

# Acquiring Grounded Representations of Words with Situated Interactive Instruction

**Shiwali Mohan, Aaron Mininger, James Kirk, John Laird**

Computer Science and Engineering  
University of Michigan

December 6, 2012

# Goal

*Long-living cognitive agents that acquire diverse types of knowledge from situated interactive instruction.*

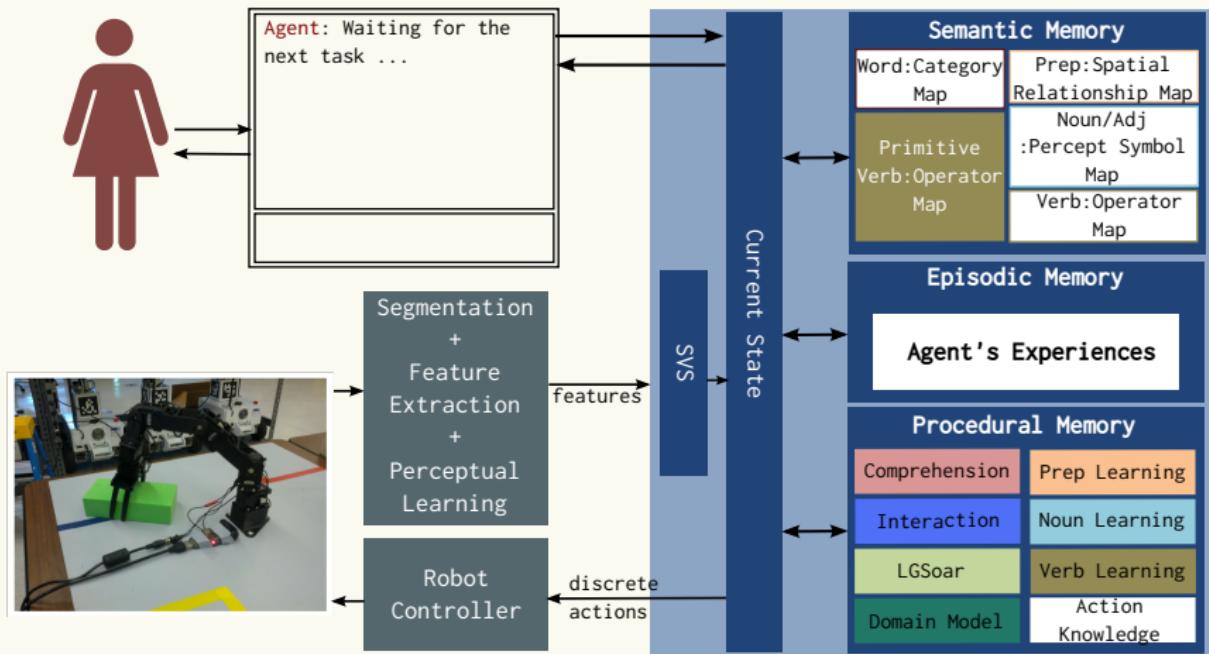
- Environment and task
  - robotic agent, human instructor embedded in real world tasks
  - knowledge-level linguistic communication
  - instruction accompanies sensory, control information
- Diverse types of knowledge
  - object classification
  - spatial relationships
  - task decomposition and execution
  - linguistic knowledge - nouns, adjectives, prepositions, verbs

# Learning from Situated Interactive Instruction

- Situated
  - instructions are provided within the context of the task
  - context is derived from structured interactions, perceptual, spatial, semantic, linguistic knowledge
- Interactive
  - flexible communication between instructor and the agent
  - onus of learning is distributed
  - agent actively asks for information
  - instructor can refine concepts
- Learning
  - incremental and assimilative, fast, online

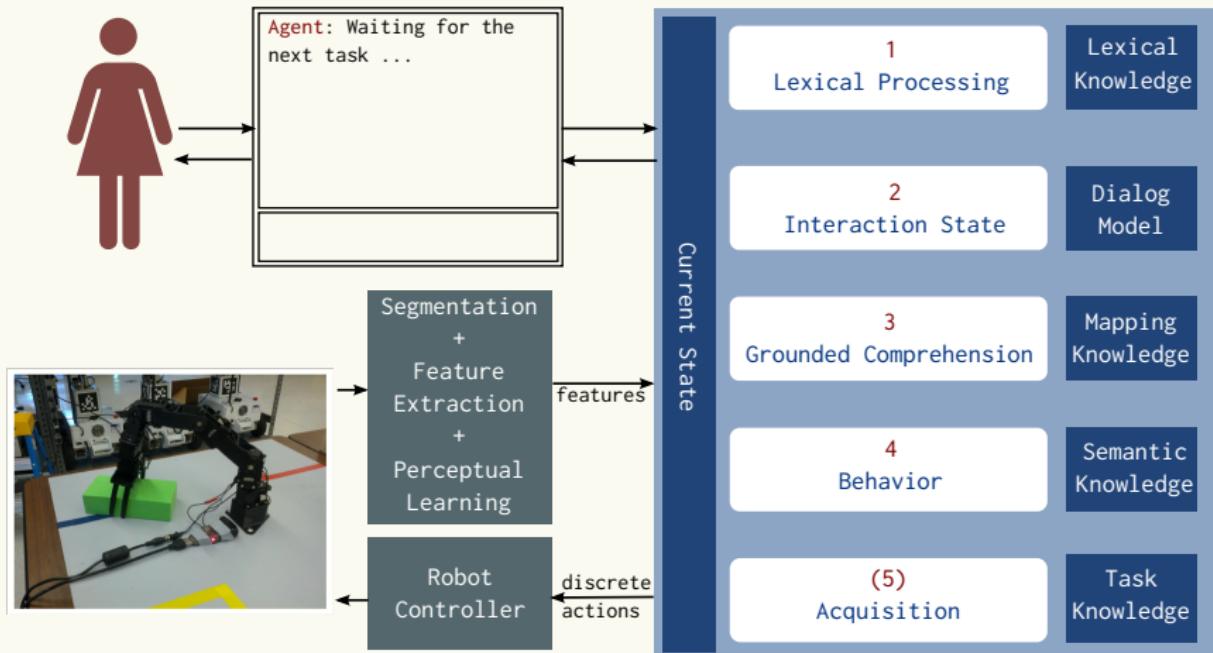
# System Overview

Agent Architecture - Soar (Laird, 2012)



# Process Overview

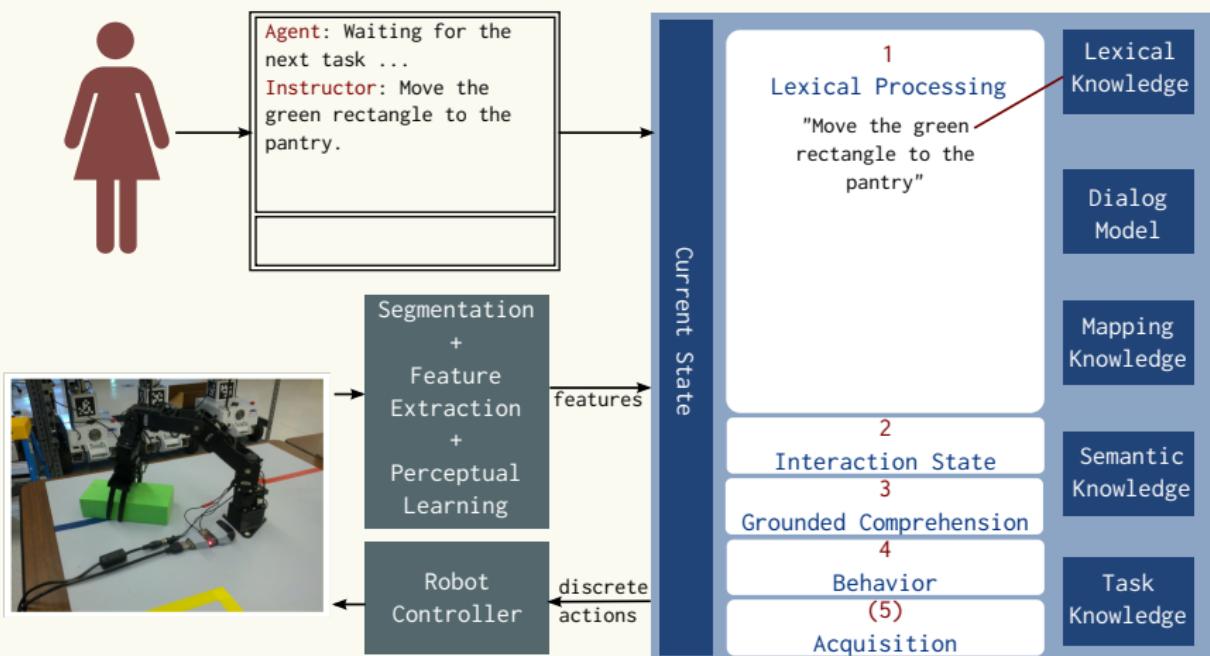
Agent Architecture - Soar (Laird, 2012)



