

Speaker (Andreas and Klein 2016)

$$\mathbf{S}^{\theta}_{\text{prag}}(msg \mid state) = \frac{\mathbf{L}^{\theta}_{0}(state \mid msg)}{\sum_{msg' \in X} \mathbf{L}^{\theta}_{0}(state \mid msg')}$$

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

Listener

$$\mathbf{L}^{\theta}_{1}(state \mid msg) \propto \mathbf{S}^{\theta}_{\text{prag}}(msg \mid state)$$

Blended neural pragmatic listener

Weighted combination of $\mathbf{L}^{\theta}_{0}$ and $\mathbf{L}^{\theta}_{1}$.

An incremental alternative

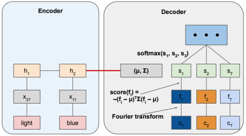
$$L_0^{\mathrm{WORD}}(state \mid [word_1 \ldots word_n], word)$$

Cohn-Gordon et al. 2018, 2019; Cohn-Gordon and Goodman 2019

An incremental alternative

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

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

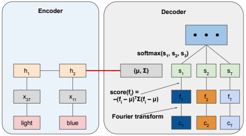


Cohn-Gordon et al. 2018, 2019; Cohn-Gordon and Goodman 2019

An incremental alternative

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

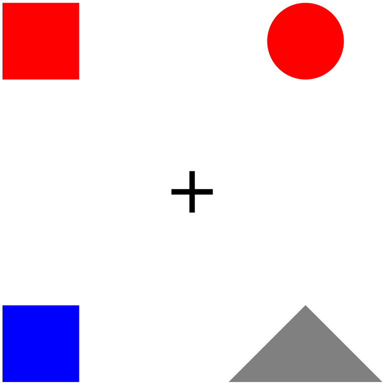
$$L_0^{\text{WORD}}(state \mid [word_1 \ldots word_n], word)$$



$$S_1^{\text{WORD}}(word \mid [word_1 \ldots word_n], state) = \frac{L_0^{\text{WORD}}(state \mid [word_1 \ldots word_n], word)}{\boxed{\sum_{word' \in \mathbf{V}}} L_0^{\text{WORD}}(state \mid [word_1 \ldots word_n], word')}$$

Cohn-Gordon et al. 2018, 2019; Cohn-Gordon and Goodman 2019

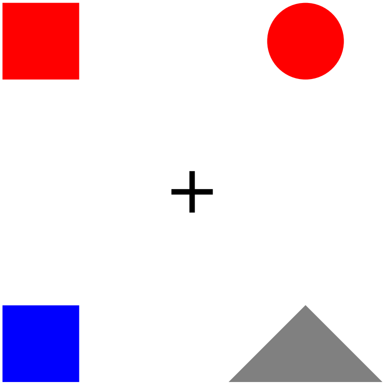
Linguistic evidence for incremental pragmatics



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

Linguistic evidence for incremental pragmatics

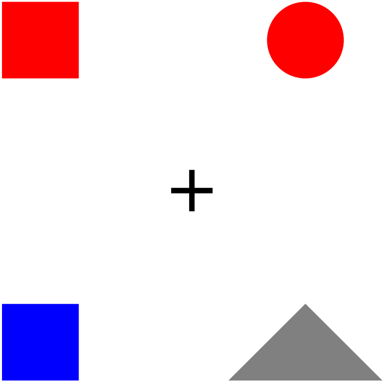
Touch



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

Linguistic evidence for incremental pragmatics

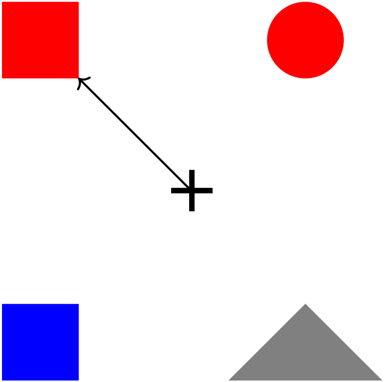
Touch
the



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

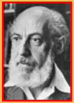
RSA learning objectives

TUNA people example

TUNA people example

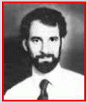
Utterance: “The bald man with a beard”

