Grounding Natural Language Instructions to Robot Behavior: A Goal-Directed View

Lawson L.S. Wong

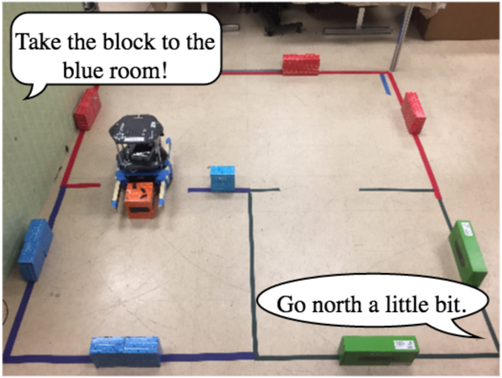
Humans to Robots Laboratory (H2R)
Department of Computer Science, Brown University

Now at College of Computer and Information Science (CCIS), Northeastern University

September 9, 2018

BROWN

Grounding natural language instructions



Spectrum of language grounding approaches

	Typical (embodied vision)	We assume (robotics)

Spectrum of language grounding approaches

	Typical (embodied vision)	We assume (robotics)
Knowledge of environment	Unseen	Known (propositional)

Spectrum of language grounding approaches

	Typical (embodied vision)	We assume (robotics)
Knowledge of environment	Unseen	Known (propositional)
Knowledge of tasks/rewards	Discovered / imitated	Known family (**goal-directed**)

Spectrum of language grounding approaches

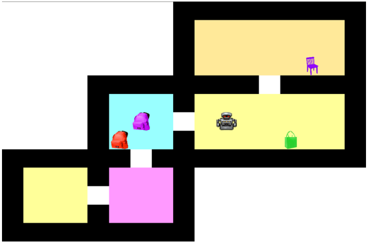
	Typical (embodied vision)	We assume (robotics)
Knowledge of environment	Unseen	Known (propositional)
Knowledge of tasks/rewards	Discovered / imitated	Known family (**goal-directed**)
Nature of feedback	Often rich, short-horizon	Sparse, long-horizon

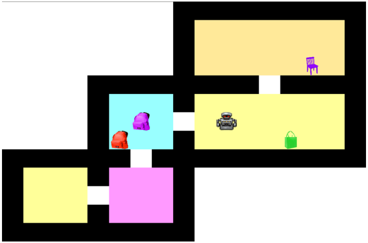
Spectrum of language grounding approaches

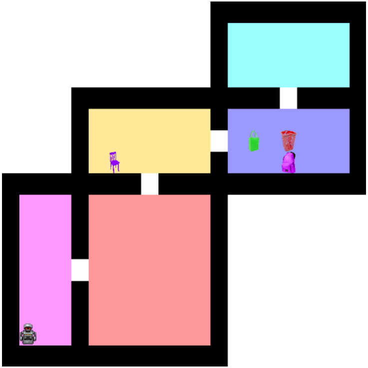
	Typical (embodied vision)	We assume (robotics)
Knowledge of environment	Unseen	Known (propositional)
Knowledge of tasks/rewards	Discovered / imitated	Known family (**goal-directed**)
Nature of feedback	Often rich, short-horizon	Sparse, long-horizon
Complexity	Learning the rich sensor $\rightarrow$ action mapping	Classification, planning

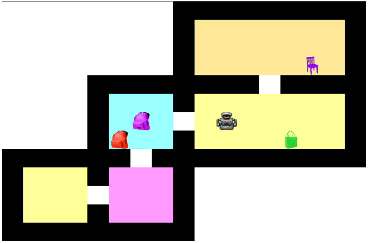
"Take the block to the blue room"

