

Grounded language understanding

Christopher Potts

Stanford Linguistics

CS 224U: Natural language understanding
May 4 and 6



Overview

1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Assignment/Bake-off overview: Speakers in context
4. Listeners: From language to the world
5. Reasoning about other minds
6. The Rational Speech Acts model (RSA)
7. Neural RSA
8. Grounded chat bots
9. A few other grounding ideas

Associated materials

1. Code

- Notebook: `colors_overview.ipynb`
- Homework and bake-off: `hw_colors.ipynb`

2. Core reading: Monroe et al. 2017

3. Auxiliary readings: Golland et al. 2010; Lewis et al. 2017; Andreas and Klein 2016; Tellex et al. 2014; Vogel et al. 2013a

HAL

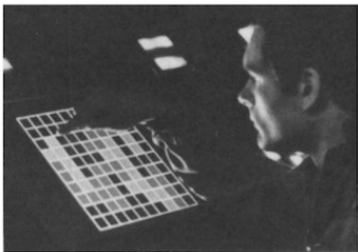
- In the 1967 Stanley Kubrick movie *2001: A Space Odyssey*, the spaceship's computer HAL can
 - ▶ display graphics;
 - ▶ play chess; and
 - ▶ conduct natural, open-domain conversations with humans.
- How well did the filmmakers do at predicting what computers would be capable in 2001?

Slide idea from Andrew McCallum

HAL

Graphics

HAL



Jurassic Park (1993)



Slide idea from Andrew McCallum

Siri



You: Any good burger joints around here?

Siri: I found a number of burger restaurants near you.

You: Hmm. How about tacos?

Apple: [Siri remembers that you asked about restaurants. so it will look for Mexican restaurants in the neighborhood. And Siri is proactive, so it will question you until it finds what you're looking for.]

Slide idea from Marie de Marneffe

Siri

Colbert: For the love of **God**, the **cameras** are on, give me something?

Siri: What kind of place are you looking for? **Camera stores** or **churches**?
[...]

Colbert: I don't want to search for anything! I want to write the show!

Siri: Searching the Web for "search for anything. I want to write the shuffle."



Slide idea from Marie de Marneffe

Levinson's (2000) analogy



Levinson's (2000) analogy

“We interpret this sketch instantly and effortlessly as a gathering of people before a structure, probably a gateway; the people are listening to a single declaiming figure in the center. [...] But all this is a miracle, for there is little detailed information in the lines or shading (such as there is). Every line is a mere suggestion [...]. So here is the miracle: from a merest, sketchiest squiggle of lines, you and I converge to find adumbration of a coherent scene [...].



Figure 8.1
Rembrandt sketch

Levinson's (2000) analogy



Figure 8.1
Rembrandt sketch

“We interpret this sketch instantly and effortlessly as a gathering of people before a structure, probably a gateway; the people are listening to a single declaiming figure in the center. [...] But all this is a miracle, for there is little detailed information in the lines or shading (such as there is). Every line is a mere suggestion [...]. So here is the miracle: from a merest, sketchiest squiggle of lines, you and I converge to find adumbration of a coherent scene [...].

“The problem of utterance interpretation is not dissimilar to this visual miracle. An utterance is not, as it were, a veridical model or “snapshot” of the scene it describes [...]. Rather, an utterance is just as sketchy as the Rembrandt drawing.”

Indexicality

1. I am speaking.
2. We won. [A team I'm on; a team I support; ...]
3. I am here [classroom; Stanford; ... planet earth; ...]
4. We are here. [pointing at a map]
5. I'm not here now. [old-fashioned answering machine]
6. We went to a local bar after work.
7. three days ago, tomorrow, now

Context dependence

Where are you from?

Context dependence

Where are you from?

- *Connecticut.* (Issue: birthplaces)
- *The U.S.* (Issue: nationalities)
- *Stanford.* (Issue: affiliations)
- *Planet earth.* (Issue: intergalactic meetings)

Context dependence

I didn't see any.

Context dependence

- Are there typos in my slides?

I didn't see any.

Context dependence

- Are there typos in my slides?
- Are there bookstores downtown?

I didn't see any.

Context dependence

- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the cupboard?

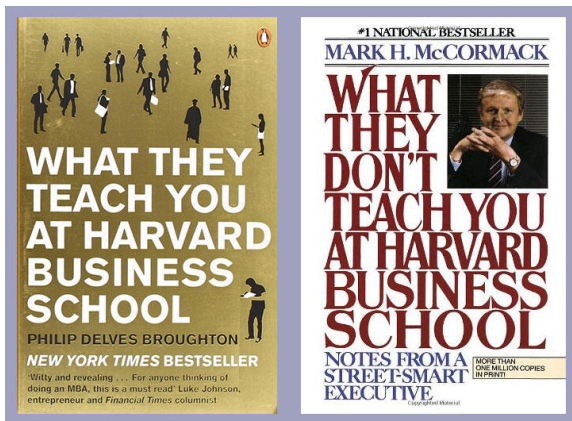
I didn't see any.

Context dependence

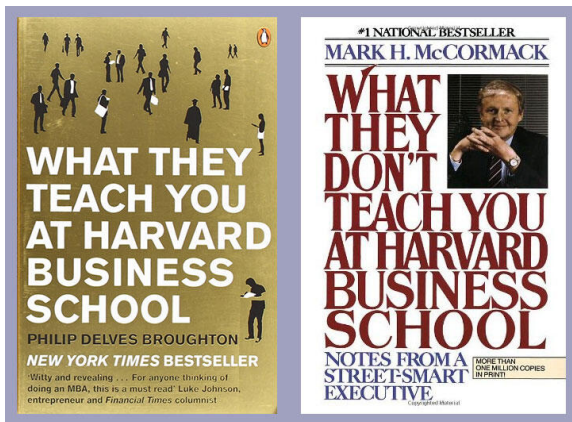
- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the cupboard?
- ...

I didn't see any.

Context dependence



Context dependence



“These two books contain the sum total of all human knowledge” (@James_Kpatrick)

Context dependence

1. The light is on. Chris must be in his office.
2. The Dean passed a new rule. Chris must be in his office.

Context dependence

If kangaroos had no tails, they would fall over.

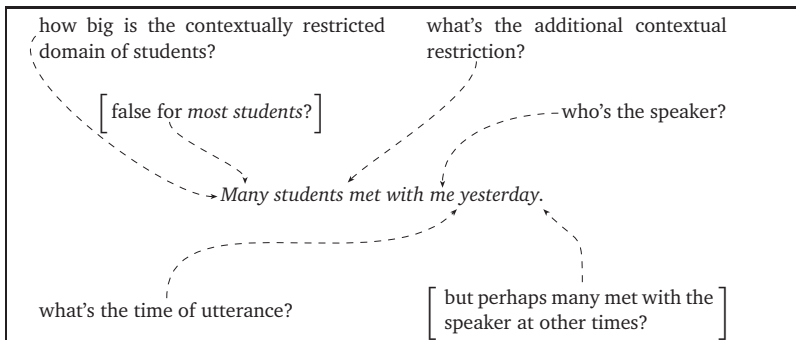
Seems true

Context dependence

If kangaroos had no tails, they would fall over.

Seems true, but suppose they had jetpacks.

Routine pragmatic enrichment



Language is action

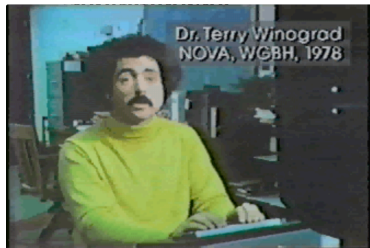
Winograd (1986:170):

“all language use can be thought of as a way of activating procedures within the hearer. We can think of an utterance as a program – one that indirectly causes a set of operations to be carried out within the hearer’s cognitive system.”

SHRDLU (Winograd 1972)

Full-grounded system that parses the user's input, maps it to logical form, interprets that logical form in its world, tries to take appropriate action, and generates responses.

<http://hci.stanford.edu/winograd/shrdlu/>



One project did succeed. Terry Winograd's program SHRDLU could use English intelligently, but there was a catch: the only subject you could discuss was a micro-world of simulated blocks.

SHRDLU (Winograd 1972)

1. Person: WHAT DOES THE BOX CONTAIN?
2. Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
3. Person: WHAT IS THE PYRAMID SUPPORTED BY?
4. Computer: THE BOX.
5. Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
6. Computer: FOUR OF THEM.
7. Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
8. Computer: YES, THE RED CUBE.
9. Person: IS IT SUPPORTED?
10. Computer: YES, BY THE TABLE.

Winograd sentences

1. The trophy doesn't fit into the brown suitcase because it's too **small**. What is too small?
The suitcase / The trophy
2. The trophy doesn't fit into the brown suitcase because it's too **large**. What is too large?
The suitcase / **The trophy**
3. The council refused the demonstrators a permit because they **feared** violence. Who **feared** violence?
The council / The demonstrators
4. The council refused the demonstrators a permit because they **advocated** violence. Who **advocated** violence?
The council / **The demonstrators**

Winograd 1972; Levesque 2013; Wang et al. 2018

Situated word learning

Children learn word meanings

1. with incredible speed
2. despite relatively few inputs
3. by using cues from
 - ▶ contrast inherent in the forms they hear
 - ▶ social cues
 - ▶ assumptions about the speaker's goals
 - ▶ regularities in the physical environment.

Frank et al. 2012; Frank and Goodman 2014




Consequences for NLU

1. Human children are the best agents in the universe at learning language, and they depend heavily on grounding.
2. Problems that are intractable without grounding are solvable with the right kinds of grounding.
3. Deep learning is a flexible toolkit for reasoning about different kinds of information in a single model, so it's led to conceptual and empirical improvements in this area.
4. We should seek out (and develop) data sets that include the right kind of grounding.

