Situated Comprehension of Action Commands

Shiwali Mohan and John E. Laird

Computer Science and Engineering
University of Michigan

June 20, 2012

# Outline

Linguistic Agents

- Use natural language for task-oriented communication.
  - Work collaboratively with humans
  - Learn from natural language communication in a complex environment

Linguistic Agents

- Use natural language for task-oriented communication.
  - Work collaboratively with humans
  - Learn from natural language communication in a complex environment
- Focus on
  - Comprehension of actions commands
    - Move the blue, large cube to the table.
  - Acquire new knowldege (linguistic/extra linguistic)
    - How actions can be performed?
    - How language is used?

Linguistic Agents

- Use natural language for task-oriented communication.
  - Work collaboratively with humans
  - Learn from natural language communication in a complex environment

- Focus on
  - Comprehension of actions commands
    - Move the blue, large cube to the table.
  - Acquire new knowldege (linguistic/extra linguistic)
    - How actions can be performed?
    - How language is used?

- NL exploits (non linguistic) context

Pick up the blue cube.
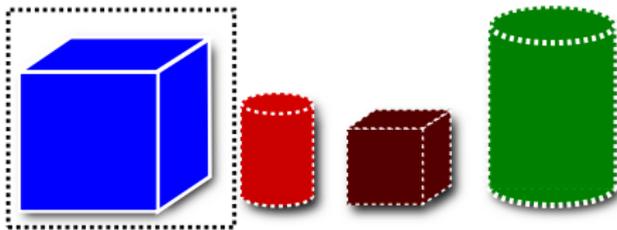
Linguistic Agents

- Use natural language for task-oriented communication.
  - Work collaboratively with humans
  - Learn from natural language communication in a complex environment

- Focus on
  - Comprehension of actions commands
    - Move the blue, large cube to the table.
  - Acquire new knowldege (linguistic/extra linguistic)
    - How actions can be performed?
    - How language is used?

- NL exploits (non linguistic) context

Pick up the large, blue cube.
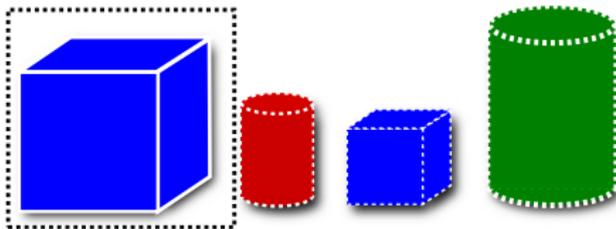
Linguistic Agents

- Use natural language for task-oriented communication.
  - Work collaboratively with humans
  - Learn from natural language communication in a complex environment

- Focus on
  - Comprehension of actions commands
    - Move the blue, large cube to the table.
  - Acquire new knowldege (linguistic/extra linguistic)
    - How actions can be performed?
    - How language is used?

- NL exploits (non linguistic) context

Pick up the cube on the left of the red cylinder.
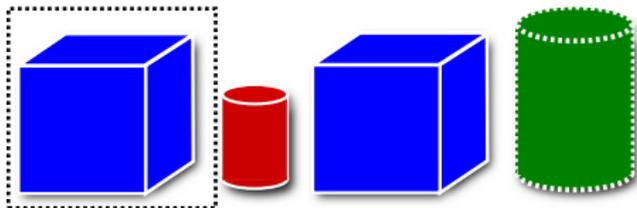
Linguistic Agents

- Use natural language for task-oriented communication.
  - Work collaboratively with humans
  - Learn from natural language communication in a complex environment
- Focus on
  - Comprehension of actions commands
    - Move the blue, large cube to the table.
  - Acquire new knowldege (linguistic/extra linguistic)
    - How actions can be performed?
    - How language is used?
- NL exploits (non linguistic) context