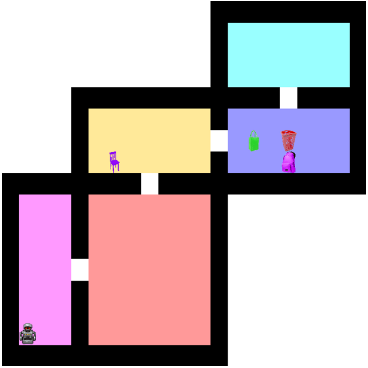
RSS 2017, ACL WS 2017, RSS 2018, Autonomous Robots 2018











For us, the role of simulators:

- Generating behavior for eliciting language data
- Model-based planning

Outline

- Paradigm
- Data
- Models

Paradigm

"Take the block to the blue room"

Paradigm

"Take the block to the blue room"

- `skill_BlockToBlueRoom`

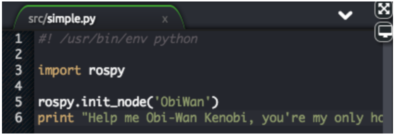
Paradigm

"Take the block to the blue room"

- `skill_BlockToBlueRoom`
- `objInRoom(block, blue)`

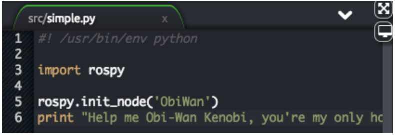
Paradigm

"Take the block to the blue room"

- `skill_BlockToBlueRoom`
- `objInRoom(block, blue)`
- 

Paradigm

"Take the block to the blue room"

- `skill_BlockToBlueRoom`
- `objInRoom(block, blue)`
- 

- "Zabierz blok do niebieskiego pokoju"

Paradigm

"Take the block to the blue room"

- `skill_BlockToBlueRoom`
- `objInRoom(block, blue)`
- src/simple.py

```python
#! /usr/bin/env python

import rospy

rospy.init_node('ObiWan')
print "Help me Obi-Wan Kenobi, you're my only hc
```

Language grounding
as
machine translation

- "Zabierz blok do niebieskiego pokoju"

Paradigm

"Take the block to the blue room"

- `skill_BlockToBlueRoom`
- `objInRoom(block, blue)`
- src/simple.py

```
#! /usr/bin/env python

import rospy

rospy.init_node('ObiWan')
print "Help me Obi-Wan Kenobi, you're my only hc
```

Language grounding
as
machine translation

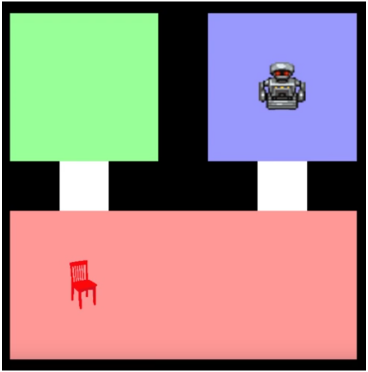
- "Zabierz blok do niebieskiego pokoju"

+ speech recognition, perception, world model, planning, control, ...
$\Rightarrow$ Robot behavior

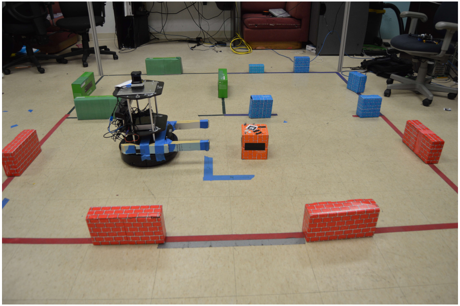
Outline

- Paradigm
- Data
- Models

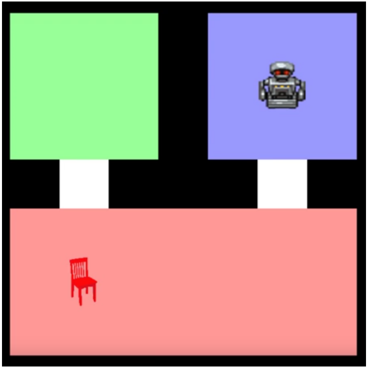
Simulation Domain



Cleanup World
[MacGlashan et al. 2015]