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

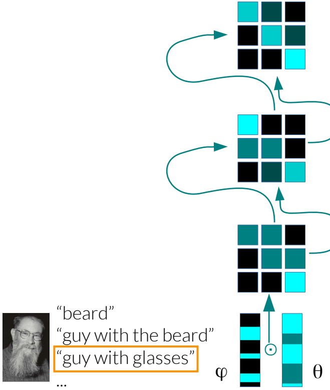
age:young
hairColour:dark
hasBeard:0
hasGlasses:0
hasHair:1
hasShirt:0
hasSuit:1
hasTie:1
type:person

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



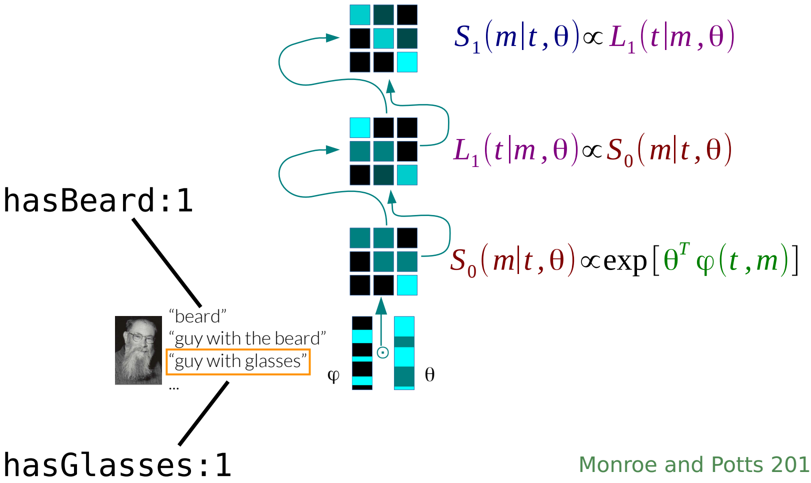
$$S_1(m|t,\theta) \propto L_1(t|m,\theta)$$

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

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

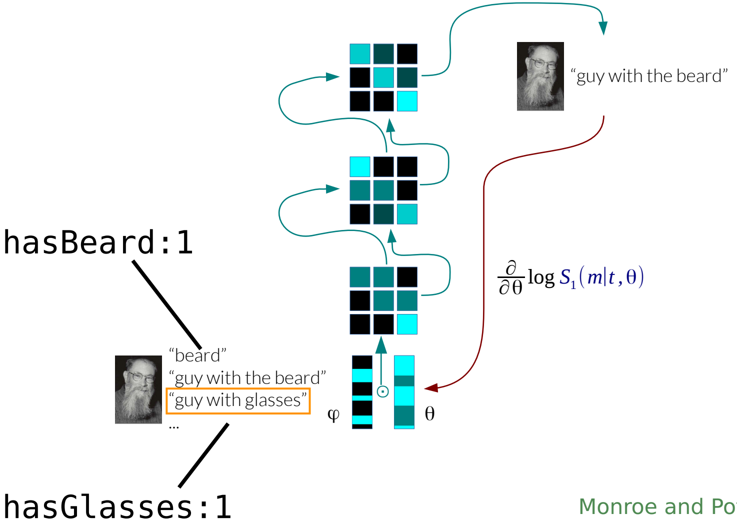
Monroe and Potts 2015

Forward inference



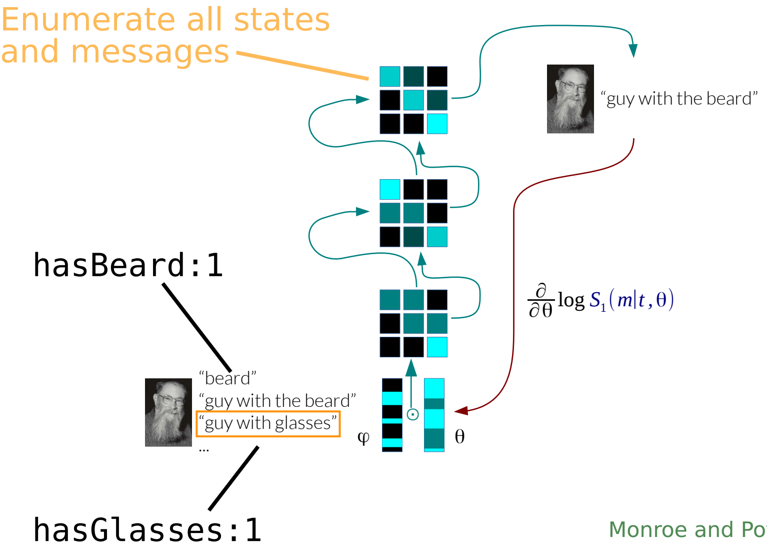
Monroe and Potts 2015

Optimization



Monroe and Potts 2015

Optimization



Monroe and Potts 2015

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

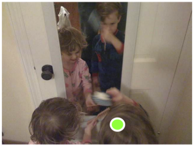
Google Refexp Dataset

S: "A yellow and black backpack"