Put down the object.
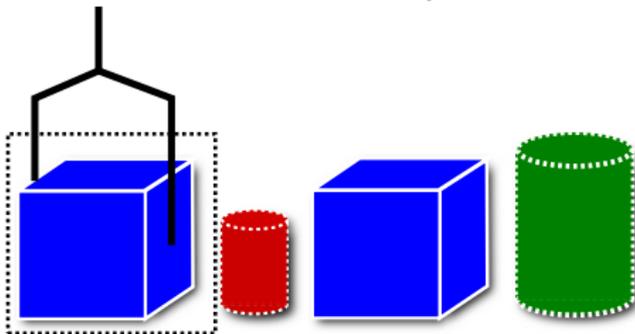
The Indexical Hypothesis for Comprehension
Glenberg and Robertson (1999)

- Sentences become meaningful by grounding their interpretation in situated action
  - *index* words and phrases to referents.
  - derive *affordances* from these referents.
  - *mesh* these affordances under syntactical constraints, physical constraints of the environment

The Indexical Hypothesis for Comprehension
Glenberg and Robertson (1999)

- Sentences become meaningful by grounding their interpretation in situated action
  - *index* words and phrases to referents.
  - derive *affordances* from these referents.
  - *mesh* these affordances under syntactical constraints, physical constraints of the environment
- In general,
  - linguistic information specifies a scene.
  - affordances specify the details for taking action.

The Indexical Hypothesis for Comprehension
Glenberg and Robertson (1999)

- Sentences become meaningful by grounding their interpretation in situated action
  - *index* words and phrases to referents.
  - derive *affordances* from these referents.
  - *mesh* these affordances under syntactical constraints, physical constraints of the environment
- In general,
  - linguistic information specifies a scene.
  - affordances specify the details for taking action.
- Indexing facilitates situated learning
  - Acquire knowledge (spatial, perceptual, procedural) through examples in the real world

The Indexical Hypothesis for Comprehension
Glenberg and Robertson (1999)

- Sentences become meaningful by grounding their interpretation in situated action
  - *index* words and phrases to referents.
  - derive *affordances* from these referents.
  - *mesh* these affordances under syntactical constraints, physical constraints of the environment
- In general,
  - linguistic information specifies a scene.
  - affordances specify the details for taking action.
- Indexing facilitates situated learning
  - Acquire knowledge (spatial, perceptual, procedural) through examples in the real world
- Our contribution
  - Formalization
  - Implementation in Soar

Using Language for Indexing

Barsalou (1999)

Using Language for Indexing

Barsalou (1999)

- Immediate Indexing
  - *"the white board"*
  - Participants are simultaneously embedded in the environment
  - Language is used to refer to objects and event in the current situation.
  - Used in learning nouns/adjectives and spatial relationships

Using Language for Indexing
### Barsalou (1999)

- Immediate Indexing
  - *"the white board"*
  - Participants are simultaneously embedded in the environment
  - Language is used to refer to objects and event in the current situation.
  - Used in learning nouns/adjectives and spatial relationships

