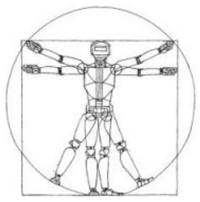
Jointly Improving Parsing and Perception for Natural Language Commands through Human-Robot Dialog

Jesse Thomason,* Aishwarya Padmakumar,† Jivko Sinapov,‡ Nick Walker,* Yuqian Jiang,† Harel Yedidsion,† Justin Hart,† Peter Stone,† and Raymond J. Mooney†

*University of Washington; †University of Texas at Austin; ‡Tufts University



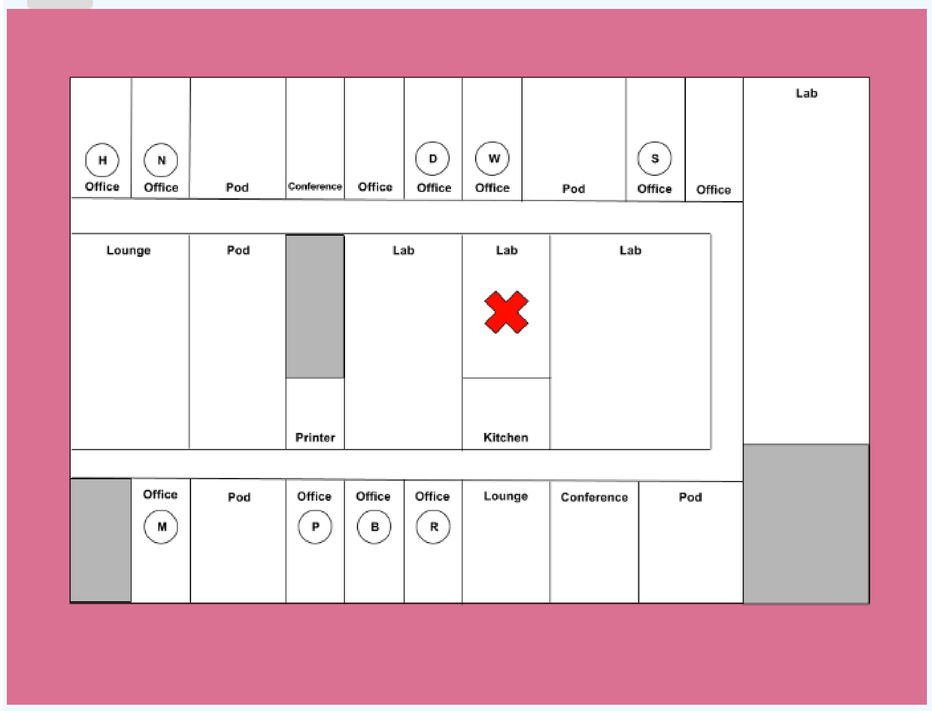
Human-Robot Dialog

Natural language understanding in robots needs to be robust to a wide-range of both human speakers and human environments. Rather than force humans to use language that robots can understand, robots in human environments should dynamically adapt—continuously learning new language constructions and perceptual concepts as they are used in context. We parse natural language to underlying meanings, and use robotic sensors to create multi-modal models of perceptual concepts.

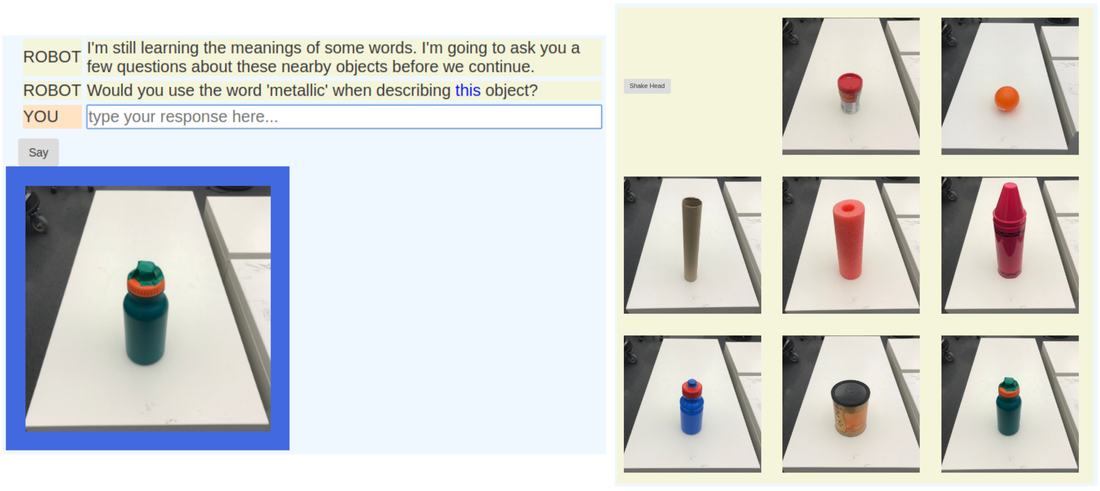
Learning From Dialogs

ROBOT	What should I do?
YOU	move a red can from the middle lab to the mallory's office
ROBOT	You want me to move an item from somewhere to somewhere (**not** give it to someone)?
YOU	yes
ROBOT	Where should I move something from on its way somewhere else?
YOU	the middle lab
ROBOT	To where should I move something from here?
YOU	type your response here...

