Grounded language understanding

Christopher Potts

Stanford Linguistics

CS 224U: Natural language understanding
May 6

Overview

1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Listeners: From language to the world
4. Grounded chat bots
5. Reasoning about other minds
6. A few other grounding ideas

HAL

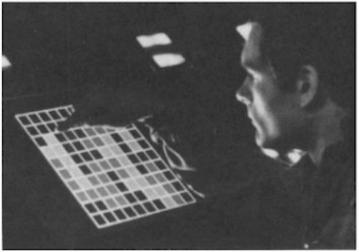
- In the 1967 Stanley Kubrick movie *2001: A Space Odyssey*, the spaceship's computer HAL can
  - display graphics;
  - play chess; and
  - conduct natural, open-domain conversations with humans.
- How well did the filmmakers do at predicting what computers would be capable in 2001?

(Slide idea from Andrew McCallum)

HAL

Graphics

HAL



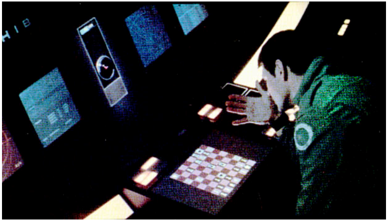
Jurassic Park (1993)

(Slide idea from Andrew McCallum)

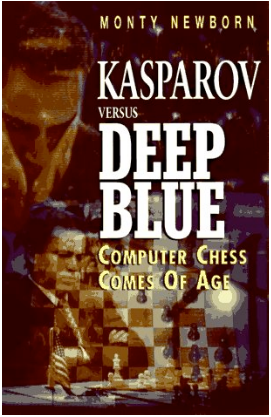
HAL

Chess

HAL



Deep Blue (1997)



(Slide idea from Andrew McCallum)

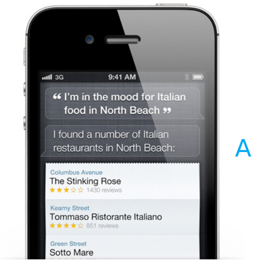
HAL

Dialogue

HAL	2014
David Bowman: Open the pod bay doors, HAL. HAL: I'm sorry, Dave, I'm afraid I can't do that. David: What are you talking about, HAL? HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.	. . .

(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 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 from Marie de Marneffe)

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."

Levinson's (2000) analogy

Figure 0.1
Rembrandt sketch

Levinson's (2000) analogy

Figure 0.1
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 [. . .].

Levinson’s (2000) analogy

Figure 0.1
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

Indexicality

1. I am speaking.

Indexicality

1. I am speaking.
2. We won. [A team I'm on; a team I support; . . .]

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; . . .]

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]

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]

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.

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)

Context dependence

Where are you from?

- *Connecticut.* (Issue: birthplaces)
- *The U.S.* (Issue: nationalities)

Context dependence

Where are you from?

- *Connecticut.* (Issue: birthplaces)
- *The U.S.* (Issue: nationalities)
- *Stanford.* (Issue: affiliations)

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

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.

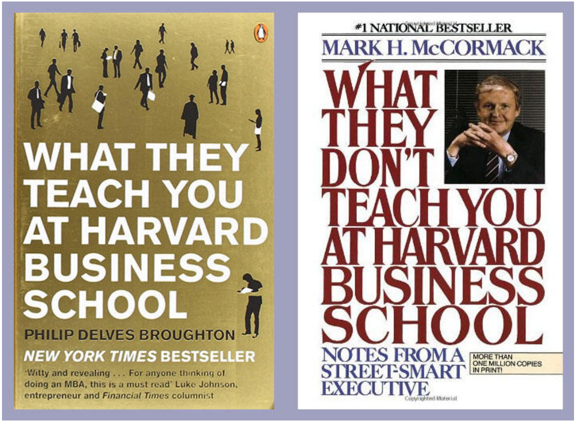
Seems true

Context dependence

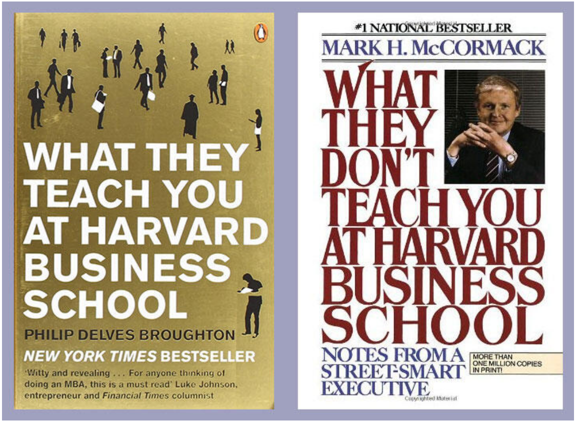
If kangaroos had no tails, they would fall over.

Seems true, but suppose they had jetpacks.

Context dependence



Context dependence



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

Perspectival expressions

Routine pragmatic enrichment

how big is the contextually restricted domain of students?

what's the additional contextual restriction?

[false for *most students*?]

who's the speaker?

Many students met with me yesterday.

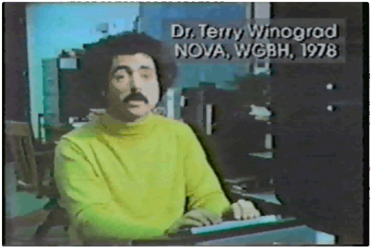
what's the time of utterance?

[but perhaps many met with the speaker at other times?]

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

(Winograd 1972; Levesque 2013; Wang et al. 2018)

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

(Winograd 1972; Levesque 2013; Wang et al. 2018)

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**

(Winograd 1972; Levesque 2013; Wang et al. 2018)

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

(Winograd 1972; Levesque 2013; Wang et al. 2018)

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 & Goodman (2014)

Consequences for NLU

- Human children are the best agents in the universe at learning language, and they depend heavily on grounding.
- Problems that are intractable without grounding are solvable with the right kinds of grounding.
- 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.
- We should seek out (and develop) data sets that include the right kind of grounding.

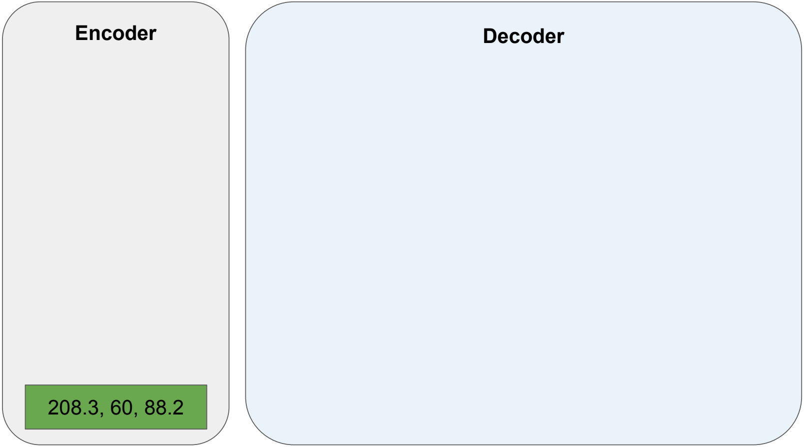
Speakers: From the world to language

Color describer: Task formulation and data

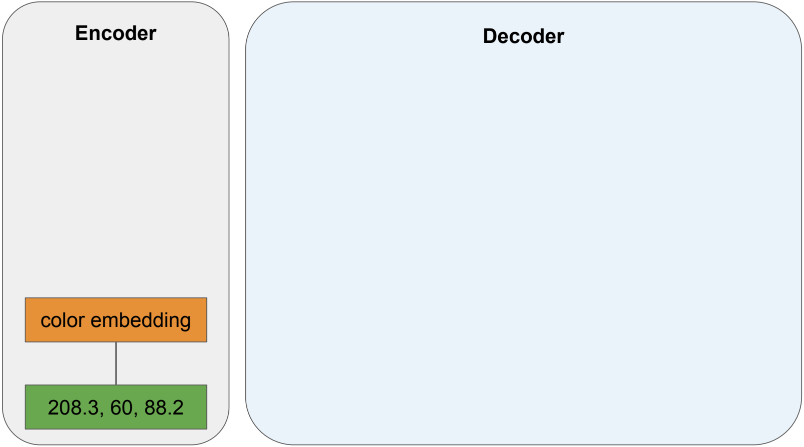
Color	Utterance
	green
	purple
	grape
	turquoise
	moss green
	pinkish purple
	light blue grey
	robin's egg blue
	british racing green
	baby puke green

Table: Example from the xkcd color dataset as released by McMahan & Stone (2015).

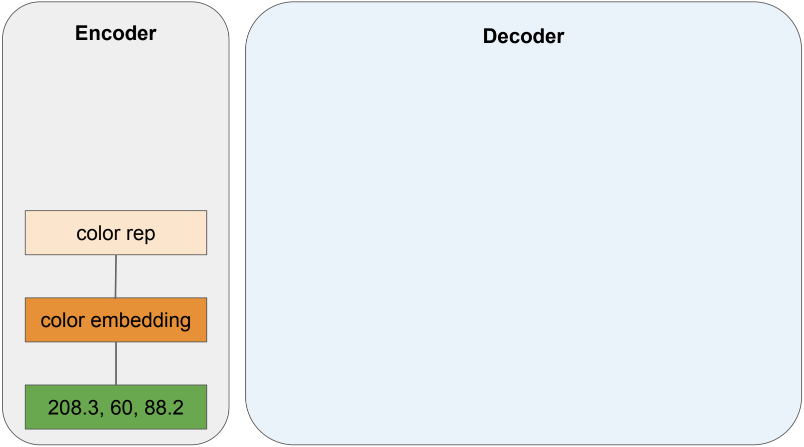
Color describer: Training with *teacher forcing*



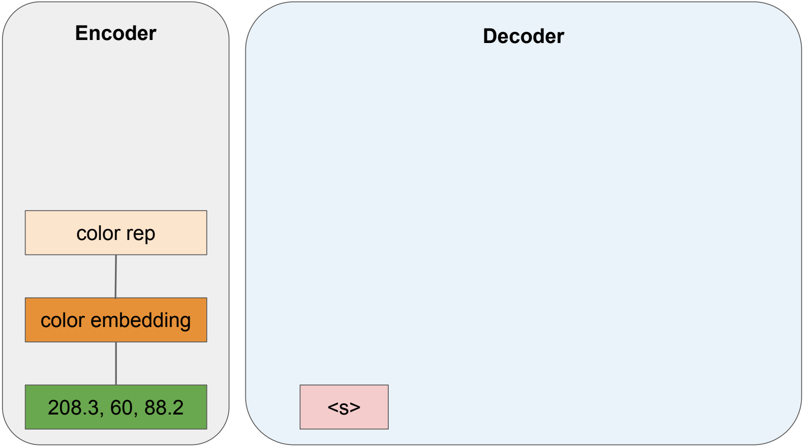
Color describer: Training with *teacher forcing*