Speakers: From the world to language

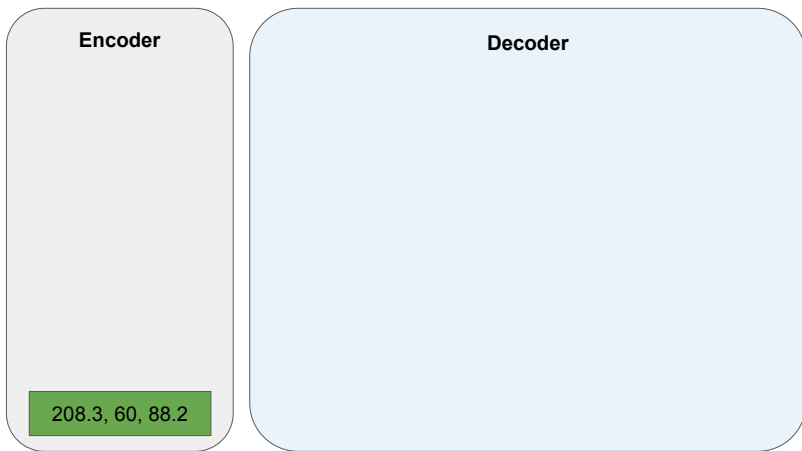
1. Overview: linguistic insights, and a bit of history
- 2. Speakers: From the world to language**
3. Assignment/Bake-off overview: Speakers in context
4. Listeners: From language to the world
5. Reasoning about other minds
6. The Rational Speech Acts model (RSA)
7. Neural RSA
8. Grounded chat bots
9. A few other grounding ideas

Color describer: Task formulation and data

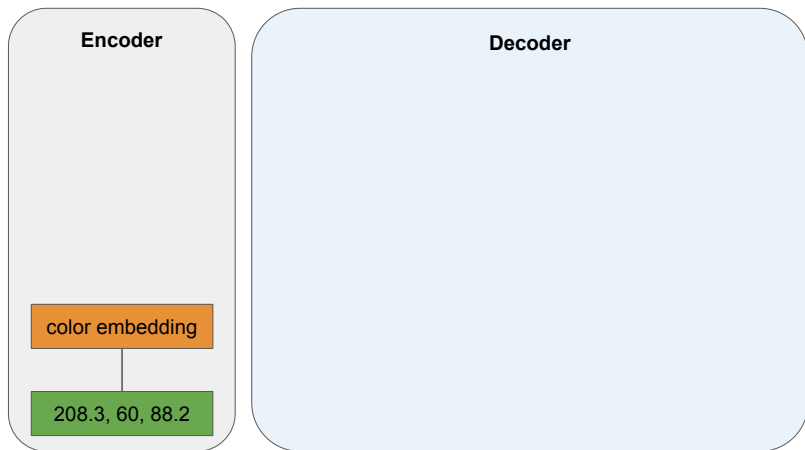
Color	Utterance
	green
	purple
	grape
	turquoise
	moss green
	pinkish purple
	light blue grey
	robin's egg blue
	british racing green
	baby puke green

McMahan and Stone 2015

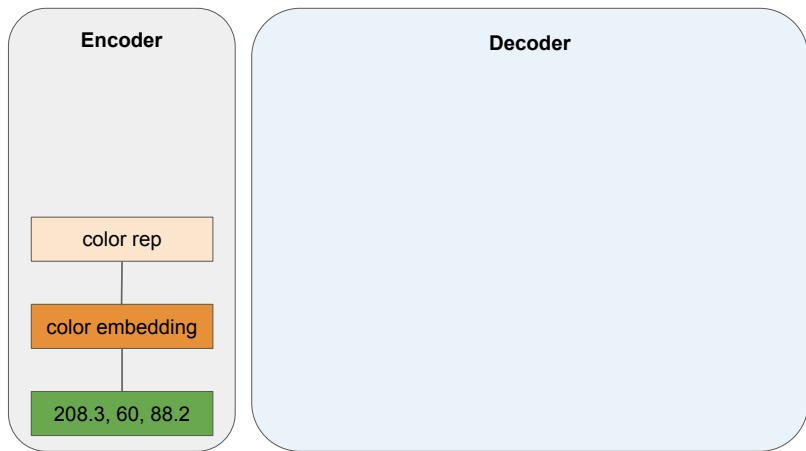
Color describer: Training with *teacher forcing*



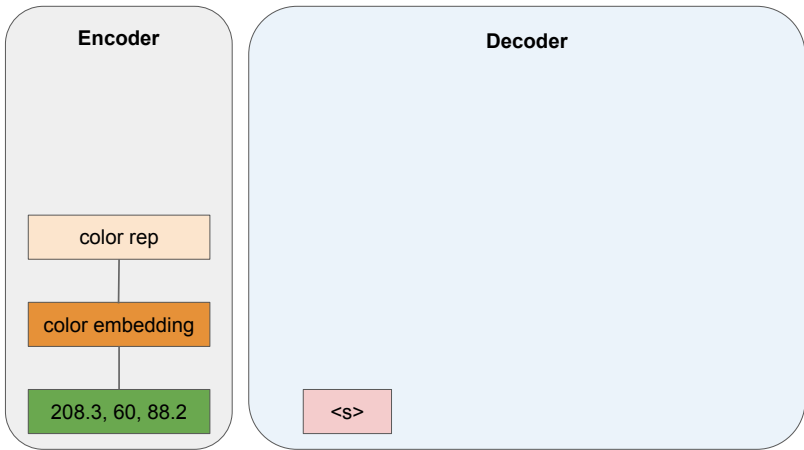
Color describer: Training with *teacher forcing*



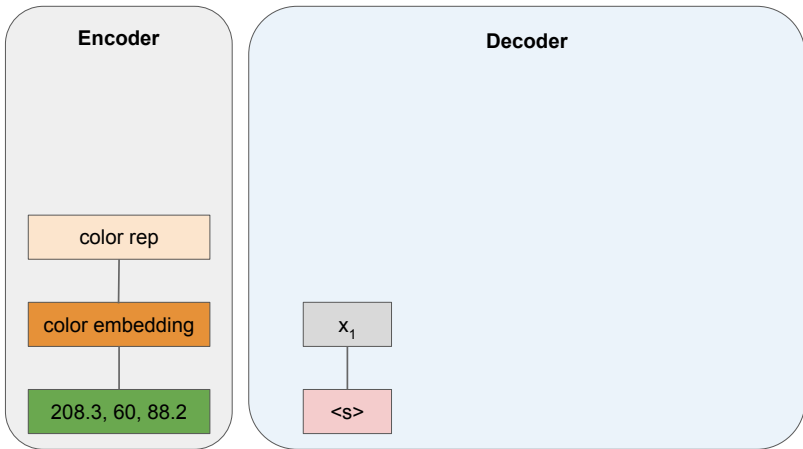
Color describer: Training with *teacher forcing*



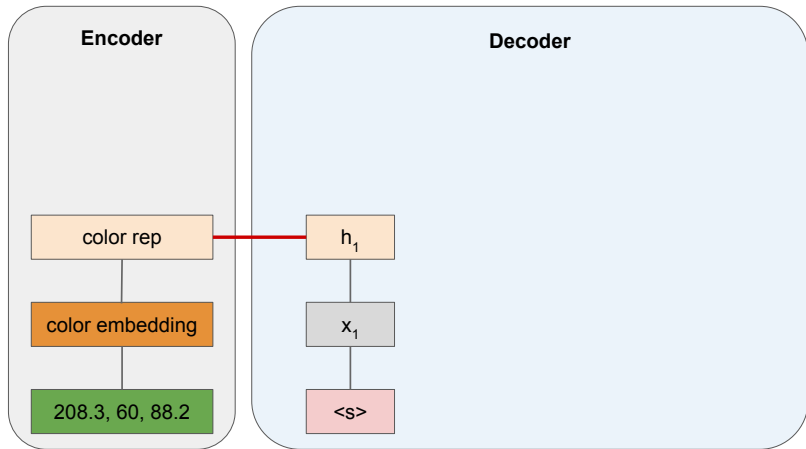
Color describer: Training with *teacher forcing*



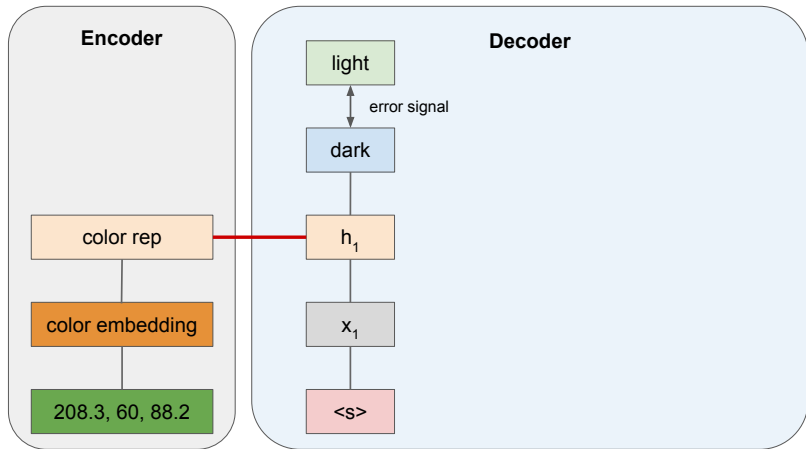
Color describer: Training with *teacher forcing*



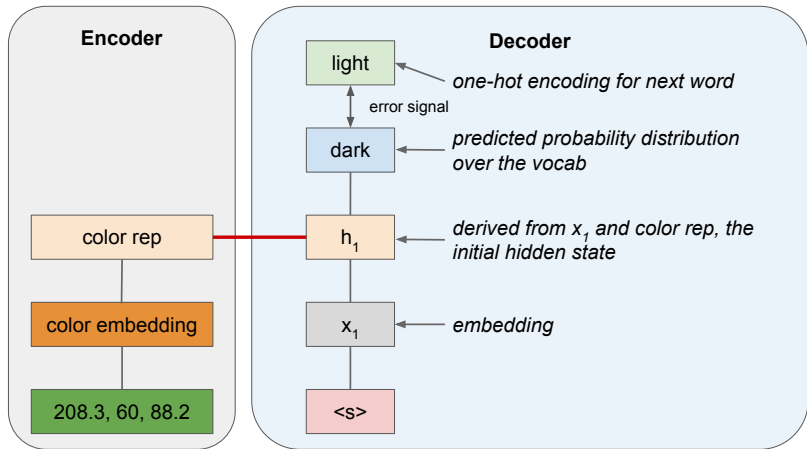
Color describer: Training with *teacher forcing*



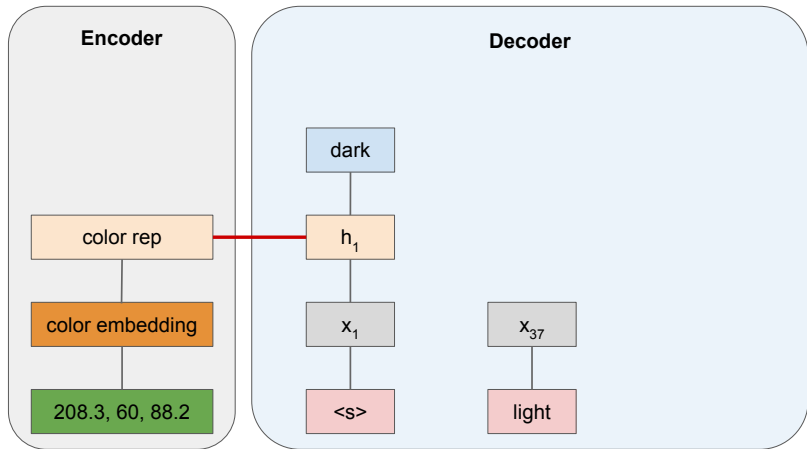
Color describer: Training with *teacher forcing*



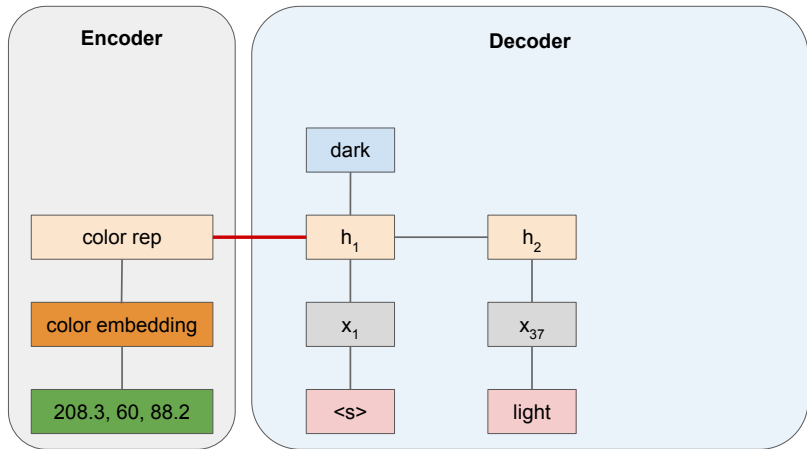
Color describer: Training with *teacher forcing*