Listener

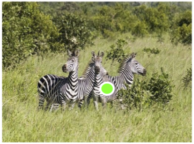
Mao et al. 2016

Google Refexp Dataset examples



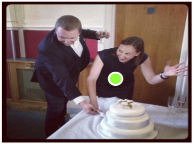
A boy brushing his hair while looking at his reflection.

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.

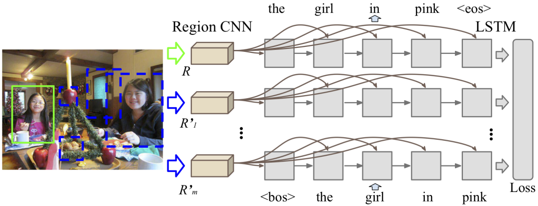
Mao et al. 2016

Maximum Mutual Information Training

Neural listener objective

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

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



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 $ent \in I$:

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

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

$$\Delta(I, state, state') = \operatorname{argmax}_{msg}\left[\lambda \log \mathbf{S}_{\text{lit}}^{\theta}(msg \mid state; I_n) + (1-\lambda)\log \frac{\mathbf{S}_{\text{lit}}^{\theta}(msg \mid state; I_n)}{\mathbf{S}_{\text{lit}}^{\theta}(msg \mid state'; I_n)}\right]$$

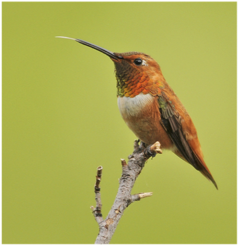
Proportional to a standard RSA $\mathbf{L}_1^{\theta}$.

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

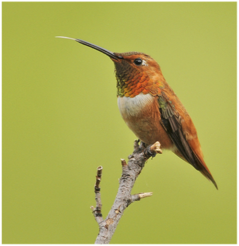



Vedantam et al. 2017

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.

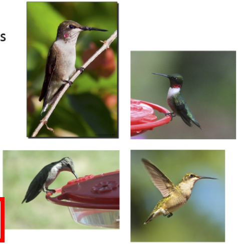
λ **= 0.50** This bird is **brown with red on its neck** and has a long , pointy beak.

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

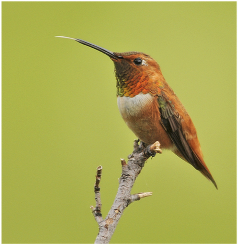


Vedantam et al. 2017

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	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

Vedantam et al. 2017

End-to-end training in White et al. (2020)

Amortized speaker

The exact RSA pragmatic speaker as the end-to-end objective, trained with a fixed, pretrained L_0:

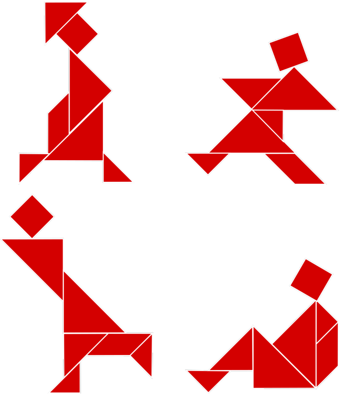
$$\mathbf{S}^{\theta}_{\text{prag}}(\mathit{msg} \mid \mathit{state}) \propto \exp\Big(\log \mathbf{L}^{\theta}_{0}(\mathit{state} \mid \mathit{msg}) - C(\mathit{msg})\Big)$$

Reinforcement learning speaker

The speaker interacts with a black-box listener, receiving positive reward iff the listener correctly infers the speaker's intended referent.

Frontiers

Continual adaptation



Round 1: All right, the next one looks like a person who's ice skating, except they're sticking their arms out in front.