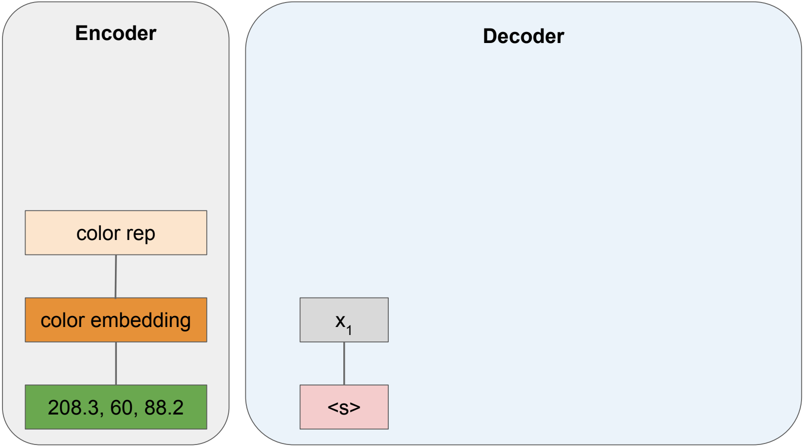
Color describer: Training with *teacher forcing*



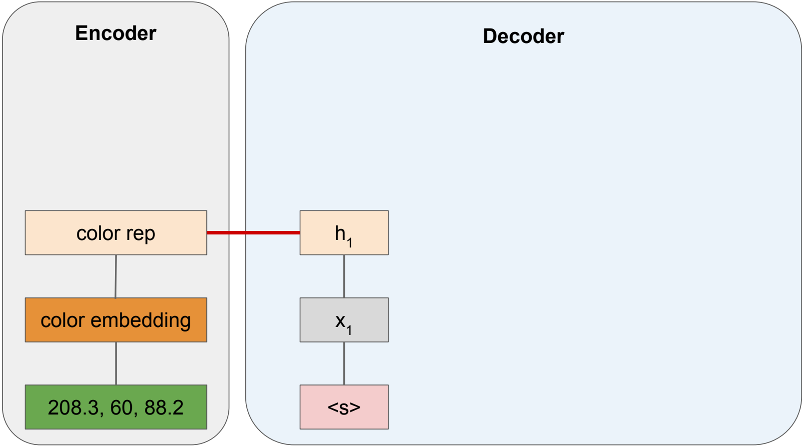
Color describer: Training with *teacher forcing*



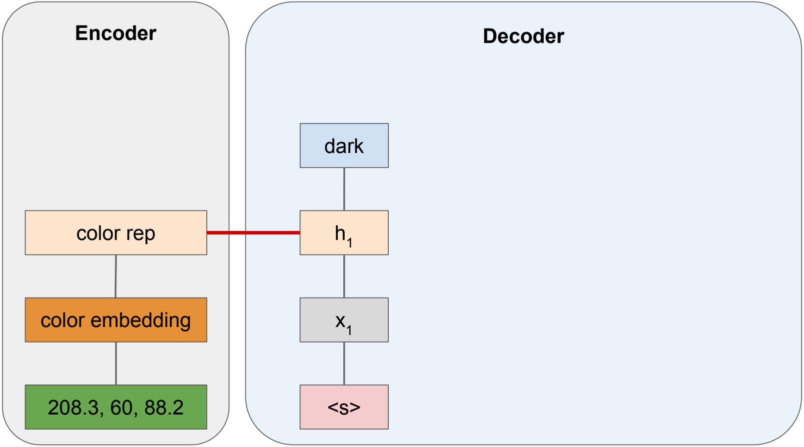
Color describer: Training with *teacher forcing*



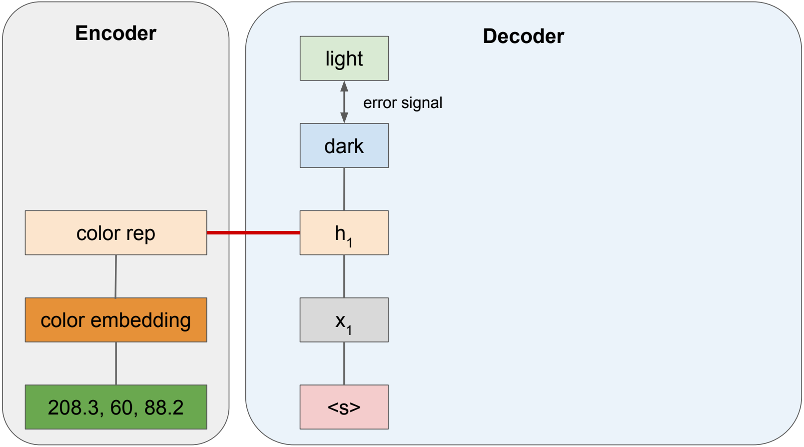
Color describer: Training with *teacher forcing*



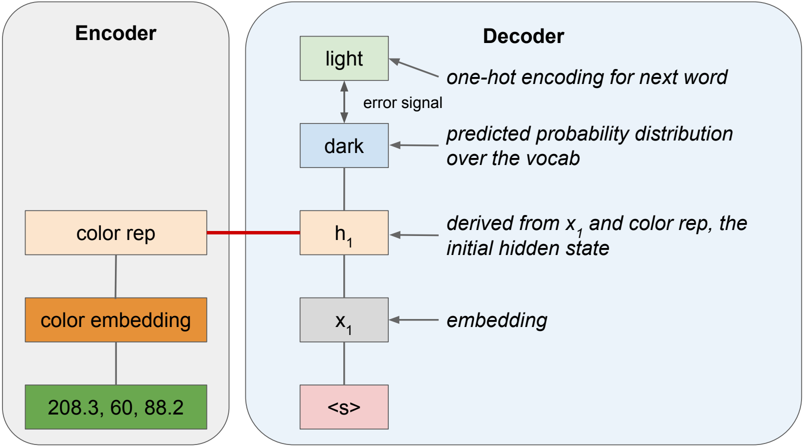
Color describer: Training with *teacher forcing*



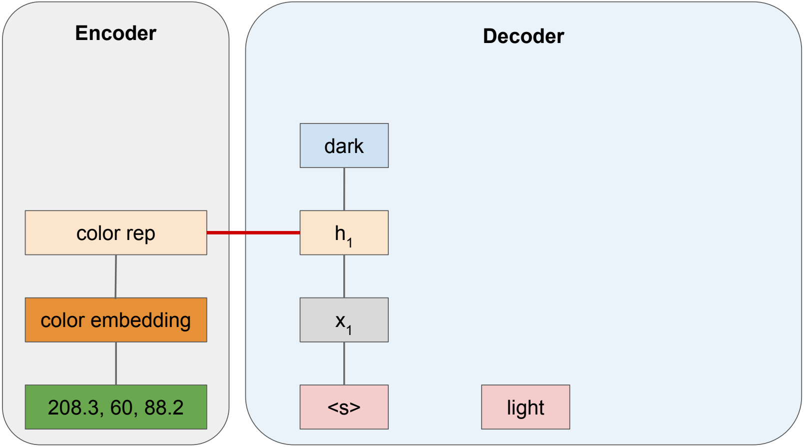
Color describer: Training with *teacher forcing*



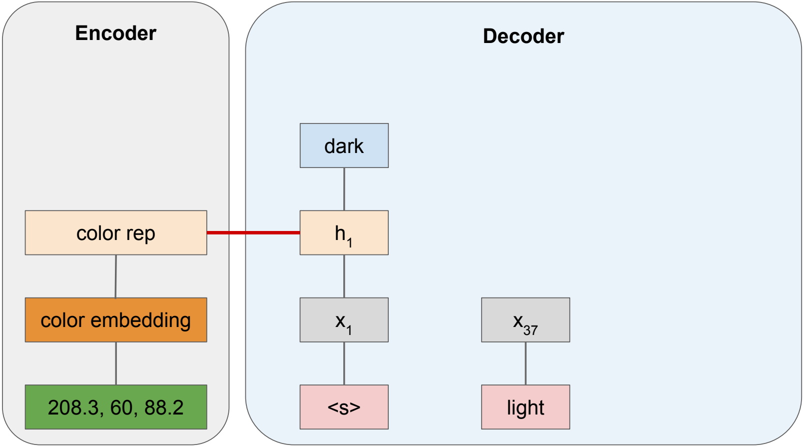
Color describer: Training with *teacher forcing*



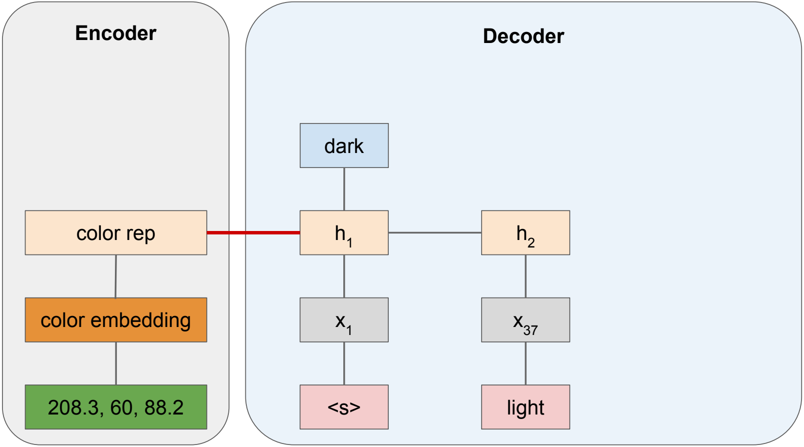
Color describer: Training with *teacher forcing*