# Demo

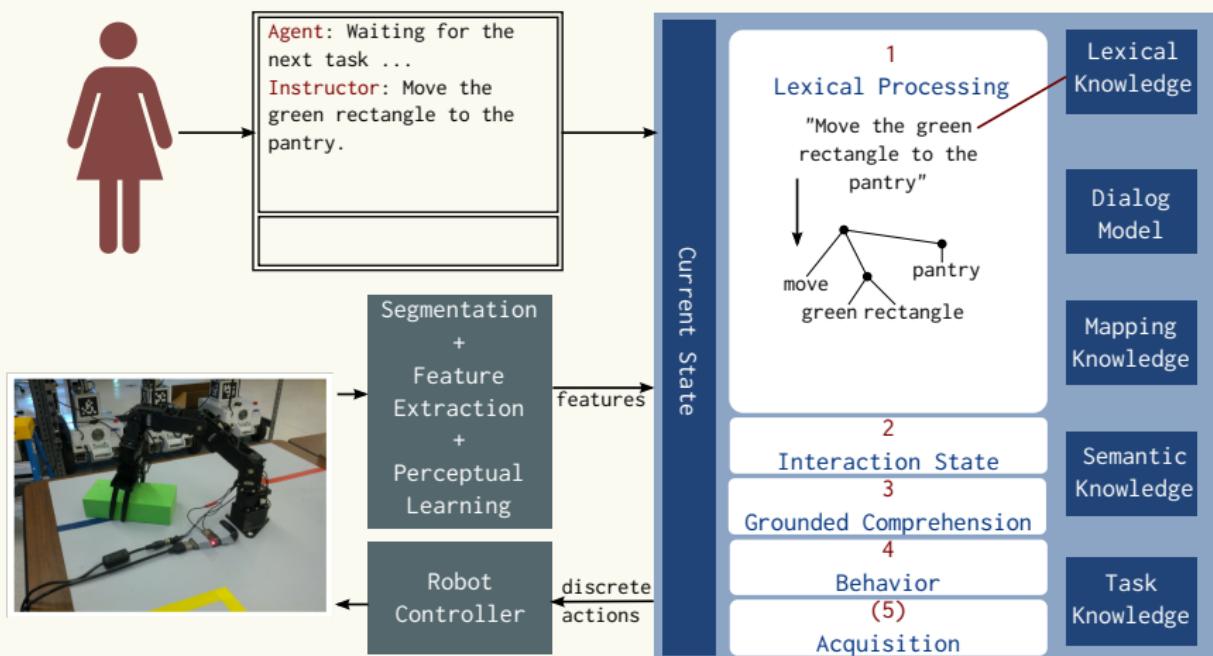
# Process Cycle

## Phase I: Lexical processing



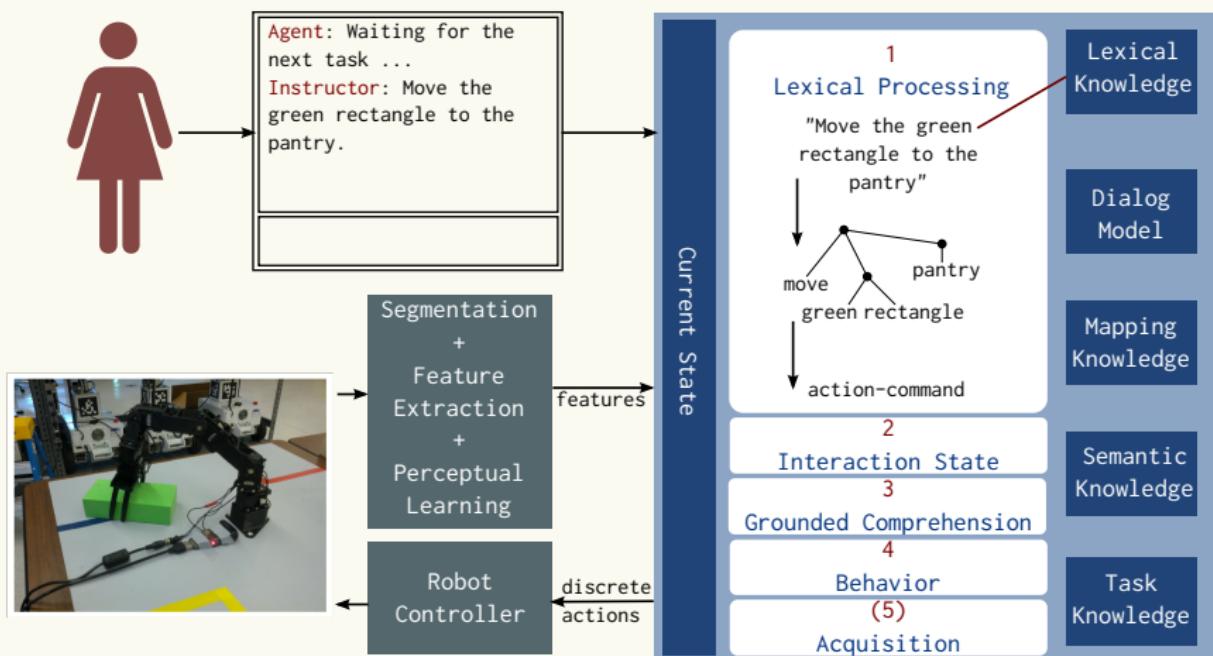
# Process Cycle

## Phase I: Lexical processing



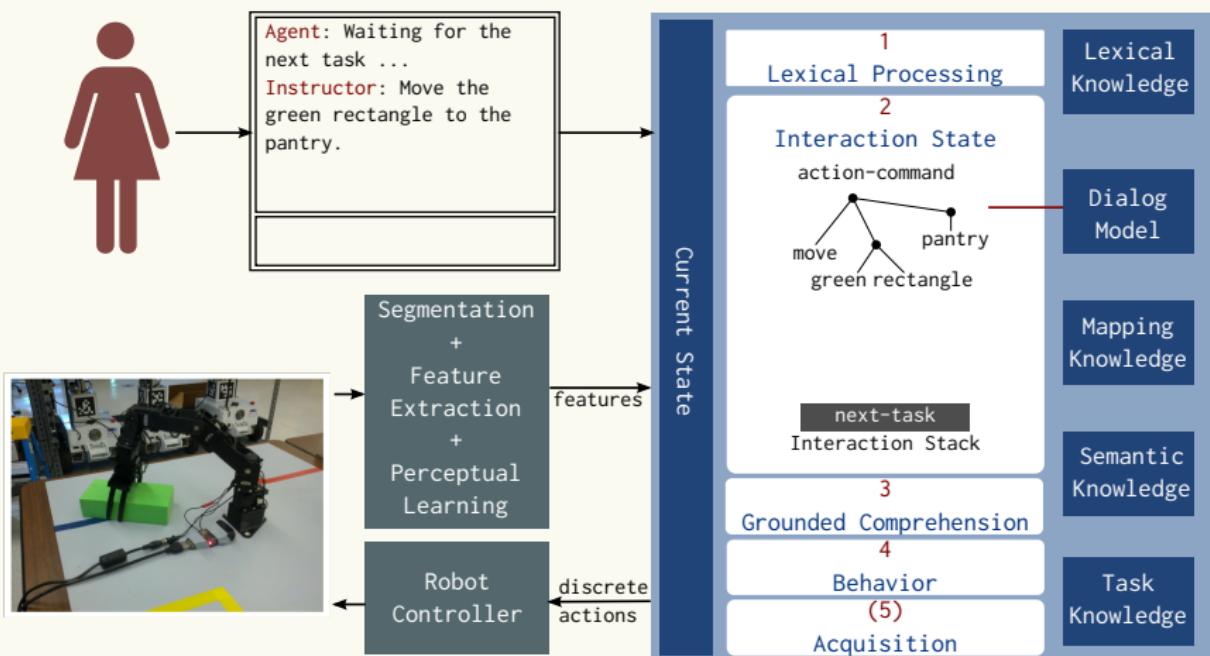
# Process Cycle

## Phase I: Lexical processing



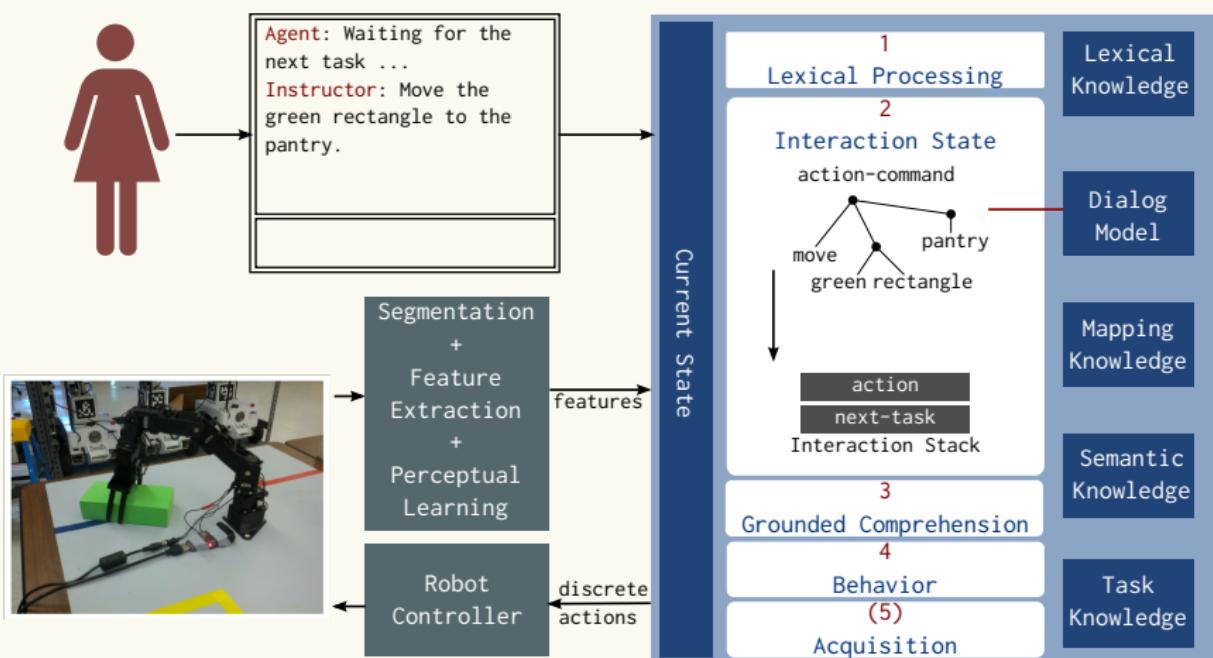
# Process Cycle

## Phase II: Interaction state management



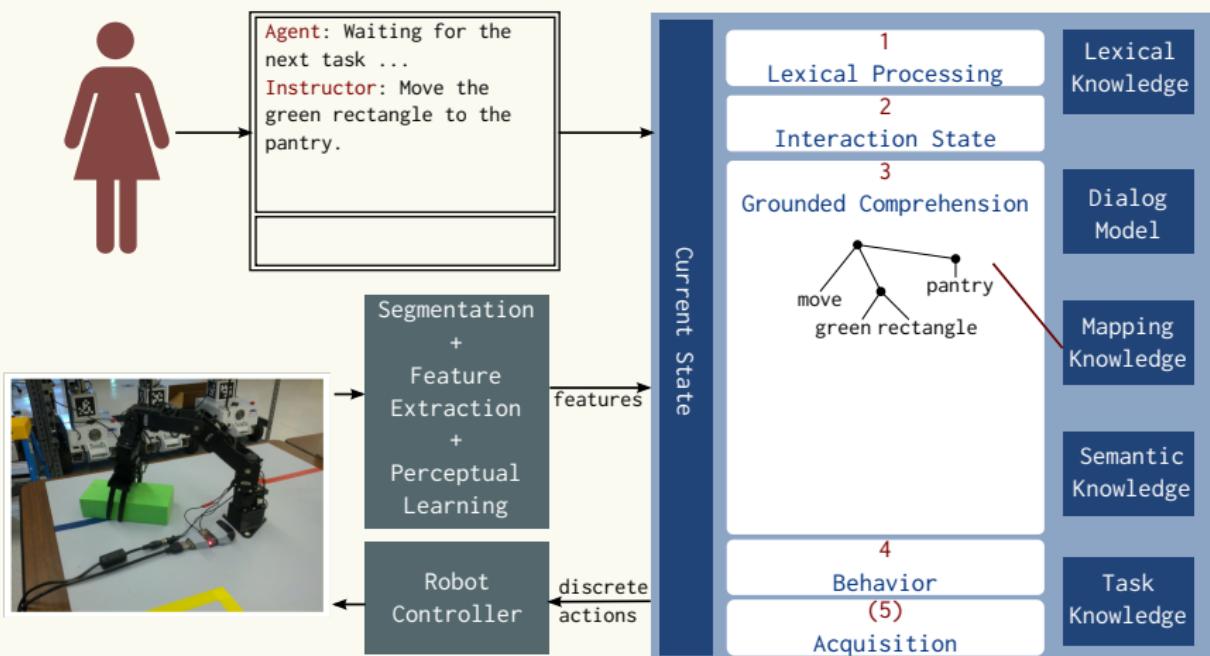
# Process Cycle

## Phase II: Interaction state management



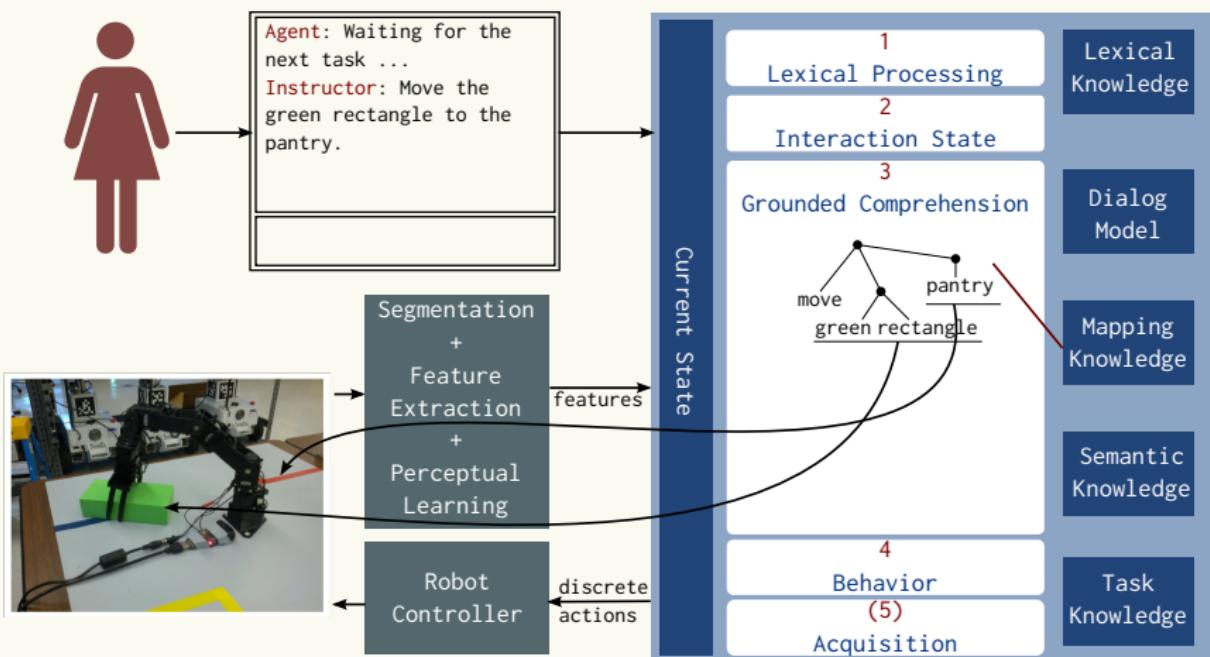
# Process Cycle

## Phase III: Grounded comprehension



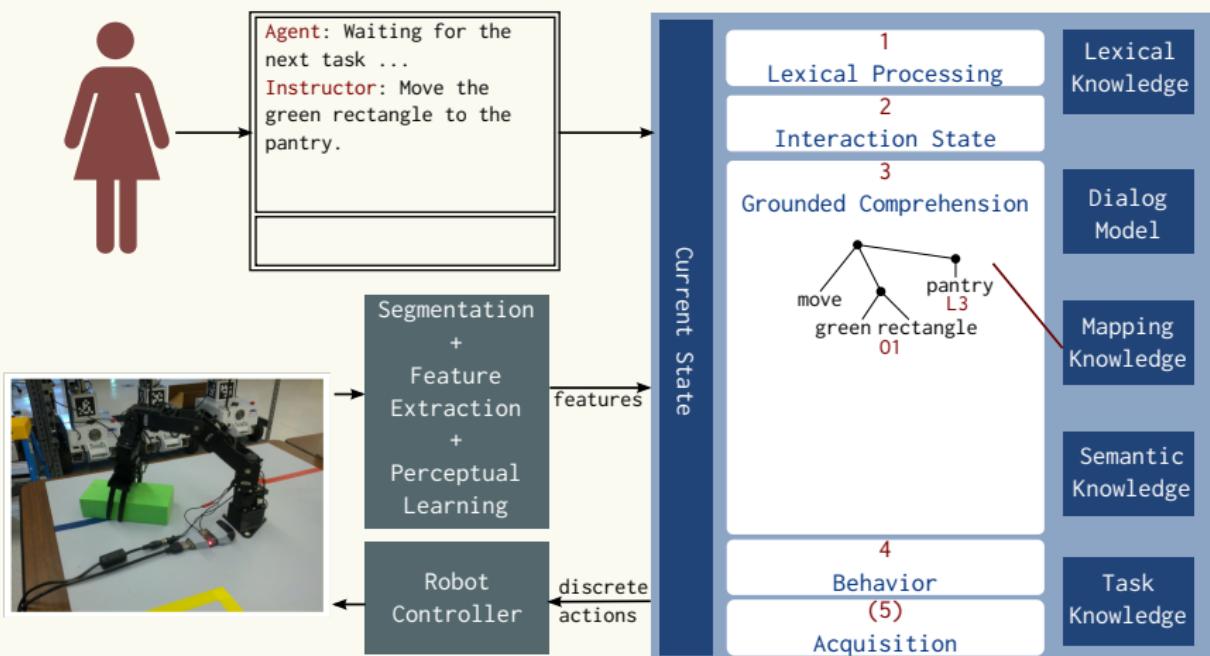
# Process Cycle

## Phase III: Grounded comprehension



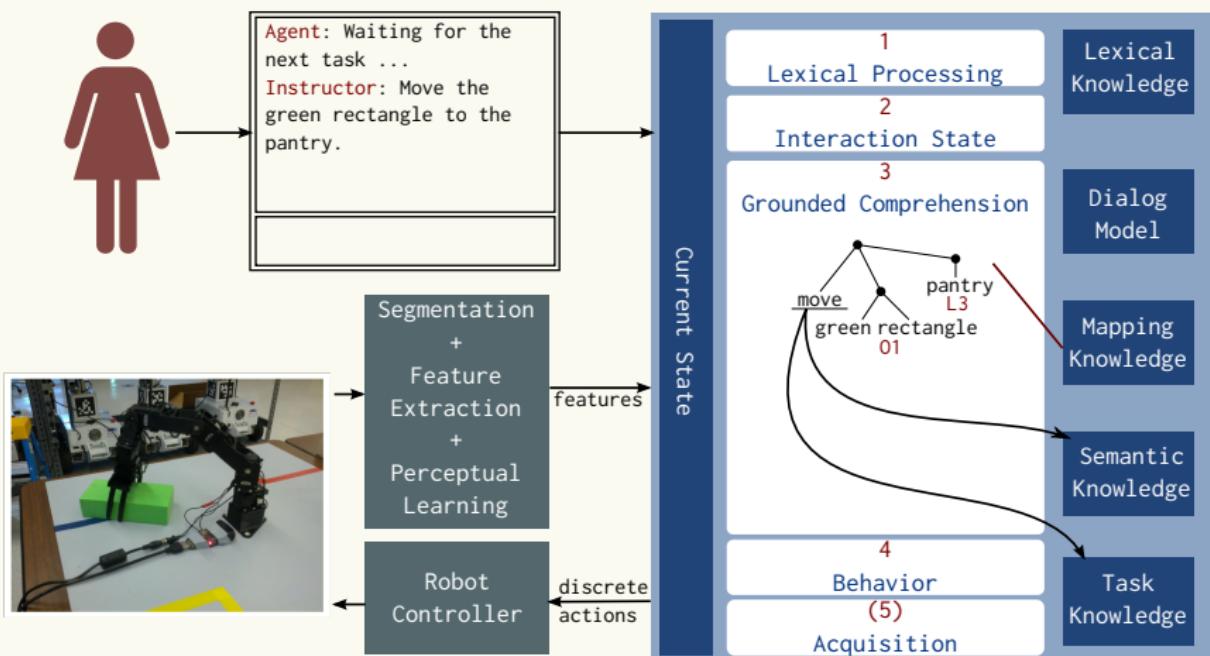