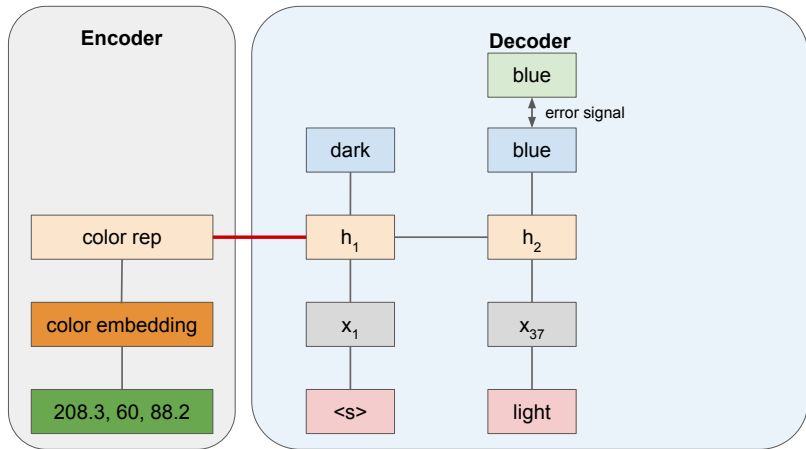
Color describer: Training with *teacher forcing*



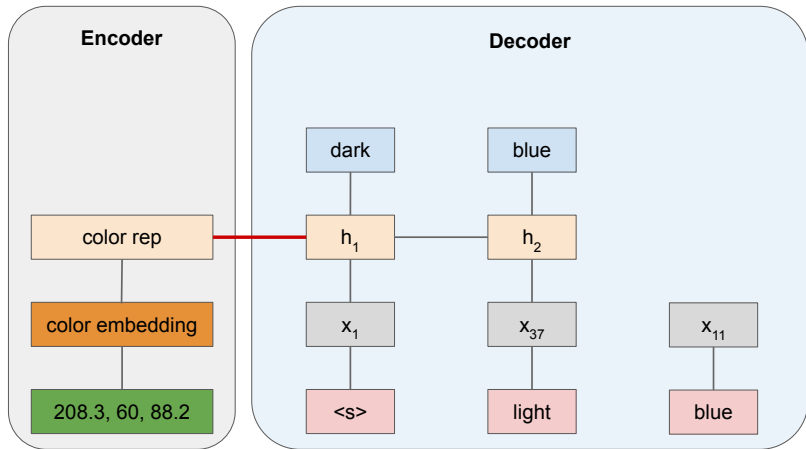
Color describer: Training with *teacher forcing*



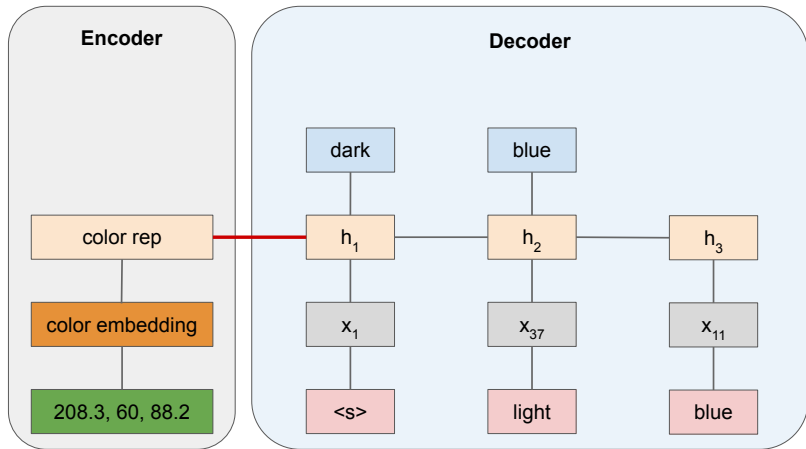
Color describer: Training with *teacher forcing*



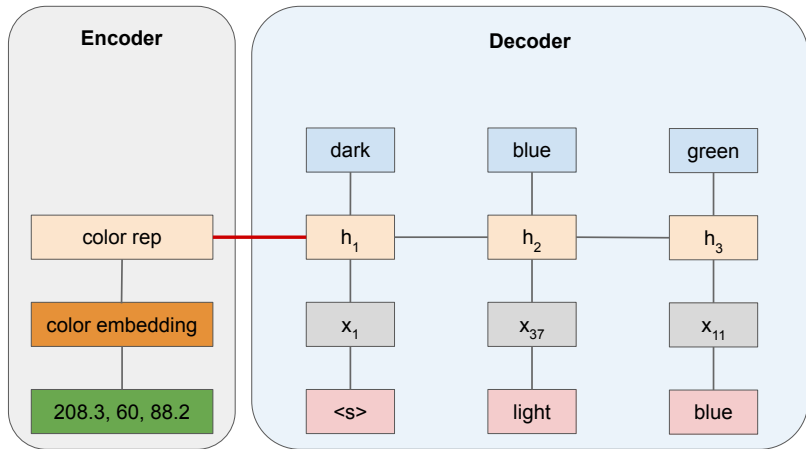
Color describer: Training with *teacher forcing*



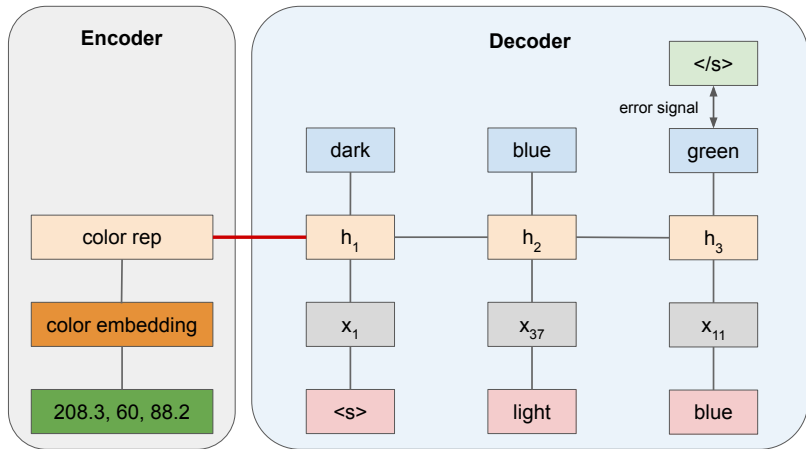
Color describer: Training with *teacher forcing*



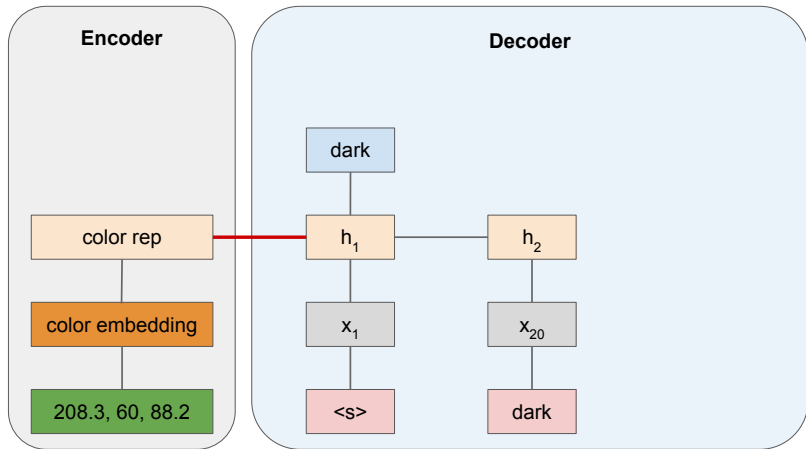
Color describer: Training with *teacher forcing*



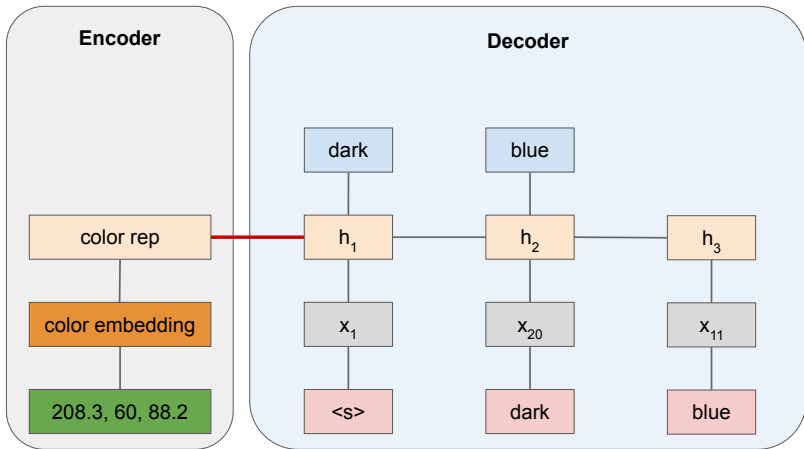
Color describer: Training with *teacher forcing*



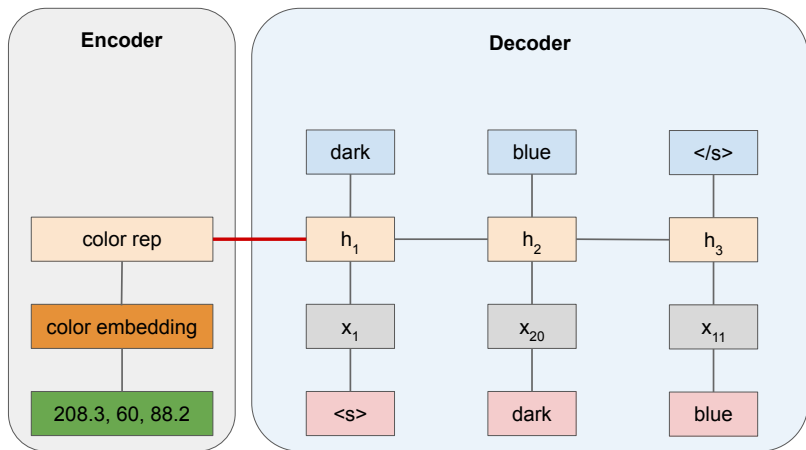
Color describer: Prediction



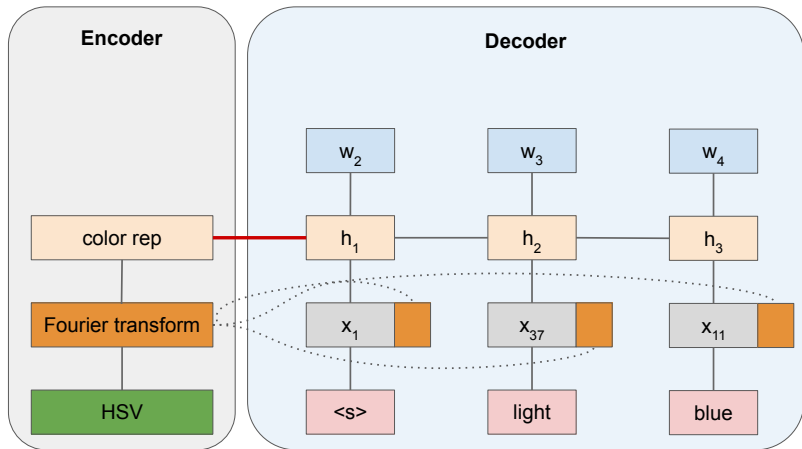
Color describer: Prediction



Color describer: Prediction



Color describer of Monroe et al. (2016)



Related tasks


Non-linguistic representation \Rightarrow Language

- Image captioning
- Scene description
- Visual Question Answering
(Image + Question-text \Rightarrow Answer-text)
- Instruction giving (State \Rightarrow Language)
- ...