- Displaced Indexing
  - *"the parking lot in front of BBB"*
  - Referents are not currently present
  - Language is used to refer to objects and events from prior experiences with the environment
  - *shared, componential, future*
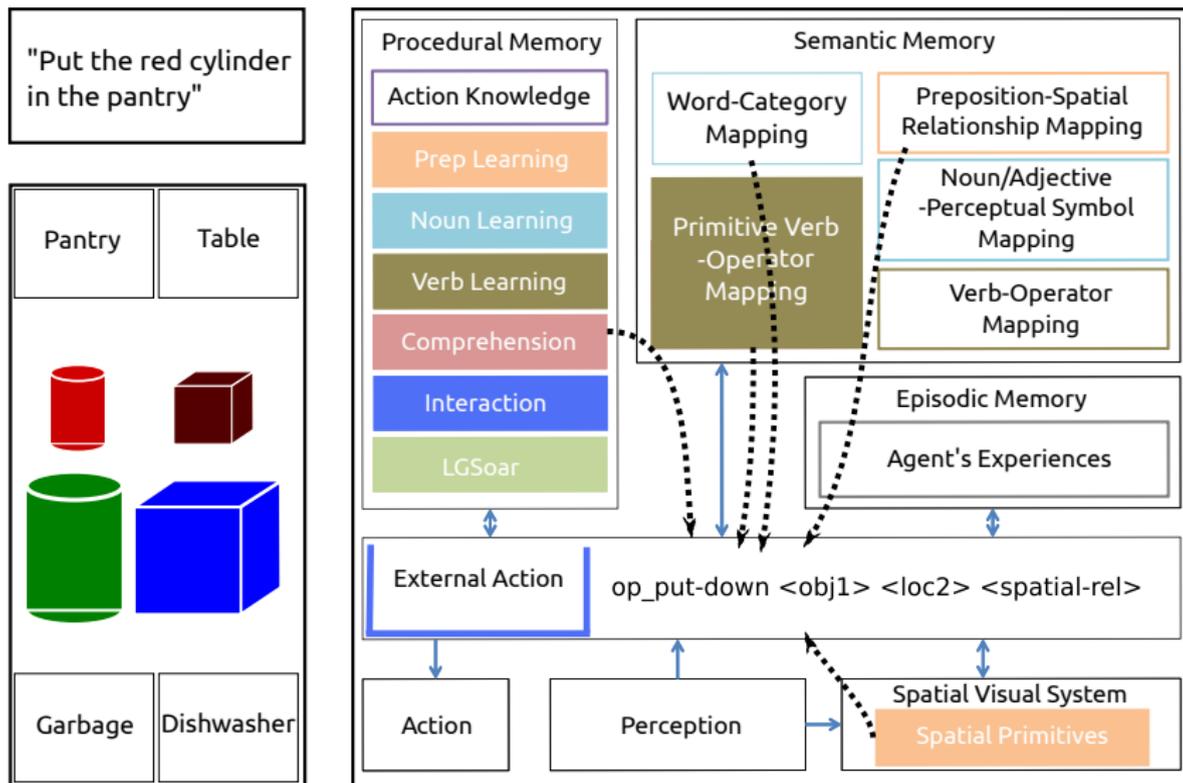
Environment and Representation

- (limited) Partial Observability
  - Distance limited sensing
  - Complete value assignment is known, if perceptible
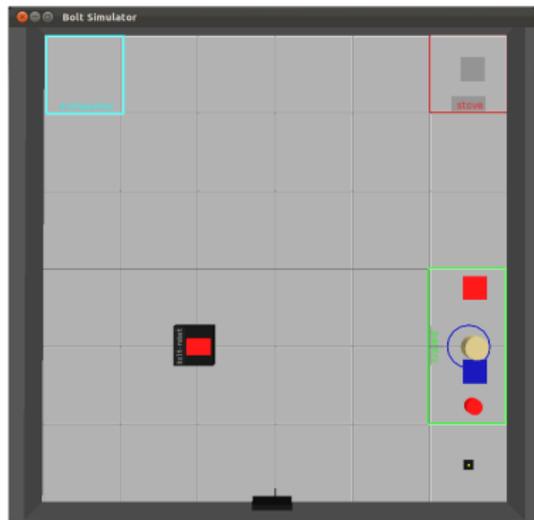
Figure: BOLT++

Environment and Representation

- (limited) Partial Observability
  - Distance limited sensing
  - Complete value assignment is known, if perceptible
- Primitive actions
  - `goto`, `pick-up(obj)`, `put-down(obj)`
  - `put-down(obj,loc)`, `put-down(obj,obj)`
  - known proposal/application/termination
  - Affordance based proposals
    - `pick-up` proposed for all 'perceptible' objects
   - `put-down` proposed for objects in the gripper
      . . .
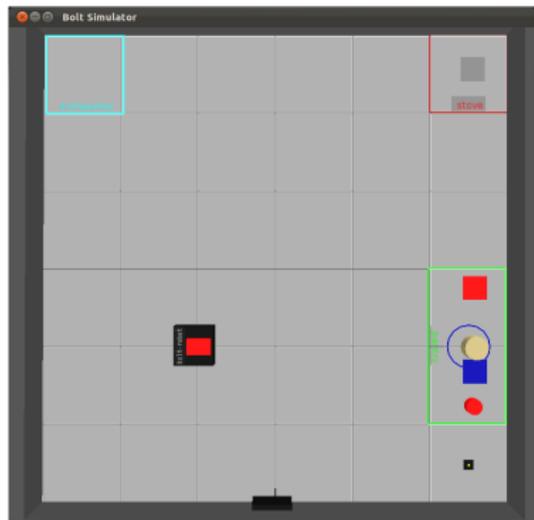
Figure: BOLT++