# Process Cycle

## Phase III: Grounded comprehension



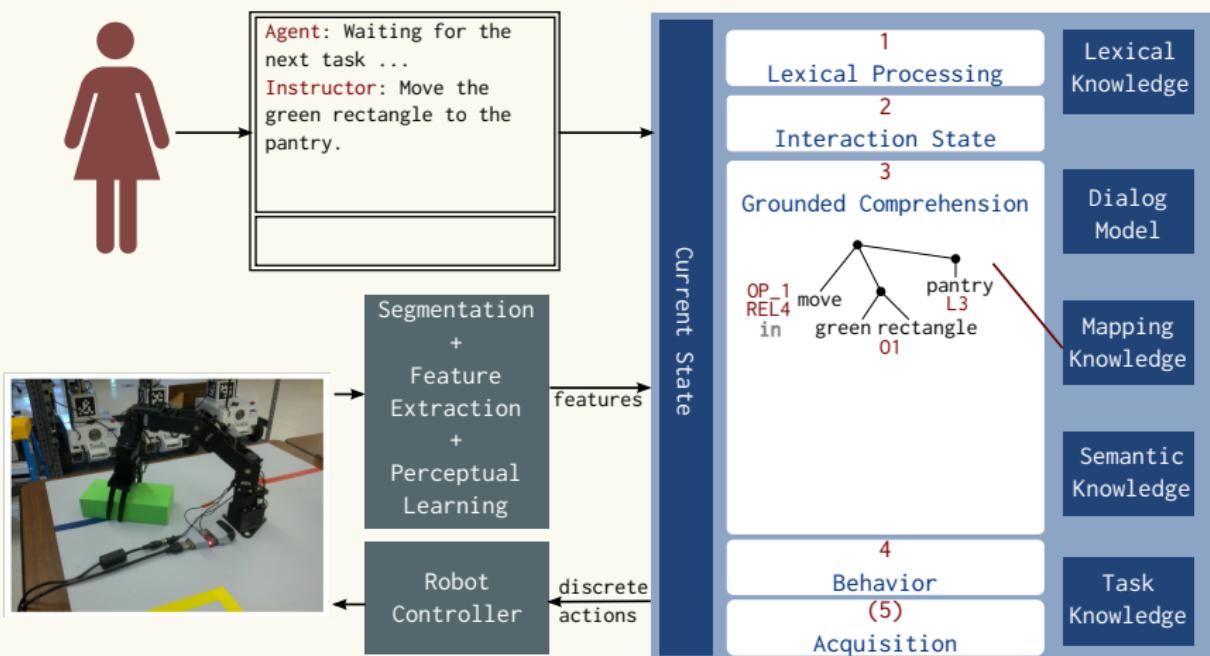
# Process Cycle

## Phase III: Grounded comprehension



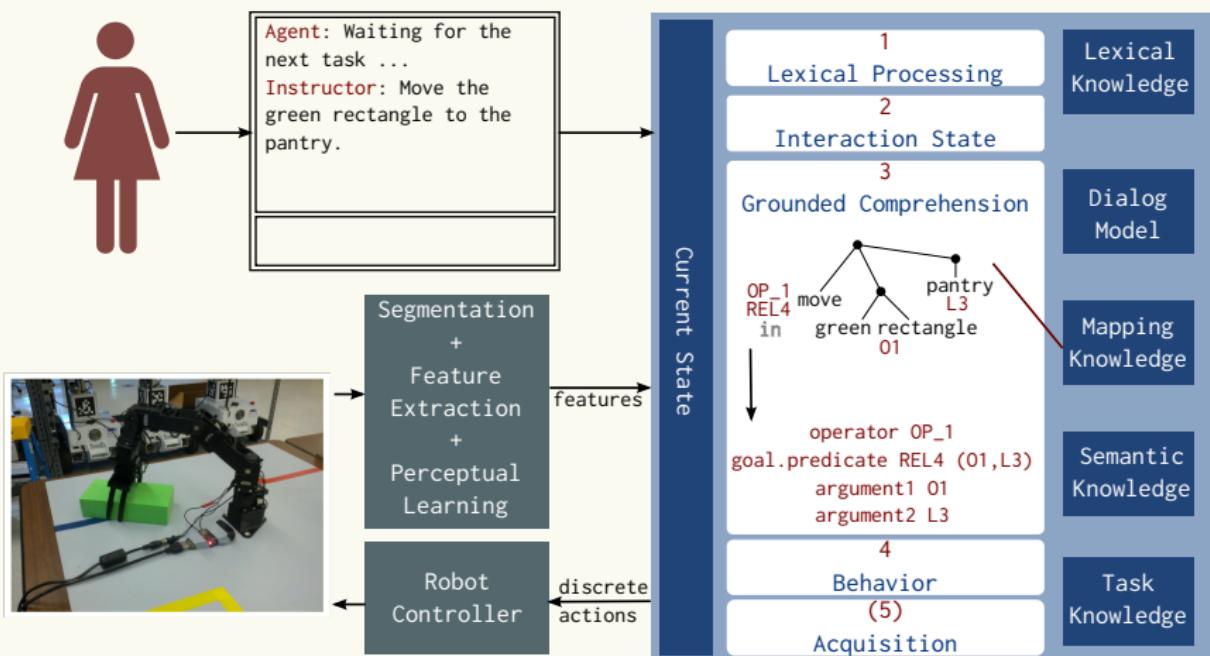
# Process Cycle

## Phase III: Grounded comprehension



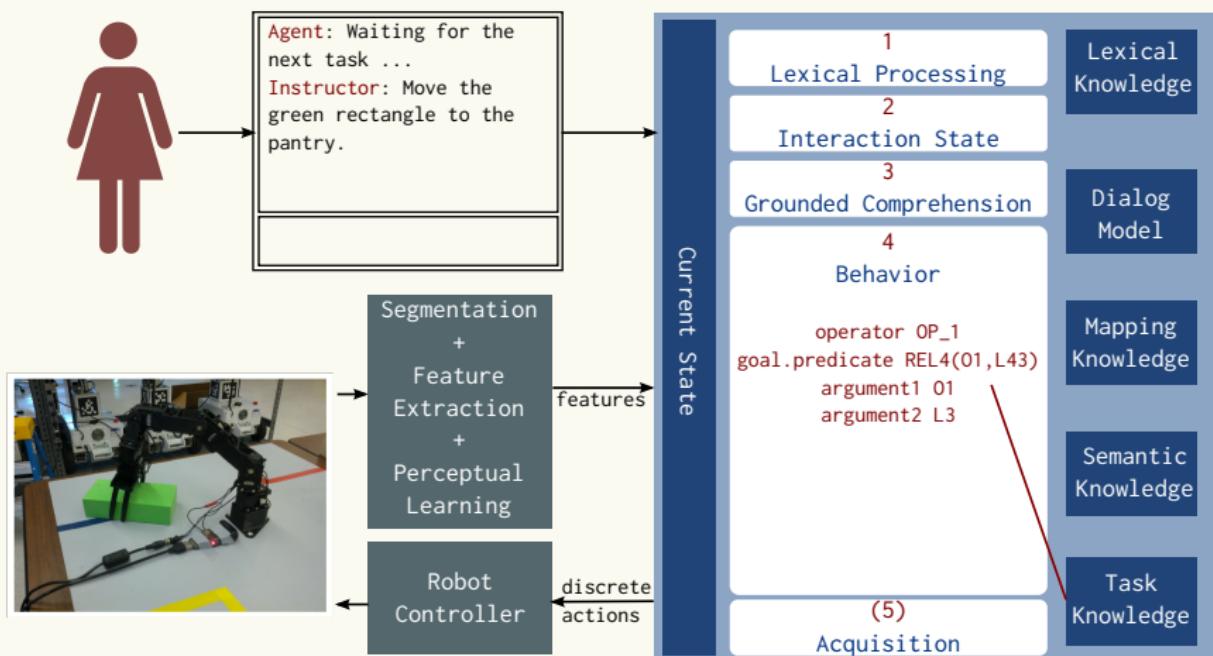
# Process Cycle

## Phase III: Grounded comprehension



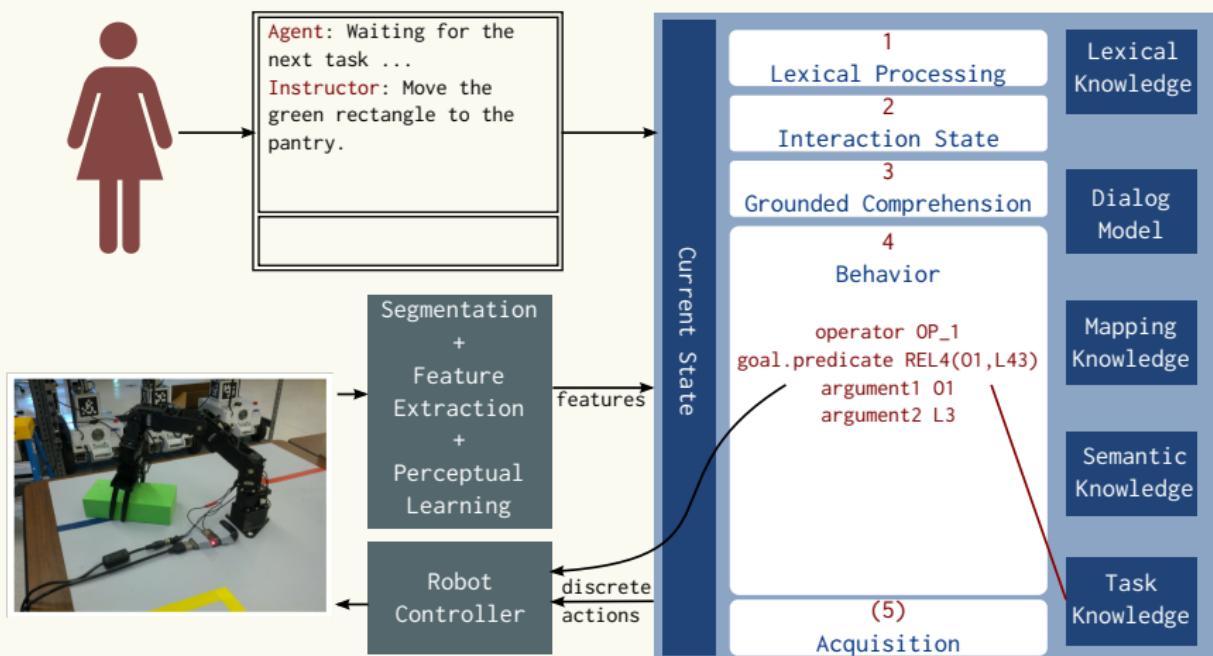
# Process Cycle

## Phase IV: Behavior



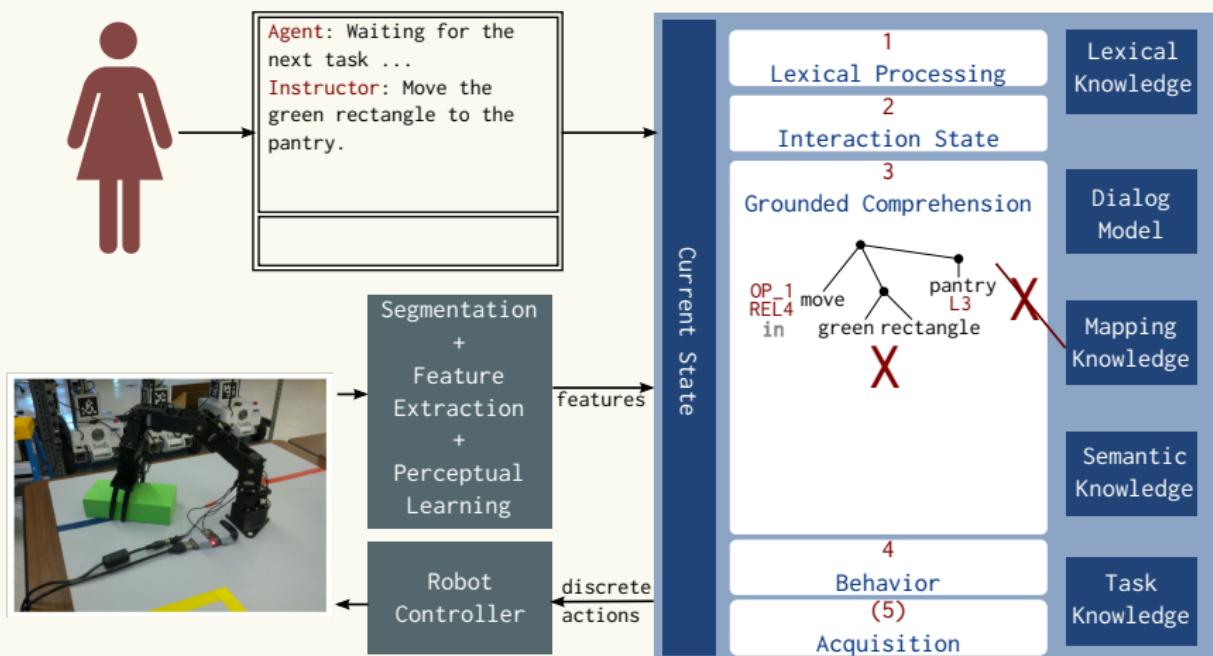
# Process Cycle

## Phase IV: Behavior



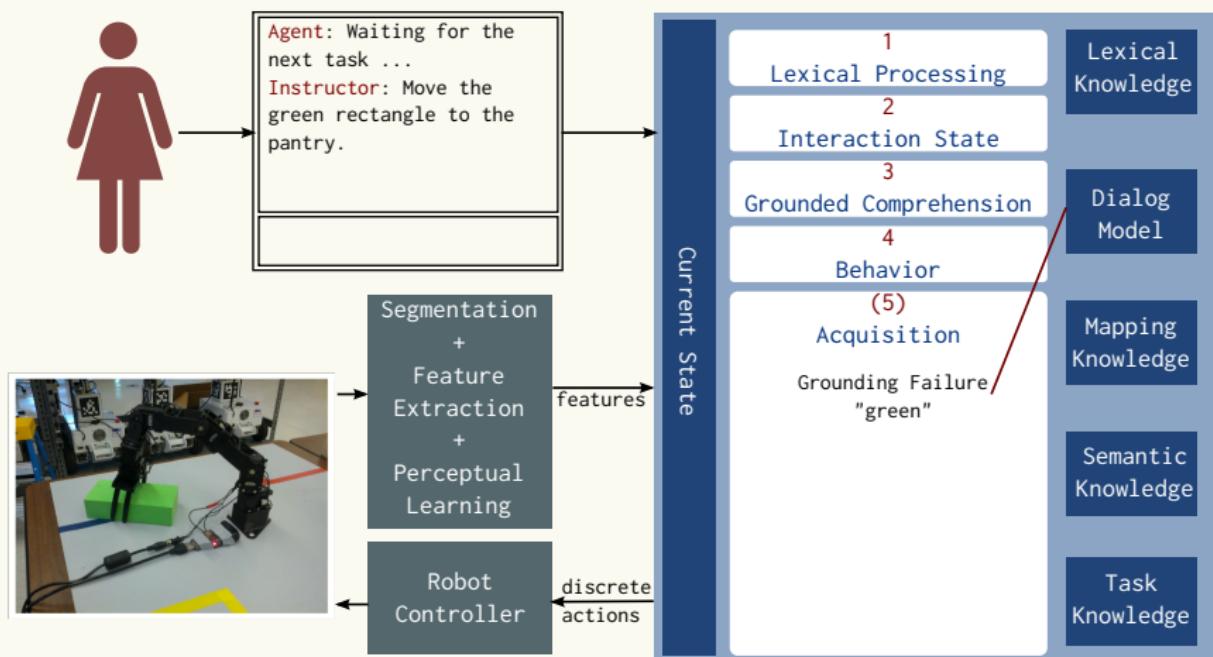
# Process Cycle

## Phase III: Grounded comprehension (failure)



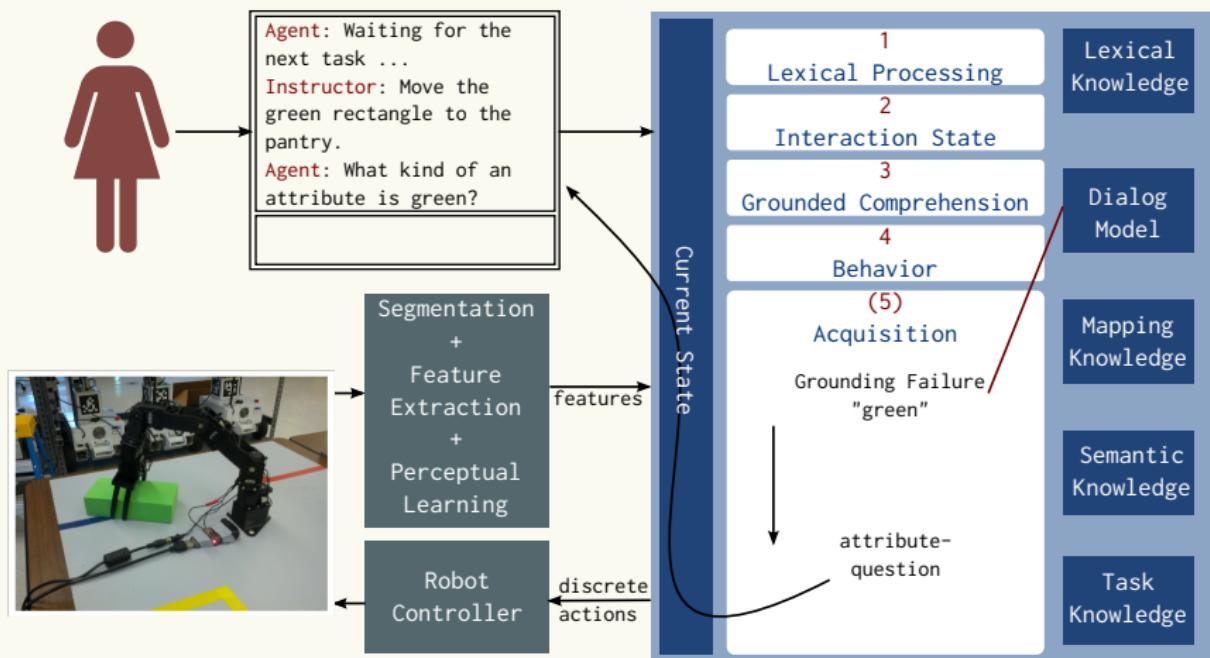
# Process Cycle

## Phase V: Acquisition (categorization)



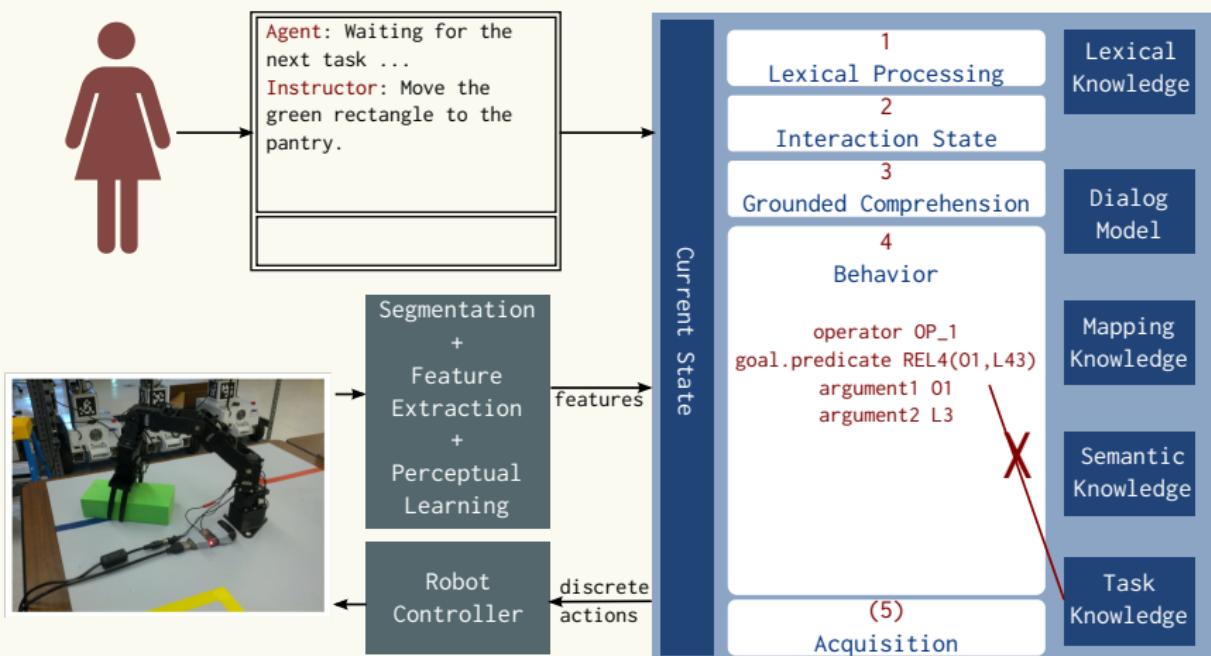
# Process Cycle

## Phase V: Acquisition (categorization)



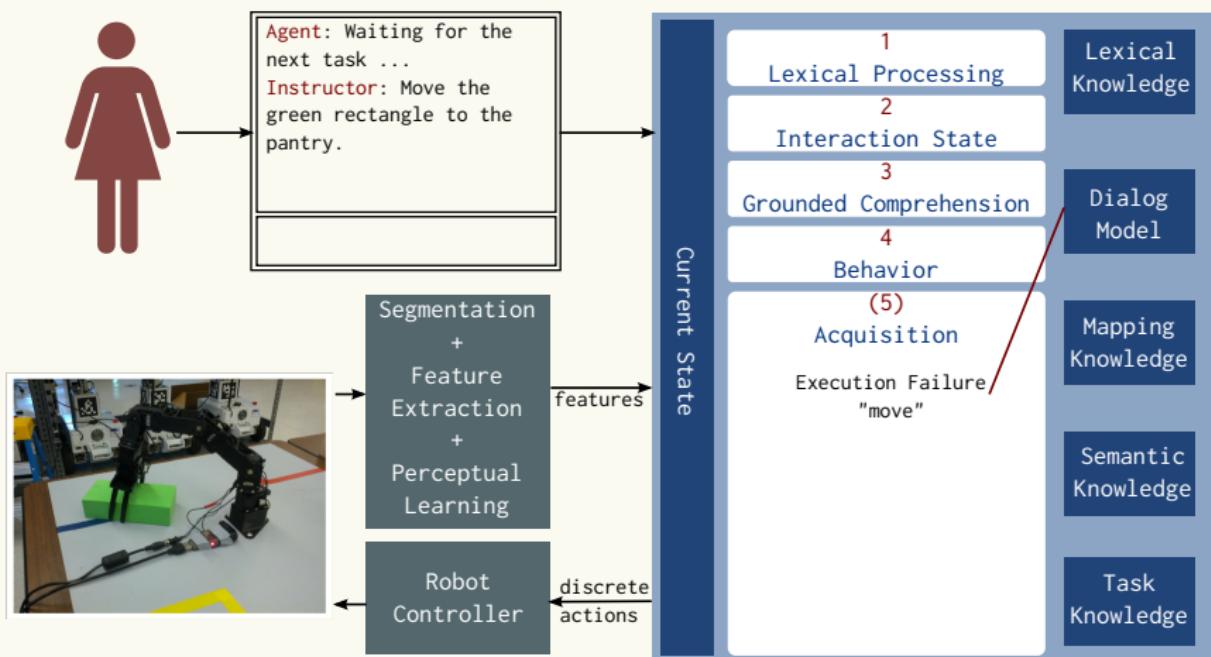