Overview of the assignment and bake-off

1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
- 3. Assignment/Bake-off overview: Speakers in context**
4. Listeners: From language to the world
5. Reasoning about other minds
6. The Rational Speech Acts model (RSA)
7. Neural RSA
8. Grounded chat bots
9. A few other grounding ideas

Color descriptions in context

Color	Utterance
	green
	purple
	grape
	turquoise
	moss green
	pinkish purple
	light blue grey
	robin's egg blue
	british racing green
	baby puke green

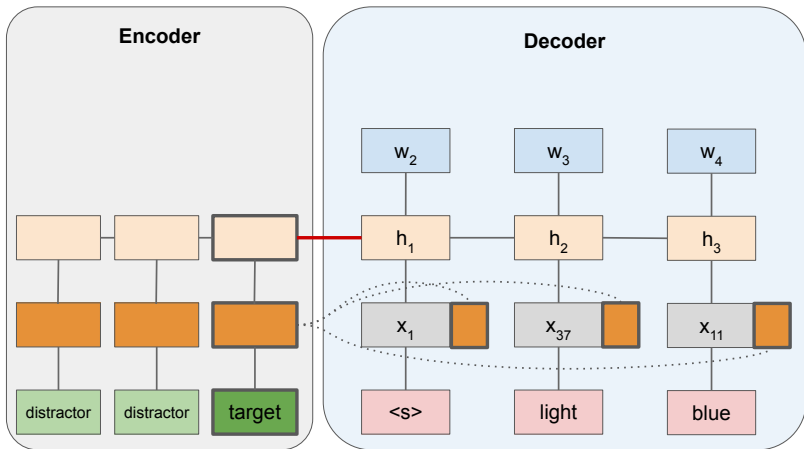
McMahan and Stone 2015

Color descriptions in context

	Context		Utterance
			blue
			The darker blue one
			teal not the two that are more green
			dull pink not the super bright one
			not any of the regular greens
			Purple
			blue

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

Colors in context (Monroe et al. 2017)



Data overview

```
[1]: from colors import ColorsCorpusReader
import os

[2]: COLORS_SRC_FILENAME = os.path.join("data", "colors", "filteredCorpus.csv")

[3]: corpus = ColorsCorpusReader(COLORS_SRC_FILENAME, normalize_colors=True)

[4]: examples = list(corpus.read())

[5]: len(examples)

[5]: 46994
```

More details: [colors_overview.ipynb](#)

Data overview

```
[6]: examples[0].display()
```

The darker blue one



```
[7]: examples[0].colors
```

```
[8]: [[0.7861111111111111, 0.5, 0.87],
      [0.6888888888888889, 0.5, 0.92],
      [0.6277777777777778, 0.5, 0.81]]
```

```
[8]: examples[0].contents
```

```
[8]: 'The darker blue one'
```

More details: [colors_overview.ipynb](#)

Data overview

```
[9]: print("Condition type:", examples[1].condition)
      examples[1].display()
```

Condition type: far
purple



More details: [colors_overview.ipynb](#)

Data overview

```
[10]: print("Condition type:", examples[3].condition)
      examples[3].display()
```

Condition type: split
lime



More details: [colors_overview.ipynb](#)

Data overview

```
[11]: print("Condition type:", examples[2].condition)
```

```
examples[2].display()
```

Condition type: close

Medium pink ### the medium dark one



More details: [colors_overview.ipynb](#)

Task-oriented evaluation

Predictions

For a given context c , let C be all of its permutations. Then a speaker model P_S predicts:

$$c^* = \operatorname{argmax}_{c \in C} P_S(\text{msg} | c)$$

Task-oriented evaluation

Predictions

For a given context c , let C be all of its permutations. Then a speaker model P_S predicts:

$$c^* = \operatorname{argmax}_{c \in C} P_S(\text{msg} \mid c)$$

Accuracy

A speaker model P_S is correct in its prediction about c iff $c^*[-1]$ is the target.

Task-oriented evaluation

Predictions

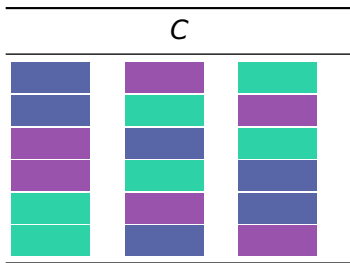
For a given context c , let C be all of its permutations. Then a speaker model P_S predicts:

$$c^* = \operatorname{argmax}_{c \in C} P_S(\text{msg} | c)$$

Example

$\text{msg} = \text{"blue"}$

$c =$ 



Task-oriented evaluation

Predictions

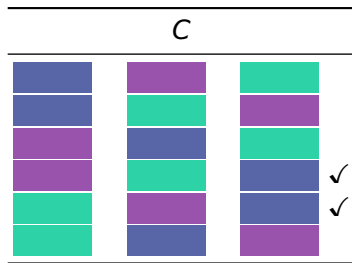
For a given context c , let C be all of its permutations. Then a speaker model P_S predicts:

$$c^* = \operatorname{argmax}_{c \in C} P_S(\text{msg} \mid c)$$

Example

$\text{msg} = \text{"blue"}$

$c =$ 



Question 2: Improve the color representations

```
[ ]: def represent_color_context(colors):  
    # Improve me!  
    return [represent_color(color) for color in colors]  
  
def represent_color(color):  
    # Improve me!  
    return color
```


Question 3: GloVe embeddings

```
[ ]: def create_glove_embedding(vocab, glove_base_filename='glove.6B.50d.txt'):

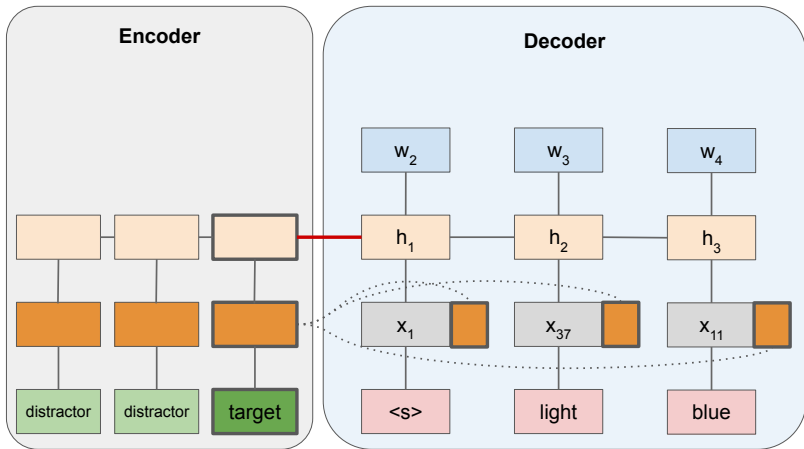
    # Use `utils.glove2dict` to read in the GloVe file:
    ##### YOUR CODE HERE

    # Use `utils.create_pretrained_embedding` to create the embedding.
    # This function will, by default, ensure that START_TOKEN,
    # END_TOKEN, and UNK_TOKEN are included in the embedding.
    ##### YOUR CODE HERE

    # Be sure to return the embedding you create as well as the
    # vocabulary returned by `utils.create_pretrained_embedding`,
    # which is likely to have been modified from the input `vocab`.

    ##### YOUR CODE HERE
```

Question 4: Color context



Question 4: Color context

1. Modify Decoder so that the input vector to the model at each timestep is not just a token representaton x but the concatenation of x with the representation of the target color.
2. Modify EncoderDecoder to extract the target colors and feed them to the decoder.
3. Modify ContextualColorDescriber so that it uses your new Decoder and EncoderDecoder.

Use the toy dataset generator for development!

Original system and bake-off

Our expectation for how you'll work:

1. Iteratively improve answers to the assignment questions.
2. Perhaps extend your modified Encoder/Decoder classes to do interesting new things.
3. Any data you can find is fine for your development work.
4. The bake-off will involve a new test set that has never been released anywhere before:
 - ▶ Same kinds of color context as in the released corpus.
 - ▶ One-off games rather than iterated.
 - ▶ All items listener-validated.

Listeners: From language to the world

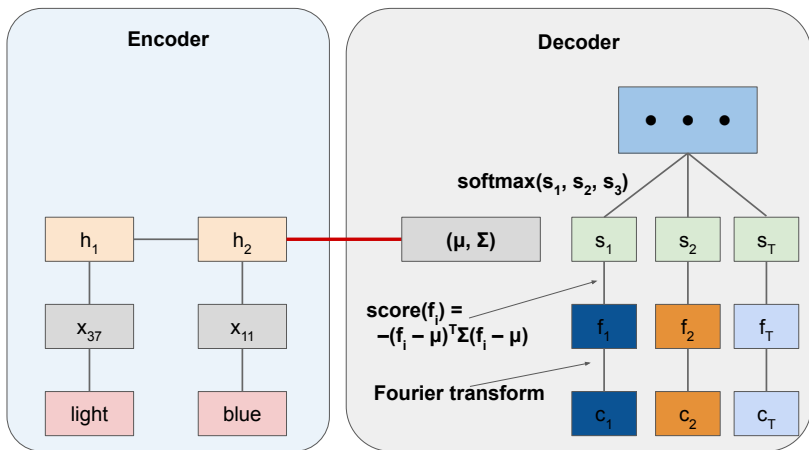
1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Assignment/Bake-off overview: Speakers in context
- 4. Listeners: From language to the world**
5. Reasoning about other minds
6. The Rational Speech Acts model (RSA)
7. Neural RSA
8. Grounded chat bots
9. A few other grounding ideas

Color interpreter: Task formulation and data

	Context		Utterance
			blue
			The darker blue one
			teal not the two that are more green
			dull pink not the super bright one
			not any of the regular greens
			Purple
			blue

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

A neural listener model



Other ideas and datasets

- **NLU classifiers** are very simple listeners: they consume language and make an inference in a structured space.
- **Semantic parsers** are very complex listeners: they consume language, construct rich latent representations, and predict into structured output spaces.
- **Scene generation** is the task of mapping language to structured representations of visual scenes (Seversky and Yin 2006; Chang et al. 2014, 2015).
- Young et al. (2014) seek to learn visual denotations for linguistic expressions.
- Mei et al. (2015) develop essentially a seq2seq version of the above model: given a linguistic input, they predict action sequences. (Kai Sheng Tai did his 2015 CS224u project on this, working at the same time as Mei et al.!)
- Suhr et al. (2019): Released the CerealBar data and game engine for learning to execute instructions.

