Reference-Centric Models for Grounded Collaborative Dialogue

Daniel Fried, Justin Chiu, and Dan Klein
A: Do you know the Thai restaurant west of the highway?
A: I know that one, sounds good!
B: I don’t, but how about the coffee place near the park?
A: I have three dots in a line with a dark one in the center.

B: I don’t have that. Do you have a cluster of three grey dots in a triangle?

A: Is there a large black dot to the left of the three grey dots?

B: Yes, let’s select the black one.

*OneCommon* [Udagawa and Aizawa, 2019 & 2020]
A: I have a large black dot with a smaller and lighter dot above it and to the right.

A: Yes, I have that. Let’s pick the large black one.

B: I don’t have that. How about a dark grey very large dot with a slightly darker dot above and slightly to its left?

OneCommon [Udagawa and Aizawa, 2019 & 2020]
A: I have a large black dot with a smaller and lighter dot above it and to the right.

A: Yes, I have that. Let’s pick the large black one.

B: I don’t have that. How about a dark grey very large dot with a slightly darker dot above and slightly to its left?

OneCommon [Udagawa and Aizawa, 2019 & 2020]
A: I have a large black dot with a smaller and lighter dot above it and to the right.

B: I don't have that. How about a dark grey very large dot with a slightly darker dot above and slightly to its left?

A: Yes, I have that. Let's pick the large black one.

OneCommon [Udagawa and Aizawa, 2019 & 2020]
Reference-Centric Models

A modular decomposition of the task:

- **Listener**: 1) Reference resolution
- **Controller**: 2) Content selection
- **Speaker**: 3) Surface realization
A: I have three dots in a line with a dark one in the center.
B: I don’t have that. Do you have a group of three grey dots?
A: Is there a large black dot to the left of the three grey dots?
B:???
Reference-Centric Models

Reference resolution

Content selection

Surface realization

\[ \begin{align*}
B: & \text{ I don’t have that. Do you have a group of three grey dots?} \\
A: & \text{ Is there a large black dot to the left of the three grey dots?} \\
B: & \text{ ???}
\end{align*} \]
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?  
A: Is there a large black dot to the left of the three grey dots?  
B: ???
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?  
A: Is there a large black dot to the left of the three grey dots?  
B: ???
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B: ???
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B: ???
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B: ???
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B: ???
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B: ????
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don’t have that. Do you have a group of three grey dots?
A: Is there a large black dot to the left of the three grey dots?
B: Yes, let’s select the black one.

Other components (see paper): a structured memory, a confirmation module, and a selection module.
Combining Work on Grounded...

Reference Resolution  
Heeman 1991; Mao et al. 2016; Yu et al. 2017; Takmaz et al. 2020

Reference Generation  

Goal-Oriented Dialogue  

Pragmatics  
Golland et al. 2010; Frank & Goodman 2012; Vogel et al. 2013; Monroe et al. 2017; Khani et al. 2018; Cohn-Gordon et al. 2018; Fried et al. 2018
A: Is there a large black dot to the left of the three grey dots?

B: Yes, let’s select the black one.
Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]
Predict individual dots in referents using graph neural net representations

Is there a large black dot to the left of the three grey dots?
RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

Predict individual dots in referents using graph neural net representations

Is there a large black dot to the left of the three grey dots?
RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]
   Predict individual dots in referents using graph neural net representations
+ Structured Neural CRF
   Add neural potentials over groups and relations

Is there a large black dot to the left of the three grey dots?
Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]
   Predict individual dots in referents using graph neural net representations
+ Structured Neural CRF
   Add neural potentials over groups and relations

Is there a large black dot to the left of the three grey dots?
Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

Predict individual dots in referents using graph neural net representations

+ Structured Neural CRF

Add neural potentials over groups and relations

Structured CRF potentials condition on encoded dot attributes and an LSTM representation of the text.

A linear chain dynamic program makes training and inference efficient.

Is there a large black dot to the left of the three grey dots?
Listener Module Evaluation

Reference Resolution Exact Match

Human Performance

RelationNet [S+ ‘17, U&A ‘20]

44

+Structured CRF

76
A: Is there a large black dot to the left of the three grey dots?
B: Yes, let's select the black one.
Pragmatic Generation

Let’s select the grey one.

Speaker

Let’s select the right one.

Listener

Let’s select the left one.

Listener

Let’s select the black one.

Controller

$P_S(r \mid u, ...)$ 0.9

Generalization of Rational Speech Acts [Frank and Goodman 2012]
See paper for ablations showing that structure and pragmatics help substantially.
Success by Human Skill Level

- Human with Human
- Human with Ours
- Human with U&A ’20

Game Success Rate vs. Human Skill Percentile

- Normal distribution observed for different skill levels.
Success by Human Skill Level

Game Success Rate

Human with Human
Human with Ours
Human with U&A '20

Human Skill Percentile
Demo

Time Remaining: 6:00

[02/12/21 08:57:44] <You entered the room.>
[02/12/21 08:57:46] <Your partner has joined the room.>

Waiting on your partner to take a turn...
Takeaways

Structure still benefits grounded neural dialogue models.

Pragmatic modeling makes grounded dialogue more effective.
Thank you!

Poster: Sunday Nov 7th,
7-9pm AST / 4-6pm Pacific

github.com/dpfried/onecommon