# Process Cycle

## Phase IV: Behavior (failure)



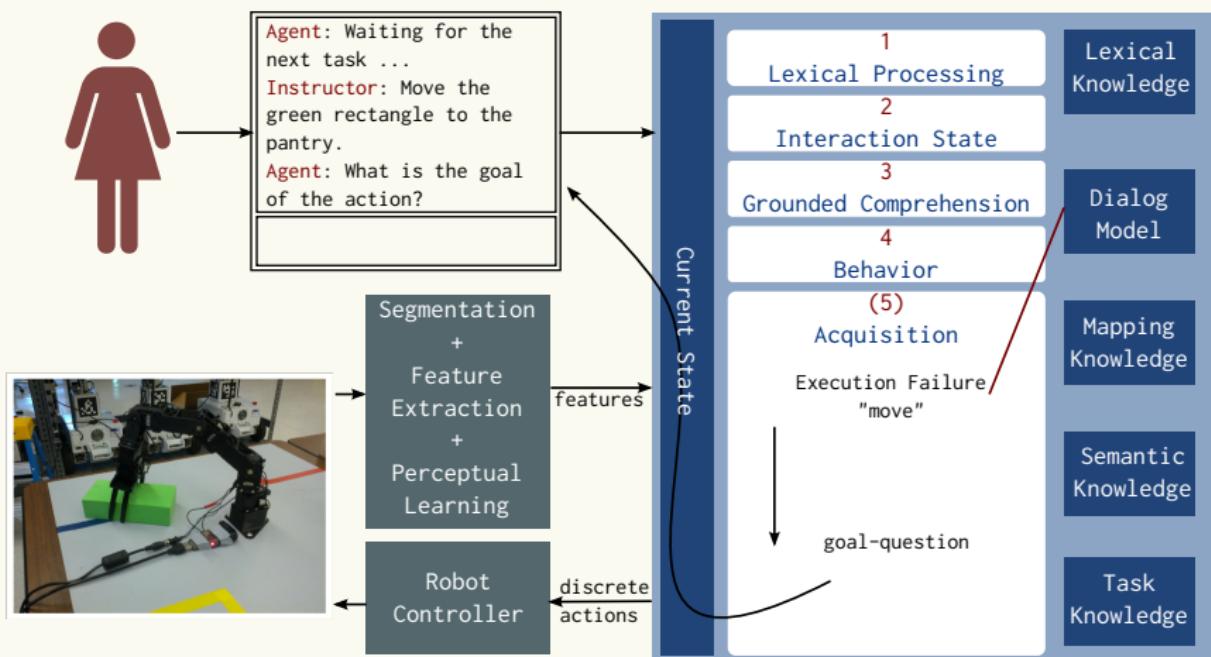
# Process Cycle

## Phase V: Acquisition (goal)



# Process Cycle

## Phase V: Acquisition (goal)



# Verb Acquisition

3 types of knowledge

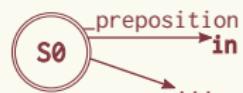
- Linguistic Map
  - the verb and the argument structure: *move the green rectangle to the pantry* → action concept network
  - extracted from the syntactic parse
  - stored in semantic memory
- Semantic
  - action concept network
  - verb termination conditions (goal predicates)
  - association with objects, prepositions
  - explicit, declarative description from the instructor: "*The goal is the green rectangle is in the pantry*".
  - semantic memory