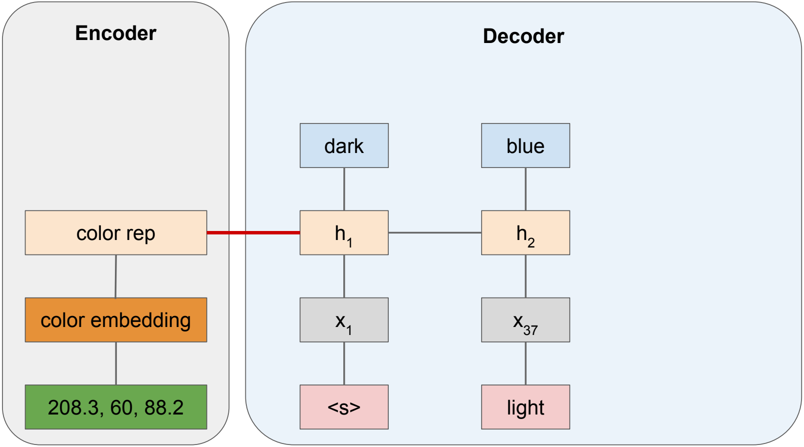
Color describer: Training with *teacher forcing*



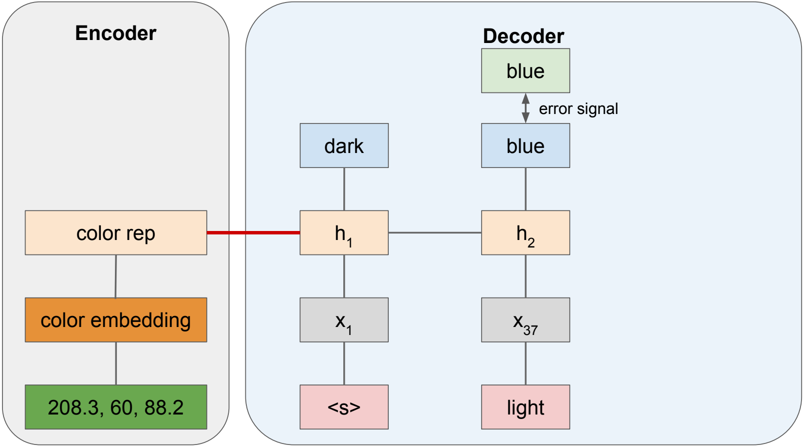
Color describer: Training with *teacher forcing*



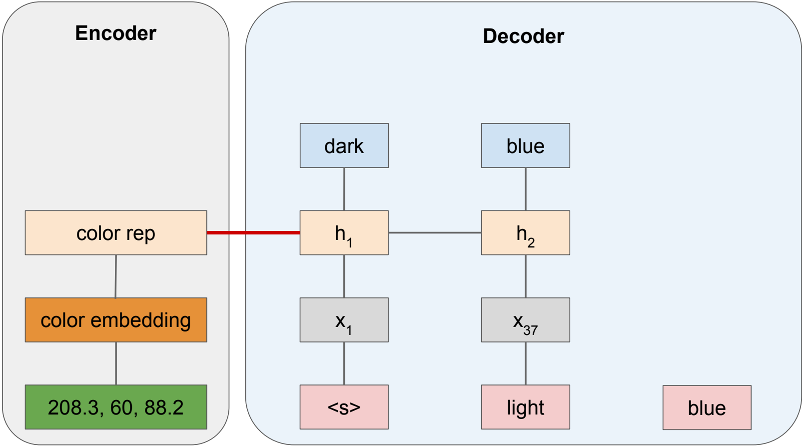
Color describer: Training with *teacher forcing*



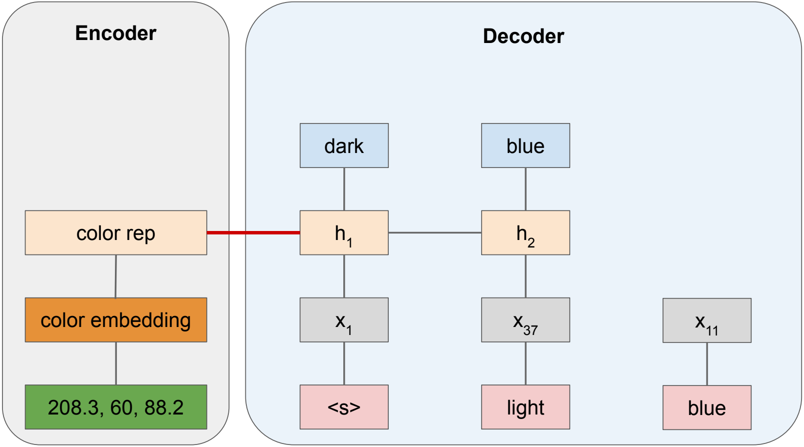
Color describer: Training with *teacher forcing*



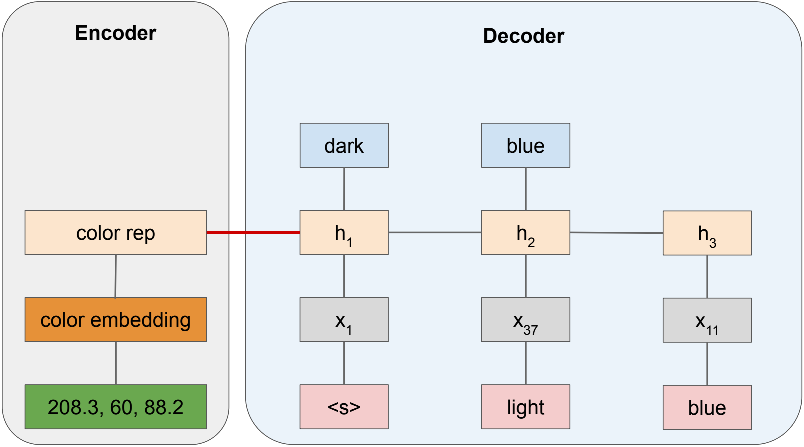
Color describer: Training with *teacher forcing*



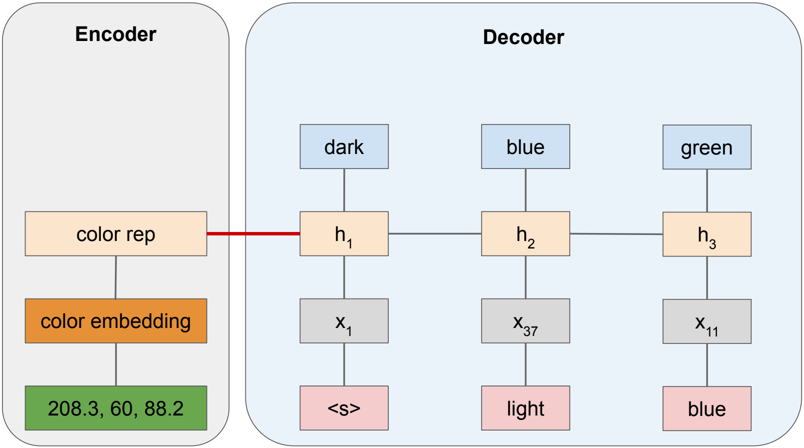
Color describer: Training with *teacher forcing*



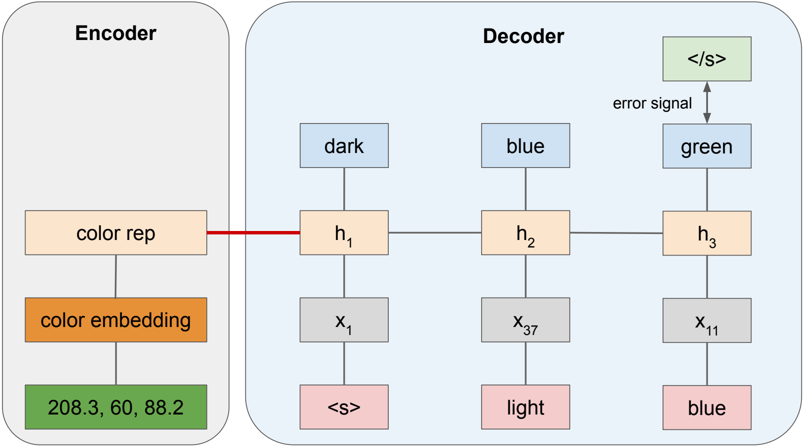
Color describer: Training with *teacher forcing*



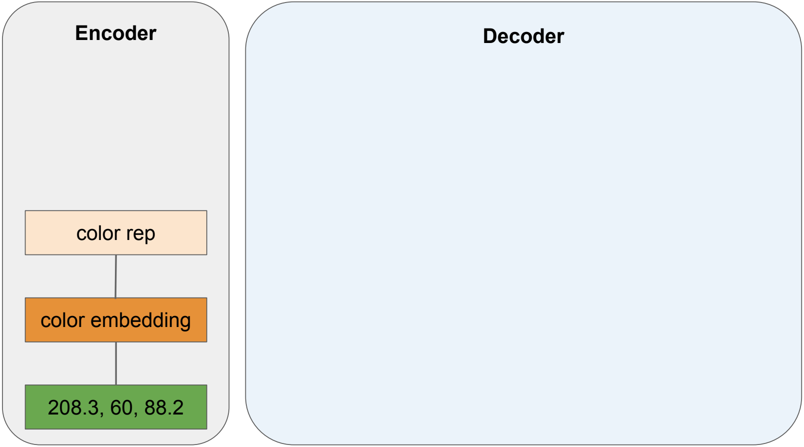
Color describer: Training with *teacher forcing*



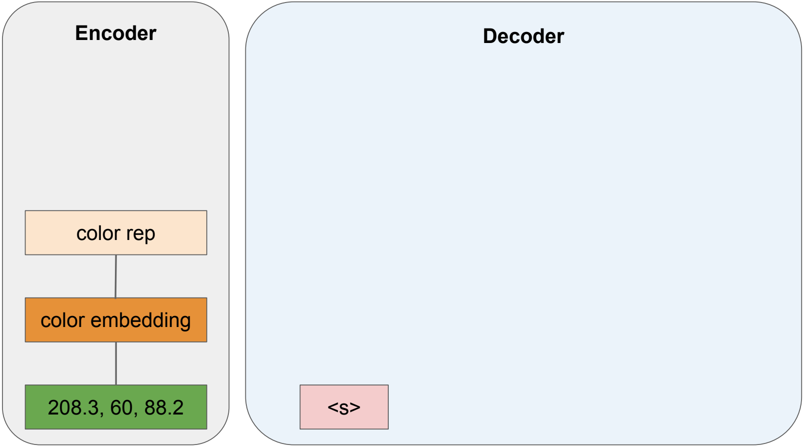
Color describer: Training with *teacher forcing*