Data collection with Amazon Mechanical Turk



Example Command	Goal
Go to the green room.	`agentInRoom(agent0, r)` $\wedge$ `roomIsGreen(r)`
Bring the chair to the blue room.	`objInRoom(chair0, r)` $\wedge$ `roomIsBlue(r)`

Outline

- Paradigm
- Data
- Models

Language grounding as machine translation

Source: Natural language (English)

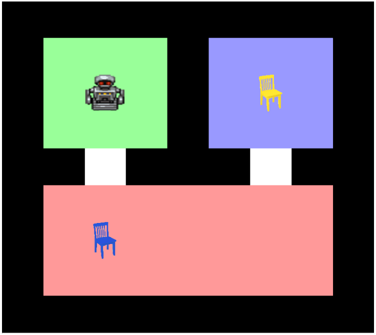
Papers	Target language / representation

Language grounding as machine translation

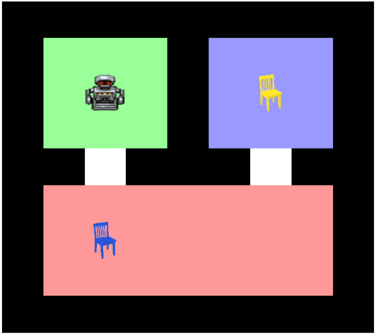
Source: Natural language (English)

Papers	Target language / representation
[MacMahon et al.], [Chen & Mooney] [Tellex et al.], [Matuszek et al.] [Artzi & Zettlemoyer], . . .	Action space

Do we really want to specify actions?

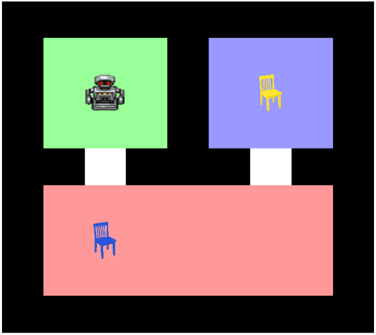


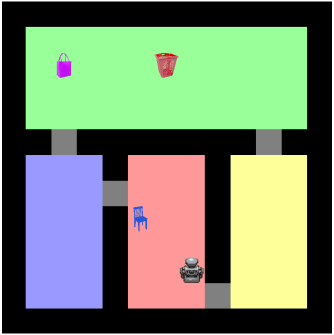
Do we really want to specify actions?



"Go to the red room" $\mapsto$ `down; down; down`

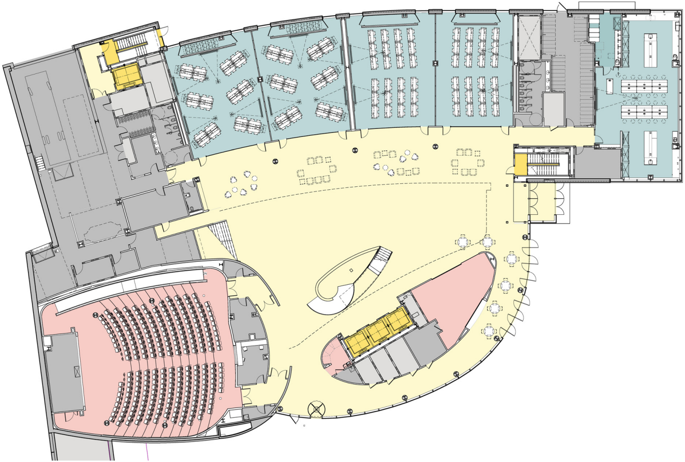
Do we really want to specify actions?





"Go to the red room" $\mapsto$ `down; down; down`

Do we really want to specify actions?