- Procedural
  - rules to select actions that execute behavior
  - acquired through explanation based generalization (using forward projection)
  - procedural memory

# Verb Acquisition

## Existing Knowledge

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

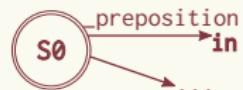
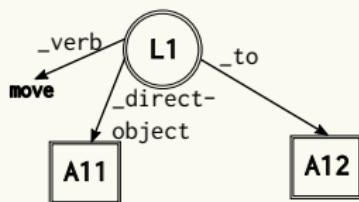


# Verb Acquisition

## Syntactical Knowledge

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

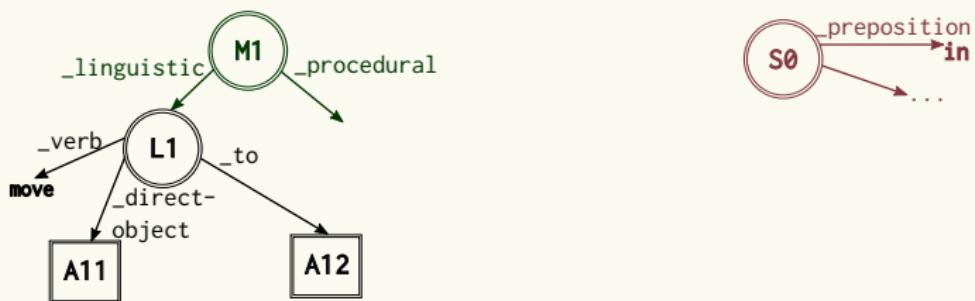


# Verb Acquisition

## Linguistic Map

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

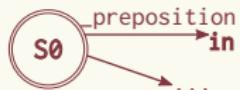
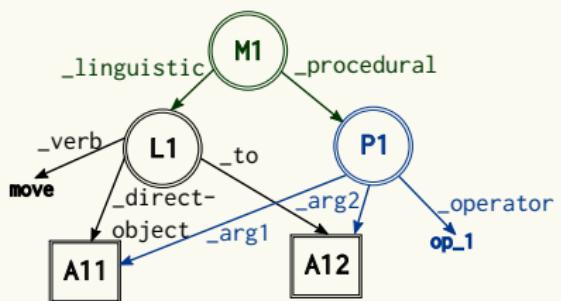


# Verb Acquisition

## Linguistic Map

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.