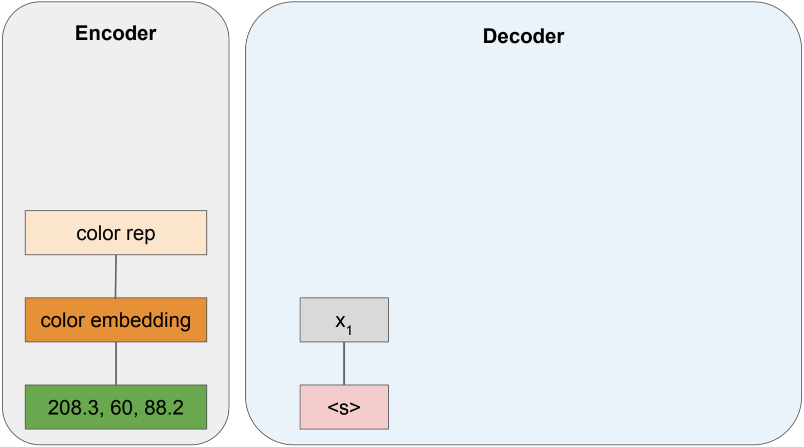
Color describer: Prediction



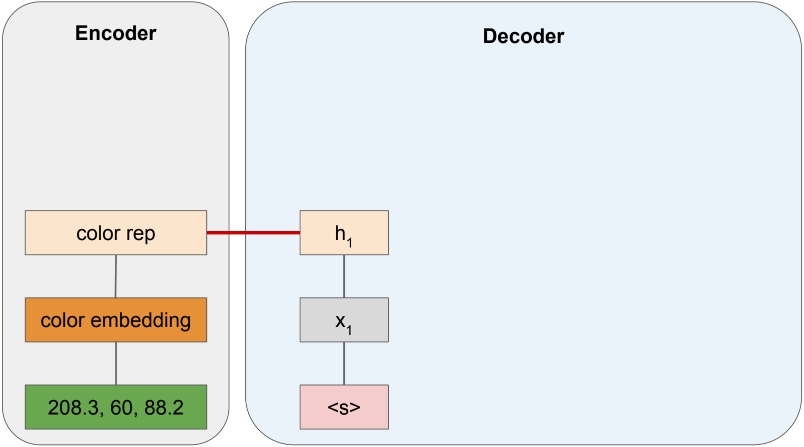
Color describer: Prediction



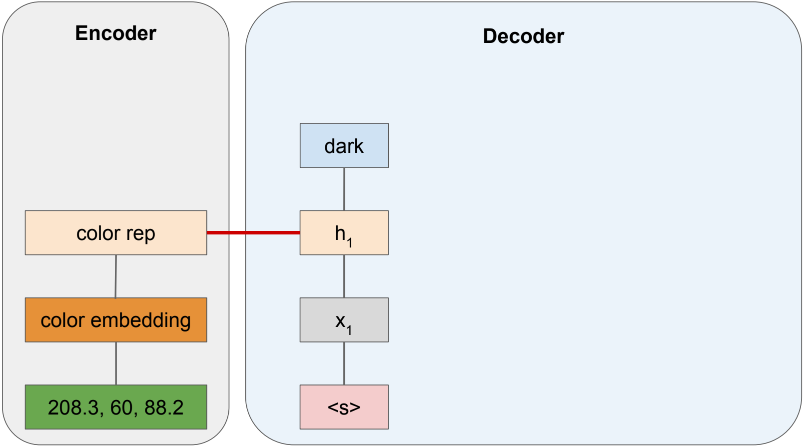
Color describer: Prediction



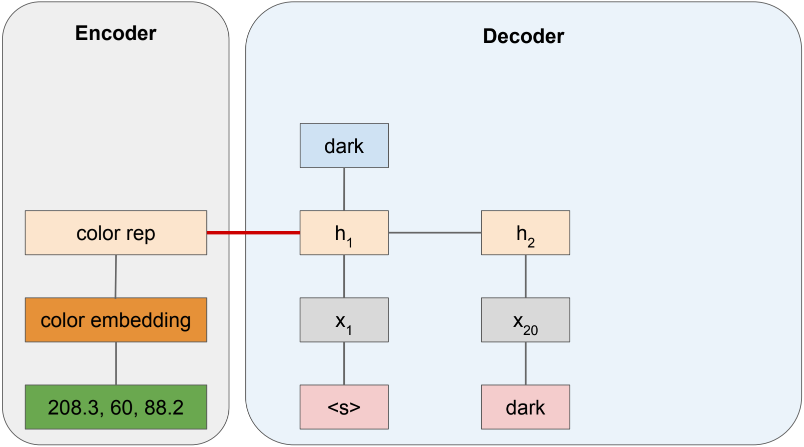
Color describer: Prediction



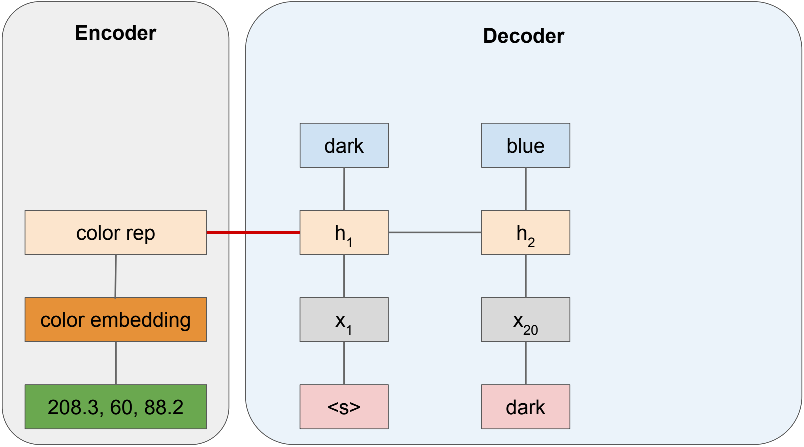
Color describer: Prediction



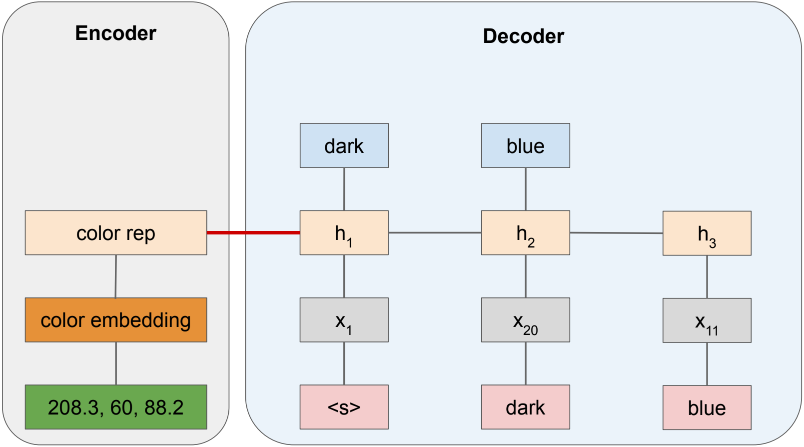
Color describer: Prediction



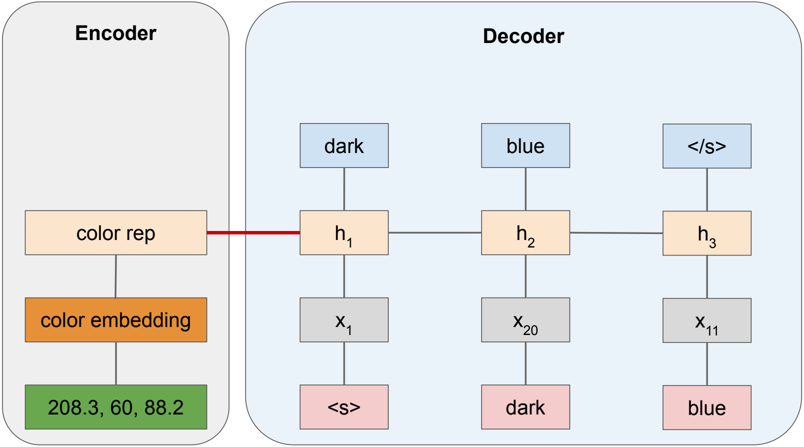
Color describer: Prediction



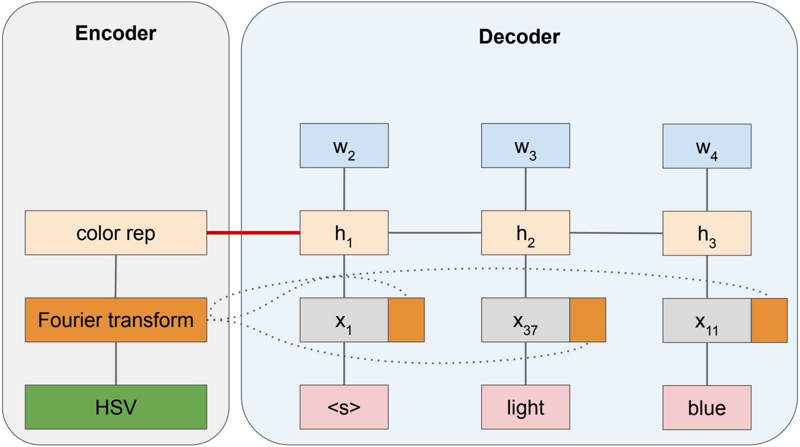
Color describer: Prediction



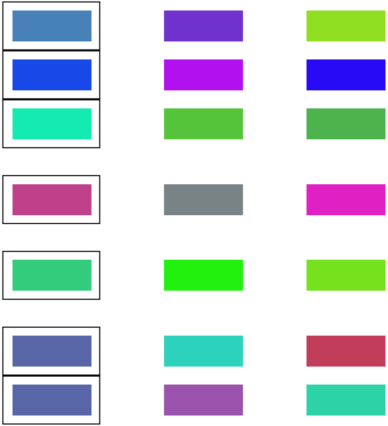
Color describer: Prediction



Color describer of Monroe et al. (2016)



