Learning Actions and Verbs from Situated Interactive Instruction for Embodied Cognitive Agents

Shiwali Mohan

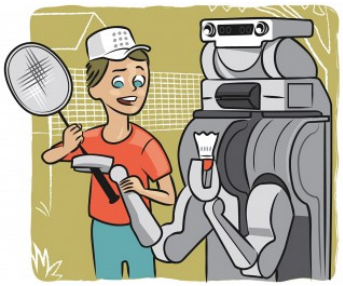
Computer Science and Engineering
University of Michigan

May 30, 2013

Autonomous Embodied Agents

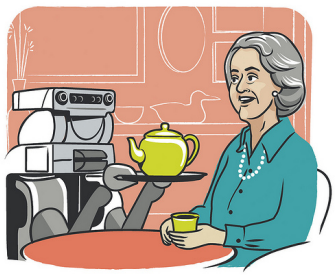
Intelligent behavior in novel environments

recognize and categorize objects

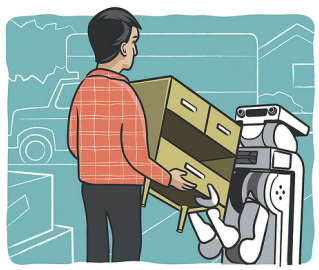


use complex skills

understand and use spatial relationships



perform complex tasks

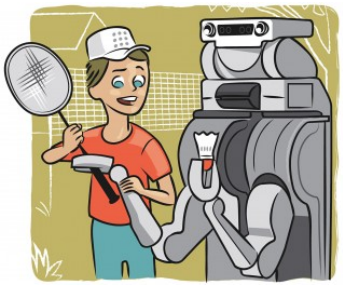


collaborate with humans

Autonomous Embodied Agents

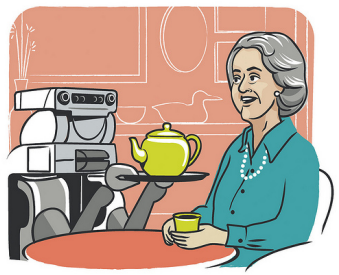
Intelligent behavior in novel environments

recognize and categorize objects

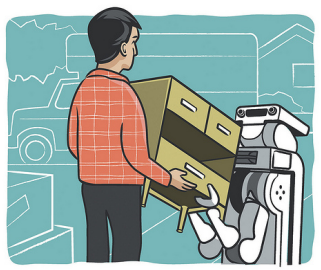


use complex skills

understand and use spatial relationships



perform complex tasks



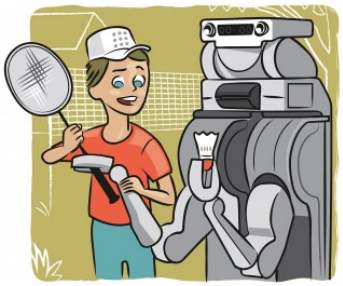
collaborate with humans

- Must be online, efficient learners
  - designers cannot pre-program behaviors

Autonomous Embodied Agents

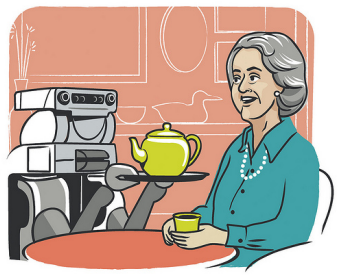
Intelligent behavior in novel environments

recognize and categorize objects

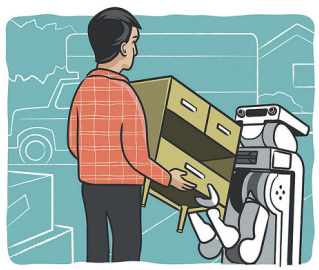


use complex skills

understand and use spatial relationships



perform complex tasks

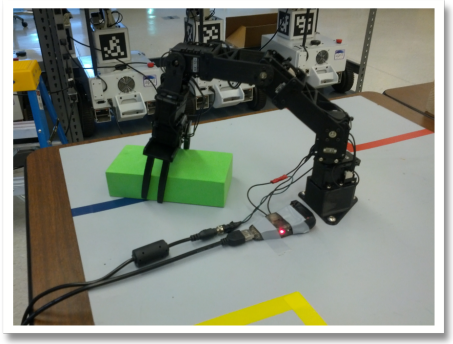


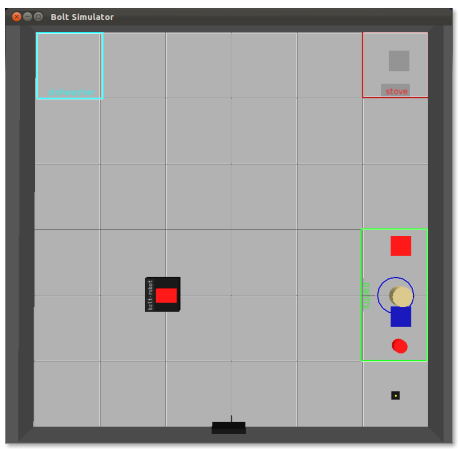
collaborate with humans

- Must be online, efficient learners
  - designers cannot pre-program behaviors
- Learn from human-agent interactions
  - Learning from demonstration
  - Embodied interactions teleoperation, robot manipulation
  - Skill acquisition - *pour*, *pick-up*, *put-down*, etc

Overview

- Study learning from *knowledge-level* human-agent interactions
- Agents embodied in real/simulated worlds
- Learn goal-oriented actions/tasks
- Compositions of primitive skills
- Grounded verb representations





Research Questions

- Mixed-initiative Interaction
  - task-oriented, flexible

Research Questions

- Mixed-initiative Interaction
  - task-oriented, flexible
- Situated Comprehension
  - natural language
  - exploit common ground
    - shared perceptions
    - shared domain knowledge
    - shared experience

Research Questions

- Mixed-initiative Interaction
  - task-oriented, flexible
- Situated Comprehension
  - natural language
  - exploit common ground
    - shared perceptions
    - shared domain knowledge
    - shared experience
- Active Learning
  - goals, policies from interactions

Research Questions

- Mixed-initiative Interaction
  - task-oriented, flexible
- Situated Comprehension
  - natural language
  - exploit common ground
    - shared perceptions
    - shared domain knowledge
    - shared experience
- Active Learning
  - goals, policies from interactions

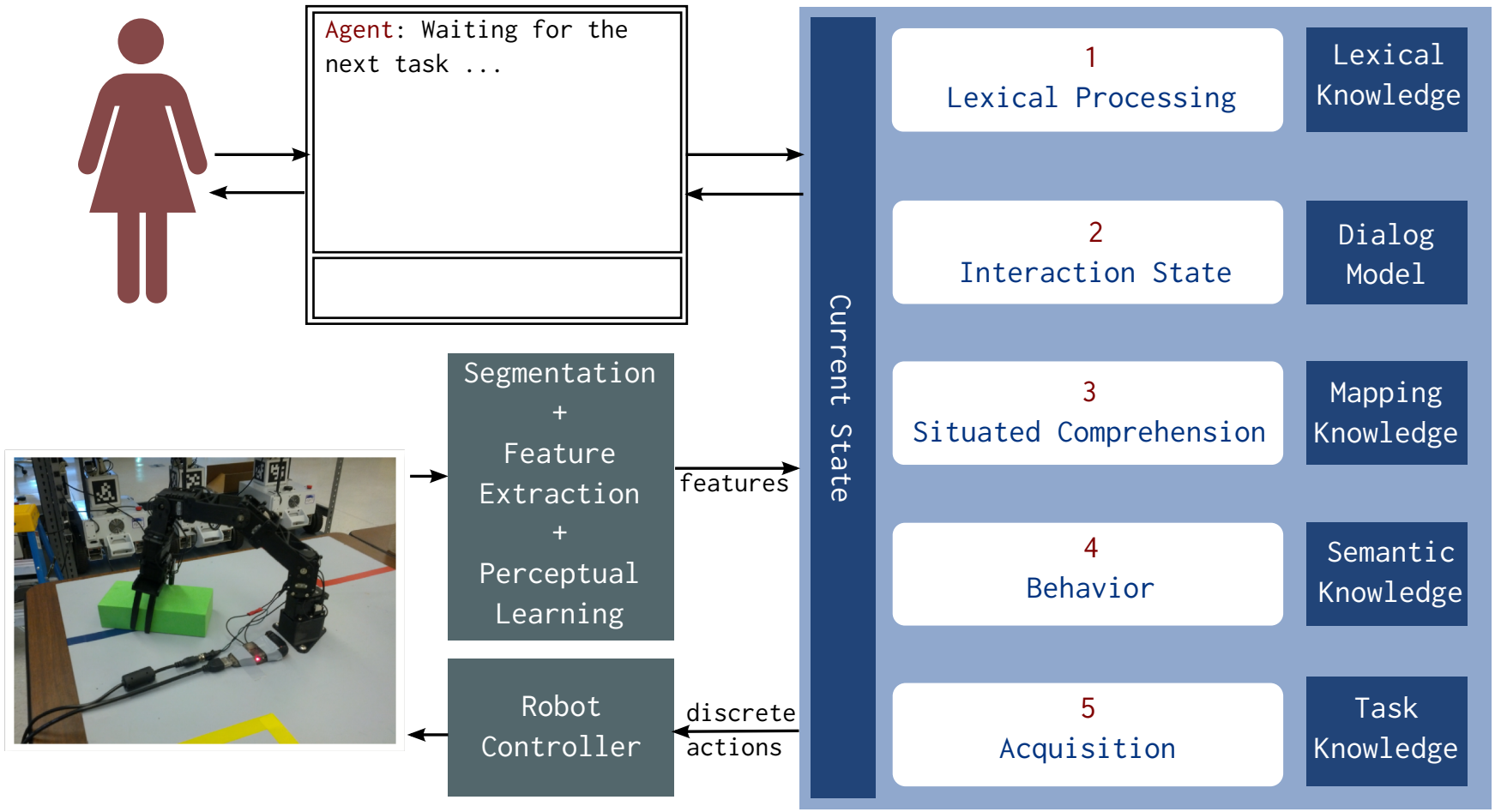
Research Questions

- Mixed-initiative Interaction [✓]
  - task-oriented, flexible
- Situated Comprehension [...]
  - natural language
  - exploit common ground
    - shared perceptions
    - shared domain knowledge
    - shared experience
- Active Learning [...]
  - goals, policies from interactions

Contributions

- A representation distributed across different types of knowledge
  - encodes action knowledge
  - grounds verbs in actions
  - is comprehensive
- An active learning model
  - integrates diverse knowledge acquisition with interaction
  - assimilates information/knowledge in several interactions
  - allows for flexibility in instruction
- A situated comprehension model
  - uses perceptions, state, domain and task knowledge, and experience

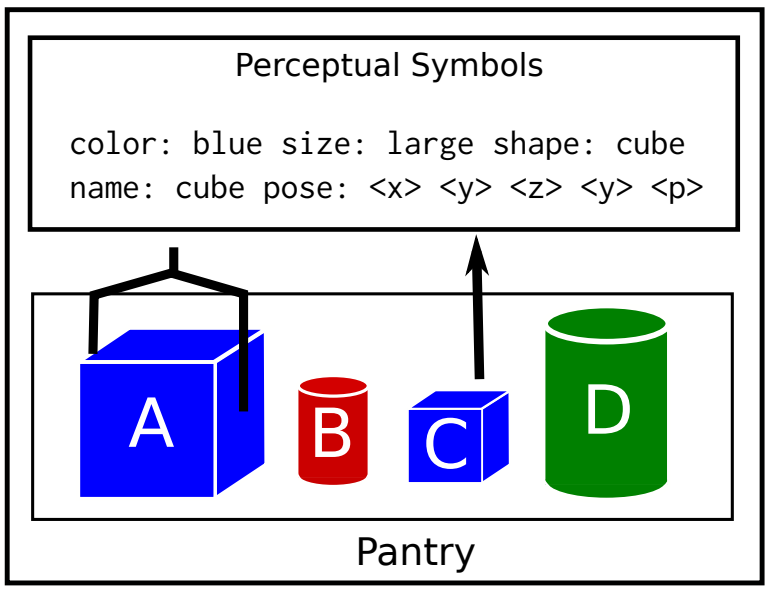
Interaction Overview



State and Interactions

- Relational, object-oriented, state description
  - Object and features (color: *red*, shape *large*, position)
  - Predicate defined over objects
    - categories of objects, `LOC(pantry)`, `BLK(C)`
    - state of blocks, `IN-GRIPPER(A)`
    - state of locations, `OPEN(pantry)`, `ON(stove)`
    - spatial relationships, `IN(C,pantry)`
- Instructions
  - Action commands: *store the green rectangle*
  - Goal definitions: *goal of store is the rectangle is in the pantry and the pantry is closed*
  - grounding functions: instruction $\rightarrow$ action, state

  ...