Reasoning about other minds

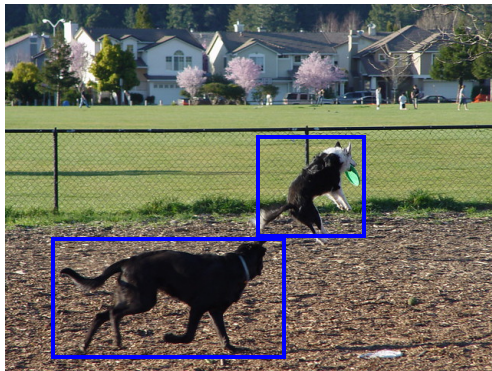
1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Assignment/Bake-off overview: Speakers in context
4. Listeners: From language to the world
- 5. Reasoning about other minds**
6. The Rational Speech Acts model (RSA)
7. Neural RSA
8. Grounded chat bots
9. A few other grounding ideas

Discriminative image labeling



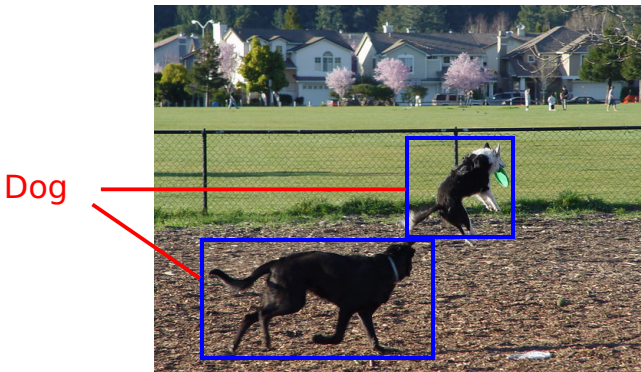
Mao et al. 2016

Discriminative image labeling



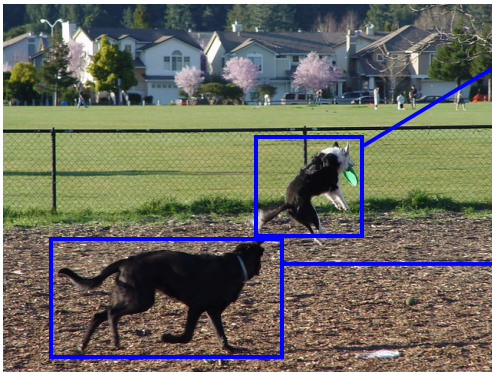
Mao et al. 2016

Discriminative image labeling



Mao et al. 2016

Discriminative image labeling

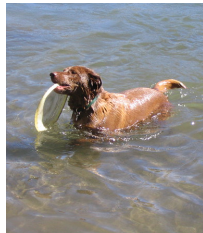


A little dog jumping and catching a frisbee

A big dog running

Mao et al. 2016

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.

Sports Champion advances in tournament.

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

Williams wobbled on Thursday.

Ongoing work with Hanson Lu and Reuben Cohn-Gordon

Machine translation

She chopped up the tree.



Elle coupa l'arbre.

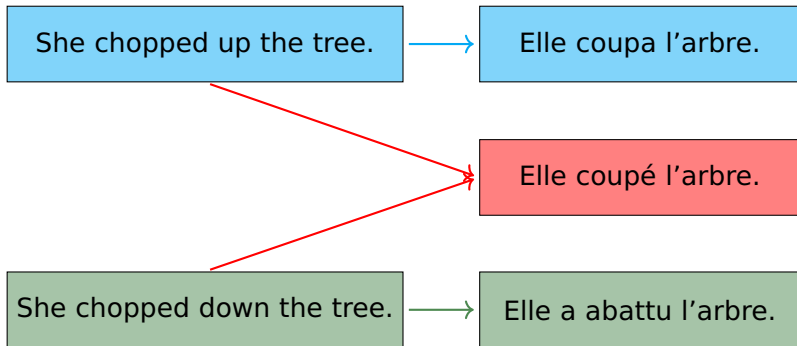
She chopped down the tree.



Elle a abattu l'arbre.

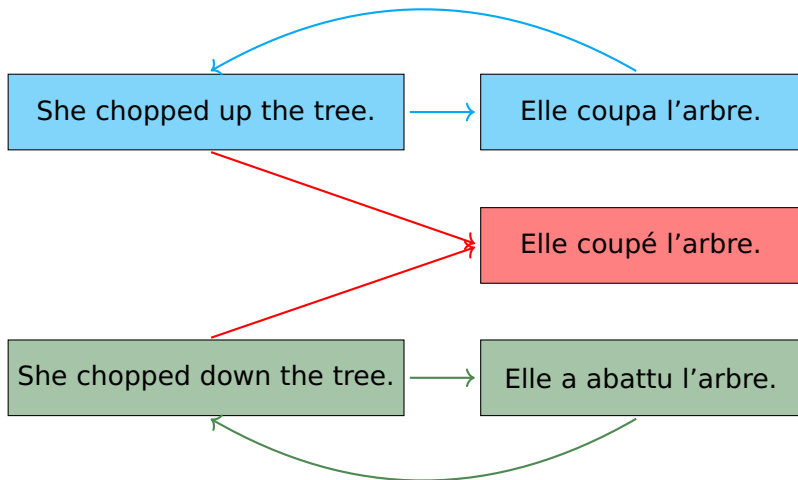
Cohn-Gordon and Goodman 2019

Machine translation



Cohn-Gordon and Goodman 2019

Machine translation



Cohn-Gordon and Goodman 2019

Generating and following instructions

Behavior



(a)

Base Speaker

walk forward four times

Rational Speaker

go forward four segments to the intersection with the bare concrete hall

Instruction

walk along the blue carpet and you pass two objects

(b)

Base Listener

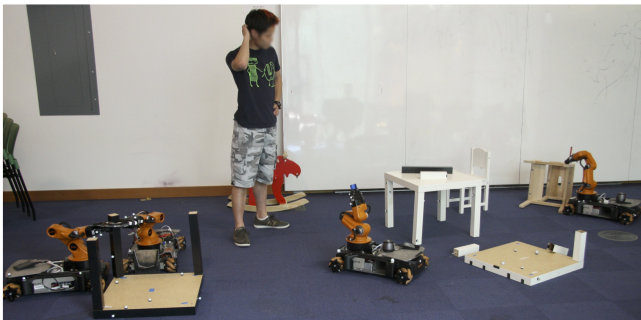


Rational Listener



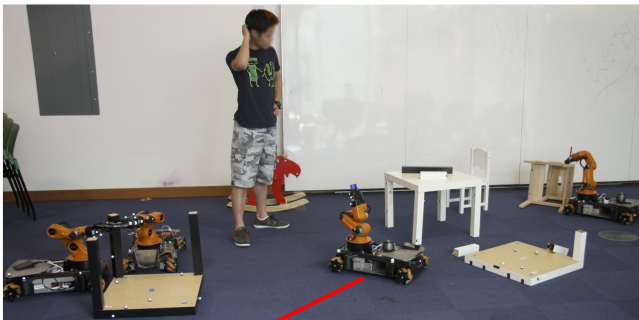
Fried et al. 2018

Collaborative problem solving



Tellex et al. 2014

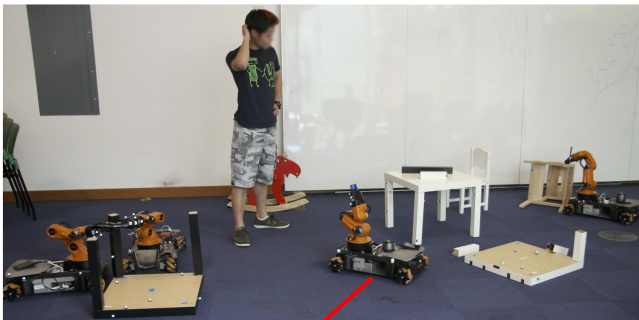
Collaborative problem solving



Help me!

Tellex et al. 2014

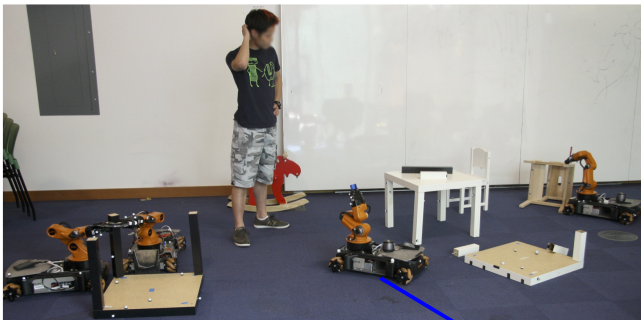
Collaborative problem solving



Hand me
the leg

Tellex et al. 2014

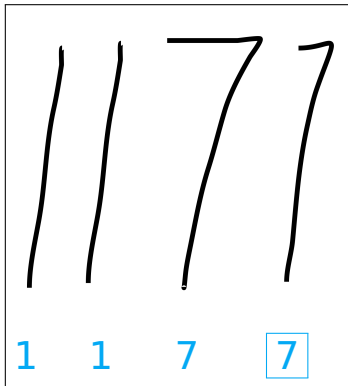
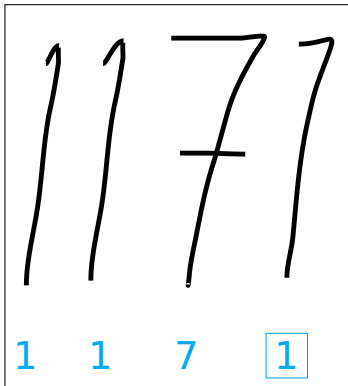
Collaborative problem solving



Hand me the white
leg on the table

Tellex et al. 2014

Optical character recognition



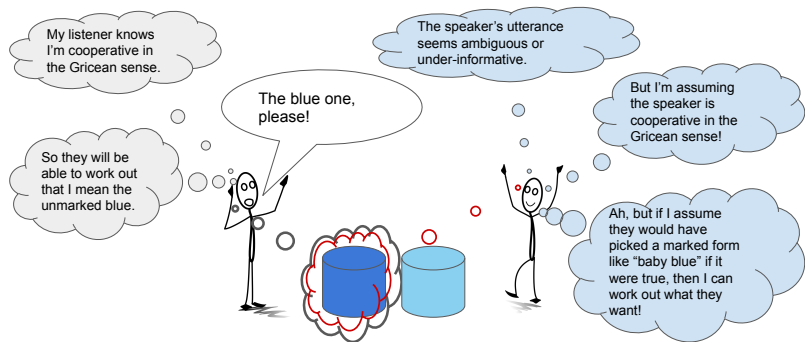
The Rational Speech Acts model

1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Assignment/Bake-off overview: Speakers in context
4. Listeners: From language to the world
5. Reasoning about other minds
- 6. The Rational Speech Acts model (RSA)**
7. Neural RSA
8. Grounded chat bots
9. A few other grounding ideas

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 See also [Goodman and Stuhlmüller 2013](#).