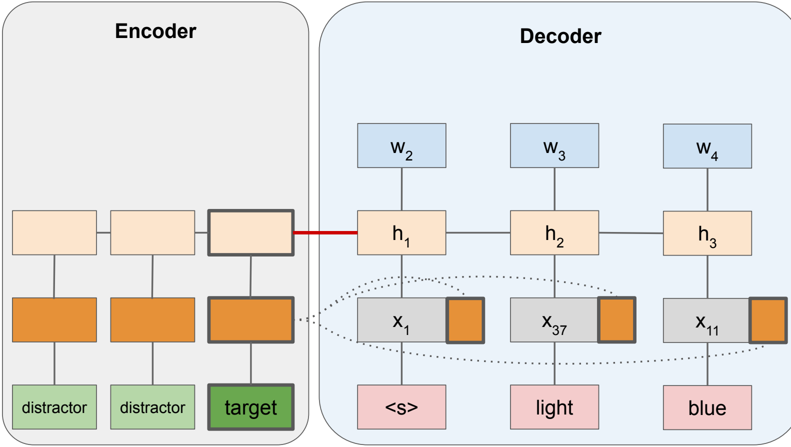
Colors in context (Monroe et al. 2017)



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

Table: Examples from the Colors in Context corpus from the Stanford Computation & Cognition Lab

Colors in context (Monroe et al. 2017)



Related ideas and tasks

- The preceding can be seen as a special case of *image captioning*, which has been revolutionized by neural methods in recent years (Karpathy & Fei-Fei 2015; Vinyals et al. 2015).
- The Encoder part of captioning models is likely to be more involved than the above, but the basic structure is the same.
- Mao et al. (2016) and Vedantam et al. (2017) explore variants of the image captioning task that are like the 'colors in context' task above.
- Visual Question Answering is a more structured variant of the problem in which an image and a question text are the inputs and the goal is to produce grounded answers.

Listeners: From language to the world

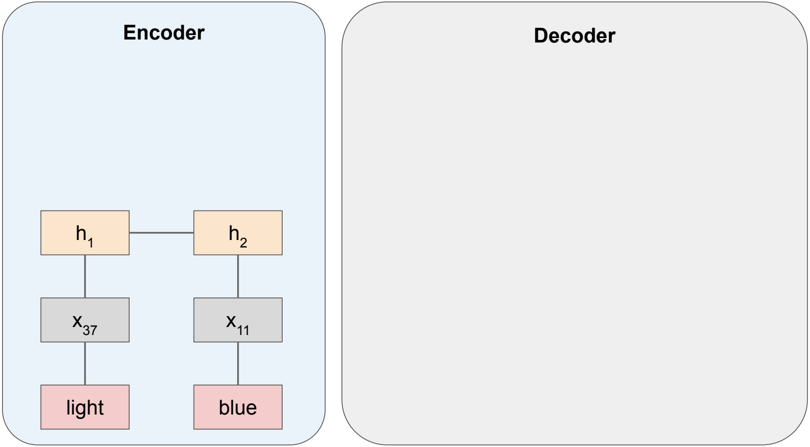
1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Listeners: From language to the world
4. Grounded chat bots
5. Reasoning about other minds
6. A few other grounding ideas

Color interpreter: Task formulation and data

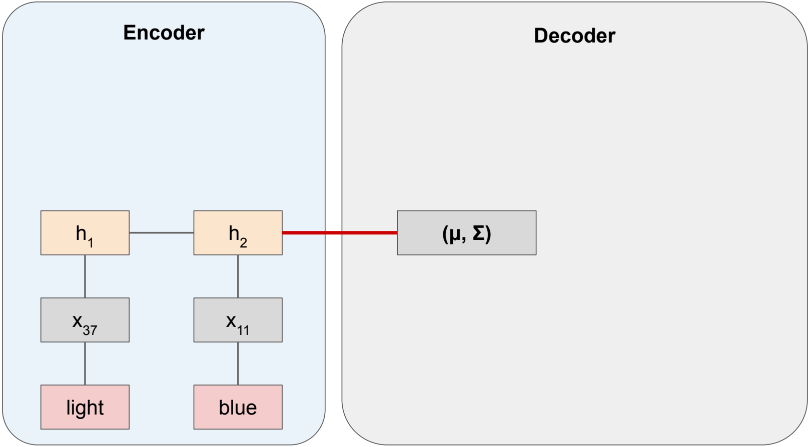
Context	Utterance
	blue
	The darker blue one
	teal not the two that are more green
	dull pink not the super bright one
	not any of the regular greens
	Purple
	blue

Table: Examples from the Colors in Context corpus from the Stanford Computation & Cognition Lab

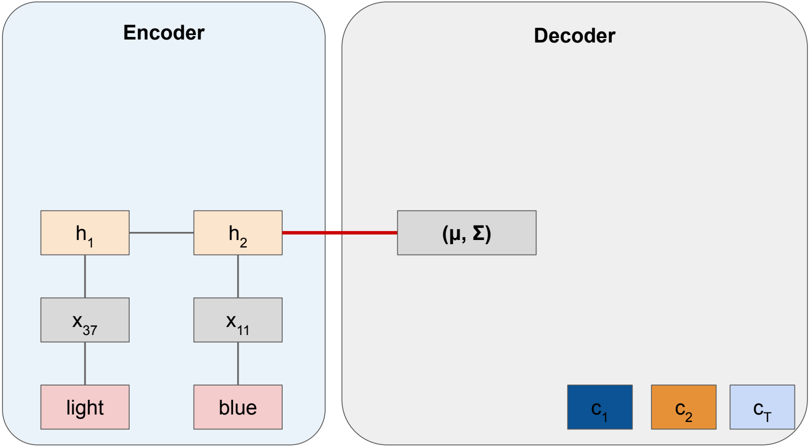
A neural listener model



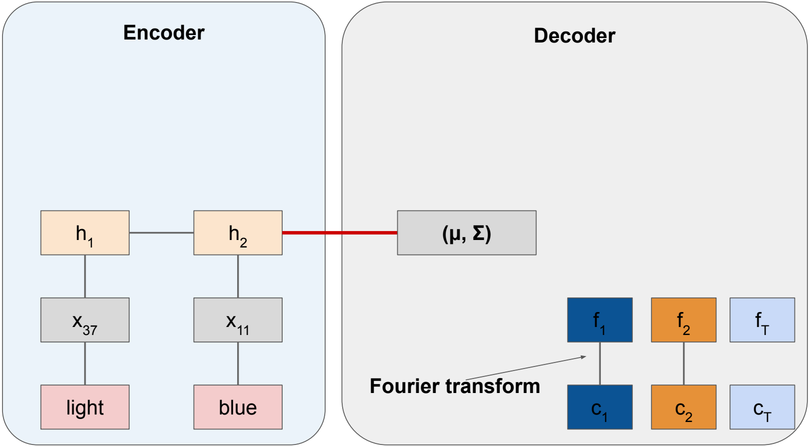
A neural listener model



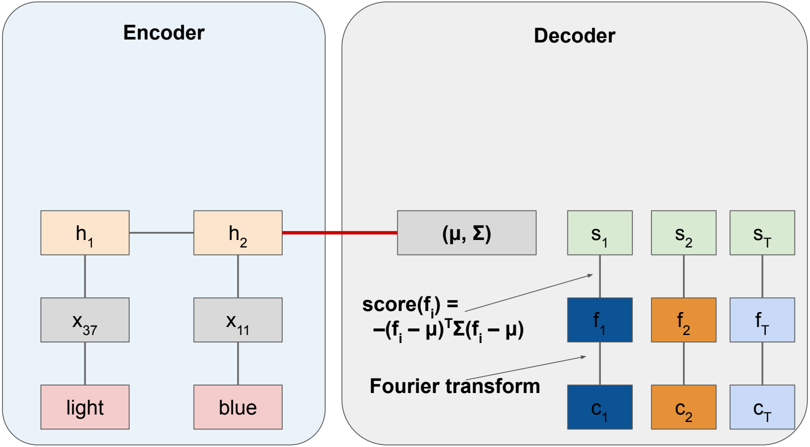
A neural listener model



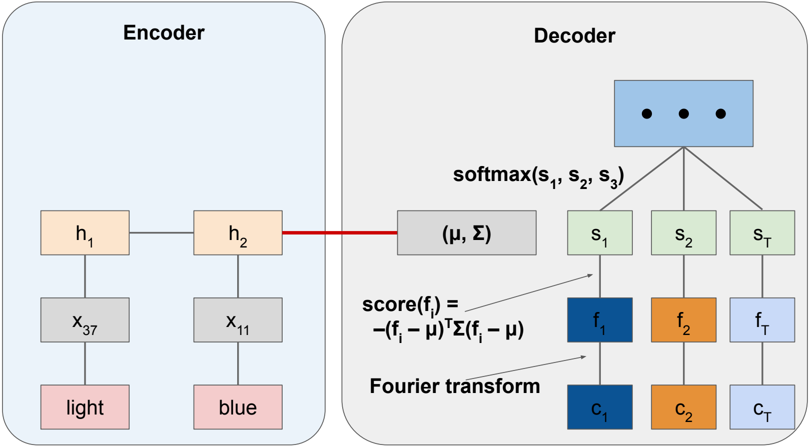
A neural listener model



A neural listener model



A neural listener model



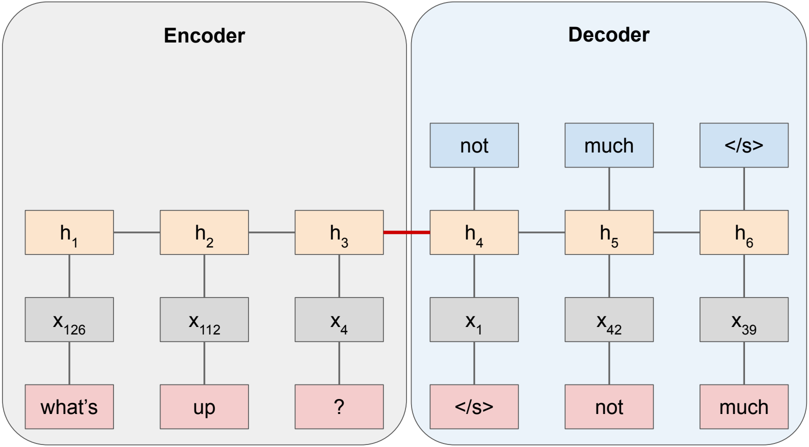
Other ideas and datasets

- 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 & Yin 2006; Chang et al. 2014, 2015).
- Young et al. (2014) seek to learn visual denotations for linguistic expressions.