Actions and Verbs
Knowledge Requirements

Given: a set of primitive actions, model
Learn: composite action *store* from interactions
(*store* → *put object in the pantry and close*)

- syntactic constraints on instantiation

  store the green rectangle → operator `OPstore(OBJ1)`
- parameters: explicit, implicit

  `OPstore([x])`: `[x]`, `pantry`
- availability conditions

  if `[x]` is an block → `OPstore([x])`
- policy

  if pantry is open and the task is `OPstore([x])` → `OPmove(IN,[x],pantry)`
- termination conditions (goal)

  `IN([x],pantry)`, `CLOSED(pantry)`
- model

  on applying `OPstore([x])` → `IN([x],pantry)`, `CLOSED(pantry)`

Approach: Representation

- Knowledge in 3 categories
- Action Concept Network (declarative knowledge)
  - verb termination conditions (goal predicates)
  - constituent composite actions, primitive actions
  - association with objects, spatial-relationships
  - acquired through interaction with declarative description, interactive experience
- Rules (procedural knowledge)
  - propose the action in appropriate states (availability conditions)
  - implement general, state-conditional policy
  - acquired through retrospective explanation based learning
- Map (syntactic+semantic constraints)
  - the verb and the argument structure: *store the green rectangle* $\rightarrow$ action concept network
  - acquired through heuristics

Approach: Active Learning

- When to ask a query?
- What is the correct question to ask?
- How to integrate information with prior knowledge?

Approach: Active Learning

- When to ask a query?
  - failures in processing
- What is the correct question to ask?
- How to integrate information with prior knowledge?

Approach: Active Learning

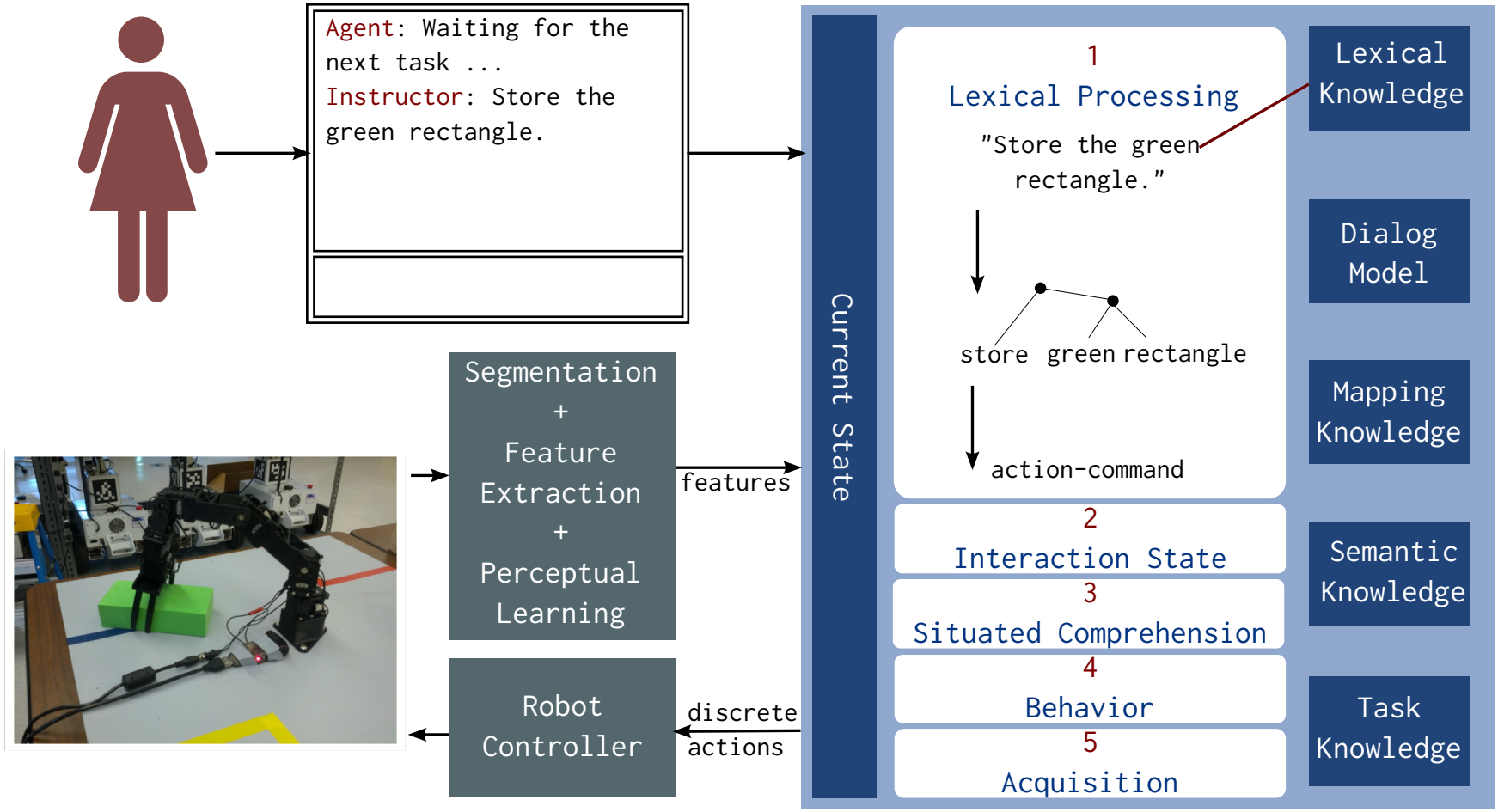
- When to ask a query?
  - failures in processing
- What is the correct question to ask?
  - processing stage of failure
- How to integrate information with prior knowledge?

Approach: Active Learning

- When to ask a query?
  - failures in processing
- What is the correct question to ask?
  - processing stage of failure
- How to integrate information with prior knowledge?
  - maintain state of knowledge search