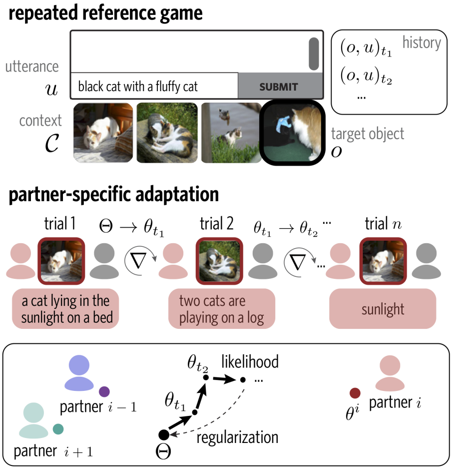
Round 2: Um, the next one's the person ice skating that has arms out.

[. . .]

Round 6: The ice skater.

Clark and Wilkes-Gibbs 1986; DeVault et al. 2005; Paetzel et al. 2014; Hawkins et al. 2017

Continual adaptation



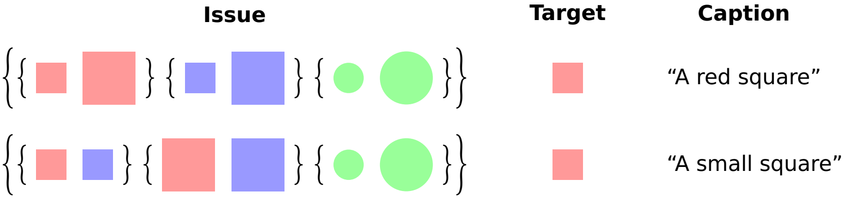
Hawkins et al. 2019

Issue sensitivity

In an article about recent campaign events:

- "Politician Joe Biden speaking on stage"
- "Elderly, gray-haired politician Joe Biden speaking on stage"

Issue sensitivity



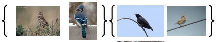
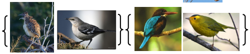
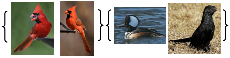
Issue sensitivity

$$U_1^{\mathbf{C}}(state, msg, \mathbf{C}) = \log\left(\sum_{state' \in \mathcal{I}} \delta_{[\mathbf{C}(state)=\mathbf{C}(state')]} L_1(state' \mid msg)\right)$$

$$U_2(msg, state, \mathbf{C}) = H(L_1(state' \mid msg) \cdot \delta_{[\mathbf{C}(state)=\mathbf{C}(state')]})$$

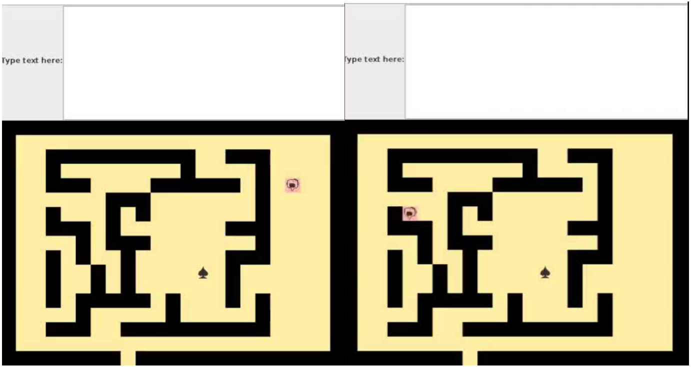
$$S_1^{\mathbf{C}+H}(msg \mid state, \mathbf{C}) \propto \exp\left(\alpha\left((1-\beta)U_1 + \beta U_2\right) - cost(msg)\right)$$

Issue sensitivity

Target	Issues	Partitions	Issue-sensitive Caption	Base Caption
	What is the **bill shape**?		a small bird with a white breast and belly brown wings and tail and **a pointed beak**	this bird has a brown crown a white eyebrow and a brown and white breast
	What is the **tail pattern**?		a small brown and white bird with a long beak and **long tail feathers**	
	What is the **belly color**?		this is a bird with a **white belly** brown back and a brown head	this is a grey bird with a red head and a red beak
	What is the **crown color**?		this is a reddish orange bird with black and white wings and a **red crown**	