Say

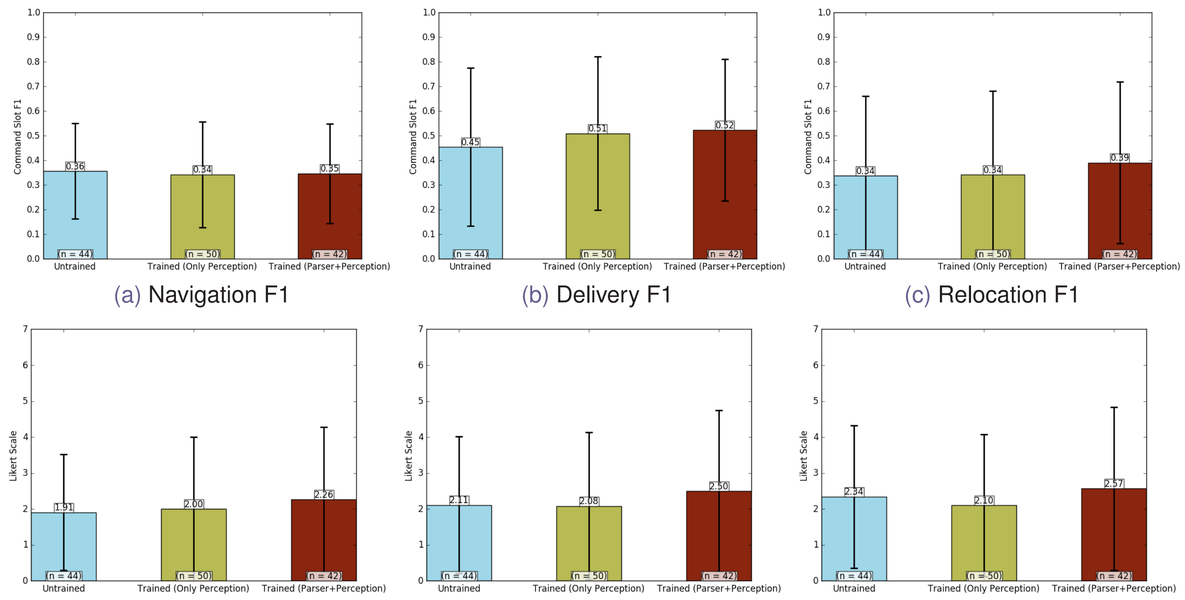


The user gives a command to the learning agent, which replies with questions to clarify the command until the user confirms that the agent has correctly understood. Each clarification generates a new training example for the agent's semantic parser, enabling retraining from conversational data.



The agent can ask whether a predicate applies to a specific object, and for positive/negative object examples. Learning these models while retraining the semantic parser connects words to concept models corresponding to physical properties of objects like color, weight, and sound.

Mechanical Turk Experiments



(a) Navigation F1 (b) Delivery F1 (c) Relocation F1

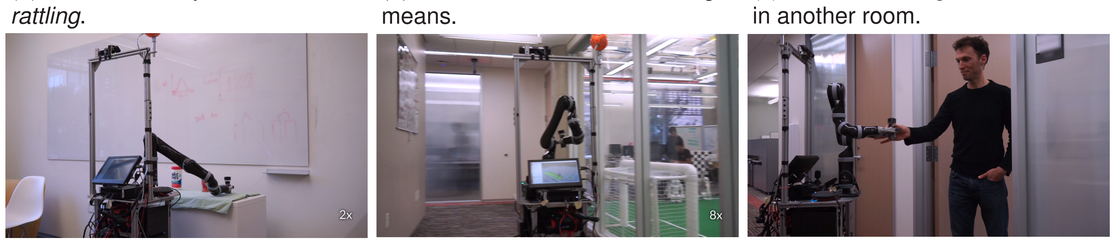
(d) Use Navigation (e) Use Delivery (f) Use Relocation

We measured the average semantic slot *f* scores between the semantic roles in the target task and the task confirmed by the user. Users responded to survey prompts about whether they would use a robot in the real world to accomplish each task. We compare an initialized agent to one that learns perceptual concepts, and one that retrains its semantic parser based on user conversations.

BWIBot Demonstration

R: What should I do?
H: Move a rattling container from the lounge by the conference room to Bob's office.
R: Does rattling refer to a property of an object?
H: Yes.
R: Does rattling mean the same thing as empty?
H: No.
R: I am still learning the meanings of some words.
R: Show me an object you could use the word rattling when describing, or say "none of them."
H: Watch.
R: I am getting the objects on the table into focus. . . Okay, I see them.
H: This one. *The human picks up a small, full pill jar.*
R: I see.
R: Show me an object you could not use the word rattling when describing, or say "all of them."
H: Watch. This one. *The human picks up a red water bottle.*
R: I see.
R: Show me an object you could use the word rattling when describing, or say "none of them."
H: Watch. This one. *The human picks up a container half-full of beans.*
R: I see.
R: Thanks. Now back to business.
R: You want me to move an item from 3.514 to somewhere?
H: Yes.
R: I should move something from 3.514 to 3.510?
H: Yes.

(a) The human says novel word *rattling*. (b) The robot learns what *rattling* means. (c) The robot navigates to the item in another room.



(d) The robot grasps a *rattling container*. (e) The robot moves to deliver the item. (f) The robot hands over the target item.

The embodied agent learns a new word, *rattling*, which requires perception using the auditory sensing modality, and is then able to complete the task.

jdtho@cs.washington.edu https://jessethomason.com/