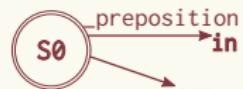
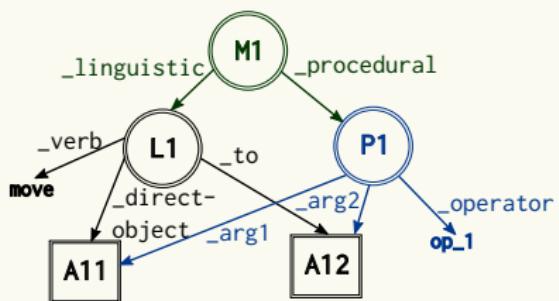
# Verb Acquisition

## Linguistic Map

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?



# Verb Acquisition

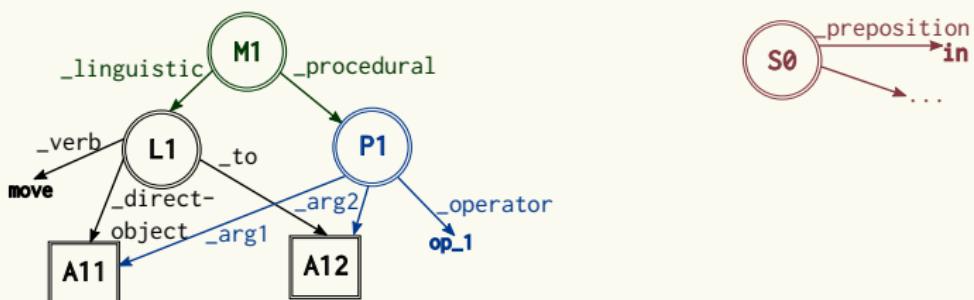
## Linguistic Map

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.



# Verb Acquisition

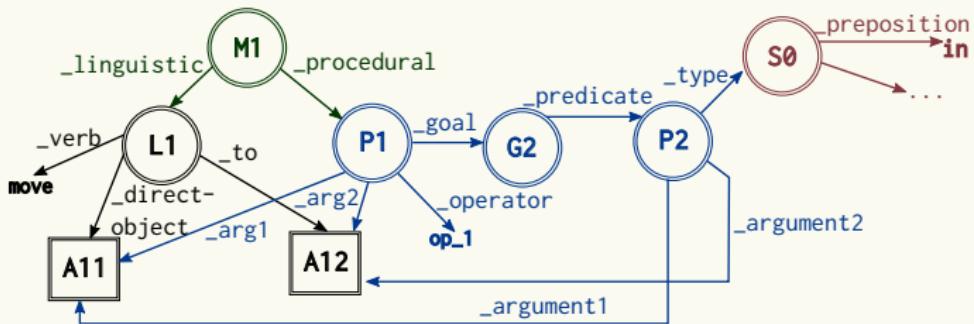
## Action Concept Network

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.



# Verb Acquisition

## Action Concept Network

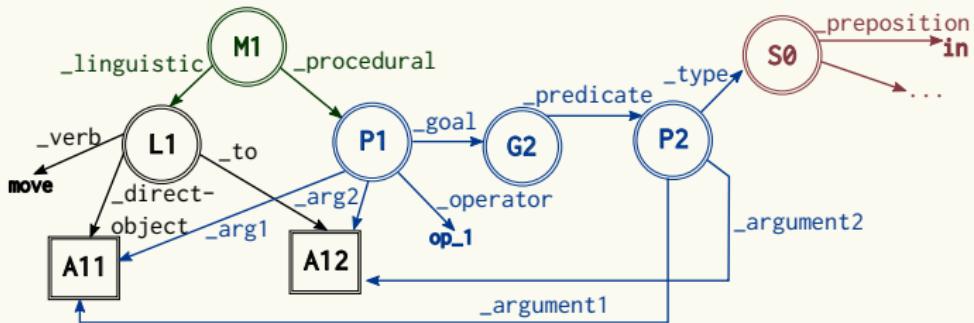
### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.

**Agent:** Which action should I take?



# Verb Acquisition

## Action Concept Network

### Interaction trace

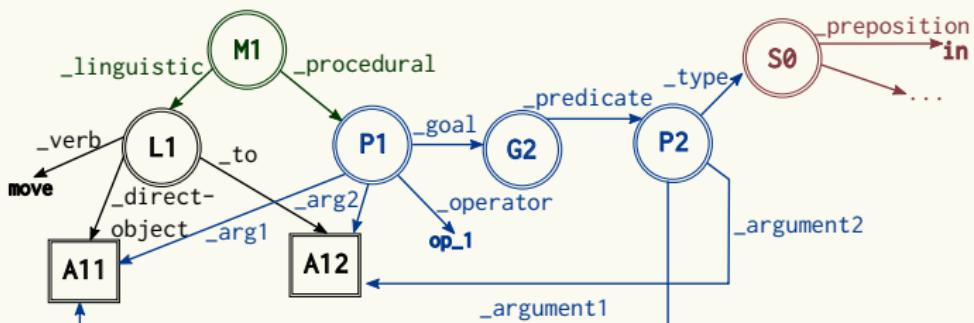
**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.

**Agent:** Which action should I take?

**Instructor:** Pick up the green rectangle.



# Verb Acquisition

## Action Concept Network

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

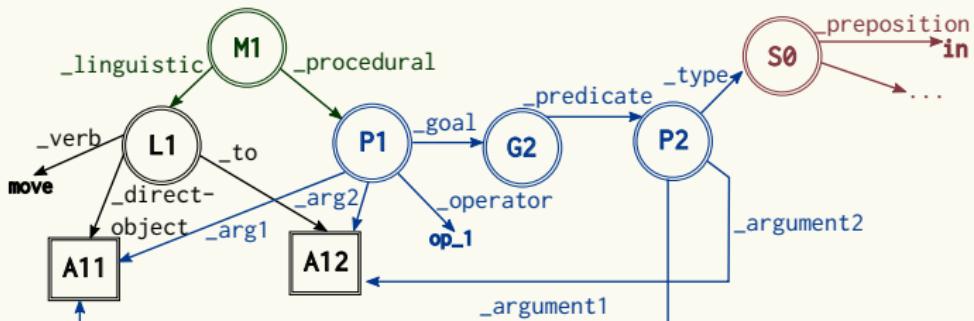
**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.

**Agent:** Which action should I take?

**Instructor:** Pick up the green rectangle.

**Agent:** Which action should I take?



# Verb Acquisition

## Action Concept Network

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

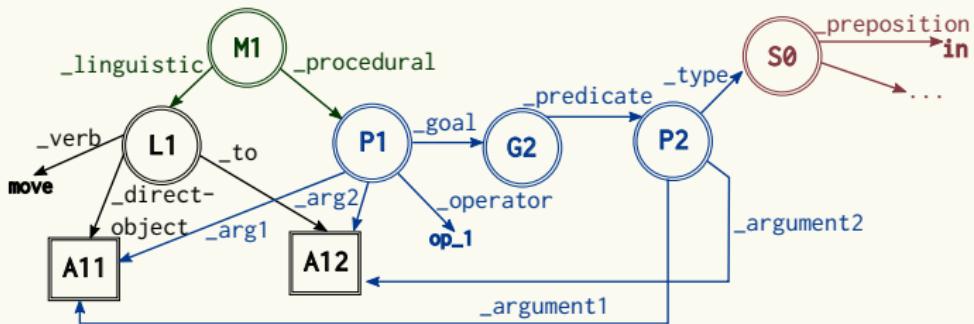
**Instructor:** The goal is the green rectangle in the pantry.

**Agent:** Which action should I take?

**Instructor:** Pick up the green rectangle.

**Agent:** Which action should I take?

**Instructor:** Put the object in the pantry.



# Verb Acquisition

## Action Concept Network

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.

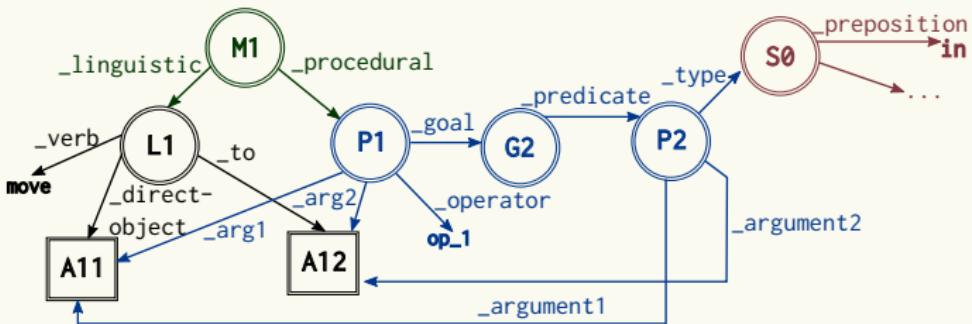
**Agent:** Which action should I take?

**Instructor:** Pick up the green rectangle.

**Agent:** Which action should I take?

**Instructor:** Put the object in the pantry.

**Agent:** Which action should I take?



# Verb Acquisition

## Action Concept Network

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.

**Agent:** Which action should I take?

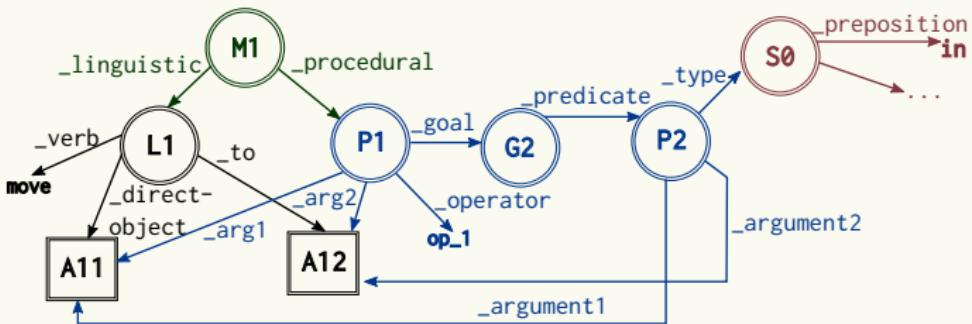
**Instructor:** Pick up the green rectangle.