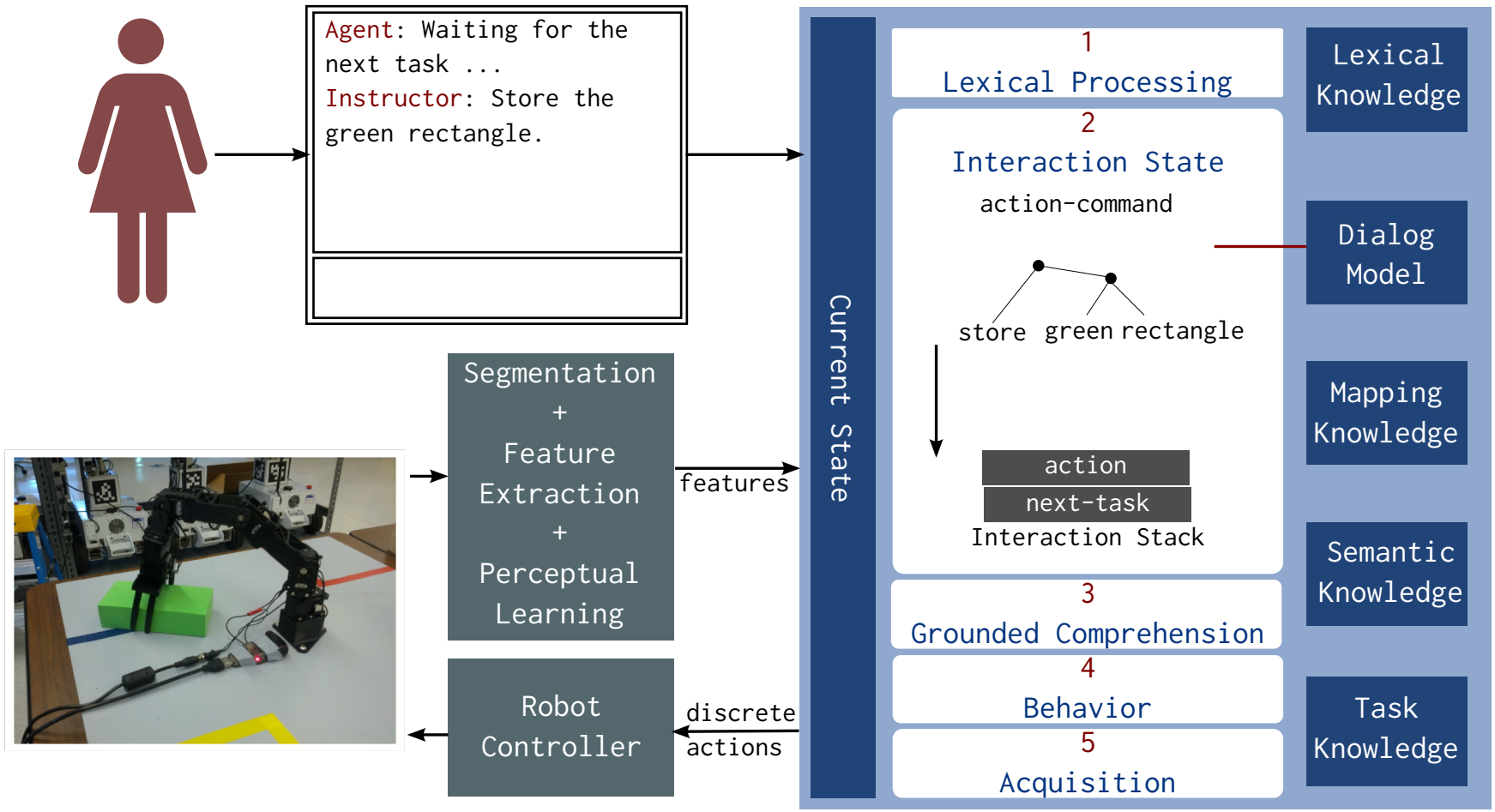
Interaction Cycle

Phase I: Lexical processing



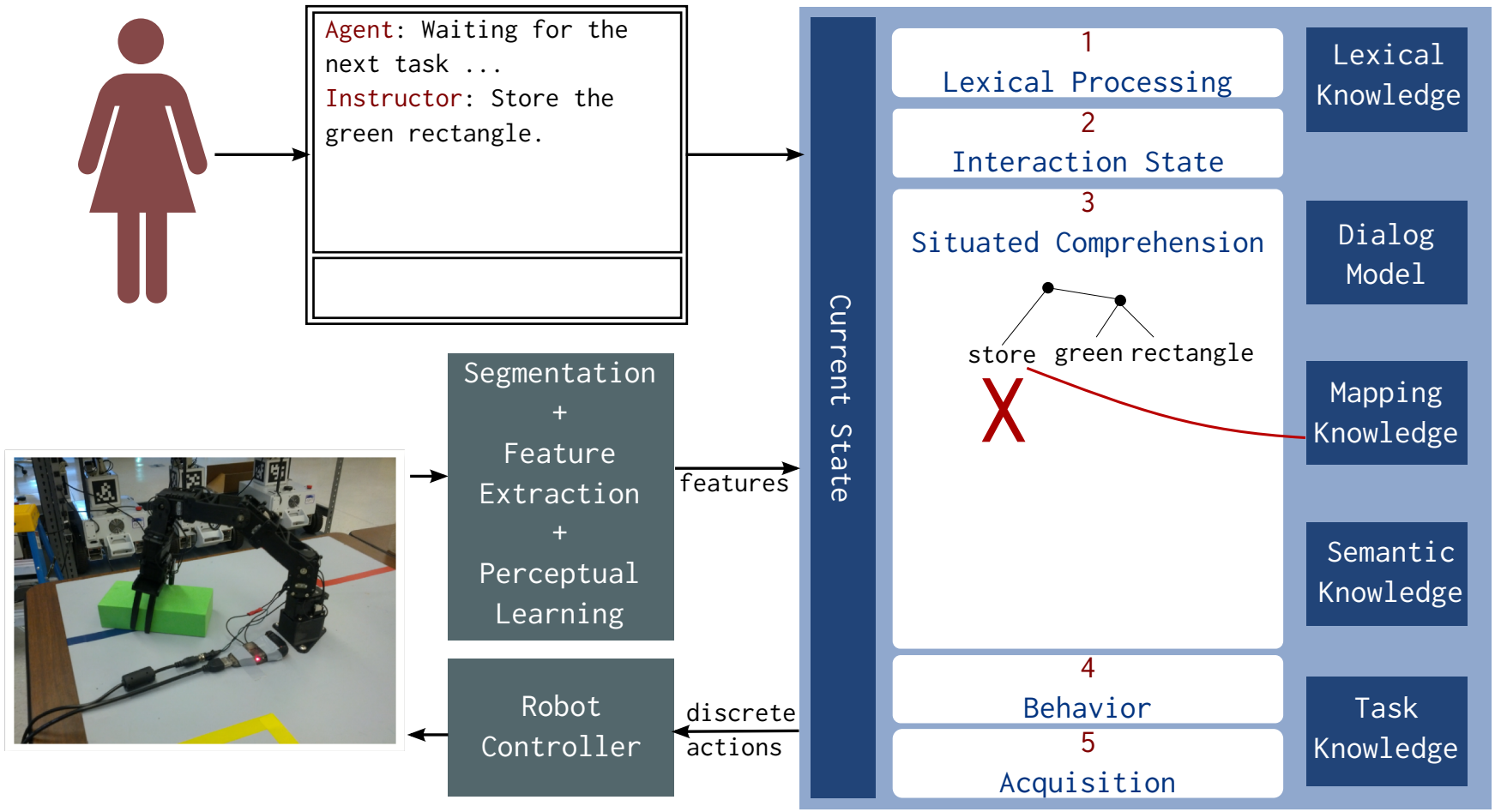
Interaction Cycle

Phase II: Interaction state management



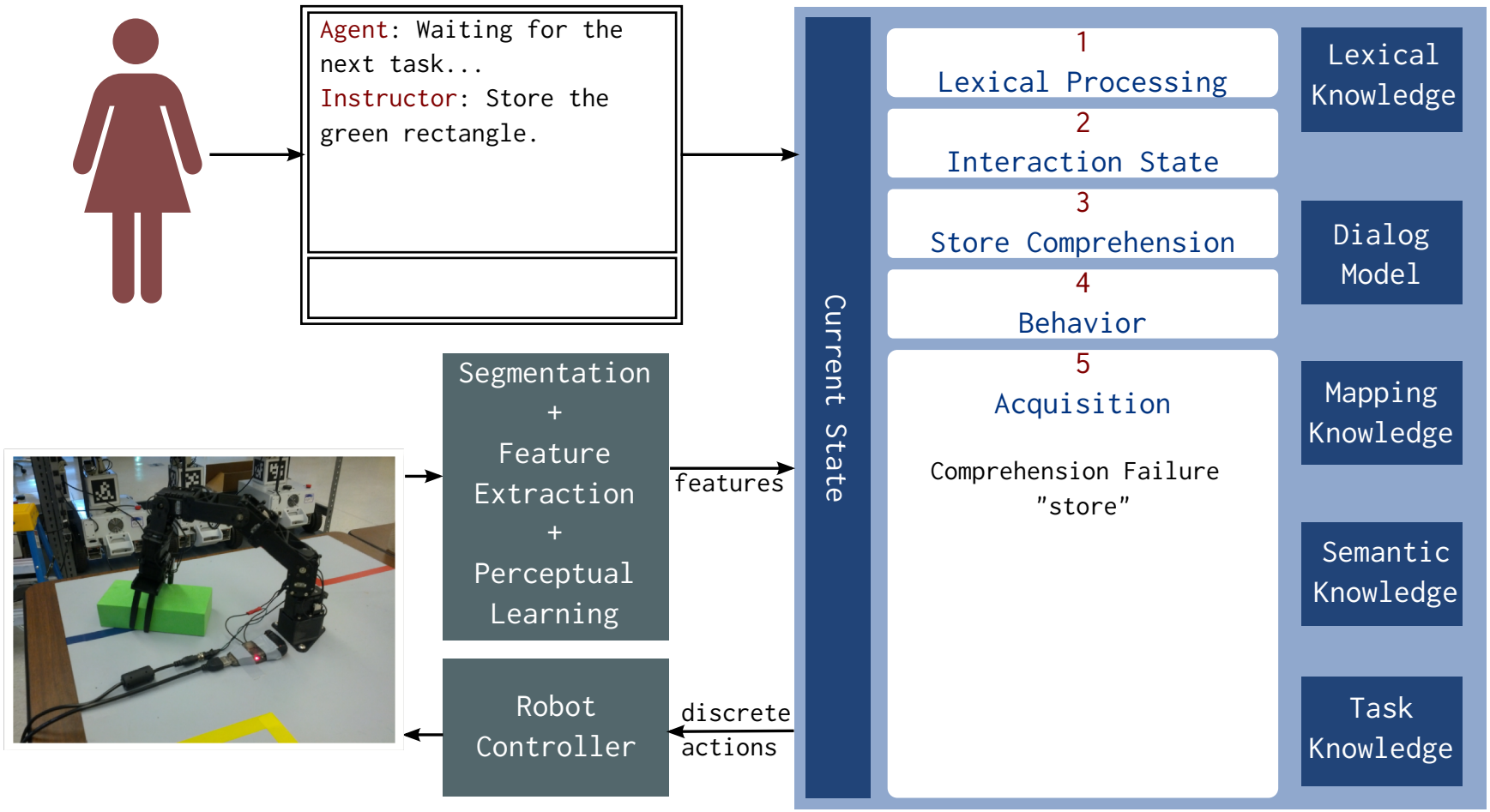
Interaction Cycle

Phase III: Situated comprehension (failure)



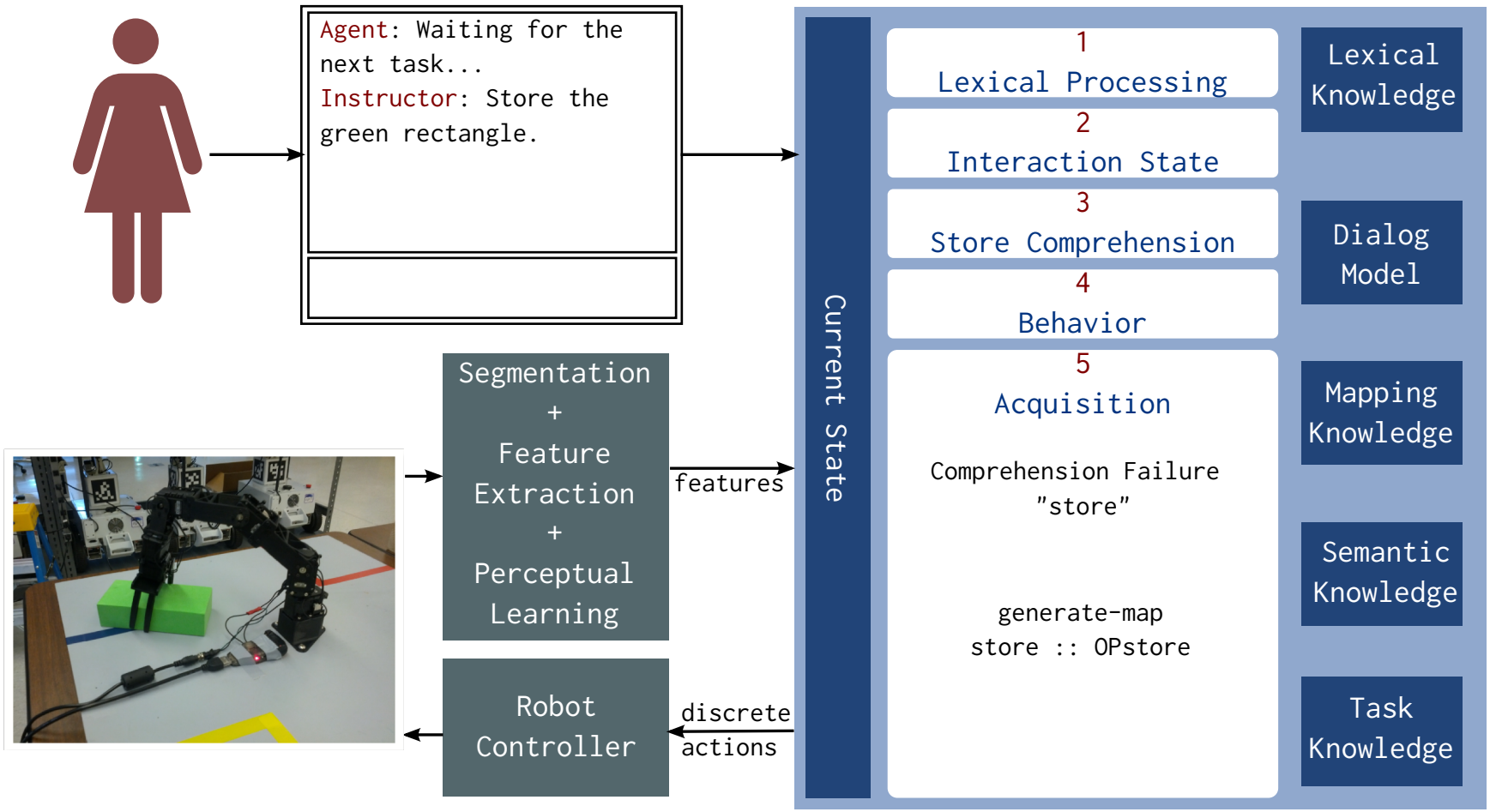
Interaction Cycle

Phase V: Acquisition (map learning)



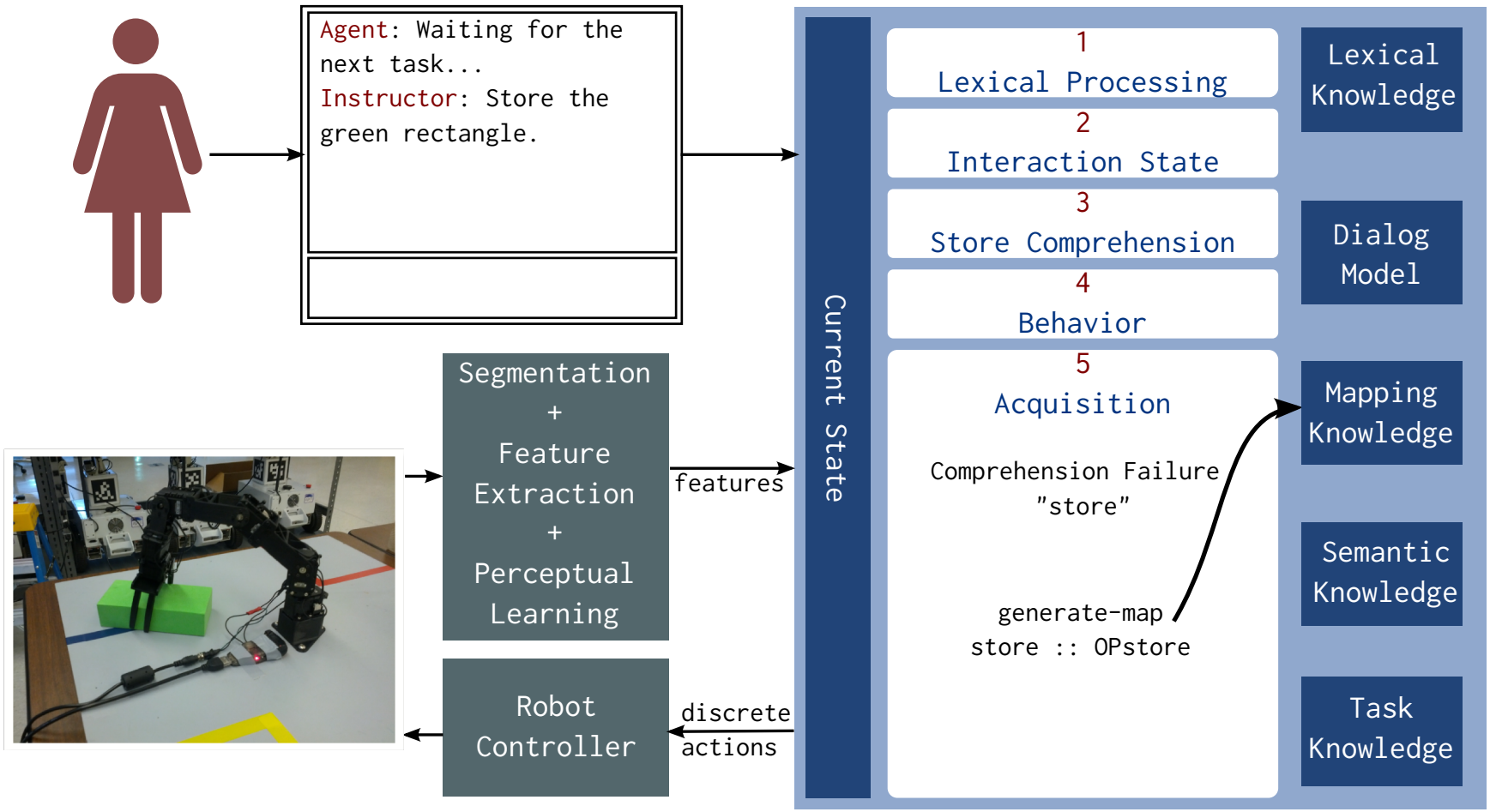
Interaction Cycle

Phase V: Acquisition (map learning)