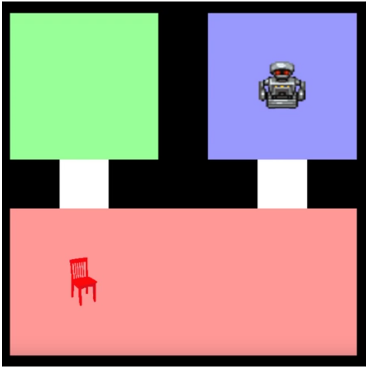
Robots need the right
semantic representation of tasks
to interact with humans effectively.

Language grounding as machine translation

Source: Natural language (English)

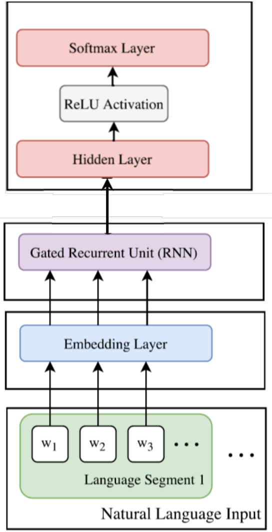
Papers	Target language / representation
[MacMahon et al.], [Chen & Mooney] [Tellex et al.], [Matuszek et al.] [Artzi & Zettlemoyer], . . .	Action space
[MacGlashan et al.], **[Arumugam et al.]**	Propositional, goal-based MDP reward function

Language grounding as machine translation



Example Command	Goal
Go to the green room.	`agentInRoom(agent0, r)` $\wedge$ `roomIsGreen(r)`
Bring the chair to the blue room.	`objInRoom(chair0, r)` $\wedge$ `roomIsBlue(r)`

Sequence classification architecture



Results

Grounding accuracy ($\approx$ 3000 sentences, $\approx$ 30 grounded tasks):

Model	Level Selection	Reward Function
IBM2 [MacGlashan et al.]	79.87%	27.26%
Single-RNN	**95.91%**	**80.46%**

RSS 2017

Results

Grounding accuracy (≈ 3000 sentences, ≈ 30 grounded tasks):

Model	Level Selection	Reward Function
IBM2 [MacGlashan et al.]	79.87%	27.26%
Single-RNN	**95.91%**	**80.46%**

Planning:

- **2-20x planning speedup**
  when grounding to appropriate hierarchy level
  in Abstract Markov Decision Process [ICAPS 2017]

RSS 2017

Results

Grounding accuracy ($\approx$ 3000 sentences, $\approx$ 30 grounded tasks):

Model	Level Selection	Reward Function
IBM2 [MacGlashan et al.]	79.87%	27.26%
Single-RNN	**95.91%**	**80.46%**

Planning:

- **2-20x planning speedup**
  when grounding to appropriate hierarchy level
  in Abstract Markov Decision Process [ICAPS 2017]
- Takes $<$ **1s on 90% of tasks**
  vs. baseline takes $>$ 20s on 50% of tasks

RSS 2017

Language grounding as machine translation

Source: Natural language (English)

Papers	Target language / representation
[MacMahon et al.], [Chen & Mooney] [Tellex et al.], [Matuszek et al.] [Artzi & Zettlemoyer], ...	Action space
[MacGlashan et al.], **[Arumugam et al.]**	Propositional, goal-based MDP reward function
[Dzifcak et al.], **[Karamcheti et al.]**	Actions <u>and</u> goals

Predicate goals

Previously: `agentInRoom(agent0, red)`

Predicate goals

Previously: `agentInRoom(agent0, red)`

Action-Oriented	Goal-Oriented
`goUp(steps)`	`agentInRoom(agent, room_attr)`
`goDown(steps)`	`objInRoom(object, room_attr)`
`goLeft(steps)`	
`goRight(steps)`	