Decision-theoretic agents

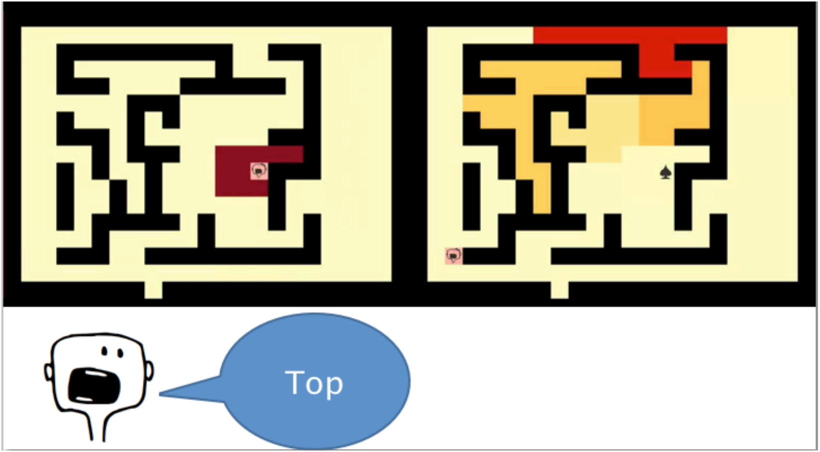
Both players must find the ace of spades. DialogBot:



Vogel et al. 2013a,b

Decision-theoretic agents

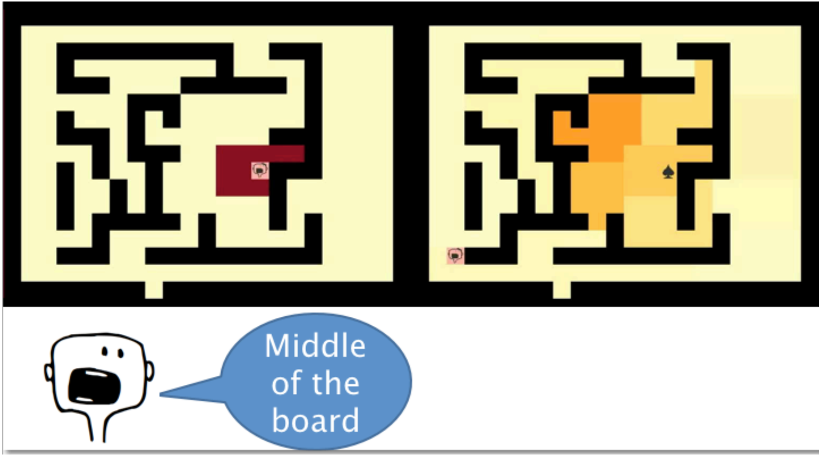
Baby DialogBots (a few hours of policy exploration)



Vogel et al. 2013a,b

Decision-theoretic agents

Grown-up DialogBots (a week of policy exploration)



Vogel et al. 2013a,b

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.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

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.

David DeVault, Natalia Kariaeva, Anubha Kothari, Iris Oved, and Matthew Stone. 2005. An information-state approach to collaborative reference. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 1–4, Ann Arbor, MI. Association for Computational Linguistics.

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

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.

Robert X. D. Hawkins, Minae Kwon, Dorsa Sadigh, and Noah D. Goodman. 2019. Continual adaptation for efficient machine communication. In *ICML Workshop on Adaptive & Multitask Learning: Algorithms & Systems*.

Robert XD Hawkins, Mike Frank, and Noah D. Goodman. 2017. Convention-formation in iterated reference games. In *Proceedings of the Cognitive Science Society*, pages 482–487, London. Cognitive Science Society.

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.

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.

Maike Paetzel, David Nicolas Racca, and David DeVault. 2014. A multimodal corpus of rapid dialogue games. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 4189–4195, Reykjavik, Iceland. European Language Resources Association (ELRA). ACL Anthology Identifier: L14-1548.

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*.

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

Adam Vogel, Max Bodoia, Christopher Potts, and Dan Jurafsky. 2013a. Emergence of Gricean maxims from multi-agent decision theory. In *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1072–1081, Stroudsburg, PA. Association for Computational Linguistics.

Adam Vogel, Christopher Potts, and Dan Jurafsky. 2013b. Implicatures and nested beliefs in approximate Decentralized-POMDPs. In *Proceedings of the 2013 Annual Conference of the Association for Computational Linguistics*, pages 74–80, Stroudsburg, PA. Association for Computational Linguistics.

Julia White, Jesse Mu, and Noah D. Goodman. 2020. Learning to refer informatively by amortizing pragmatic reasoning. In *Proceedings of the Cognitive Science Society*.

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*.