**Agent:** Which action should I take?

**Instructor:** Put the object in the pantry.

**Agent:** Which action should I take?

**Instructor:** Your are done.



# Verb Acquisition

## Action Concept Network

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.

**Agent:** Which action should I take?

**Instructor:** Pick up the green rectangle.

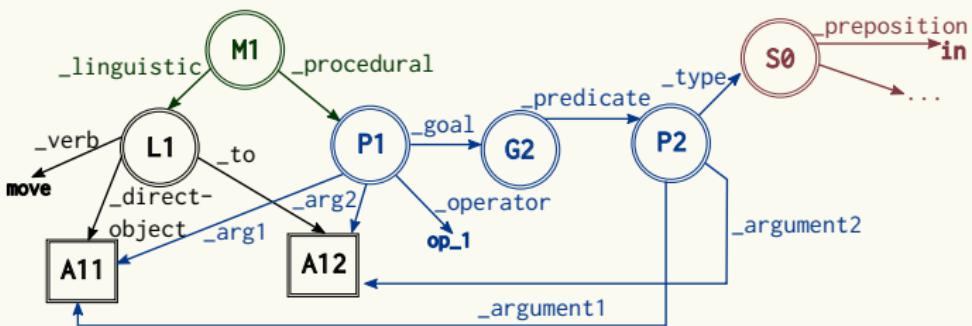
**Agent:** Which action should I take?

**Instructor:** Put the object in the pantry.

**Agent:** Which action should I take?

**Instructor:** Your are done.

**Agent:** forward  
projection  
EBG



# Verb Acquisition

## Procedural Knowledge

### Interaction trace

**Instructor:** Move the green rectangle to the pantry.

**Agent:** What is the goal of the action?

**Instructor:** The goal is the green rectangle in the pantry.

**Agent:** Which action should I take?

**Instructor:** Pick up the green rectangle.

**Agent:** Which action should I take?

**Instructor:** Put the object in the pantry.

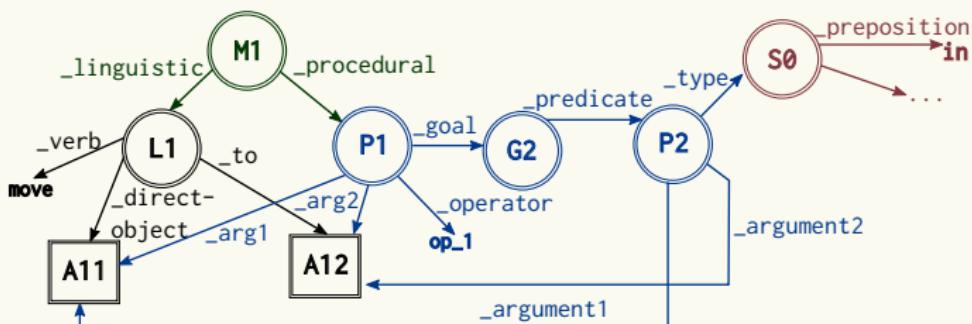
**Agent:** Which action should I take?

**Instructor:** Your are done.

**Agent:** forward

projection

EBG



### rule-1

If command is 'move' with direct-object [A11] and to-object [A12], and spatial-relationship is S0  
 --> execute pick-up([A11])

### rule-2

If command is 'move' with direct-object [A11] and to-object [A12], and spatial-relationship is S0 and holding [A11]  
 --> execute put-down([A12],S0)

# Verb Acquisition Analysis

## Generality of instruction

*Reasons about and removes instructions (actions) that do not occur in the causal link*

**Move the red block to the pantry.**

The goal is red block in the pantry.

Pick up the red block.

Put the block in the garbage.

Pick up the red block.

Put the block in the pantry.

# Verb Acquisition Analysis

## Generality of instruction

*Reasons about and removes instructions (actions) that do not occur in the causal link*

**Move the red block to the pantry.**

The goal is red block in the pantry.

Pick up the red block.

~~Put the block in the garbage.~~

~~Pick up the red bloek.~~

Put the block in the pantry.

# Verb Acquisition Analysis

Generality of situation

*Reasons about the situation, rather than memorizing steps*

## Scenario 1

State: -Holding

Command: *Move the red block to the pantry.*

Actions: Pick up the red block, Put the block in the pantry.

## Scenario 2

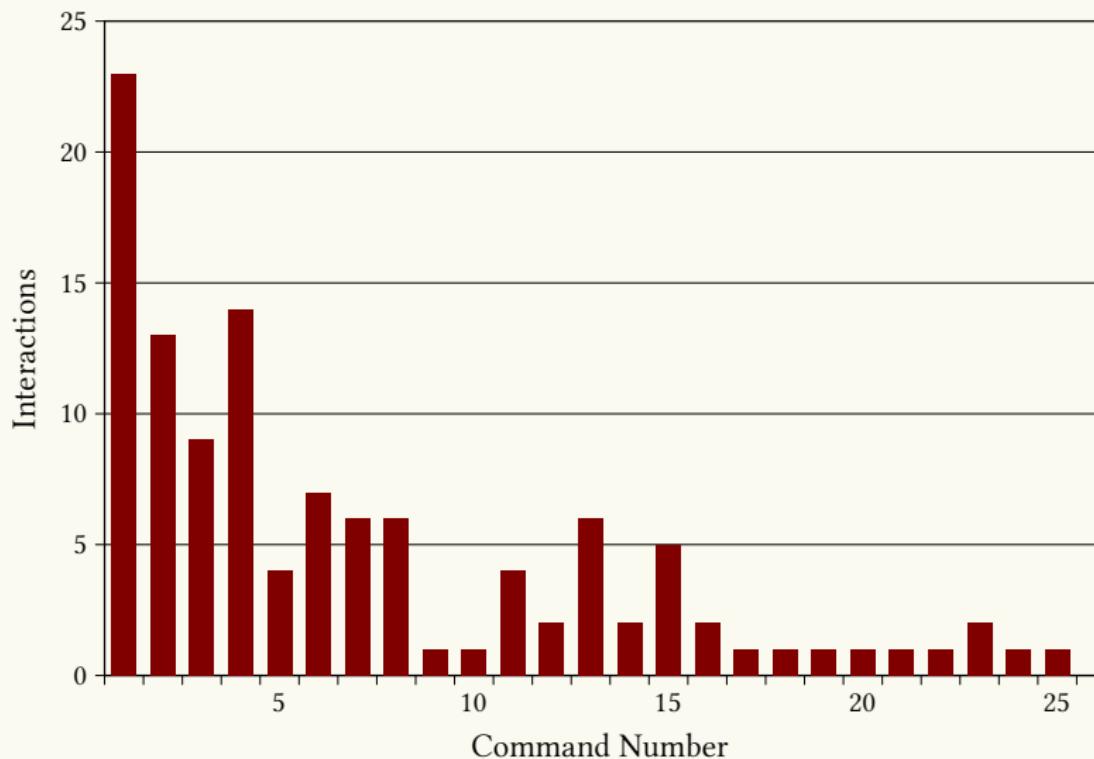
State: Holding(red block)

Command: *Move the red block to the pantry.*

Actions: Put the red block in the pantry.

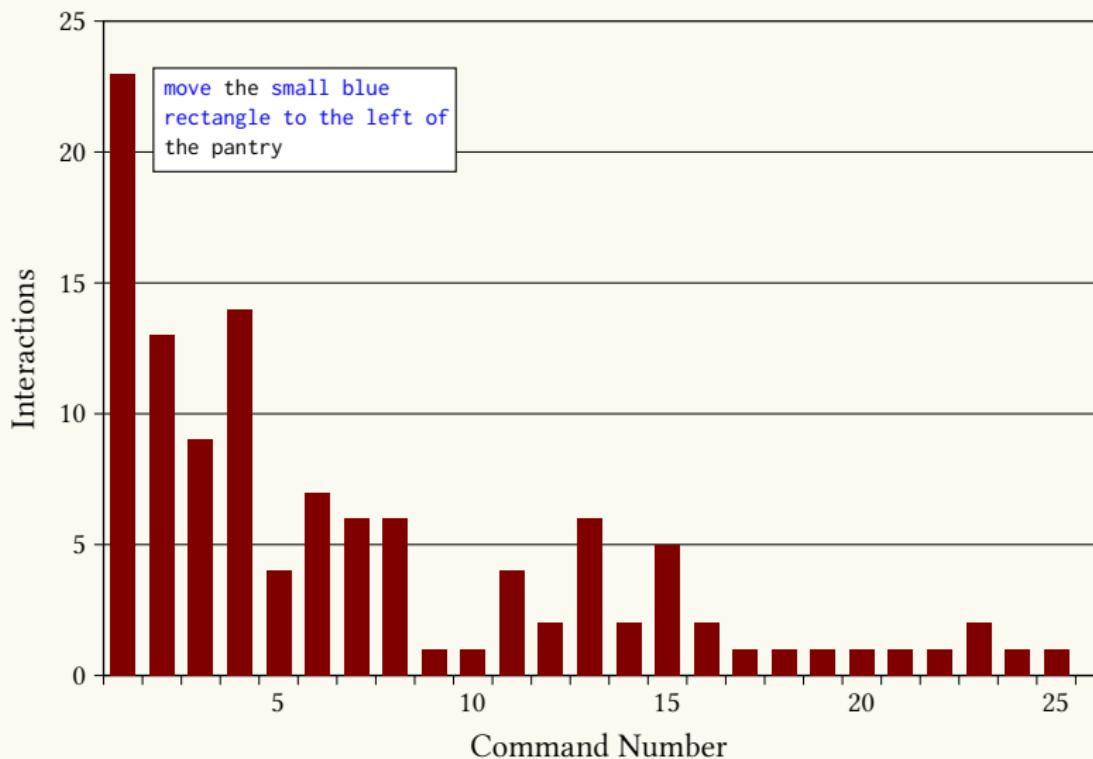
# Full System Evaluation

9 nouns/adjectives, 3 prepositions, 3 verbs (100% correct execution)



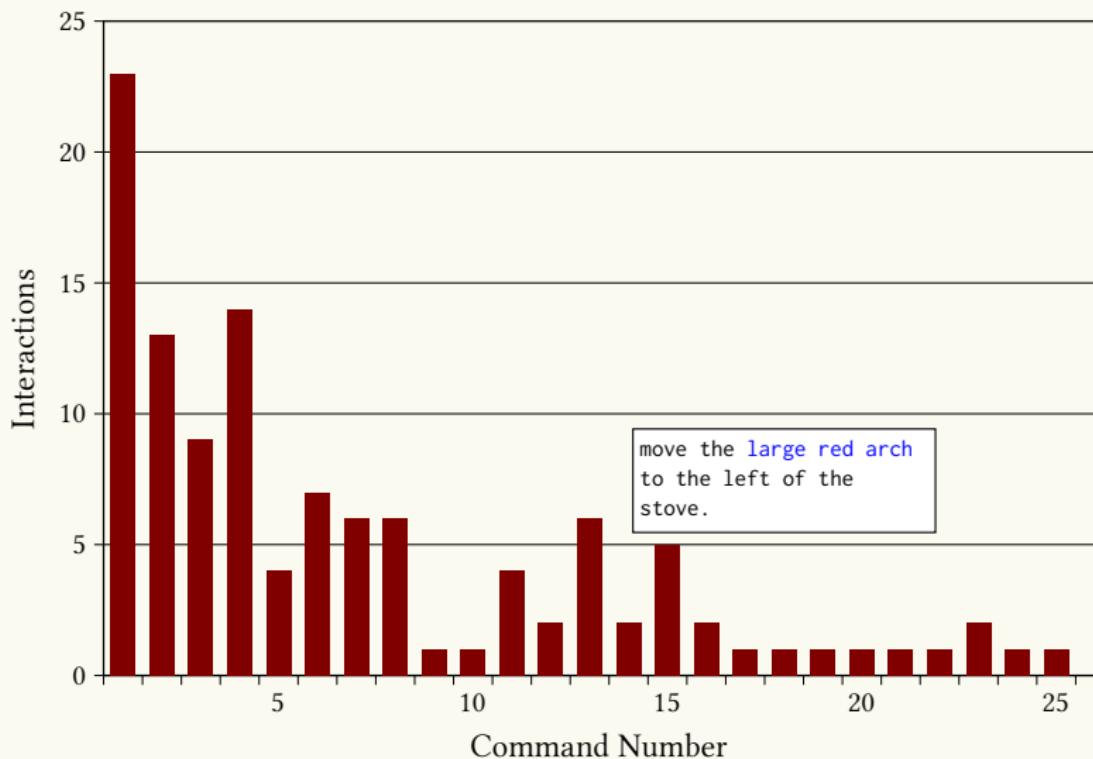
# Full System Evaluation

9 nouns/adjectives, 3 prepositions, 3 verbs (100% correct execution)