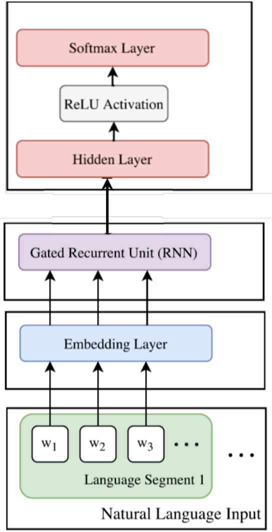
Predicate goals

Previously: `agentInRoom(agent0, red)`

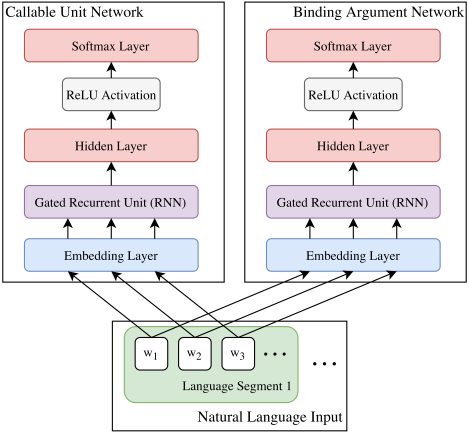
Action-Oriented	Goal-Oriented
`goUp(steps)`	`agentInRoom(agent, room_attr)`
`goDown(steps)`	`objInRoom(object, room_attr)`
`goLeft(steps)`	
`goRight(steps)`	

Natural Language	Callable Unit	Arguments
Go to the red room.	`agentInRoom`	`agent0, red`
Put the block in the green room.	`objInRoom`	`chair0, green`
Go up three spaces.	`goUp`	3

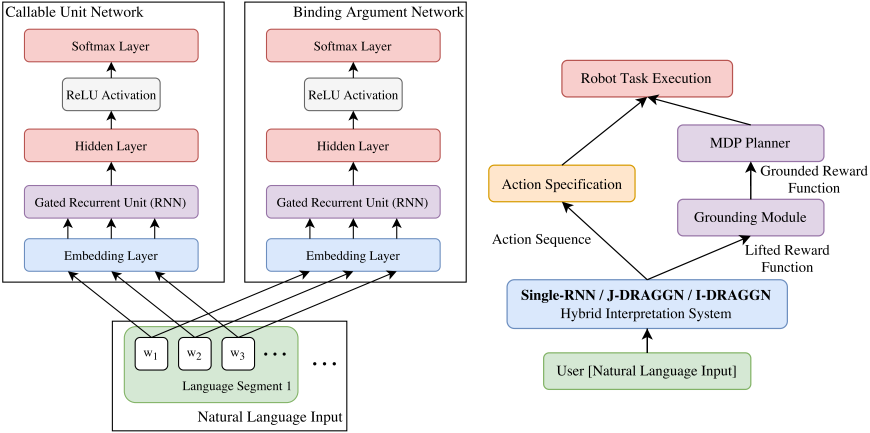
Recall: Sequence classification architecture



Factored output space



Factored output space



Language grounding as machine translation

Source: Natural language (English)

Papers	Target language / representation
[MacMahon et al.], [Chen & Mooney] [Tellex et al.], [Matuszek et al.] [Artzi & Zettlemoyer], ...	Action space
[MacGlashan et al.], **[Arumugam et al.]**	Propositional, goal-based MDP reward function
[Dzifcak et al.], **[Karamcheti et al.]**	Actions and goals
[Dzifcak et al.], [Artzi & Zettlemoyer] [Gopalan et al.]	Semantic parse (CCG)
[Raman & Kress-Gazit], **[Gopalan et al.]**	Linear temporal logic