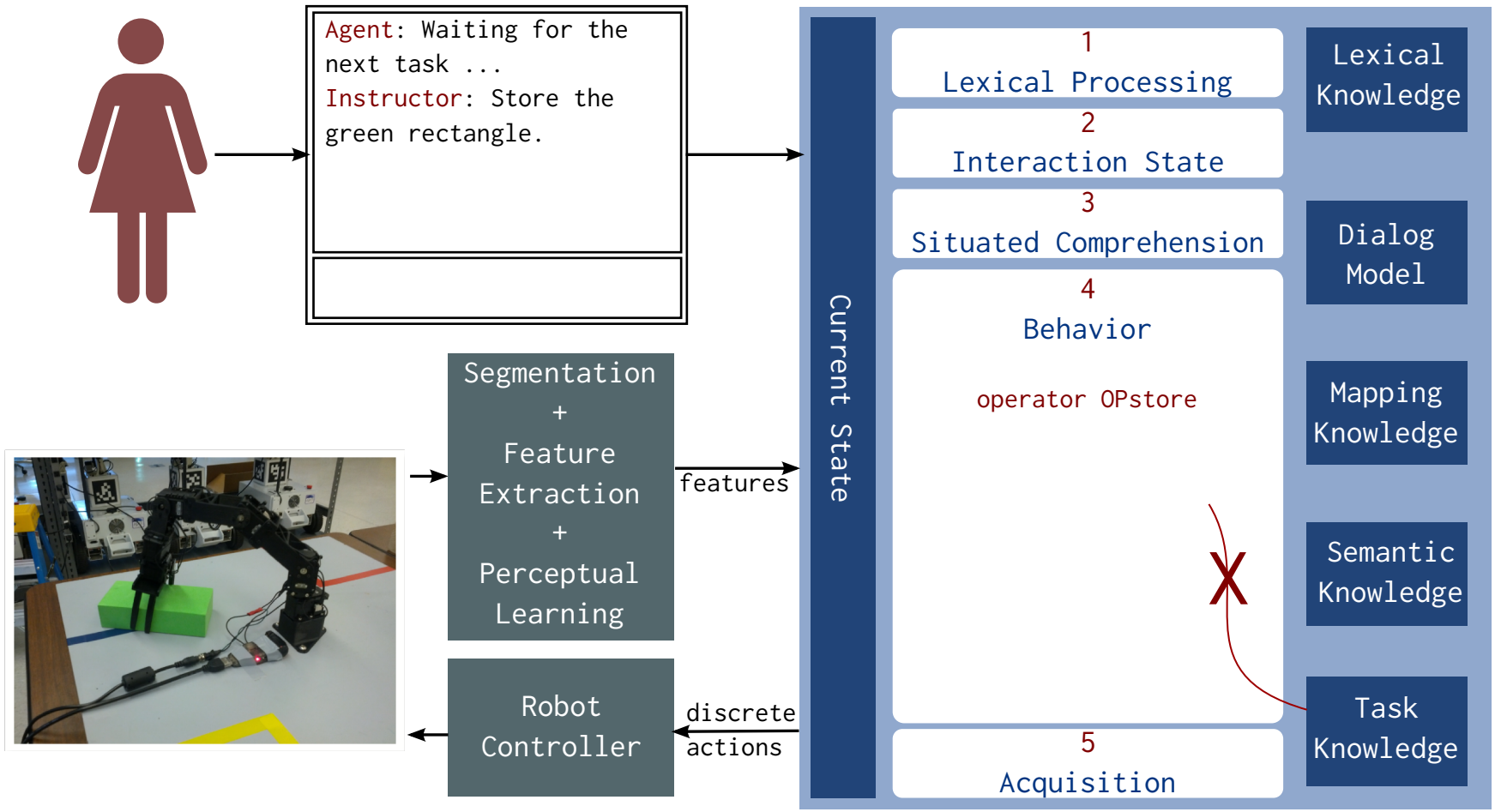
Interaction Cycle

Phase V: Acquisition (map learning)



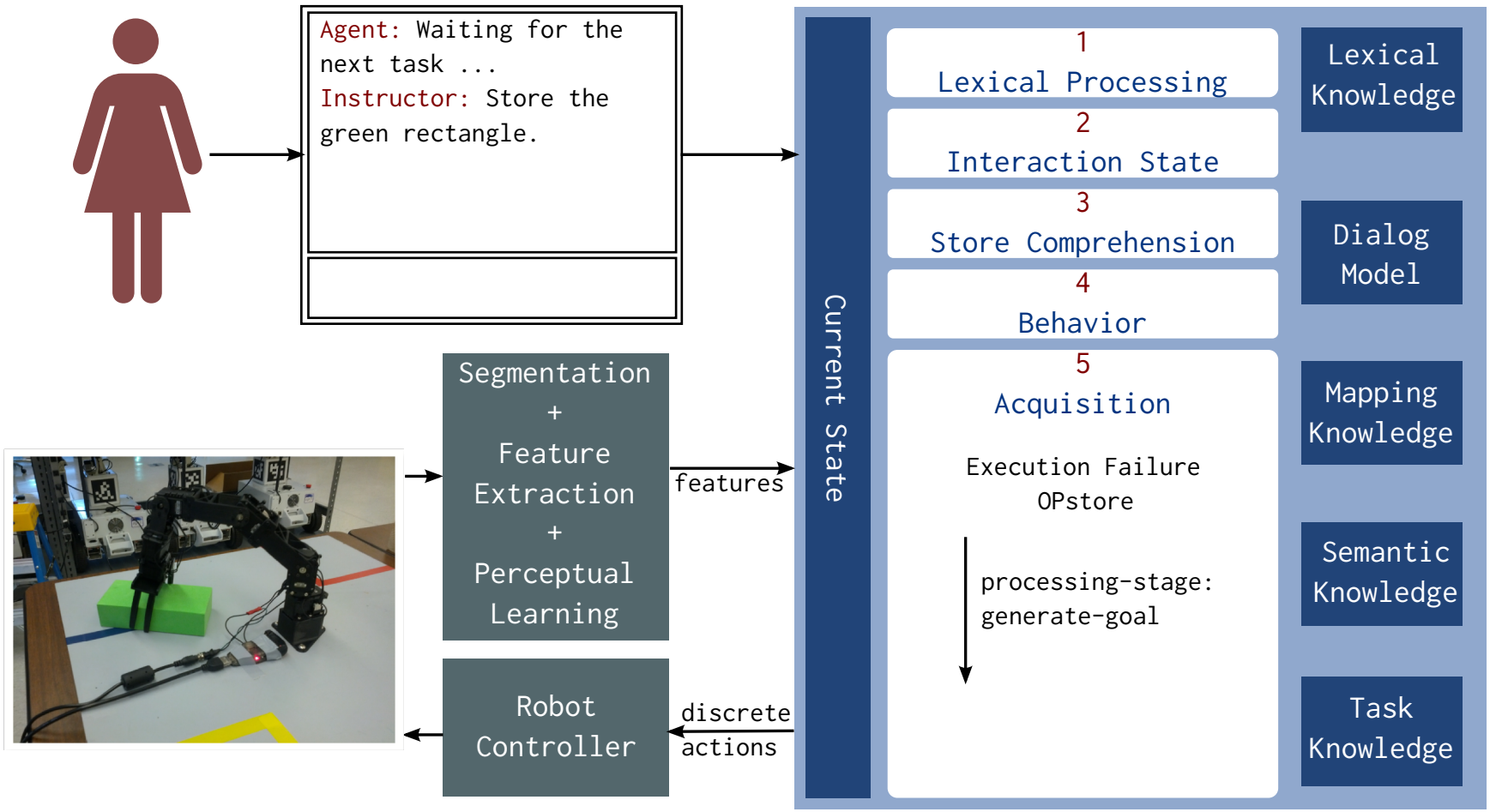
Interaction Cycle

Phase IV: Behavior (failure)



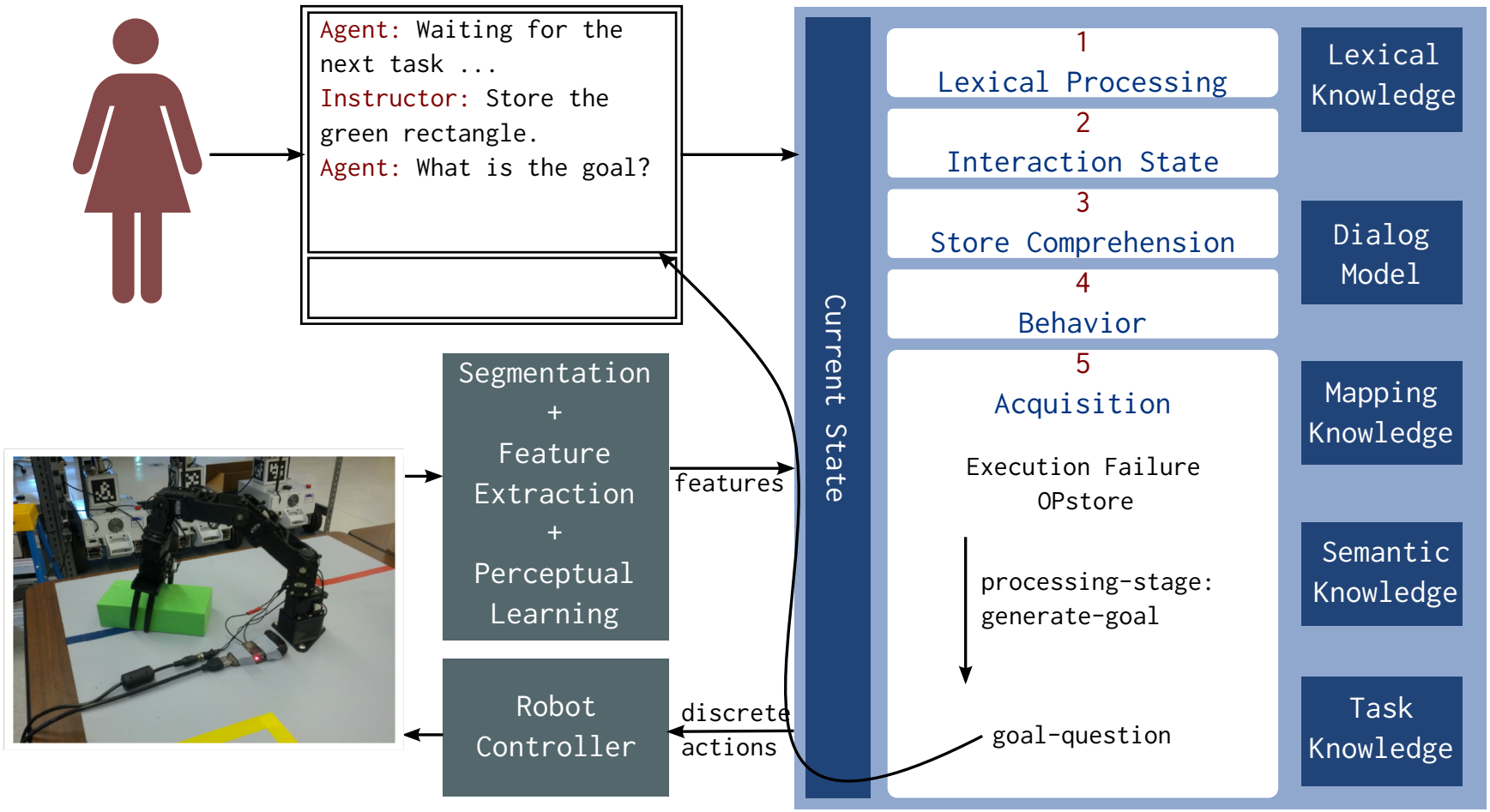
Interaction Cycle

Phase V: Acquisition (goal)



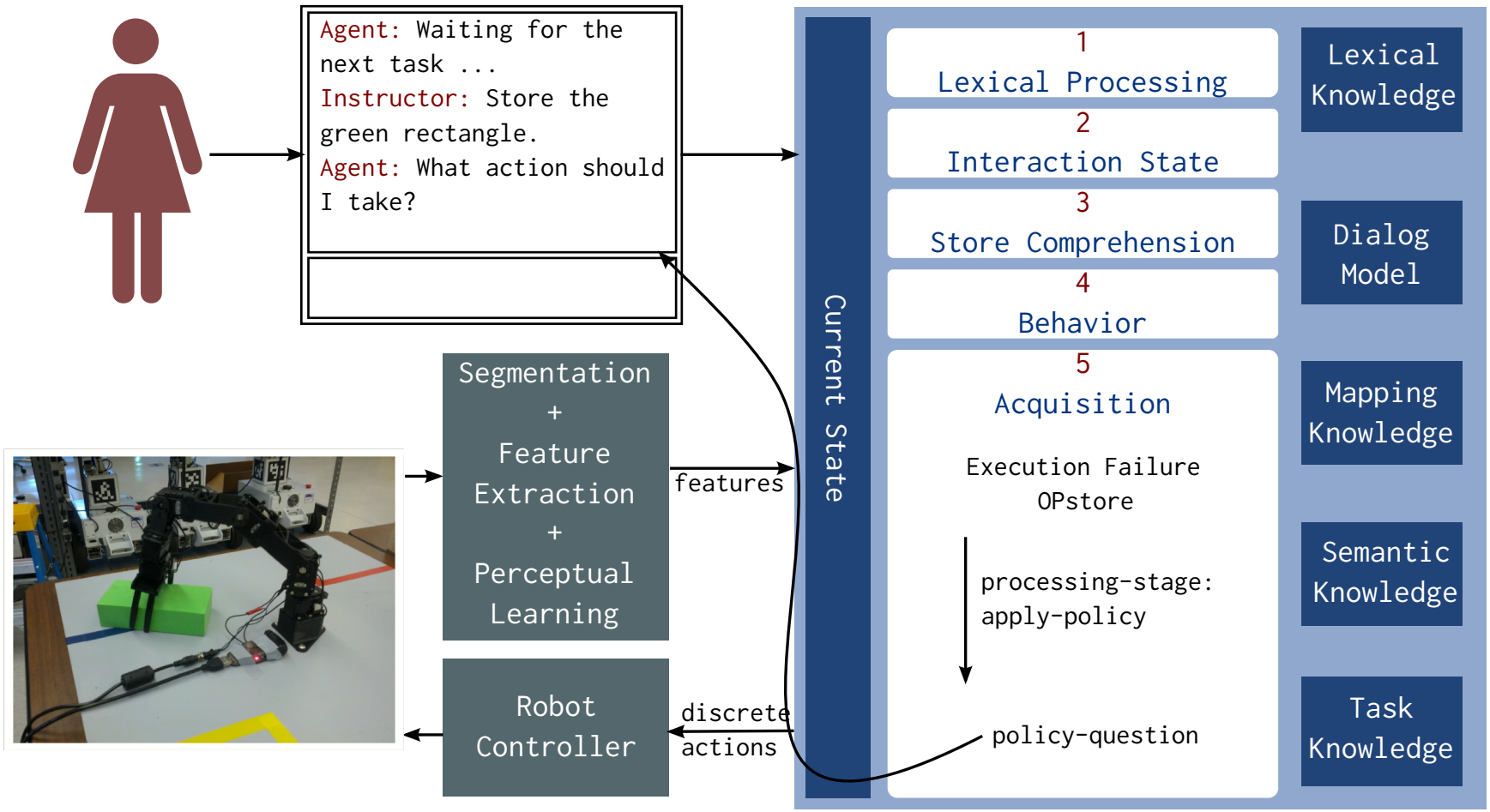
Interaction Cycle

Phase V: Acquisition (goal)