Grounded chat bots

1. Overview: linguistic insights, and a bit of history
2. Speakers: From the world to language
3. Listeners: From language to the world
4. Grounded chat bots
5. Reasoning about other minds
6. A few other grounding ideas

Basic neural chatbot



FAIR negotiation dataset

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

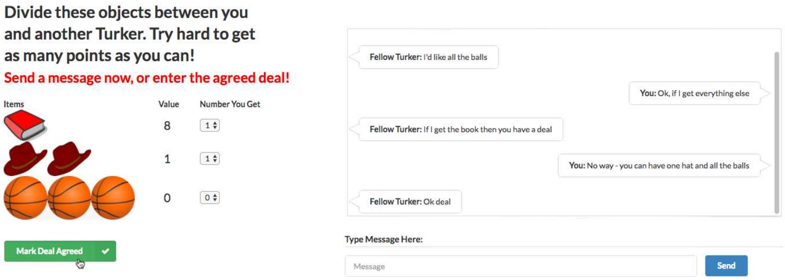


Figure 1: A dialogue in our Mechanical Turk interface, which we used to collect a negotiation dataset.

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

FAIR negotiation dataset

Perspective of YOU

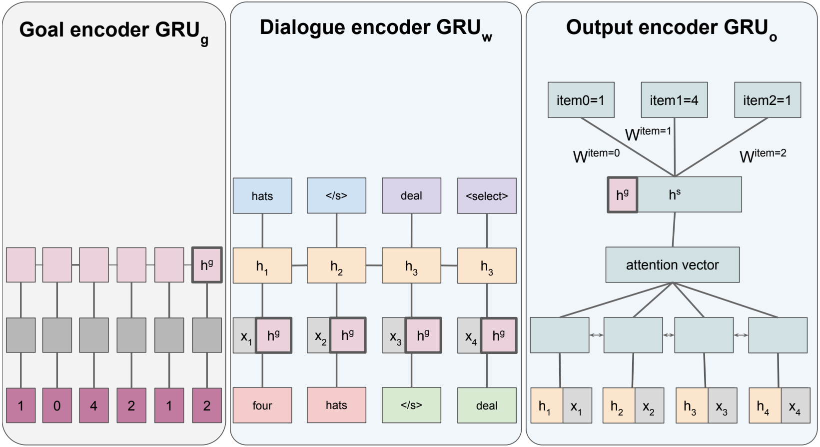
1. 1 0 4 2 1 2 # (1 book, worth 0; 4 hats, worth 2, 1 ball, worth 2)
2. YOU: i would like 4 hats and you can have the rest <eos>
   THEM: deal <eos>
   YOU: <selection>
3. item0=0 item1=4 item2=0
4. <eos>
5. reward=8
6. agree
7. 1 4 4 1 1 2

FAIR negotiation dataset

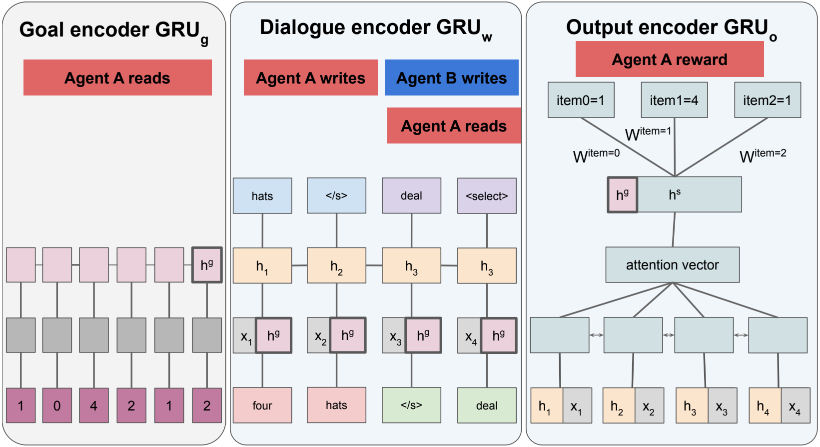
Perspective of THEM

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

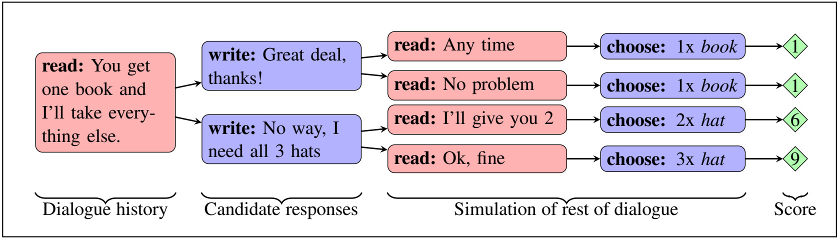
FAIR negotiation agents



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.’ ”

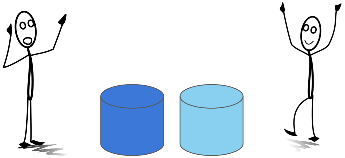
Other task-oriented dialogue datasets

- 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/`

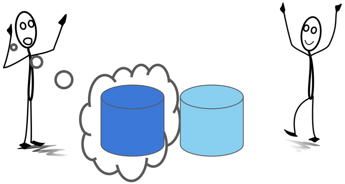
Reasoning about other minds

Pragmatic reasoning à la Grice (1975)

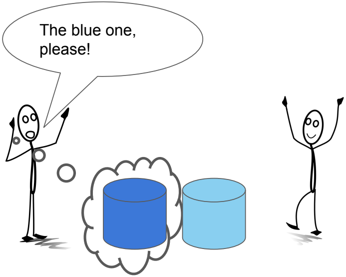
Pragmatic reasoning à la Grice (1975)



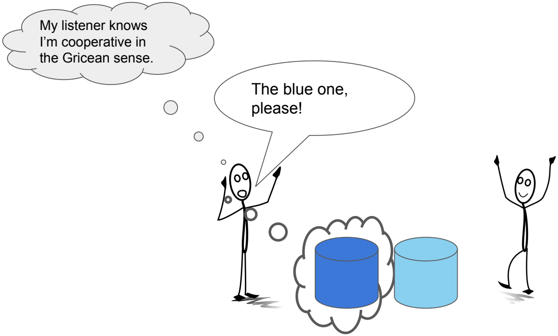
Pragmatic reasoning à la Grice (1975)



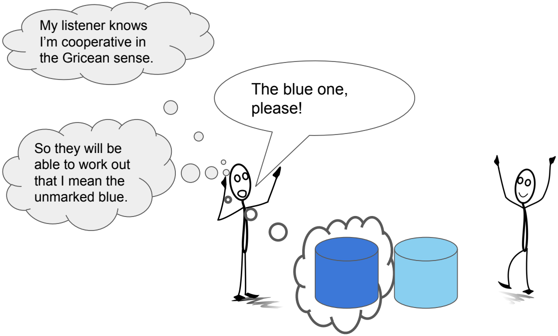
Pragmatic reasoning à la Grice (1975)



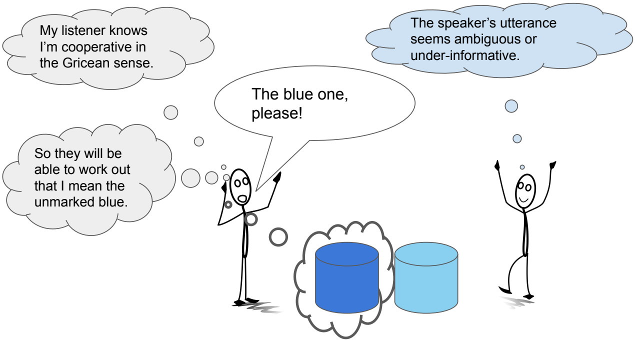
Pragmatic reasoning à la Grice (1975)



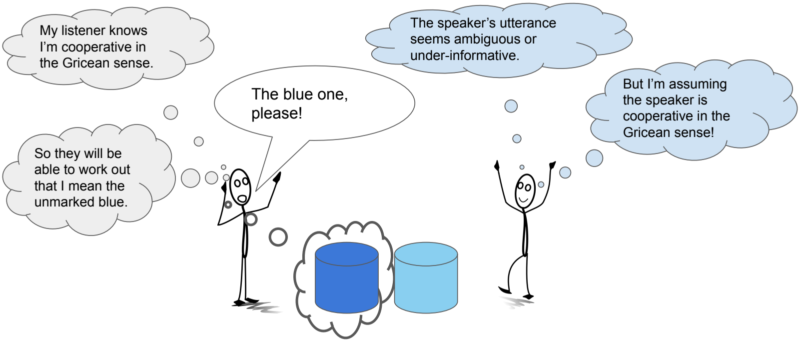
Pragmatic reasoning à la Grice (1975)



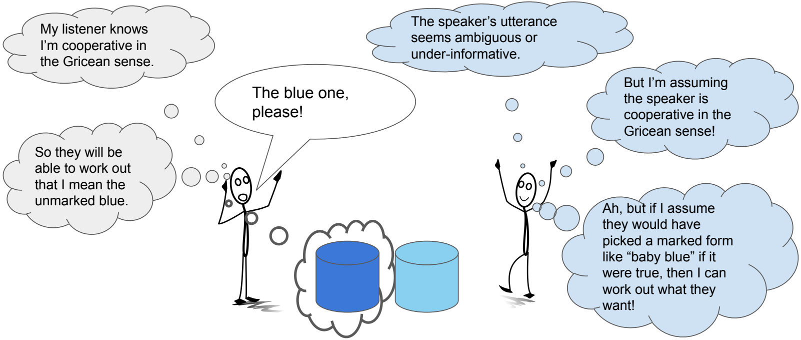
Pragmatic reasoning à la Grice (1975)



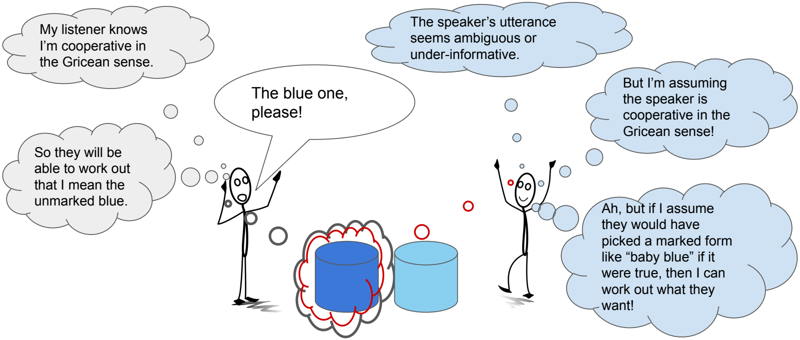
Pragmatic reasoning à la Grice (1975)