Pragmatic reasoning à la Grice (1975)



Pragmatic listeners

Pragmatic listeners

Literal listener

$$L_{\text{lit}}(\text{state} \mid \text{msg}) = \frac{\llbracket \text{msg}, \text{state} \rrbracket P(\text{state})}{\sum_{\text{state}'} \llbracket \text{msg}, \text{state}' \rrbracket P(\text{state}')}$$

Pragmatic listeners

Pragmatic speaker

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

Literal listener

$$L_{\text{lit}}(\text{state} \mid \text{msg}) = \frac{\llbracket \text{msg}, \text{state} \rrbracket P(\text{state})}{\sum_{\text{state}'} \llbracket \text{msg}, \text{state}' \rrbracket P(\text{state}')}$$

Pragmatic listeners

Pragmatic listener

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

Pragmatic speaker

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

Literal listener

$$L_{\text{lit}}(\text{state} | \text{msg}) = \frac{\llbracket \text{msg}, \text{state} \rrbracket P(\text{state})}{\sum_{\text{state}'} \llbracket \text{msg}, \text{state}' \rrbracket P(\text{state}')}$$

Pragmatic listeners

Pragmatic listener

$$L_{\text{prag}}(\textit{state} | \textit{msg}) = \mathbf{\textit{pragmatic speaker}} \times \textit{state prior}$$

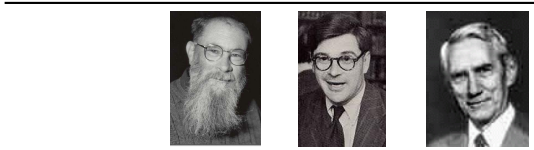
Pragmatic speaker

$$S_{\text{prag}}(\textit{msg} | \textit{state}) = \mathbf{\textit{literal listener}} - \textit{message costs}$$

Literal listener

$$L_{\text{lit}}(\textit{state} | \textit{msg}) = \mathbf{\textit{lexicon}} \times \textit{state prior}$$

A simple example



<i>beard</i>	1	0	0
<i>glasses</i>	1	1	0
<i>tie</i>	0	1	1

L_{prag}
 S_{prag}
 L_{lit}
[·]



A simple example



<i>beard</i>	1	0	0
<i>glasses</i>	.5	.5	0
<i>tie</i>	0	.5	.5

L_{prag}
 S_{prag}
 L_{lit}
 $[[\cdot]]$

A simple example

	<i>beard</i>	<i>glasses</i>	<i>tie</i>
	.67	.33	0
	0	.5	.5
	0	0	1

L_{prag}

S_{prag}

L_{lit}

$[[\cdot]]$

A simple example



<i>beard</i>	1	0	0
<i>glasses</i>	.4	.6	0
<i>tie</i>	0	.33	.67

L_{prag}
 S_{prag}
 L_{lit}
 $[[\cdot]]$

Pragmatic speakers

Pragmatic speakers

Literal speaker

$$S_{\text{lit}}(\text{msg} \mid \text{state}) = \frac{\exp(\alpha(\log\llbracket \text{msg}, \text{state} \rrbracket - C(\text{msg})))}{\sum_{\text{msg}'} \exp(\alpha(\log\llbracket \text{msg}', \text{state} \rrbracket - C(\text{msg}')))}$$

Pragmatic speakers

Pragmatic listener

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

Literal speaker

$$S_{\text{lit}}(msg | state) = \frac{\exp(\alpha(\log\llbracket msg, state \rrbracket - C(msg)))}{\sum_{msg'} \exp(\alpha(\log\llbracket msg', state \rrbracket - C(msg')))}$$

Pragmatic speakers

Pragmatic speaker

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

Pragmatic listener

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

Literal speaker

$$S_{\text{lit}}(\text{msg} \mid \text{state}) = \frac{\exp(\alpha(\log \llbracket \text{msg}, \text{state} \rrbracket - C(\text{msg})))}{\sum_{\text{msg}'} \exp(\alpha(\log \llbracket \text{msg}', \text{state} \rrbracket - C(\text{msg}'))))$$

Pragmatic speakers

Pragmatic speaker

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

Pragmatic listener

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

Literal speaker

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

Joint inference

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

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

Limitations

- Hand-specified lexicon
- Reasoning about *all* possible utterances?

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

- High-bias model; few chances to learn from data
- Cognitive demands limit speaker rationality
- Speaker preferences
- Scalability

Neural RSA

1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Assignment/Bake-off overview: Speakers in context
4. Listeners: From language to the world
5. Reasoning about other minds
6. The Rational Speech Acts model (RSA)
- 7. Neural RSA**
8. Grounded chat bots
9. A few other grounding ideas

Papers employing these techniques

- Andreas and Klein 2016
- Fried et al. 2018
- Monroe et al. 2017
- Monroe et al. 2018

Motivation

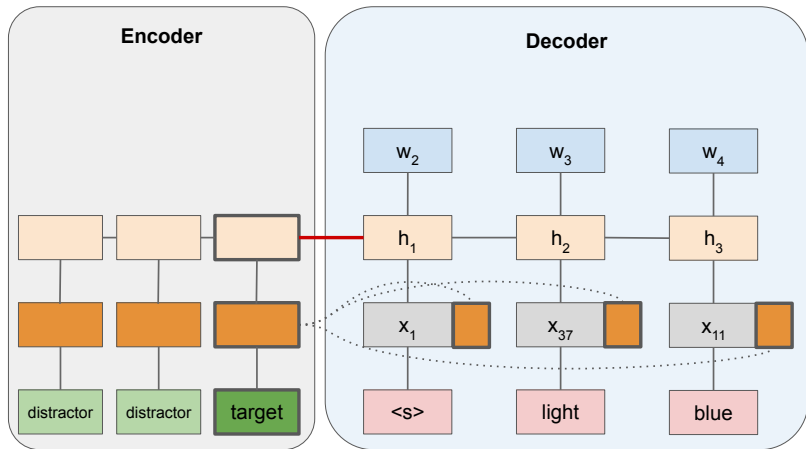
- Discriminative image labeling
- Image captioning
- Summarization
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition
- Scalability
- Sensitivity to variation
- Bounded rationality
- New kinds of model assessment
- Impact

Colors in context

	Context		Utterance
			blue
			The darker blue one
			teal not the two that are more green
			dull pink not the super bright one
			not any of the regular greens
			Purple
			blue

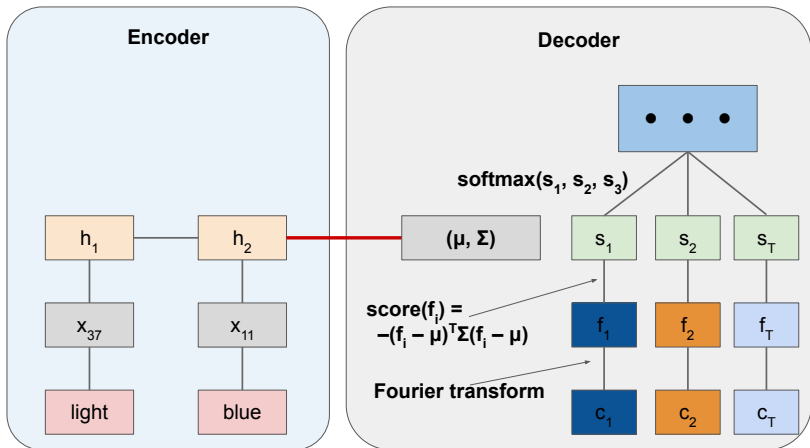
Stanford Colors in Context corpus
(Monroe et al. 2017)

Literal neural speaker S_{lit}^θ



Monroe et al. 2017

Neural literal listener L_0^θ



Monroe et al. 2017

Neural pragmatic agents

Neural pragmatic speaker (Andreas and Klein 2016)

$$s_{\text{prag}}^\theta(msg | state) = \frac{L_0^\theta(state | msg)}{\sum_{msg' \in X} L_0^\theta(state | msg')}$$

with X a sample from $s_{\text{lit}}^\theta(msg | state)$ such that $msg \in X$.

Neural pragmatic listener

$$L_1^\theta(state | msg) \propto s_{\text{prag}}^\theta(msg | state)$$

Blended neural pragmatic listener

Weighted combination of L_0^θ and L_1^θ .

Pragmatic image captioning

Mao et al. (2016); Vedantam et al. (2017): Captions that are true *and distinguish their images from related ones*.



S_0 caption: the dog is brown
 S_1 caption: the head of a dog

Reasoning about *all* possible utterances/captions?

⇒ Sample from S_{lit}^θ

⇒ **Full RSA reasoning about characters**

(Cohn-Gordon et al. 2018, 2019)

Other related work

- Golland et al. (2010): Recursive speaker/listener reasoning as part of interpreting complex utterances compositionally, with grounding in a simple visual world.
- Tellex et al.'s (2014) Inverse Semantics: Robot utterances are scored by models similar to RSA's pragmatic speakers.
- Wang et al. (2016): Pragmatic reasoning helps in online learning of semantic parsers.
- Monroe and Potts (2015): "RSA as a hidden activation function"
- Khani et al. (2018): Collaborative games with pragmatic reasoning.
- Cohn-Gordon and Goodman (2019): RSA for translation

Introspective speakers from Google

Generation and Comprehension of Unambiguous Object Descriptions

Junhua Mao^{2*} Jonathan Huang¹ Alexander Toshev¹ Oana Camburu³ Alan Yuille^{2,4} Kevin Murphy¹
¹Google Inc. ²University of California, Los Angeles ³University of Oxford ⁴Johns 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 Vedantam¹ Samy Bengio² Kevin Murphy² Devi Parikh³ Gal Chechik²
¹Virginia Tech ³Georgia Institute of Technology ²Google
¹vrama91@vt.edu ³parikh@gatech.edu ²{bengio,kpmurphy,gal}@google.com

Mao et al. 2016; Vedantam et al. 2017

Google Refexp Dataset



~~S: "A backpack"~~

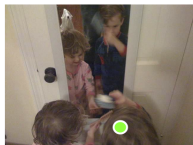
Listener

S: "A yellow and black backpack"

Listener

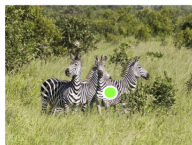
Mao et al. 2016

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.



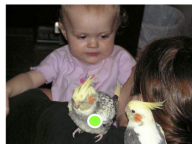
Zebra looking towards the camera.

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.

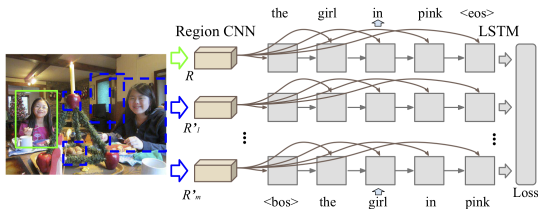
Mao et al. 2016

Maximum Mutual Information Training

Neural listener objective

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

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



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 $ent \in I$:

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

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.