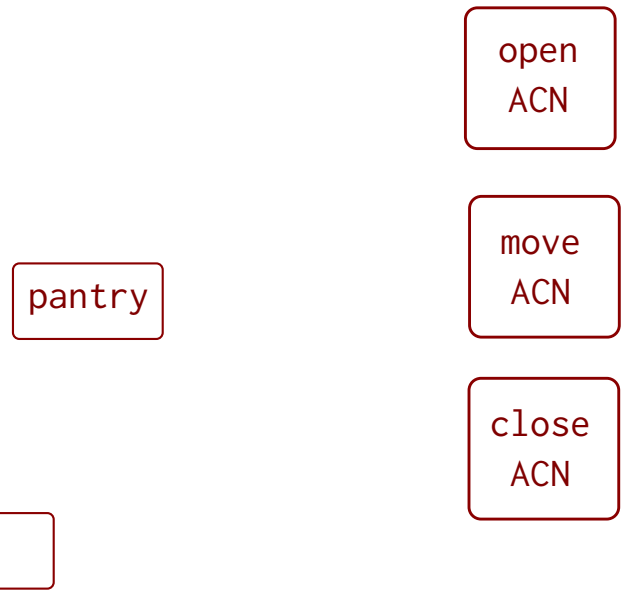
Interaction Cycle

Phase V: Acquisition (policy)



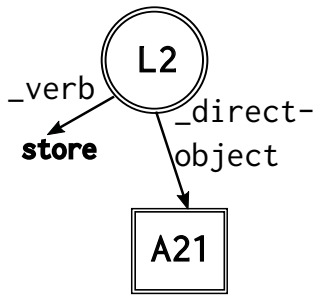
Learning

Interaction trace
Instructor: Store the green rectangle.



Learning

Interaction trace
Instructor: Store the green rectangle.



open ACN

move ACN

close ACN

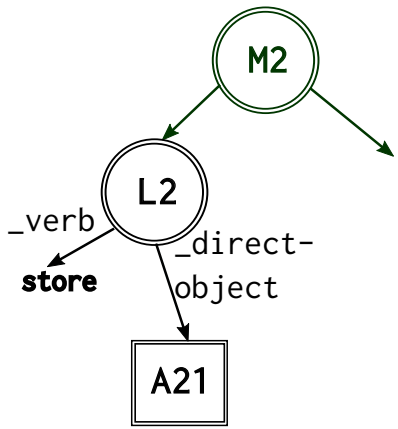
pantry

in

Learning

Interaction trace

Instructor: Store the green rectangle.



open ACN

move ACN

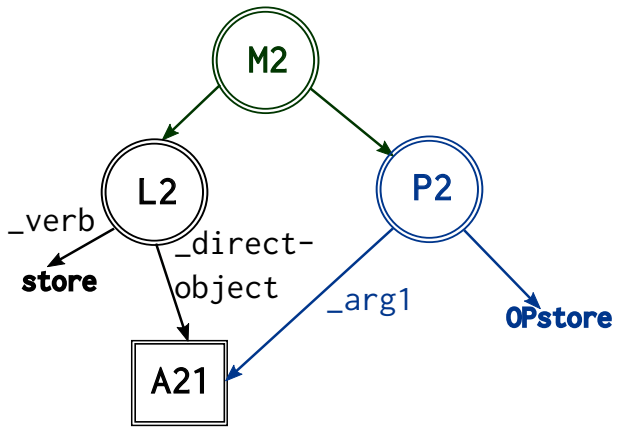
close ACN

pantry

in

Learning

Interaction trace
Instructor: Store the green rectangle.



open ACN

move ACN

close ACN

pantry

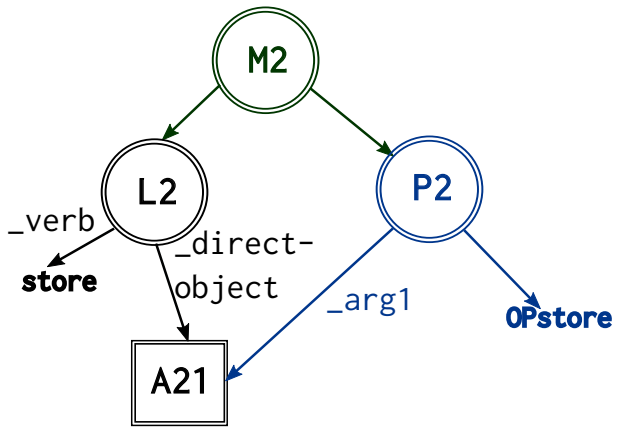
in

Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?



open ACN

move ACN

close ACN

pantry

in

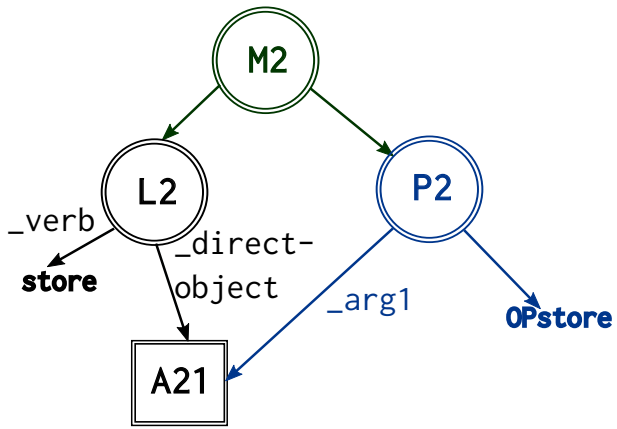
Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.



open ACN

move ACN

close ACN

pantry

in

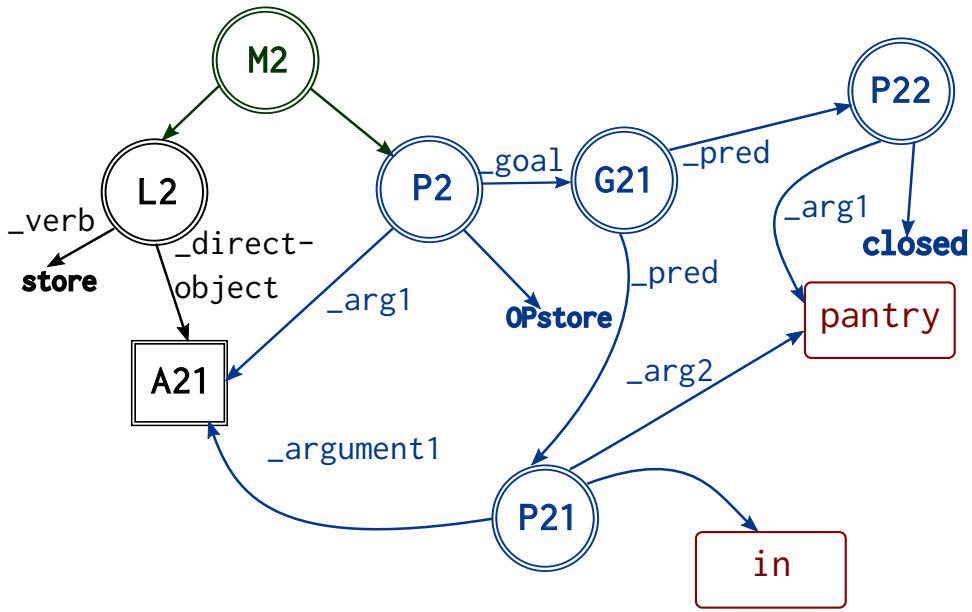
Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.



open ACN

move ACN

close ACN

Learning

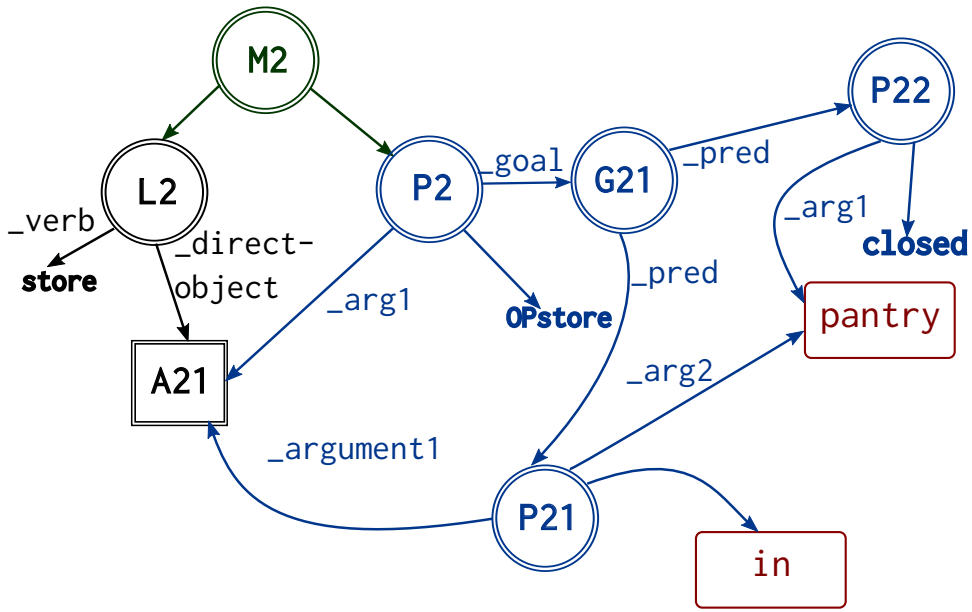
Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?



open ACN

move ACN

close ACN

Learning

Interaction trace

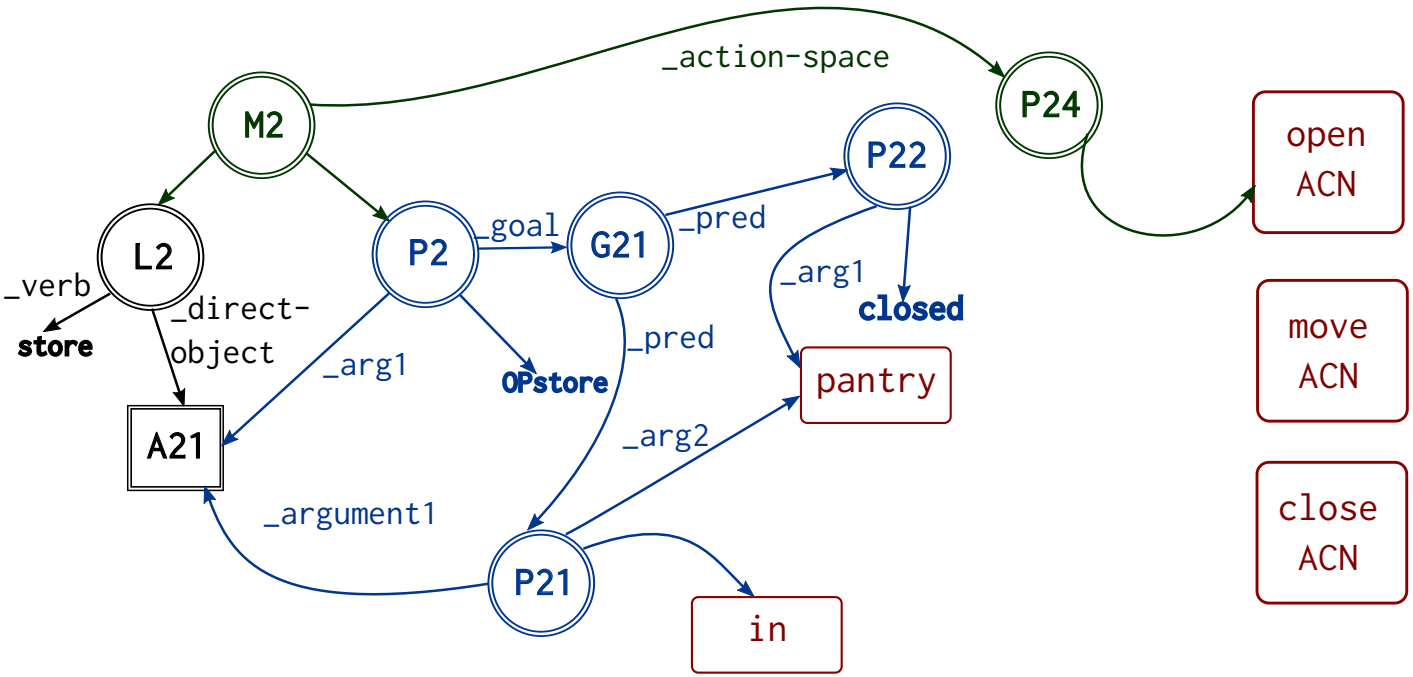
Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

Instructor: Open the pantry.