Formal language: Linear temporal logic

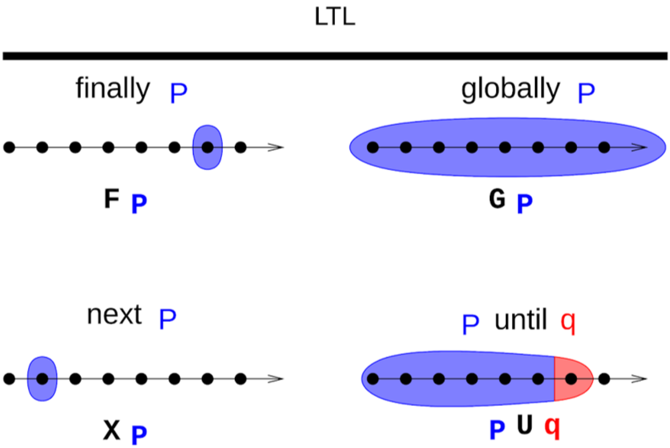
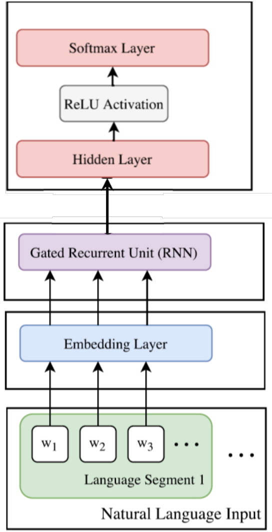


Figure by M. Pistore and M. Roveri, *Symbolic Model Checking*

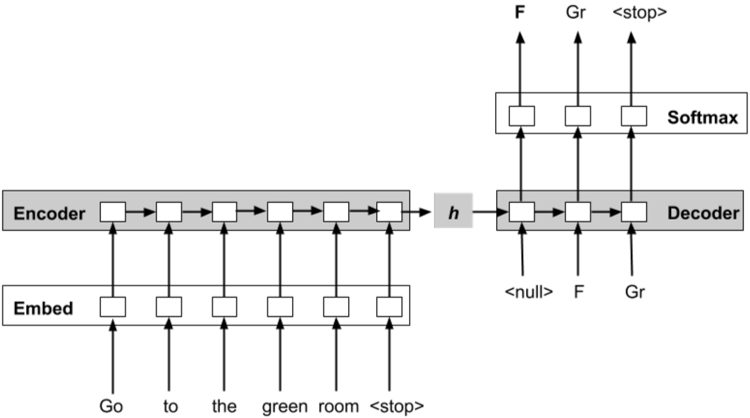
Grounding to linear temporal logic (LTL)

Example Command	Geometric LTL Expression
Go to the green room.	$\mathbf{F}Gr$
Go into the red room.	$\mathbf{F}R$
Enter blue room via green room.	$\mathbf{F}(Gr \wedge \mathbf{F}B)$
Go through the yellow or red room, and enter the blue room	$\mathbf{F}((R \vee Y) \wedge \mathbf{F}B)$
Go to the blue room but avoid the red room.	$\mathbf{F}B \wedge \mathbf{G}\neg R$
While avoiding yellow navigate to green.	$\mathbf{F}Gr \wedge \mathbf{G}\neg Y$
Scan for blocks and insert any found into bin.	$\mathbf{G}((S\mathbf{U}\neg A) \wedge \mathbf{F}A)$
Look for and pick up any non red cubes and put them in crate.	$\mathbf{G}((S\mathbf{U}\neg R) \wedge \mathbf{F}R)$