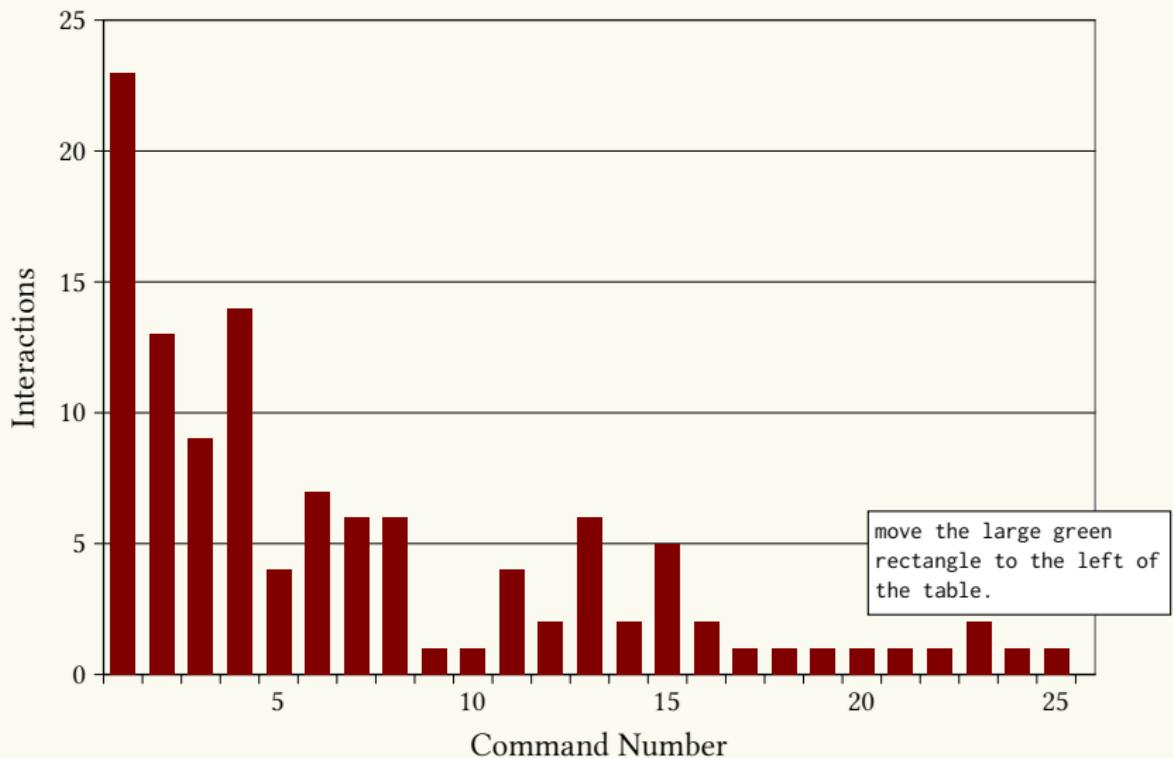
# Full System Evaluation

9 nouns/adjectives, 3 prepositions, 3 verbs (100% correct execution)



# Full System Evaluation

9 nouns/adjectives, 3 prepositions, 3 verbs (100% correct execution)



# Conclusions

- System capabilities
  - integrates communication, execution, and learning
  - learns perceptual nouns/adjectives, spatial prepositions, action verbs
- Interactive instruction
  - guides the agent to ask useful questions
  - assimilates novel information with pre-existing knowledge
- Diverse knowledge
  - provides context for extracting grounded examples from linguistic instructions

# Future Work

- Determine usefulness of instruction
- Task Complexity
  - Physical, non-visual properties such as *weight*
  - ternary relationships such as *between*, multi object relationships such as *clear*
  - hierarchical verb execution
- Robustness
  - Corrective instruction
- Perceptual Uncertainty
  - Information gathering actions
  - Use of long-term memories to augment perception
- Instruction
  - Hypothetical

# Noun/Adjective Acquisition

- Nouns + adjectives
  - Linguistic references to objects

# Noun/Adjective Acquisition

- Nouns + adjectives
  - Linguistic references to objects
- Three perceptual properties
  - *color, shape, size (red, triangle, large)*

# Noun/Adjective Acquisition

- Nouns + adjectives
  - Linguistic references to objects
- Three perceptual properties
  - *color, shape, size (red, triangle, large)*
- Perceptual Learning
  - Gaussian KNN classifier for every property
  - Every class → perceptual symbol (color:red → R43)

# Noun/Adjective Acquisition

- Nouns + adjectives
  - Linguistic references to objects
- Three perceptual properties
  - *color, shape, size (red, triangle, large)*
- Perceptual Learning
  - Gaussian KNN classifier for every property
  - Every class → perceptual symbol (color:red → R43)
- Linguistic Learning
  - Linguistic symbol to perceptual symbol mapping (*red* → R43)
  - Categorization (R43 is a color)

# Noun/Adjective Acquisition

- Nouns + adjectives
  - Linguistic references to objects
- Three perceptual properties
  - *color, shape, size (red, triangle, large)*
- Perceptual Learning
  - Gaussian KNN classifier for every property
  - Every class → perceptual symbol (color:red → R43)
- Linguistic Learning
  - Linguistic symbol to perceptual symbol mapping (*red* → R43)
  - Categorization (R43 is a color)
- Bi-directional access

Pick up the **red** triangle. → red: R43

This is a **red** object. ← red: R43



# Preposition Acquisition

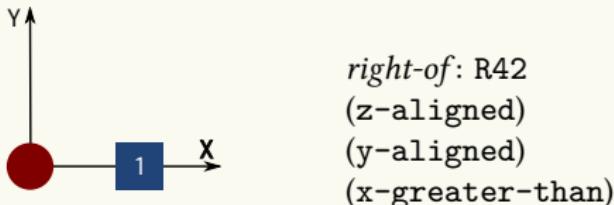
- Prepositions
  - Spatial references to objects
  - Goal tracking of actions

# Preposition Acquisition

- Prepositions
  - Spatial references to objects
  - Goal tracking of actions
- Characterization
  - Primitives: direction, distance
  - Composition
    - Conjunctions along different axes ( $x\text{-greater-than} \wedge z\text{-aligned}$ )
    - Disjunctions along same axis ( $x\text{-aligned} \vee x\text{-greater-than}$ )
    - Distribution over distances of examples

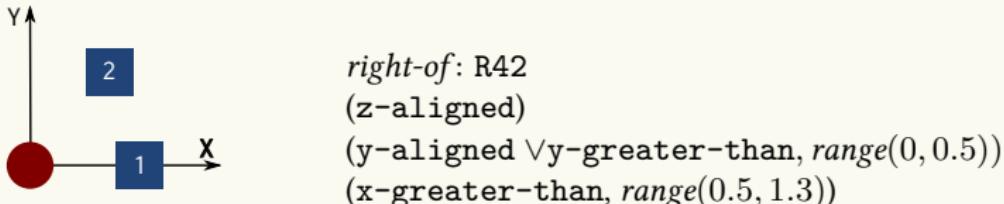
# Preposition Acquisition

- Prepositions
  - Spatial references to objects
  - Goal tracking of actions
- Characterization
  - Primitives: direction, distance
  - Composition
    - Conjunctions along different axes ( $x\text{-greater-than} \wedge z\text{-aligned}$ )
    - Disjunctions along same axis ( $x\text{-aligned} \vee x\text{-greater-than}$ )
    - Distribution over distances of examples
- Learning (linguistic *map*, spatial *configuration*)



# Preposition Acquisition

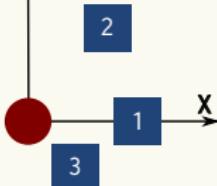
- Prepositions
  - Spatial references to objects
  - Goal tracking of actions
- Characterization
  - Primitives: direction, distance
  - Composition
    - Conjunctions along different axes ( $x\text{-greater-than} \wedge z\text{-aligned}$ )
    - Disjunctions along same axis ( $x\text{-aligned} \vee x\text{-greater-than}$ )
    - Distribution over distances of examples
- Learning (linguistic *map*, spatial *configuration*)



# Preposition Acquisition

- Prepositions
  - Spatial references to objects
  - Goal tracking of actions
- Characterization
  - Primitives: direction, distance
  - Composition
    - Conjunctions along different axes ( $x\text{-greater-than} \wedge z\text{-aligned}$ )
    - Disjunctions along same axis ( $x\text{-aligned} \vee x\text{-greater-than}$ )
    - Distribution over distances of examples
- Learning (linguistic *map*, spatial *configuration*)

Y



*right-of: R42*

(*z-aligned*)

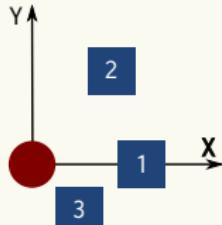
(*y-aligned*  $\vee$  *y-greater-than*  $\vee$

*y-less-than, range*( $-0.5, 0.5$ )

(*x-greater-than, range*( $0.5, 1.3$ ))

# Preposition Acquisition

- Prepositions
  - Spatial references to objects
  - Goal tracking of actions
- Characterization
  - Primitives: direction, distance
  - Composition
    - Conjunctions along different axes ( $x\text{-greater-than} \wedge z\text{-aligned}$ )
    - Disjunctions along same axis ( $x\text{-aligned} \vee x\text{-greater-than}$ )
    - Distribution over distances of examples
- Learning (linguistic *map*, spatial *configuration*)



*right-of: R42*  
*(z-aligned)*  
*(y-aligned  $\vee$  y-greater-than  $\vee$*   
*y-less-than, range(-0.5, 0.5))*  
*(x-greater-than, range(0.5, 1.3))*

- Bi-directional access

# Verb Acquisition Analysis

## Characterization

- Parameterization
  - explicit arguments: *move the red large triangle to the pantry*
  - implicit arguments: *store the red large triangle*
- Verb termination conditions
  - spatial predicates
  - achievement goals
- Action composition
  - causal chain of primitive actions

# Component Evaluation

Table: Noun/adjective Evaluation (12 objects)

Perceptual property	Classes	Examples	Coverage (%)
Color	4	1	100.00
Shape	2	1.5	100.00
Size	4	12.9	96.00

Table: Preposition Evaluation

# Prepositions	Spatial Arrangements	Examples	Coverage (%)
6	144	3.17	93.98

Table: Verb Evaluation

# Action Commands	Average Instantiations	Examples	Coverage (%)
5	35.2	1.26	100.00

## Timing Data