Background Knowledge

(Noun/Adj:Perceptual Symbols, Preposition:Spatial Relationship)

- Domain Semantic Knowledge
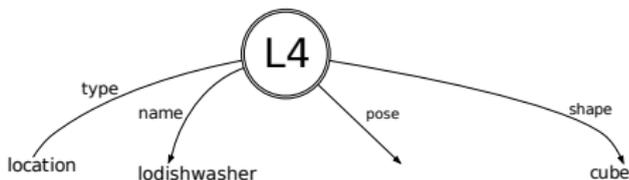
- May be acquired from previous experiences
- Allows the agent to communicate non-perceptible locations

Background Knowledge

(Noun/Adj:Perceptual Symbols, Preposition:Spatial Relationship)

- Domain Semantic Knowledge
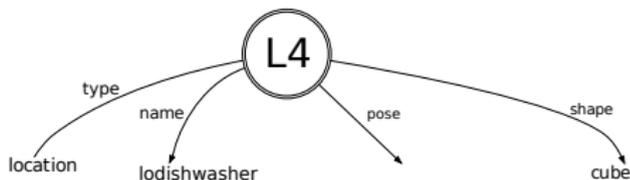
  - May be acquired from previous experiences
  - Allows the agent to communicate non-perceptible locations

- Verb-Action Mapping
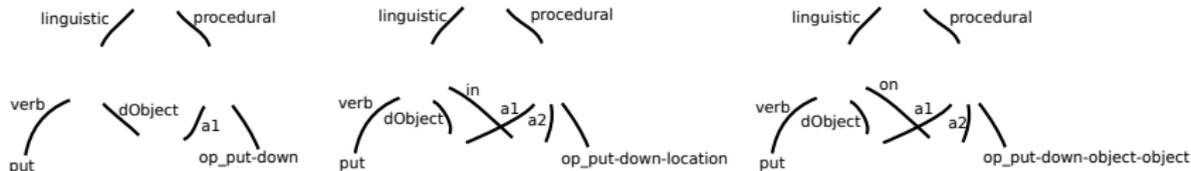
Background Knowledge

(Noun/Adj:Perceptual Symbols, Preposition:Spatial Relationship)

- Domain Semantic Knowledge
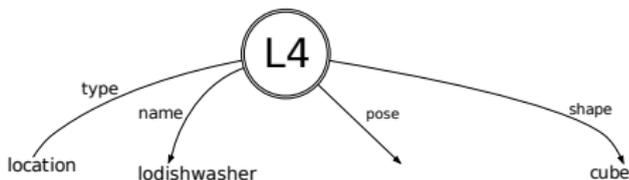
  - May be acquired from previous experiences
  - Allows the agent to communicate non-perceptible locations

- Verb-Action Mapping
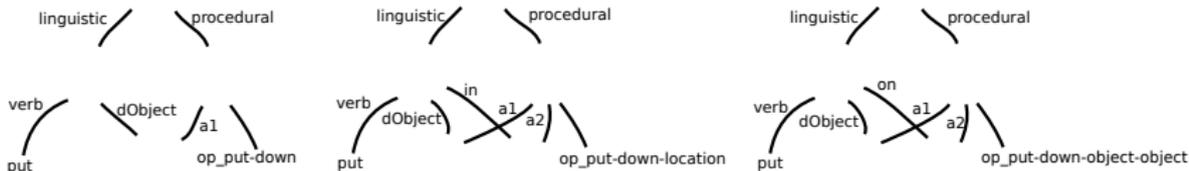
- Action Models
  - Changes in the world for primitive actions