Recall: Sequence classification architecture

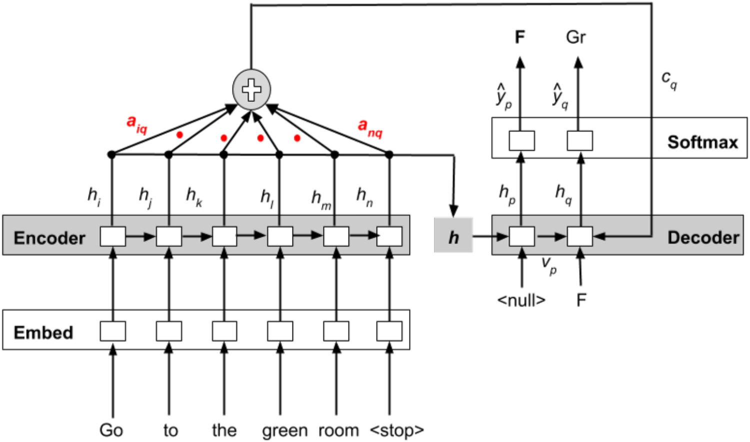


Sequence-to-sequence translation architecture

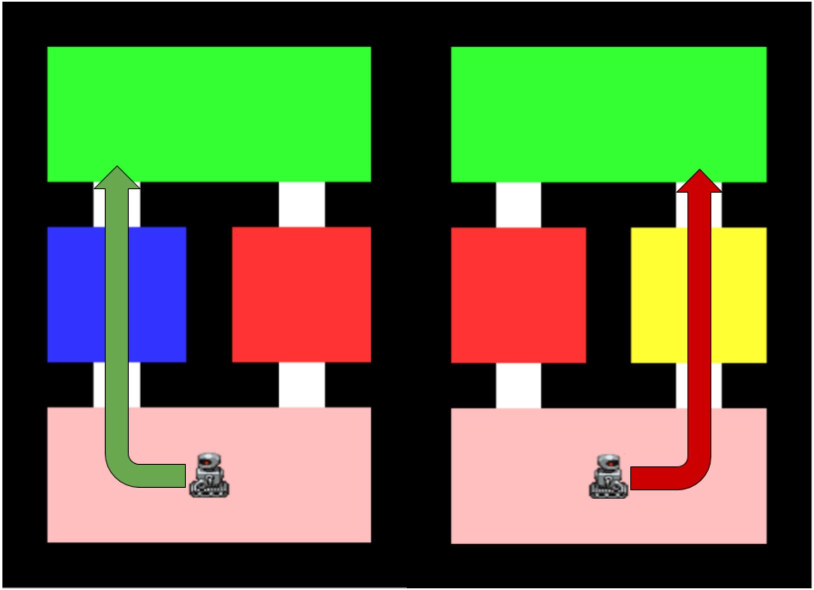


[Sutskever et al. 2014, Cho et al. 2014]
Figure adapted from S. Merity's webpage

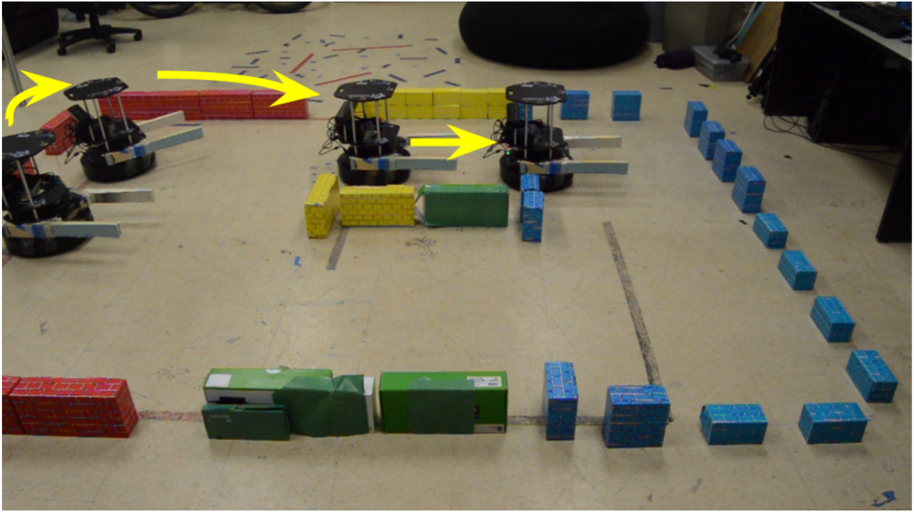
Sequence-to-sequence with **attention**



[Sutskever et al. 2014, Cho et al. 2014, Bahdanau et al. 2014]
Figure adapted from S. Merity's webpage