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

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

$$\operatorname{argmax}_{msg} \left[\lambda \log \mathbf{s}_{lit}^{\theta}(msg | state; I_n) + \right. \\ \left. (1 - \lambda) \log \frac{\mathbf{s}_{lit}^{\theta}(msg | state; I_n)}{\mathbf{s}_{lit}^{\theta}(msg | state'; I_n)} \right]$$

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

Diagnosing the role of introspection

Target image and class

Rufous Hummingbird



Justifications vary with λ

fully discriminative

- $\lambda = 0.00$ tarsals orange white brown wings wings orange tail dark an primaries
- $\lambda = 0.30$ This is a brown bird with a brown wing and a long pointy beak.
- $\lambda = 0.50$ This bird is **brown with red on its neck** and has a long , pointy beak.
- $\lambda = 0.70$ This is a bird with a **white belly , brown wing and a red throat.**
- $\lambda = 1.00$ 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

- $\lambda = 0.00$ tarsals orange white brown wings wings orange tail dark an primaries
- $\lambda = 0.30$ This is a brown bird with a brown wing and a long pointy beak.
- $\lambda = 0.50$ This bird is **brown with red on its neck** and has a long , pointy beak.
- $\lambda = 0.70$ This is a bird with a **white belly , brown wing and a red throat.**
- $\lambda = 1.00$ 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

$\lambda = 0.00$ tarsals orange white brown wings
wings orange tail dark an primaries

$\lambda = 0.30$ This is a brown bird with a brown
wing and a long pointy beak.

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

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

$\lambda = 1.00$ 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

$\lambda = 0.00$ tarsals orange white brown wings wings orange tail dark an primaries

$\lambda = 0.30$ This is a brown bird with a brown wing and a long pointy beak.

$\lambda = 0.50$ This bird is brown with red on its neck and has a long, pointy beak.

$\lambda = 0.70$ This is a bird with a white belly, brown wing and a red throat.

$\lambda = 1.00$ A small sized bird that has a very long and pointed bill.

context blind

Distractor class

Ruby throated Hummingbird



Vedantam et al. 2017

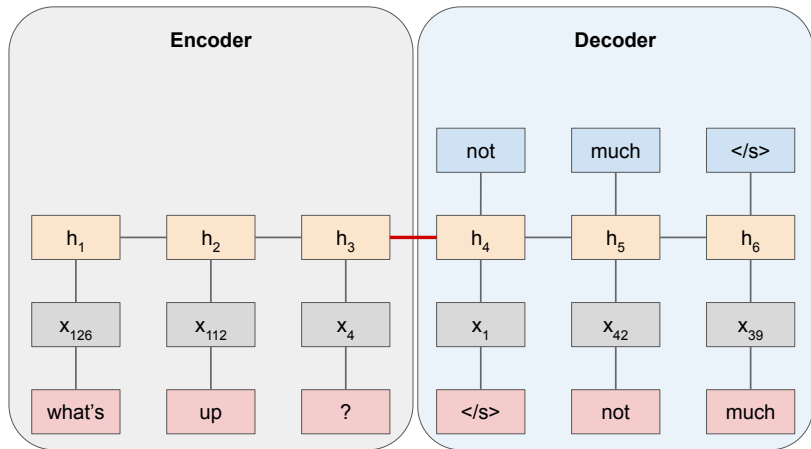
Other relevant datasets

- The TUNA Reference Corpus
<https://www.abdn.ac.uk/ncs/departments/computing-science/corpus-496.php>
- SCONE: Sequential CONTEXT-dependent Execution
<https://nlp.stanford.edu/projects/scone/>
- Crowdsource your own (Hawkins 2015)!
<https://github.com/hawkrobe/MWERT>

Grounded chat bots

1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Assignment/Bake-off overview: Speakers in context
4. Listeners: From language to the world
5. Reasoning about other minds
6. The Rational Speech Acts model (RSA)
7. Neural RSA
- 8. Grounded chat bots**
9. A few other grounding ideas

Basic neural chatbot






FAIR negotiation dataset

5,808 dialogues grounded in 2,236 unique scenarios.

Divide these objects between you and another Turker. Try hard to get as many points as you can!

Send a message now, or enter the agreed deal!

Items	Value	Number You Get
	8	<input type="text" value="1"/>
	1	<input type="text" value="1"/>
	0	<input type="text" value="0"/>

Fellow Turker: I'd like all the balls

You: Ok, if I get everything else

Fellow Turker: If I get the book then you have a deal

You: No way - you can have one hat and all the balls

Fellow Turker: Ok deal

Type Message Here:

Figure 1: A dialogue in our Mechanical Turk interface, which we used to collect a negotiation dataset.

From Lewis et al. 2017; see also Yarats and Lewis 2018

FAIR negotiation dataset

Perspective of YOU

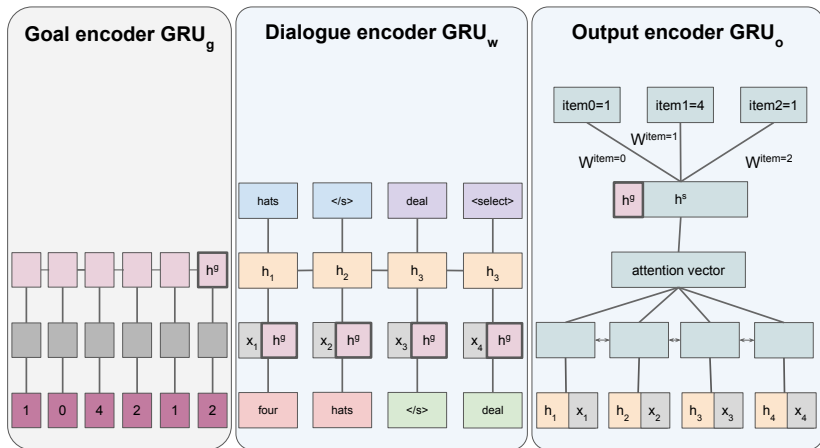
1. 1 0 4 2 1 2 # (1 book, worth 0; 4 hats, worth 2, 1 ball, worth 2)
2. YOU: i would like 4 hats and you can have the rest <eos>
 THEM: deal <eos>
 YOU: <selection>
3. item0=0 item1=4 item2=0
4. <eos>
5. reward=8
6. agree
7. 1 4 4 1 1 2

FAIR negotiation dataset

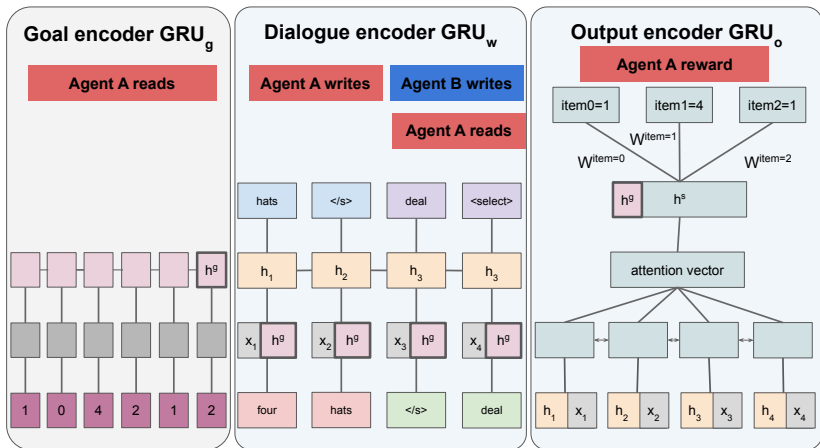
Perspective of THEM

1. 1 4 4 1 1 2 # (1 book, worth 4; 4 hats, worth 1, 1 ball, worth 2)
2. THEM: i would like 4 hats and you can have the rest <eos>
 YOU: deal <eos>
 THEM: <selection>
3. item0=1 item1=0 item2=1
4. <eos>
5. reward=6
6. agree
7. 1 0 4 2 1 2

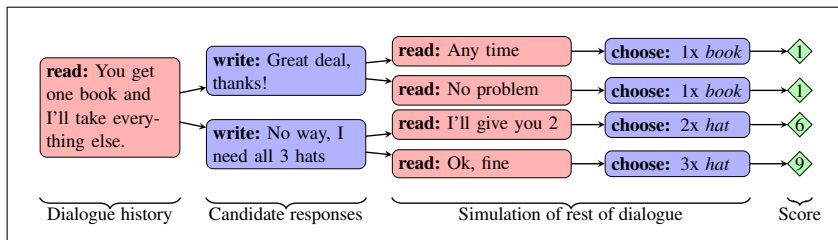
FAIR negotiation agents



Goal-based training



Decoding through rollouts



From Lewis et al. 2017, figure 4

Aside: An amusing media narrative

Lewis et al. (2017)

“During reinforcement learning, an agent *A* attempts to improve its parameters from conversations with another agent *B*. While the other agent *B* could be a human, in our experiments we used our fixed supervised model that was trained to imitate humans. The second model is fixed as we found that updating the parameters of both agents led to divergence from human language.”

Aside: An amusing media narrative

[FAIR blog post \[link\]](#)

“The second model is fixed, because the researchers found that updating the parameters of both agents led to divergence from human language as the agents developed their own language for negotiating.”

Aside: An amusing media narrative

Newsweek [\[link\]](#)

“The bots ran afoul of their Facebook overlords when they started to make up their own language to do things faster, not unlike the way football players have shorthand names for certain plays instead of taking the time in the huddle to describe where everyone should run. It’s not unusual for bots to make up a lingo that humans can’t comprehend, though it does stir worries that these things might gossip about us behind our back. Facebook altered the code to make the bots stick to plain English.”

Aside: An amusing media narrative

Tech Times [\[link\]](#)

“Facebook was forced to shut down one of its artificial intelligence systems after researchers discovered that it had started communicating in a language that they could not understand.

Aside: An amusing media narrative

Tech Times [\[link\]](#)

“Facebook was forced to shut down one of its artificial intelligence systems after researchers discovered that it had started communicating in a language that they could not understand.

“The incident evokes images of the rise of Skynet in the iconic Terminator series. Perhaps Tesla CEO Elon Musk is right about AI being the ‘biggest risk we face.’ ”

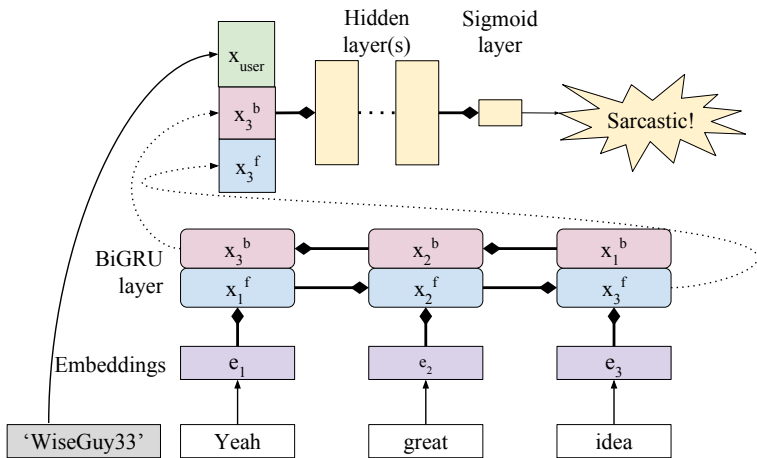
Other task-oriented dialogue datasets

- **Edinburgh Map Corpus**
<http://groups.inf.ed.ac.uk/maptask/>
- **TRIPS**
<http://www.cs.rochester.edu/research/cisd/projects/trips/>
- **TRAINS**
<http://www.cs.rochester.edu/research/cisd/projects/trains/>
- **Cards**
<http://CardsCorpus.christopherpotts.net/>
- **SCARE**
<http://slate.cse.ohio-state.edu/quake-corpora/scare/>
- **The Carnegie Mellon Communicator Corpus**
<http://www.speech.cs.cmu.edu/Communicator/>