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Index Arguments**
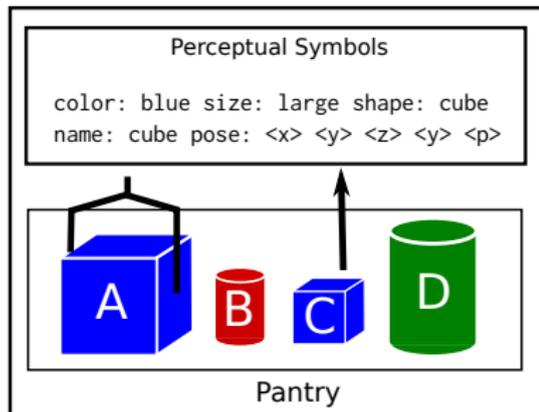
Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Index Arguments**

- Immediate Indexing
  - match description to perceptible objects
  - add all matching objects to arg-candidate set $CA_{dO} = \{A, C\}$
  - if nothing matches, displaced indexing

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Index Arguments**

- Immediate Indexing
  - match description to perceptible objects
  - add all matching objects to arg-candidate set $CA_{dO} = \{A, C\}$
  - if nothing matches, displaced indexing

- Displaced Indexing
  - match description to semantic objects
  - add all to arg-candidate set $CA_{in} = \{dishwasher\}$

Perceptual Symbols

color: blue size: large shape: cube
name: cube pose: <x> <y> <z> <y> <p>

A   B   C   D

Pantry

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Index Arguments**

- Immediate Indexing
  - match description to perceptible objects
  - add all matching objects to arg-candidate set $CA_{dO} = \{A, C\}$
  - if nothing matches, displaced indexing

- Displaced Indexing
  - match description to semantic objects
  - add all to arg-candidate set $CA_{in} = \{dishwasher\}$

- If $CA = \phi$, communicate

Perceptual Symbols

color: blue size: large shape: cube
name: cube pose: <x> <y> <z> <y> <p>

A   B   C   D

Pantry

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Index Arguments**

- Immediate Indexing
  - match description to perceptible objects
  - add all matching objects to arg-candidate set $CA_{dO} = \{A, C\}$
  - if nothing matches, displaced indexing

- Displaced Indexing
  - match description to semantic objects
  - add all to arg-candidate set $CA_{in} = \{dishwasher\}$

- If $CA = \phi$, communicate

- Displaced indexing into episodic memory?