"Go to the blue room through the yellow room" $\mapsto$ $\mathbf{F}(Y \wedge \mathbf{F}B)$



RSS 2018: 93% accuracy on trained tasks, 60% accuracy on novel tasks ($\approx$ 4000 sentences, $\approx$ 40 grounded tasks)

Language grounding as machine translation

Source: Natural language (English)

Papers	Target language / representation
[MacMahon et al.], [Chen & Mooney] [Tellex et al.], [Matuszek et al.] [Artzi & Zettlemoyer], ...	Action space
[MacGlashan et al.], **[Arumugam et al.]**	Propositional, goal-based MDP reward function
[Dzifcak et al.], **[Karamcheti et al.]**	Actions <u>and</u> goals
[Dzifcak et al.], [Artzi & Zettlemoyer] [Gopalan et al.]	Semantic parse (CCG)
[Raman & Kress-Gazit], **[Gopalan et al.]**	Linear temporal logic

... ?

References

- Accurately and Efficiently Interpreting Human-Robot Instructions of Varying Granularities.
  Dilip Arumugam*, Siddharth Karamcheti*, Nakul Gopalan, Lawson L.S. Wong, Stefanie Tellex.
  Robotics: Science and Systems (RSS), 2017.
- A tale of two DRAGGNs: A hybrid approach for interpreting action-oriented and goal-oriented instructions.
  Siddharth Karamcheti, Edward C. Williams, Dilip Arumugam, Mina Rhee, Nakul Gopalan, Lawson L.S. Wong, Stefanie Tellex.
  Annual Meeting of the Association for Computational Linguistics (ACL) Workshop on Language Grounding for Robotics, 2017.
- Sequence-to-Sequence Language Grounding of Non-Markovian Task Specifications.
  Nakul Gopalan*, Dilip Arumugam*, Lawson L.S. Wong, Stefanie Tellex.
  Robotics: Science and Systems (RSS), 2018.
- Grounding Natural Language Instructions to Semantic Goal Representations for Abstraction and Generalization.
  Dilip Arumugam*, Siddharth Karamcheti*, Nakul Gopalan, Edward C. Williams, Mina Rhee, Lawson L.S. Wong, Stefanie Tellex.
  Autonomous Robots, 2018 (in press).

* denotes equal contribution

Acknowledgments

Nakul Gopalan

Dilip Arumugam

Siddharth Karamcheti

Stefanie Tellex

Michael L. Littman
James MacGlashan

Mina Rhee
Edward Clem Williams

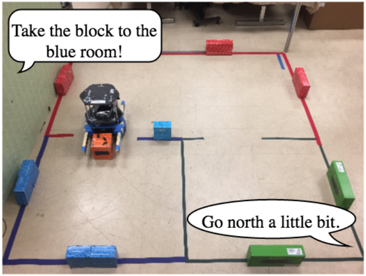
NSF
NASA
DARPA

Croucher Foundation
裘 槎 基 金 會

Conclusion

Language grounding

- Paradigm
- Data
- Representations and Models
  - Propositional goals – Sequence classification
  - Predicate goals – Factored output space
  - Linear temporal logic – Sequence-to-sequence translation



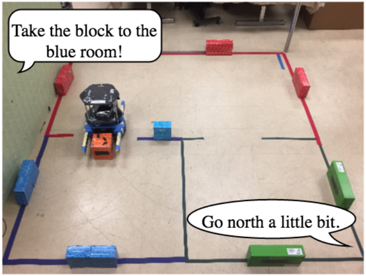
"People want to talk to the robot about everything the robot can see, and everything the robot can do."

– Stefanie Tellex

Conclusion

Language grounding

- Paradigm
- Data
- Representations and Models
    - Propositional goals –
      Sequence classification
    - Predicate goals –
      Factored output space
    - Linear temporal logic –
      Sequence-to-sequence translation



More sophisticated simulation $\Rightarrow$ Greater robustness to diversity in language?