Learning

Interaction trace

Instructor: Store the green rectangle.

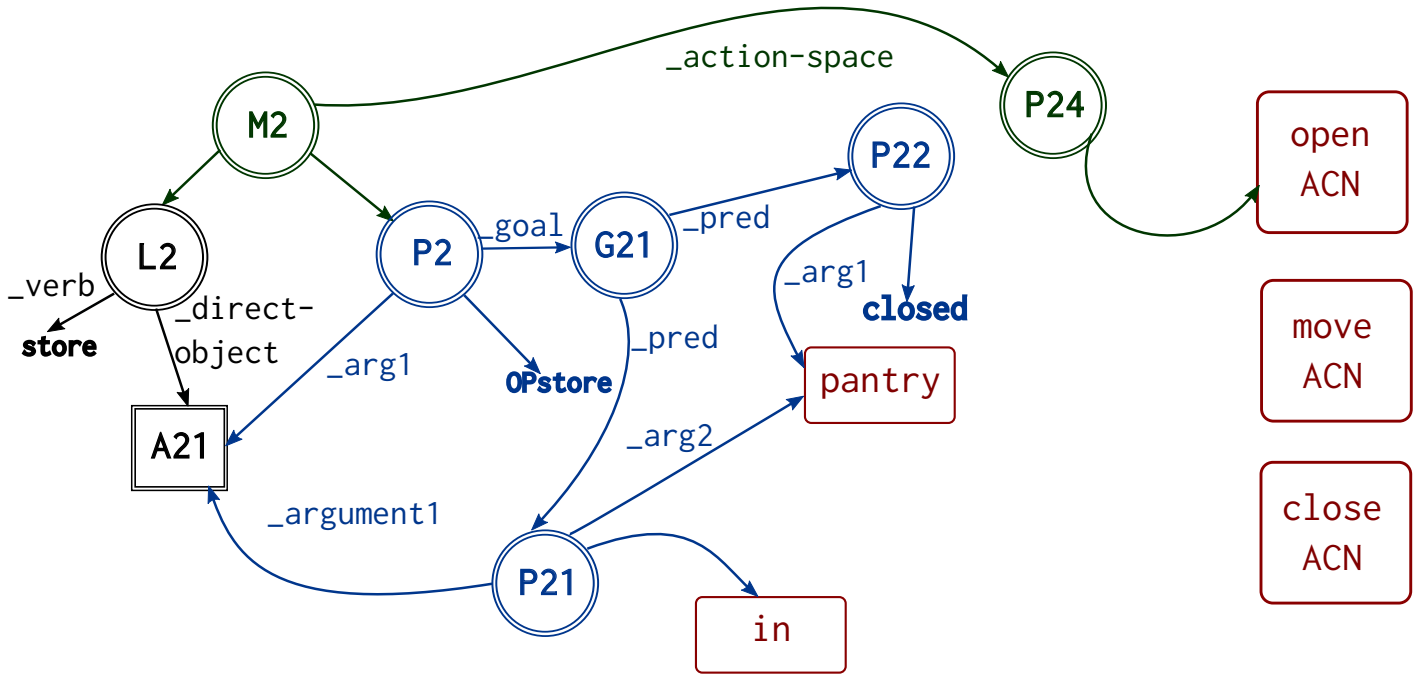
Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

Instructor: Open the pantry.

Agent: Which action should I take?



Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

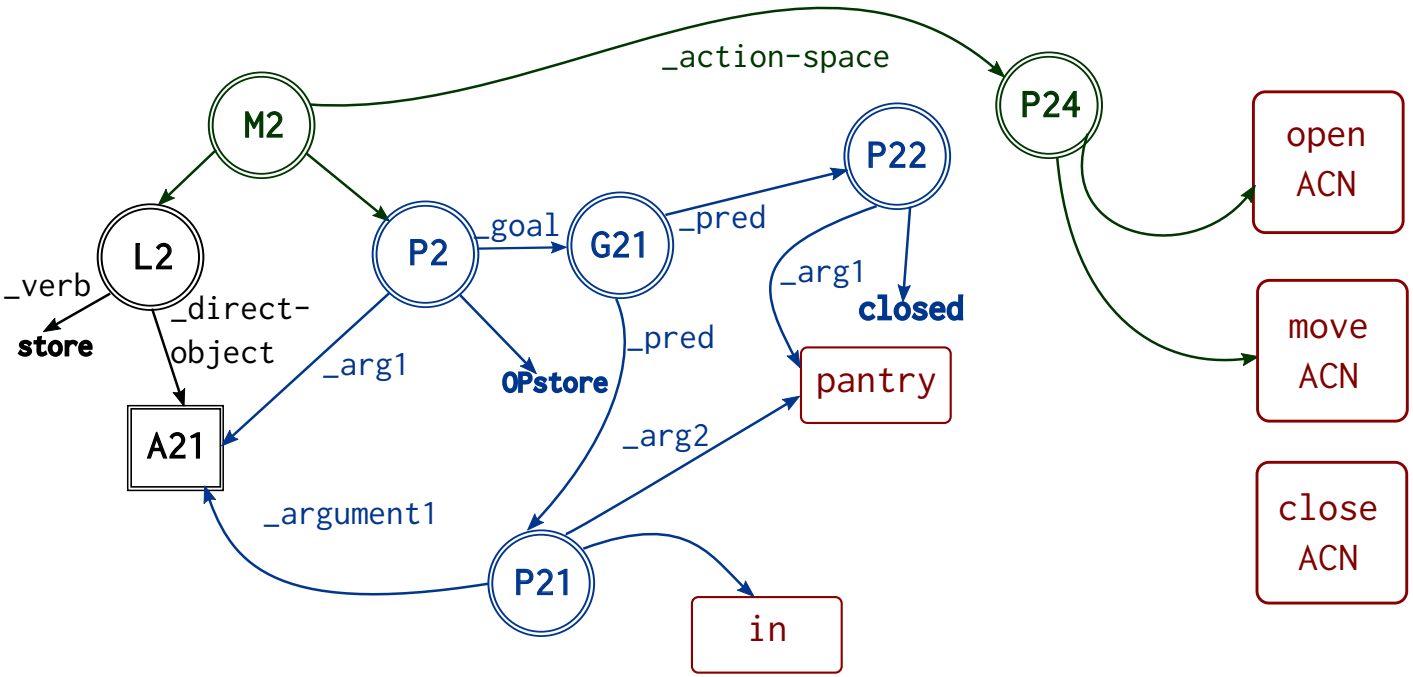
Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

Instructor: Open the pantry.

Agent: Which action should I take?

Instructor: Move the object to the pantry.



Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

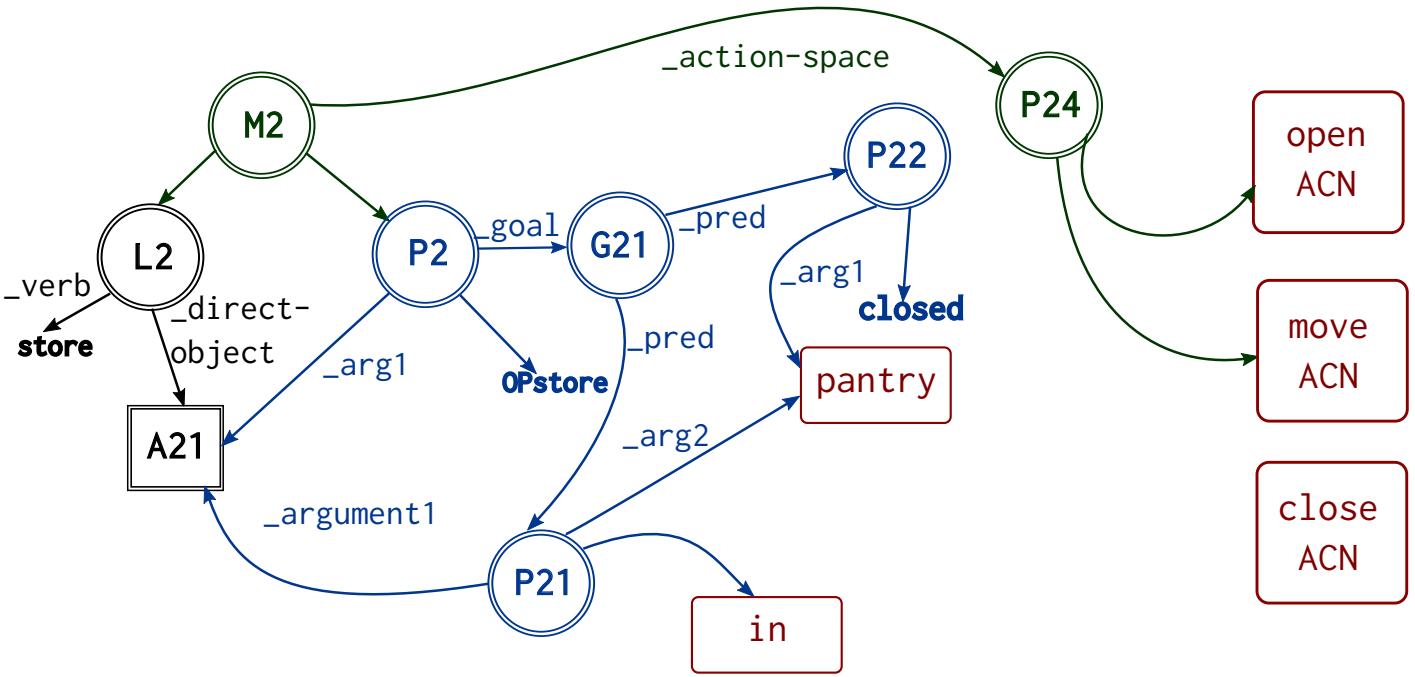
Agent: Which action should I take?

Instructor: Open the pantry.

Agent: Which action should I take?

Instructor: Move the object to the pantry.

Agent: Which action should I take?



Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

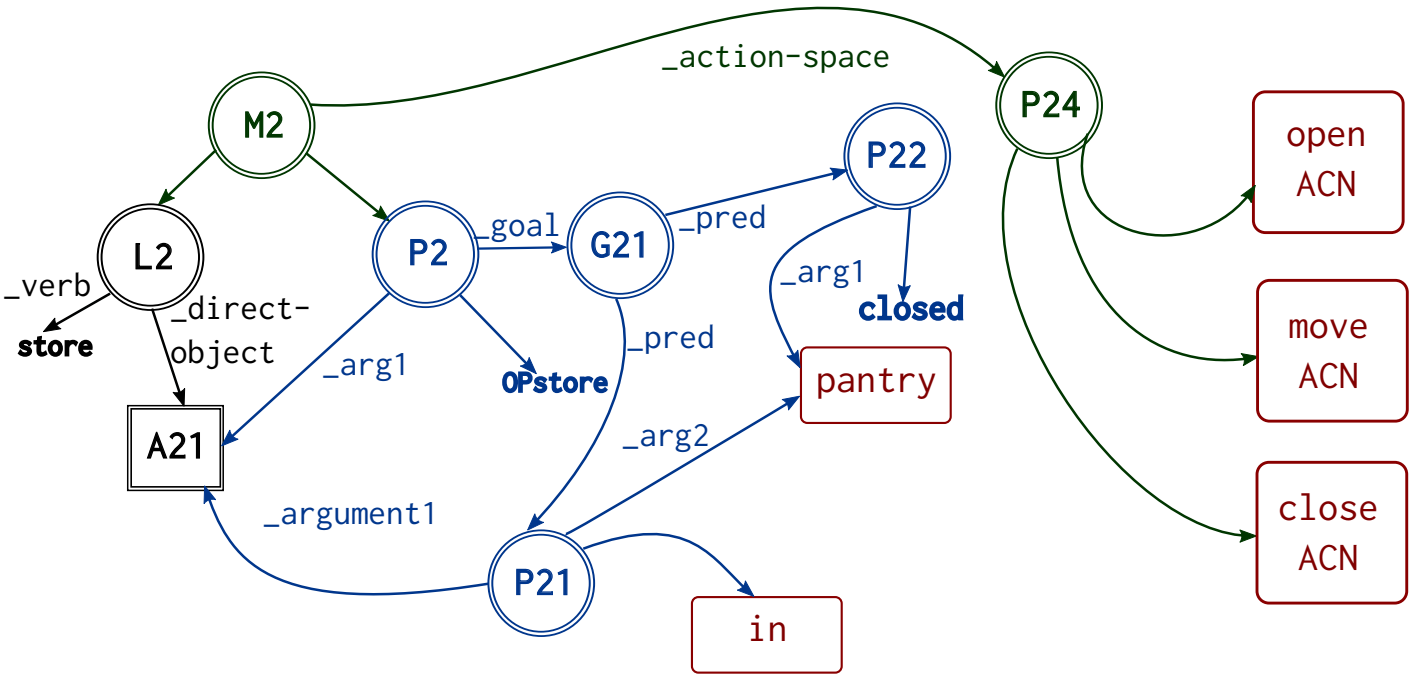
Instructor: Open the pantry.

Agent: Which action should I take?

Instructor: Move the object to the pantry.