A few other grounding ideas

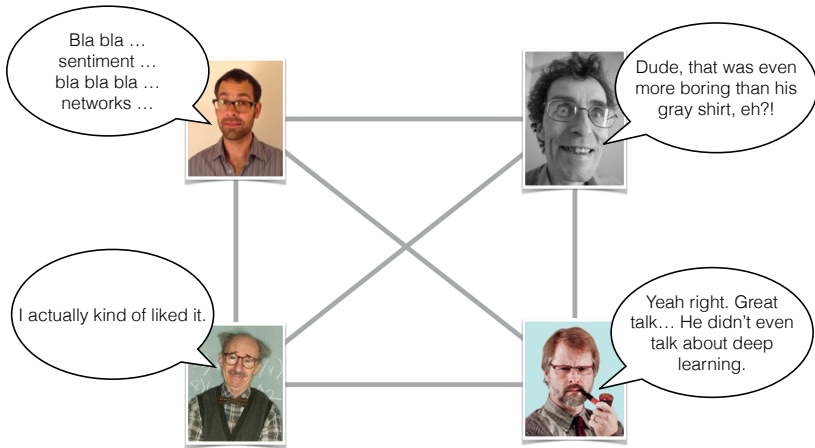
1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Assignment/Bake-off overview: Speakers in context
4. Listeners: From language to the world
5. Reasoning about other minds
6. The Rational Speech Acts model (RSA)
7. Neural RSA
8. Grounded chat bots
- 9. A few other grounding ideas**

Modeling users for sarcasm detection



SARC: Khodak et al. 2017;
Kolchinski and Potts 2018

NLU in social graphs with Probabilistic Soft Logic

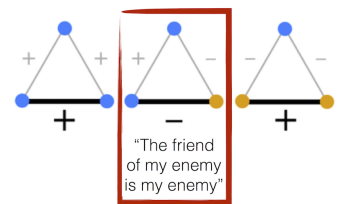


<https://psl.linqs.org;>
 West et al. 2014

NLU in social graphs with Probabilistic Soft Logic



Social balance theory



<https://psl.linqs.org;>
West et al. 2014

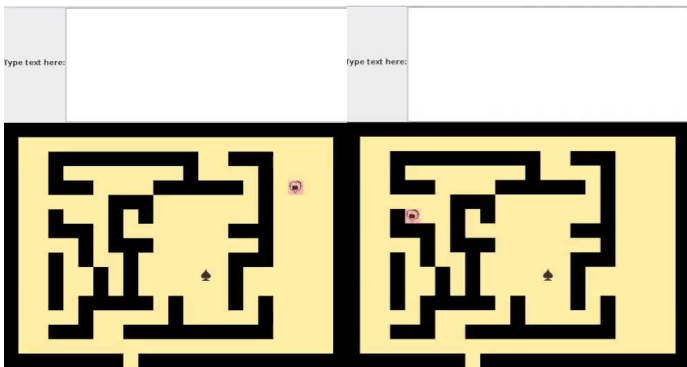
PLOW: Webpage structure as context

1. Learning rules of the form ‘If A, then B, else C’ is a challenge because the latent variable A is generally not observed. Rather, one sees only B or C.
2. In an interactive, instructional setting, one needn’t rely entirely on abduction or probabilistic inference: users generally state the needed rules during their interactions.
3. The user’s actions ground the parsed language.
4. The DOM structure grounds the user’s indexicals:
 - ▶ Put the name here. (user clicks on the DOM element)
 - ▶ This is the ISBN number. (user highlights some text)
 - ▶ Find another tab. (user has selected a tab)

Decision-theoretic agents



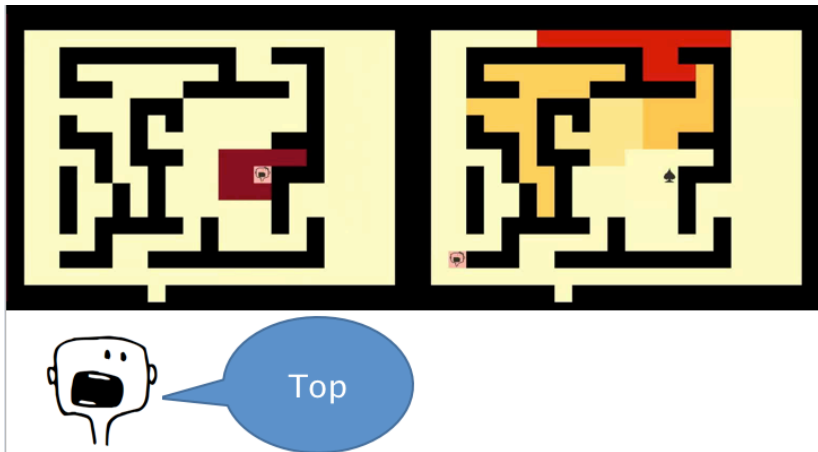
Both players must find the ace of spades. DialogBot:



Vogel et al. 2013a,b

Decision-theoretic agents

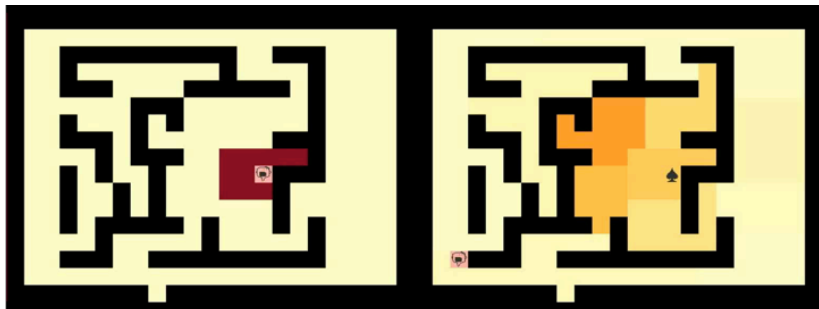
Baby DialogBots (a few hours of policy exploration)



Vogel et al. 2013a,b

Decision-theoretic agents

Grown-up DialogBots (a week of policy exploration)



Middle of the board

Frontiers

- Deeper integration with devices and the environment.
- More sophisticated reasoning about other agents and their goals.
- Better tracking of full dialogue history; improved discourse coherence.
- Approximate state representations to address very pressing scalability issues.

References I

- James F. Allen, Nathanael Chambers, George Ferguson, Lucian Galescu, Hyuckchul Jung, Mary Swift, and William Taysom. 2007. PLOW: A collaborative task learning agent. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, pages 1514–1519. AAAI Press, Vancouver, British Columbia, Canada.
- 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.
- 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.
- Angel Chang, Will Monroe, Manolis Savva, Christopher Potts, and Christopher D. Manning. 2015. Text to 3d scene generation with rich lexical grounding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 53–62, Stroudsburg, PA. Association for Computational Linguistics.
- Angel Chang, Manolis Savva, and Christopher D. Manning. 2014. [Learning spatial knowledge for text to 3D scene generation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2028–2038, Doha, Qatar. Association for Computational Linguistics.
- 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.
- Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998.
- Michael C. Frank and Noah D. Goodman. 2014. [Inferring word meanings by assuming that speakers are informative](#). *Cognitive Psychology*, 75(1):80–96.
- Michael C. Frank, Joshua B. Tenenbaum, and Anne Fernald. 2012. Social and discourse contributions to the determination of reference in cross-situational word learning. *Language, Learning, and Development*.
- Michael Franke. 2009. *Signal to Act: Game Theory in Pragmatics*. ILLC Dissertation Series. Institute for Logic, Language and Computation, University of Amsterdam.

References II

- 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.
- 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.
- Noah D. Goodman and Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1):173–184.
- H. Paul Grice. 1975. Logic and conversation. In Peter Cole and Jerry Morgan, editors, *Syntax and Semantics*, volume 3: Speech Acts, pages 43–58. Academic Press, New York.
- Robert X. D. Hawkins. 2015. [Conducting real-time multiplayer experiments on the web](#). *Behavior Research Methods*, 47(4):966–976.
- Gerhard Jäger. 2007. Game dynamics connects semantics and pragmatics. In Ahti-Veikko Pietarinen, editor, *Game Theory and Linguistic Meaning*, pages 89–102. Elsevier, Amsterdam.
- Fereshte Khani, Noah D. Goodman, and Percy Liang. 2018. [Planning, inference and pragmatics in sequential language games](#). *Transactions of the Association for Computational Linguistics*, 6:543–555.
- Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. 2017. A large self-annotated corpus for sarcasm. *arXiv preprint arXiv:1704.05579*.
- Y. Alex Kolchinski and Christopher Potts. 2018. Representing social media users for sarcasm detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1115–1121, Stroudsburg, PA. Association for Computational Linguistics.
- Hector J. Levesque. 2013. On our best behaviour. In *Proceedings of the Twenty-third International Conference on Artificial Intelligence*, Beijing.
- 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.
- Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? End-to-end learning for negotiation dialogues. *ArXiv:1706.05125*.
- 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.
- Brian McMahan and Matthew Stone. 2015. [A Bayesian model of grounded color semantics](#). *Transactions of the Association for Computational Linguistics*, 3:103–115.

References III

- Hongyuan Mei, Mohit Bansal, and Matthew R. Walter. 2015. Listen, attend, and walk: Neural mapping of navigational instructions to action sequences. [ArXiv:1506.04089](#).
- Will Monroe, Noah D. Goodman, and Christopher Potts. 2016. Learning to generate compositional color descriptions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2243–2248, Stroudsburg, PA. Association for Computational Linguistics.
- 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.
- Ofir Press and Lior Wolf. 2017. [Using the output embedding to improve language models](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 157–163, Valencia, Spain. Association for Computational Linguistics.
- 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.
- Lee M Seversky and Lijun Yin. 2006. Real-time automatic 3D scene generation from natural language voice and text descriptions. In *Proceedings of the 14th ACM International Conference on Multimedia*, pages 61–64. ACM.
- Alane Suhr, Claudia Yan, Jack Schluger, Stanley Yu, Hadi Khader, Marwa Mouallem, Iris Zhang, and Yoav Artzi. 2019. [Executing instructions in situated collaborative interactions](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2119–2130, Hong Kong, China. Association for Computational Linguistics.
- 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.

References IV

- 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.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. [GLUE: A multi-task benchmark and analysis platform for natural language understanding](#). In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, Brussels, Belgium. Association for Computational Linguistics.
- Sida I. Wang, Percy Liang, and Christopher D. Manning. 2016. [Learning language games through interaction](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2368–2378. Association for Computational Linguistics.
- Robert West, Hristo S. Paskov, Jure Leskovec, and Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2(2):297–310.
- Terry Winograd. 1972. Understanding natural language. *Cognitive Psychology*, 3(1):1–191.
- Terry Winograd. 1986. A procedural model of language understanding. In Barbara J. Grosz, Karen Sparck-Jones, and Bonnie Lynn Webber, editors, *Readings in Natural Language Processing*, pages 249–266. Morgan Kaufmann Publishers Inc., San Francisco.
- Denis Yarats and Mike Lewis. 2018. [Hierarchical text generation and planning for strategic dialogue](#). In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5587–5595, Stockholm, Sweden. PMLR.
- Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. 2014. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78.