Perceptual Symbols

color: blue size: large shape: cube
name: cube pose: <x> <y> <z> <y> <p>

A   B   C   D

Pantry
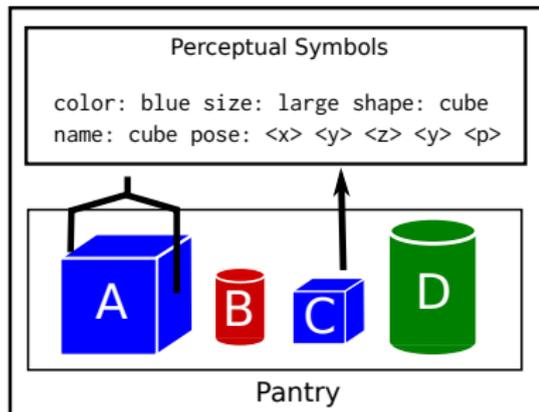
Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Index Verb**
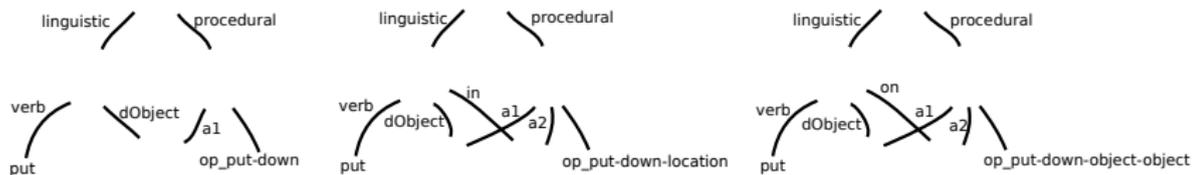
Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Index Verb**
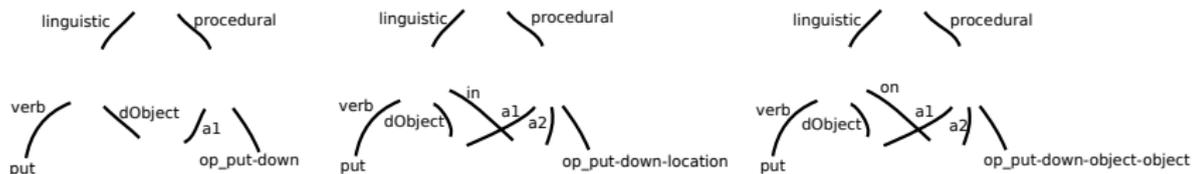
- Possible Interpretations CI

  ```
  op_put-down-object-location [A] [dishwasher]
  op_put-down-object-location [C] [dishwasher]
  ```

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Meshing**

- Generate all possible interpretations CI
  - Using all candidate objects, CA for all arguments of the verb

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Meshing**

- Generate all possible interpretations CI
  - Using all candidate objects, CA for all arguments of the verb
- Generate all possible actions PA
  - Using affordances, physical constraints
  - Soar proposals
  - `op_put-down-object-location [A] [dishwasher]`

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Meshing**

- Generate all possible interpretations CI
  - Using all candidate objects, CA for all arguments of the verb
- Generate all possible actions PA
  - Using affordances, physical constraints
  - Soar proposals
  - `op_put-down-object-location [A] [dishwasher]`
- Compute the intersection
  - $CI \cap PA$

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Meshing**

- Generate all possible interpretations CI
  - Using all candidate objects, CA for all arguments of the verb
- Generate all possible actions PA
  - Using affordances, physical constraints
  - Soar proposals
  - `op_put-down-object-location [A] [dishwasher]`
- Compute the intersection
  - $CI \cap PA$
- Excute action if only one element
  - Ground preposition: predicate projection, tracking

Comprehension of Action Commands

"Put a blue cube in the dishwasher"
Phase: **Meshing**

- Generate all possible interpretations CI
  - Using all candidate objects, CA for all arguments of the verb
- Generate all possible actions PA
  - Using affordances, physical constraints
  - Soar proposals
  - `op_put-down-object-location [A] [dishwasher]`
- Compute the intersection
  - $CI \cap PA$
- Excute action if only one element
  - Ground preposition: predicate projection, tracking
- Communicate if empty or multiple elements

Linguistic Capabilities

Linguistic Capabilities

- Situated Referent Resolution
  - Using perceptual information
    - Use of most distinctive description given the perceptual state
  - Using semantic knowledge
    - mapping *the dishwasher* to a semantic object
  - Using procedural knowledge
    - Can only `put down` the object in gripper

Linguistic Capabilities

- Situated Referent Resolution
  - Using perceptual information
    - Use of most distinctive description given the perceptual state
  - Using semantic knowledge
    - mapping *the dishwasher* to a semantic object
  - Using procedural knowledge
    - Can only `put down` the object in gripper

- Situated Action Resolution
  - Using the argument structure

Future Work

- Resolve PP phrase ambiguities using context
- Use episodic memory for displaced indexing
- Co reference resolution
- Exploiting other context
  - linguistic, interaction, procedural, perceptual, semantic, episodic

Nuggets and Coal

- Nuggets
  - Proposed a scheme for comprehending action commands
  - Uses real-world context
- Coal
  - Only action-commands!
  - Evaluation?
    - Information theoritic analysis