Agent: Which action should I take?

Instructor: Close the pantry.



Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

Instructor: Open the pantry.

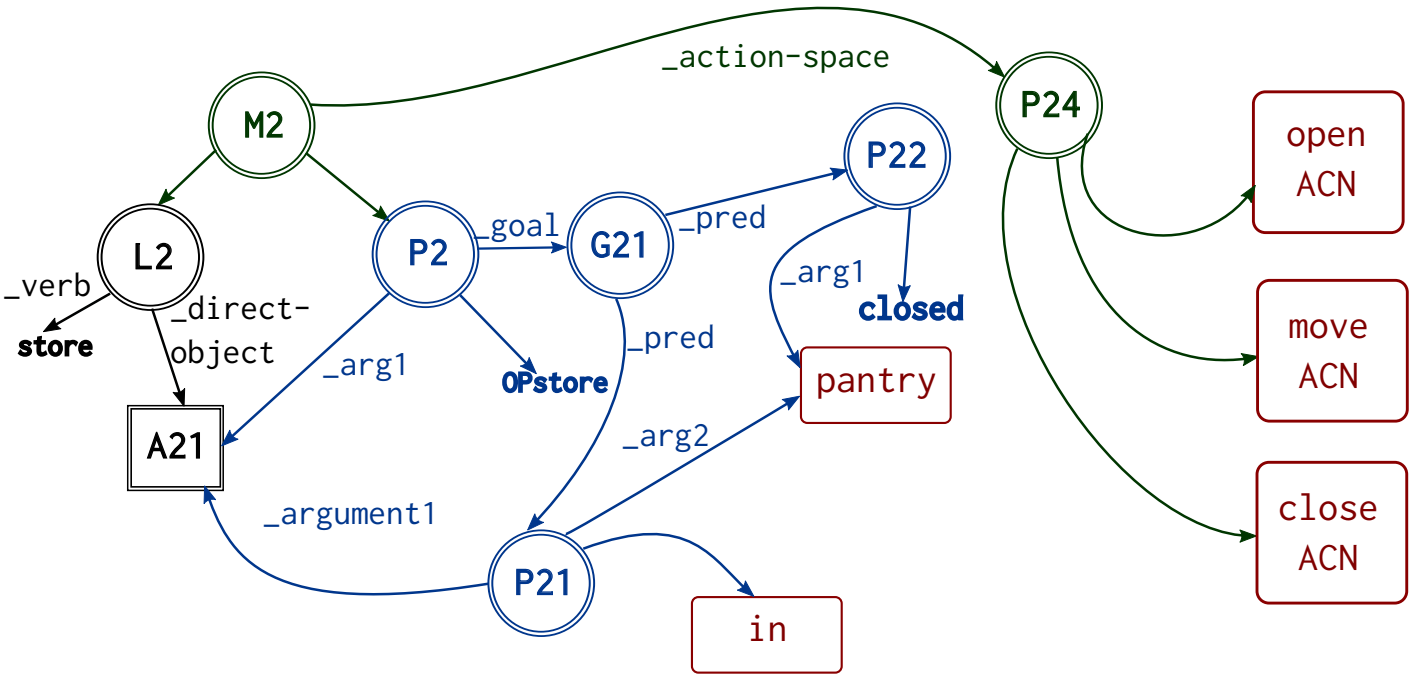
Agent: Which action should I take?

Instructor: Move the object to the pantry.

Agent: Which action should I take?

Instructor: Close the pantry.

Agent: Which action should I take?



Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

Instructor: Open the pantry.

Agent: Which action should I take?

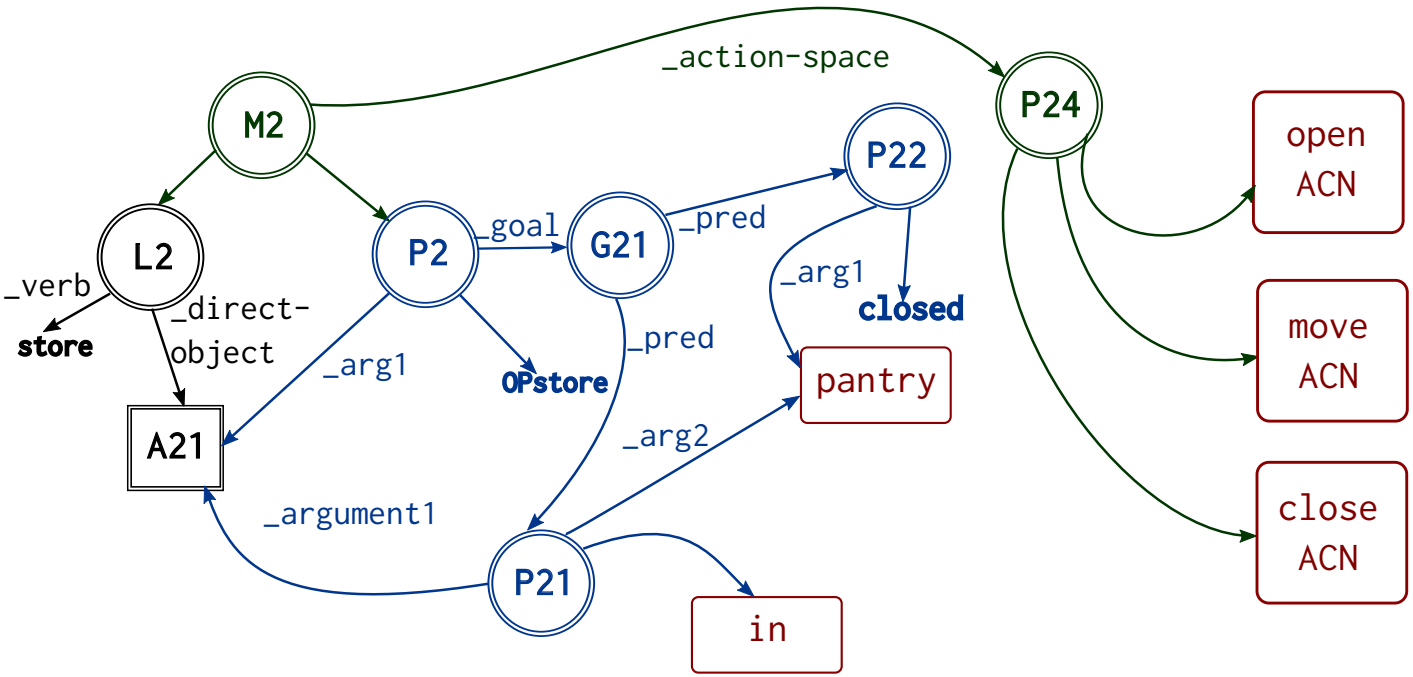
Instructor: Move the object to the pantry.

Agent: Which action should I take?

Instructor: Close the pantry.

Agent: Which action should I take?

Instructor: You are done.



Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

Instructor: Open the pantry.

Agent: Which action should I take?

Instructor: Move the object to the pantry.

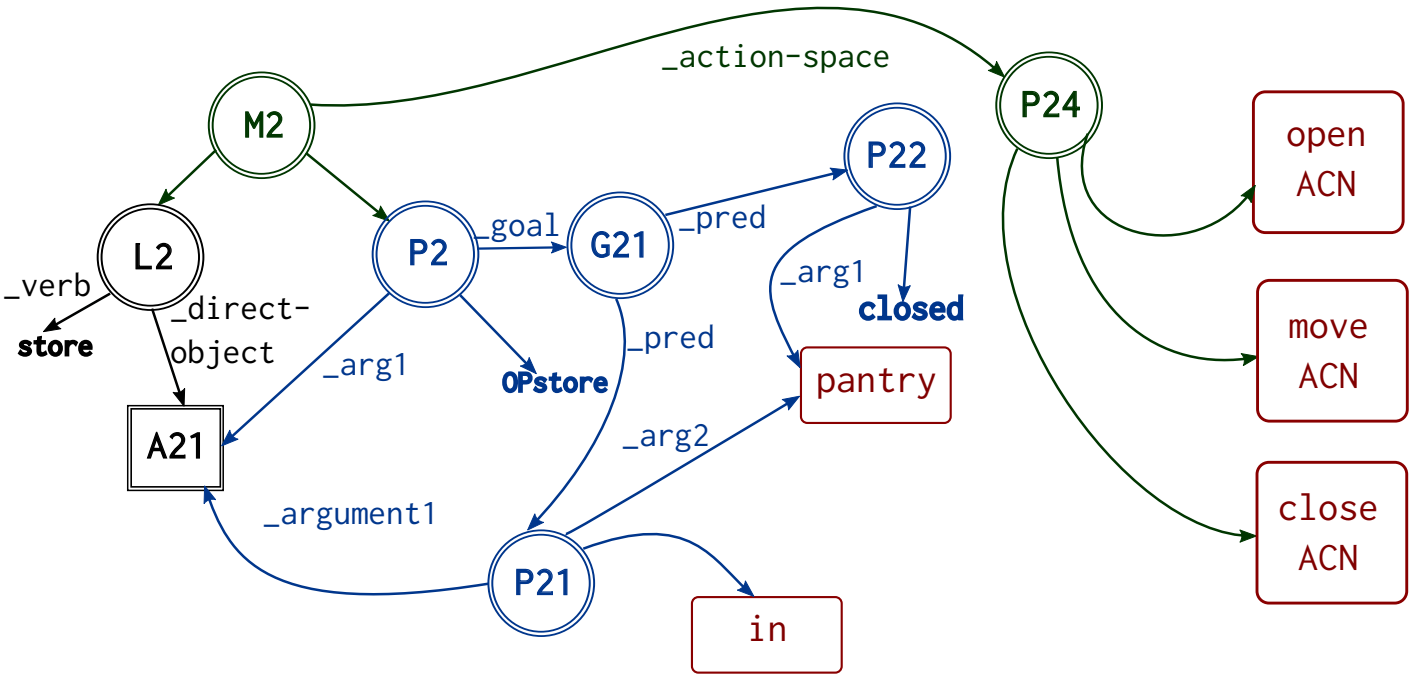
Agent: Which action should I take?

Instructor: Close the pantry.

Agent: Which action should I take?

Instructor: You are done.

Agent: `forward projection EBG`



Learning

Interaction trace

Instructor: Store the green rectangle.

Agent: What is the goal of the action?

Instructor: The goal is the green rectangle in the pantry and the pantry is closed.

Agent: Which action should I take?

Instructor: Open the pantry.

Agent: Which action should I take?

Instructor: Move the object to the pantry.

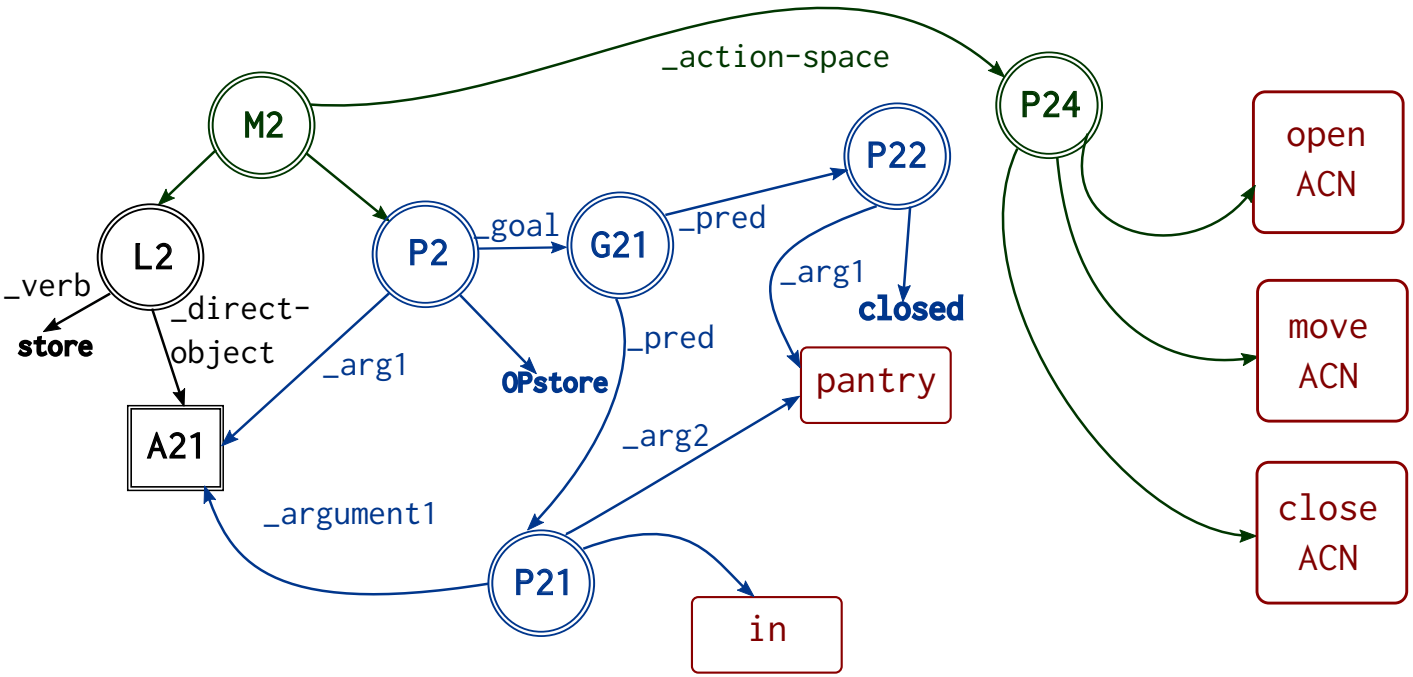
Agent: Which action should I take?