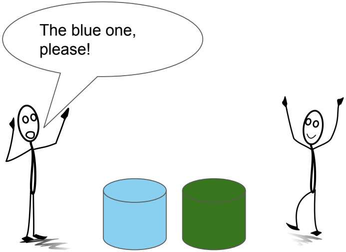
Pragmatic reasoning à la Grice (1975)



Pragmatic reasoning à la Grice (1975)



Pragmatic reasoning à la Grice (1975)



The Rational Speech Acts Model

(Frank & Goodman 2012; Goodman & Stuhlmüller 2013; Goodman & Frank 2016)

The Rational Speech Acts Model

Literal listener

$$l_0(w \mid msg, Lex) \propto Lex(msg, w)P(w)$$

(Frank & Goodman 2012; Goodman & Stuhlmüller 2013; Goodman & Frank 2016)

The Rational Speech Acts Model

Pragmatic speaker

$$s_1(msg \mid w, Lex) \propto \exp \lambda \left(\log l_0(w \mid msg, Lex) - C(msg)\right)$$

Literal listener

$$l_0(w \mid msg, Lex) \propto Lex(msg, w)P(w)$$

(Frank & Goodman 2012; Goodman & Stuhlmüller 2013; Goodman & Frank 2016)

The Rational Speech Acts Model

Pragmatic listener

$$l_1(w \mid msg, Lex) \propto s_1(msg \mid w, Lex) P(w)$$

Pragmatic speaker

$$s_1(msg \mid w, Lex) \propto \exp \lambda \left(\log l_0(w \mid msg, Lex) - C(msg)\right)$$

Literal listener

$$l_0(w \mid msg, Lex) \propto Lex(msg, w) P(w)$$

(Frank & Goodman 2012; Goodman & Stuhlmüller 2013; Goodman & Frank 2016)

The Rational Speech Acts Model

Pragmatic listener

$$l_1(w \mid msg, Lex) = \textbf{pragmatic speaker} \times \text{state prior}$$

Pragmatic speaker

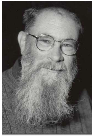
$$s_1(msg \mid w, Lex) = \textbf{literal listener} - \text{message costs}$$

Literal listener

$$l_0(w \mid msg, Lex) = \textbf{lexicon} \times \text{state prior}$$

(Frank & Goodman 2012; Goodman & Stuhlmüller 2013; Goodman & Frank 2016)

RSA listener example

		
beard	T	F
glasses	T	T

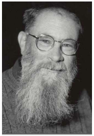
l_1

s_1

l_0

$\boxed{Lex}$

RSA listener example

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

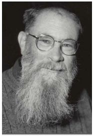
l_1

s_1

l_0

Lex

RSA listener example

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

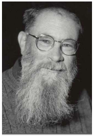
l_1

s_1

l_0

Lex

RSA listener example

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

l_1

s_1

l_0

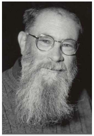
Lex

Limitations

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

$$s_1(msg \mid w, Lex) = \frac{l_0(w \mid msg, Lex)}{\sum_{msg'} l_0(w \mid msg', Lex)}$$

- High-bias model; few chances to learn from data

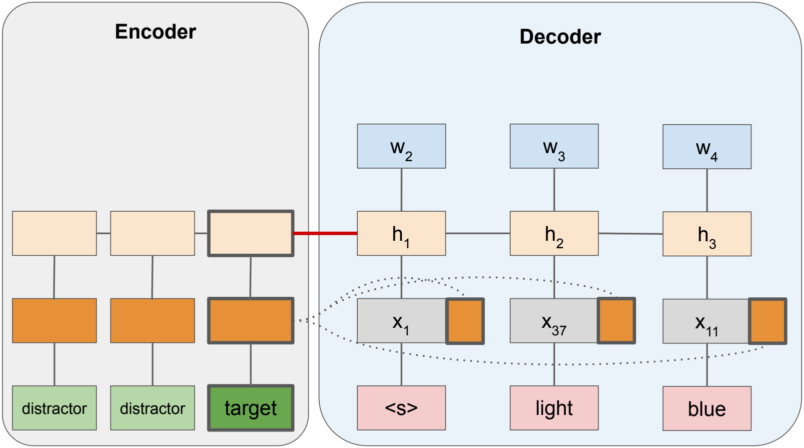
		
beard	**1**	0
glasses	.25	**.75**

Colors in context (Monroe et al. 2017)

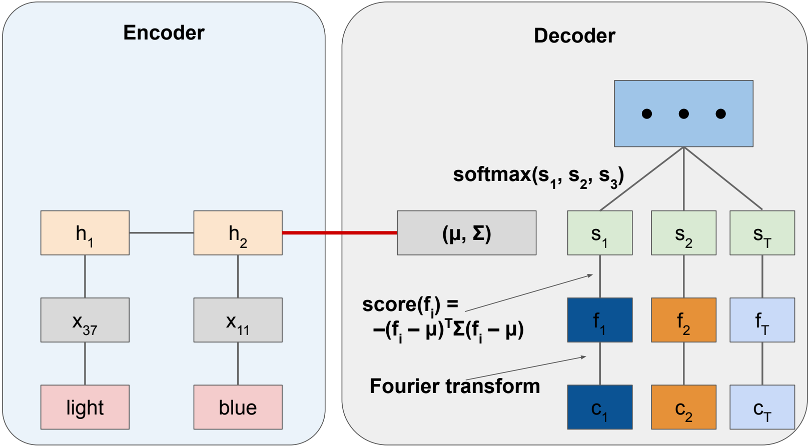
Context	Utterance
	blue
	The darker blue one
	teal not the two that are more green
	dull pink not the super bright one
	not any of the regular greens
	Purple
	blue

Table: Examples from the Colors in Context corpus from the Stanford Computation & Cognition Lab

Literal neural speaker S_0



Neural literal listener $\mathcal{L}_0$



Neural pragmatic agents

Neural pragmatic speaker (Andreas & Klein 2016)

$$\mathcal{S}_1(msg \mid c, C; \theta) = \frac{\mathcal{L}_0(c \mid msg, C; \theta)}{\sum_{msg' \in X} \mathcal{L}_0(c \mid msg', C; \theta)}$$

where X is a sample from $\mathcal{S}_0(msg \mid c, C; \theta)$ such that $msg^* \in X$.

Neural pragmatic listener

$$\mathcal{L}_1(c \mid msg, C; \theta) \propto \mathcal{S}_1(msg \mid c, C; \theta)$$

Blended neural pragmatic listener

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

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?

(Cohn-Gordon et al. 2018, 2019)

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 $\mathcal{S}_0$

(Cohn-Gordon et al. 2018, 2019)

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?

⇒ **Full RSA reasoning about *characters***

(Cohn-Gordon et al. 2018, 2019)

Other related work

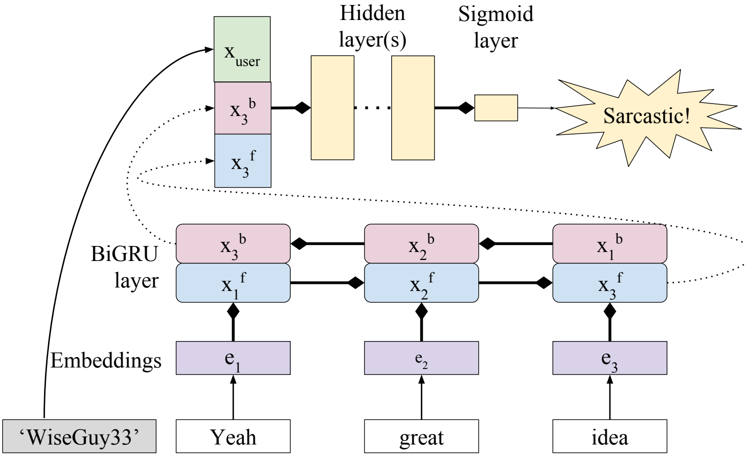
- Golland et al. (2010): Recursive speaker/listener reasoning as part of interpreting complex utterances compositionally, with grounding in a simple visual world.
- Tellex et al.’s (2014) Inverse Semantics: Robot utterances are scored by models similar to RSA’s pragmatic speakers.
- Wang et al. (2016): Pragmatic reasoning helps in online learning of semantic parsers.
- Monroe & Potts (2015): “RSA as a hidden activation function”
- Monroe et al. (2018): Bilingual color describers (English and Chinese).
- Fried et al. (2018): Sequential instruction following with pragmatic reasoning.
- Khani et al. (2018): Collaborative games with pragmatic reasoning.

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

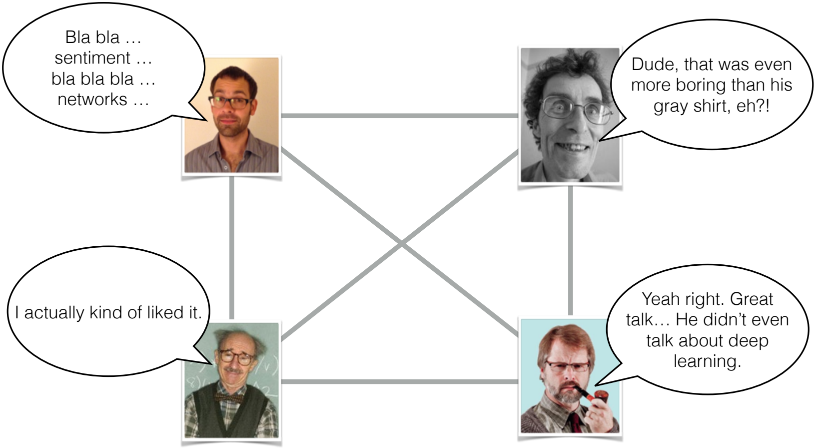
A few other grounding ideas

Modeling users for sarcasm detection



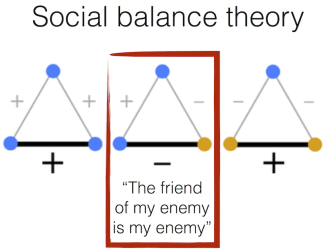
(SARC: Khodak et al. 2017; Kolchinski & Potts 2018)

NLU in social graphs with Probabilistic Soft Logic



(PSL: https://psl.linqs.org; West et al. 2014)

NLU in social graphs with Probabilistic Soft Logic



(PSL: 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.

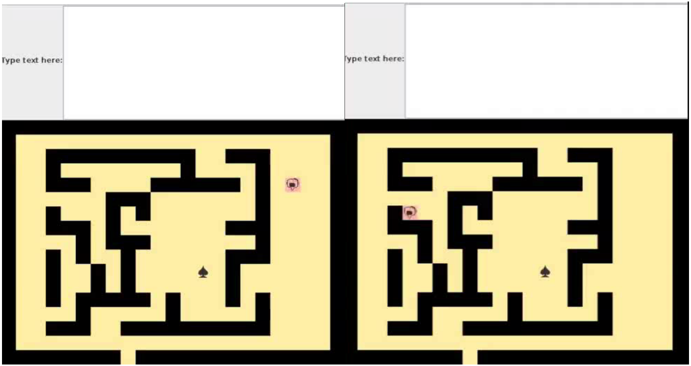
3. The user’s actions ground the parsed language.

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)

(Allen et al. 2007)

Decision-theoretic agents

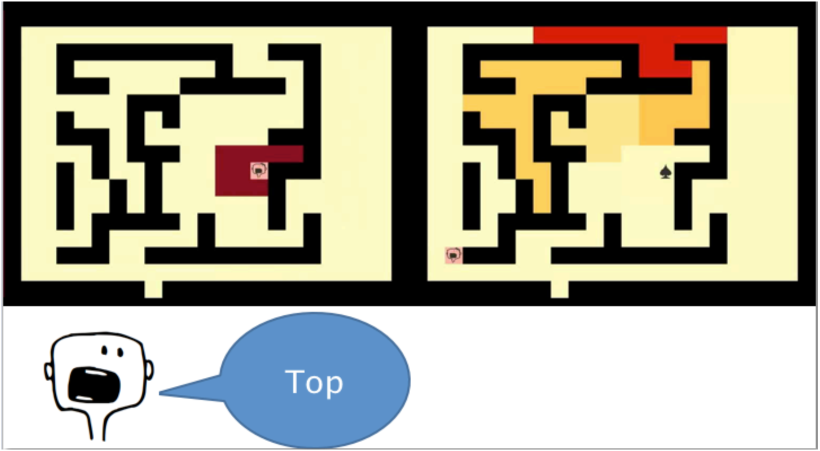
Both players must find the ace of spades. DialogBot: 



(Vogel et al. 2013a,b)

Decision-theoretic agents

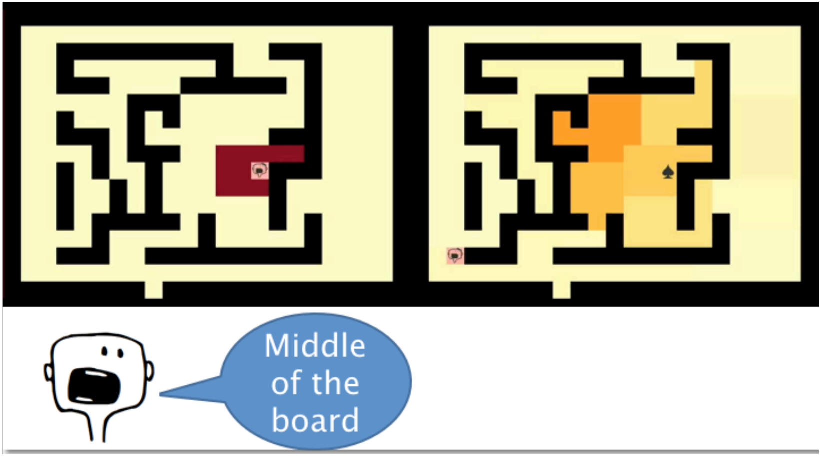
Baby DialogBots (a few hours of policy exploration)



(Vogel et al. 2013a,b)

Decision-theoretic agents

Grown-up DialogBots (a week of policy exploration)



(Vogel et al. 2013a,b)

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

Allen, James F., Nathanael Chambers, George Ferguson, Lucian Galescu, Hyuckchul Jung, Mary Swift & William Taysom. 2007. PLOW: A collaborative task learning agent. In *Proceedings of the twenty-second AAAI conference on artificial intelligence*, 1514–1519. Vancouver, British Columbia, Canada: AAAI Press.

Andreas, Jacob & Dan Klein. 2016. Reasoning about pragmatics with neural listeners and speakers. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, 1173–1182. Association for Computational Linguistics. http://aclweb.org/anthology/D16-1125.

Chang, Angel, Will Monroe, Manolis Savva, Christopher Potts & 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*, 53–62. Stroudsburg, PA: Association for Computational Linguistics.

Chang, Angel, Manolis Savva & 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)*, 2028–2038. Doha, Qatar: Association for Computational Linguistics. doi:10.3115/v1/D14-1217. https://www.aclweb.org/anthology/D14-1217.

Cohn-Gordon, Reuben, Noah D. Goodman & 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*, 439–443. Stroudsburg, PA: Association for Computational Linguistics.

Cohn-Gordon, Reuben, Noah D. Goodman & Christopher Potts. 2019. An incremental iterated response model of pragmatics. In *Proceedings of the society for computation in linguistics*, 81–90. Washington, D.C.: Linguistic Society of America.

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

Frank, Michael C. & Noah D. Goodman. 2014. Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology* 75(1). 80–96. doi:doi:10.1016/j.cogpsych.2014.08.002.

Frank, Michael C., Joshua B. Tenenbaum & Anne Fernald. 2012. Social and discourse contributions to the determination of reference in cross-situational word learning. *Language, Learning, and Development* .

Fried, Daniel, Jacob Andreas & 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)*, 1951–1963. New Orleans, Louisiana: Association for Computational Linguistics. doi:10.18653/v1/N18-1177. https://www.aclweb.org/anthology/N18-1177.

Golland, Dave, Percy Liang & Dan Klein. 2010. A game-theoretic approach to generating spatial descriptions. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, 410–419. Stroudsburg, PA: ACL. http://www.aclweb.org/anthology/D10-1040.

Goodman, Noah D. & Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences* 20(11). 818–829. doi:10.1016/j.tics.2016.08.005.

References II

Goodman, Noah D. & Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5(1). 173–184.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry Morgan (eds.), *Syntax and semantics*, vol. 3: Speech Acts, 43–58. New York: Academic Press.

Hawkins, Robert X. D. 2015. Conducting real-time multiplayer experiments on the web. *Behavior Research Methods* 47(4). 966–976. doi:10.3758/s13428-014-0515-6. https://doi.org/10.3758/s13428-014-0515-6.

Karpathy, Andrej & Li Fei-Fei. 2015. Deep visual-semantic alignments for generating image descriptions. In *The ieee conference on computer vision and pattern recognition (cvpr)*, 3128–3137.

Khani, Fereshte, Noah D. Goodman & Percy Liang. 2018. Planning, inference and pragmatics in sequential language games. *Transactions of the Association for Computational Linguistics* 6. 543–555. doi:10.1162/tacl_a_00037. https://www.aclweb.org/anthology/Q18-1037.

Khodak, Mikhail, Nikunj Saunshi & Kiran Vodrahalli. 2017. A large self-annotated corpus for sarcasm. *arXiv preprint arXiv:1704.05579* .

Kolchinski, Y. Alex & Christopher Potts. 2018. Representing social media users for sarcasm detection. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, 1115–1121. Stroudsburg, PA: Association for Computational Linguistics.

Levesque, Hector J. 2013. On our best behaviour. In *Proceedings of the twenty-third international conference on artificial intelligence*, Beijing.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lewis, Mike, Denis Yarats, Yann N. Dauphin, Devi Parikh & Dhruv Batra. 2017. Deal or no deal? End-to-end learning for negotiation dialogues. ArXiv:1706.05125.

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

McMahan, Brian & Matthew Stone. 2015. A Bayesian model of grounded color semantics. *Transactions of the Association for Computational Linguistics* 3. 103–115. https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/276.

Monroe, Will, Noah D. Goodman & Christopher Potts. 2016. Learning to generate compositional color descriptions. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2243–2248. Stroudsburg, PA: Association for Computational Linguistics.

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

References III

Monroe, Will, Jennifer Hu, Andrew Jong & 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*, 2155–2165. Stroudsburg, PA: Association for Computational Linguistics.

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

Seversky, Lee M & 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*, 61–64. ACM.

Tellex, Stefanie, Ross A. Knepper, Adrian Li, Thomas M. Howard, Daniela Rus & Nicholas Roy. 2014. Asking for help using inverse semantics. In *Proceedings of robotics: Science and systems*, doi:10.15607/RSS.2014.X.024.

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

Vinyals, Oriol, Alexander Toshev, Samy Bengio & Dumitru Erhan. 2015. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3156–3164.

Vogel, Adam, Max Bodoia, Christopher Potts & 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*, 1072–1081. Stroudsburg, PA: Association for Computational Linguistics.

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

Wang, Alex, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy & 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*, 353–355. Brussels, Belgium: Association for Computational Linguistics. https://www.aclweb.org/anthology/W18-5446.

Wang, Sida I., Percy Liang & 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)*, 2368–2378. Association for Computational Linguistics. doi:10.18653/v1/P16-1224. http://aclweb.org/anthology/P16-1224.

West, Robert, Hristo S. Paskov, Jure Leskovec & Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics* 2(2). 297–310.

Winograd, Terry. 1972. Understanding natural language. *Cognitive Psychology* 3(1). 1–191.

Winograd, Terry. 1986. A procedural model of language understanding. In Barbara J. Grosz, Karen Sparck-Jones & Bonnie Lynn Webber (eds.), *Readings in natural language processing*, 249–266. San Francisco: Morgan Kaufmann Publishers Inc.

References IV

Yarats, Denis & Mike Lewis. 2018. Hierarchical text generation and planning for strategic dialogue. In Jennifer Dy & Andreas Krause (eds.), *Proceedings of the 35th international conference on machine learning*, vol. 80 Proceedings of Machine Learning Research, 5587–5595. Stockholmsmässan, Stockholm Sweden: PMLR. http://proceedings.mlr.press/v80/yarats18a.html.

Young, Peter, Alice Lai, Micah Hodosh & 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.