Instructor: Close the pantry.

Agent: Which action should I take?

Instructor: You are done.

Agent: `forward projection EBG`



rule-1
If operator is `OPstore` and `CLOSED(pantry)`
--> execute `open(pantry)`
rule-2
If operator is `OPstore` and `arg1` is `[A21]` and `OPEN(pantry)`
--> execute `move([A21],pantry)`
rule-3
If operator is `OPstore` and `arg` is `[A21]`, and `IN([A21],pantry)` and `OPEN(pantry)`
--> execute `CLOSE(pantry)`

Situated Comprehension

Store the green rectangle.
The goal is the rectangle is in the pantry.
...
Pick it up.
...

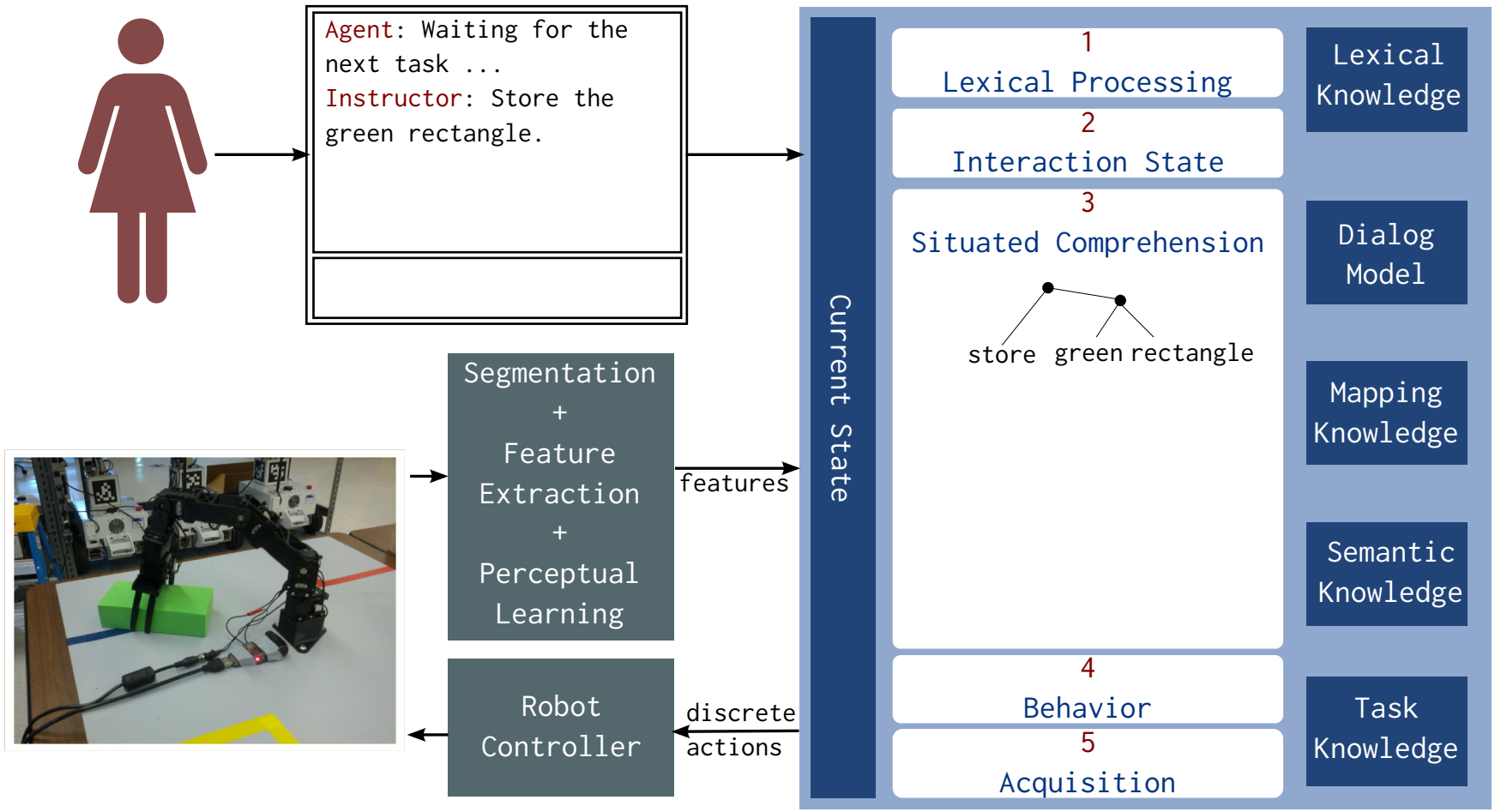
- Ideal speaker employs shared context to generate language
  - efficient, effective
- Agent collaborators should be able to use context for comprehension

Indexical Model of Comprehension

- Language features are cues to search perceptions, task knowledge, past experiences
  - language features $\rightarrow$ referents
  - *Referring expressions* $\rightarrow$ objects
  - *Verb + syntactical structure* $\rightarrow$ action-concept-network
  - *Prepositions* $\rightarrow$ spatial-relationships
- Domain knowledge is associated with referents
- Compose referents to generate a grounded representation
  - constraints: syntactic, domain knowledge, current state

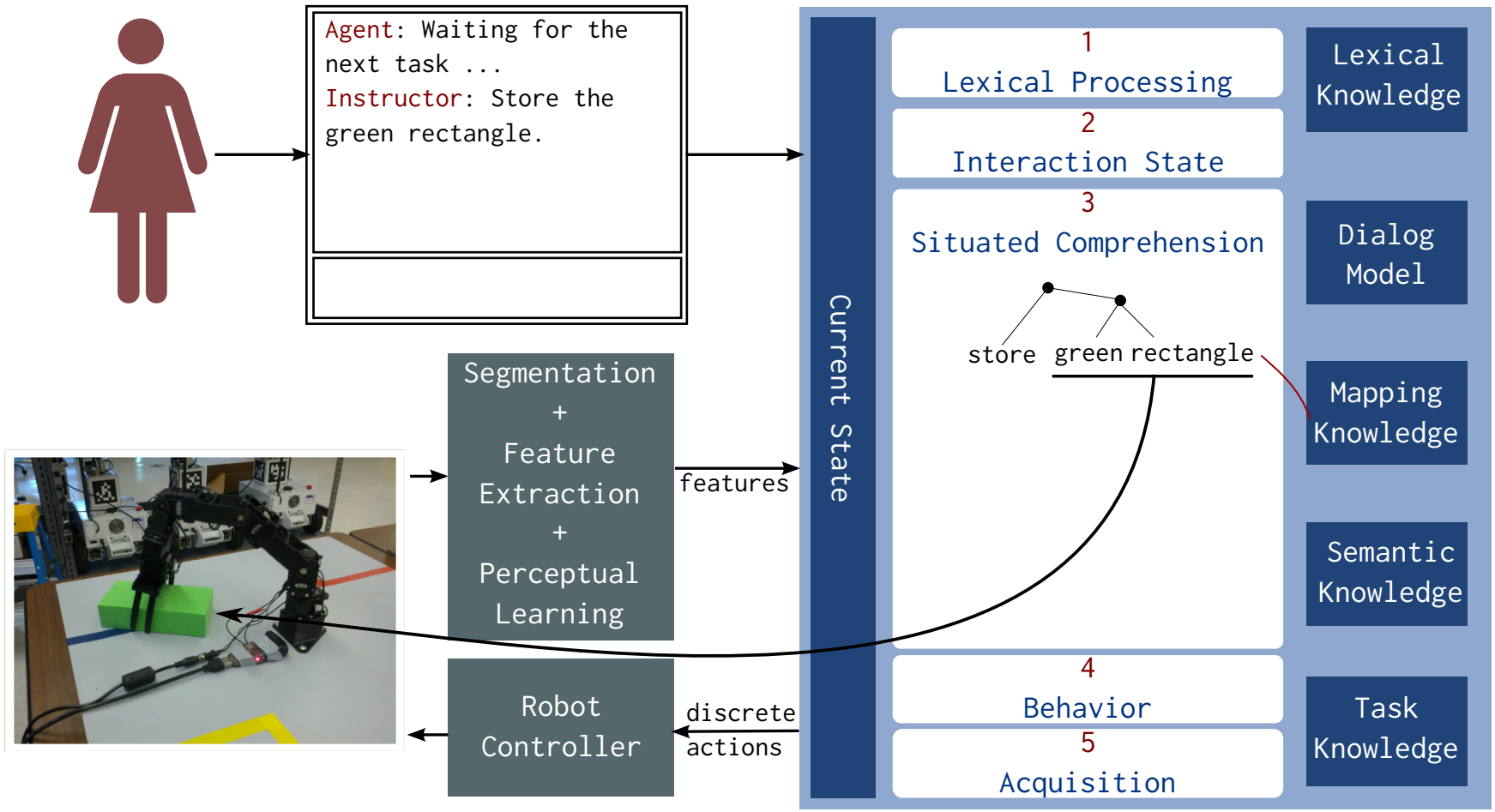
Interaction Cycle

Situated Comprehension



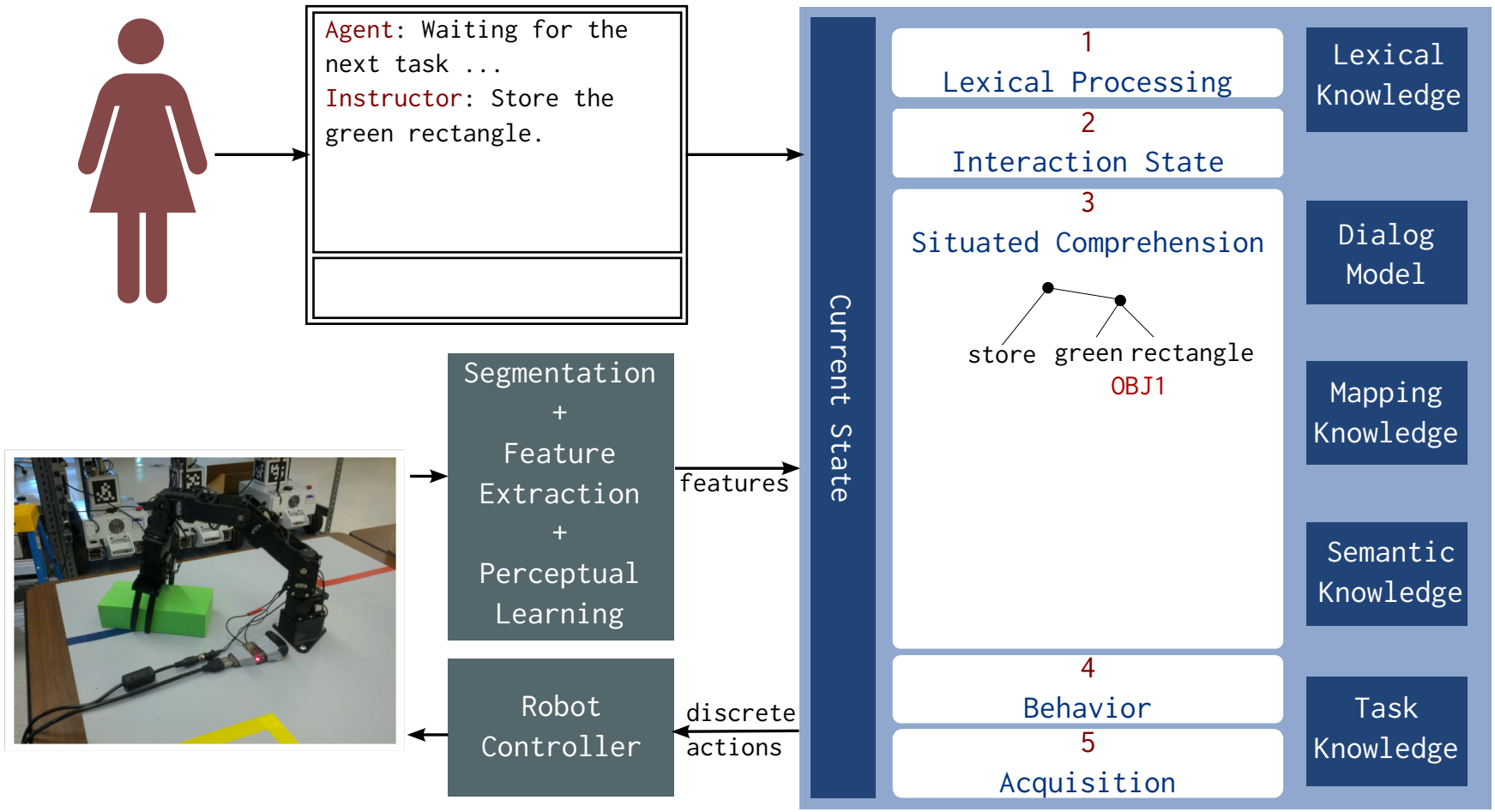
Interaction Cycle

Situated Comprehension



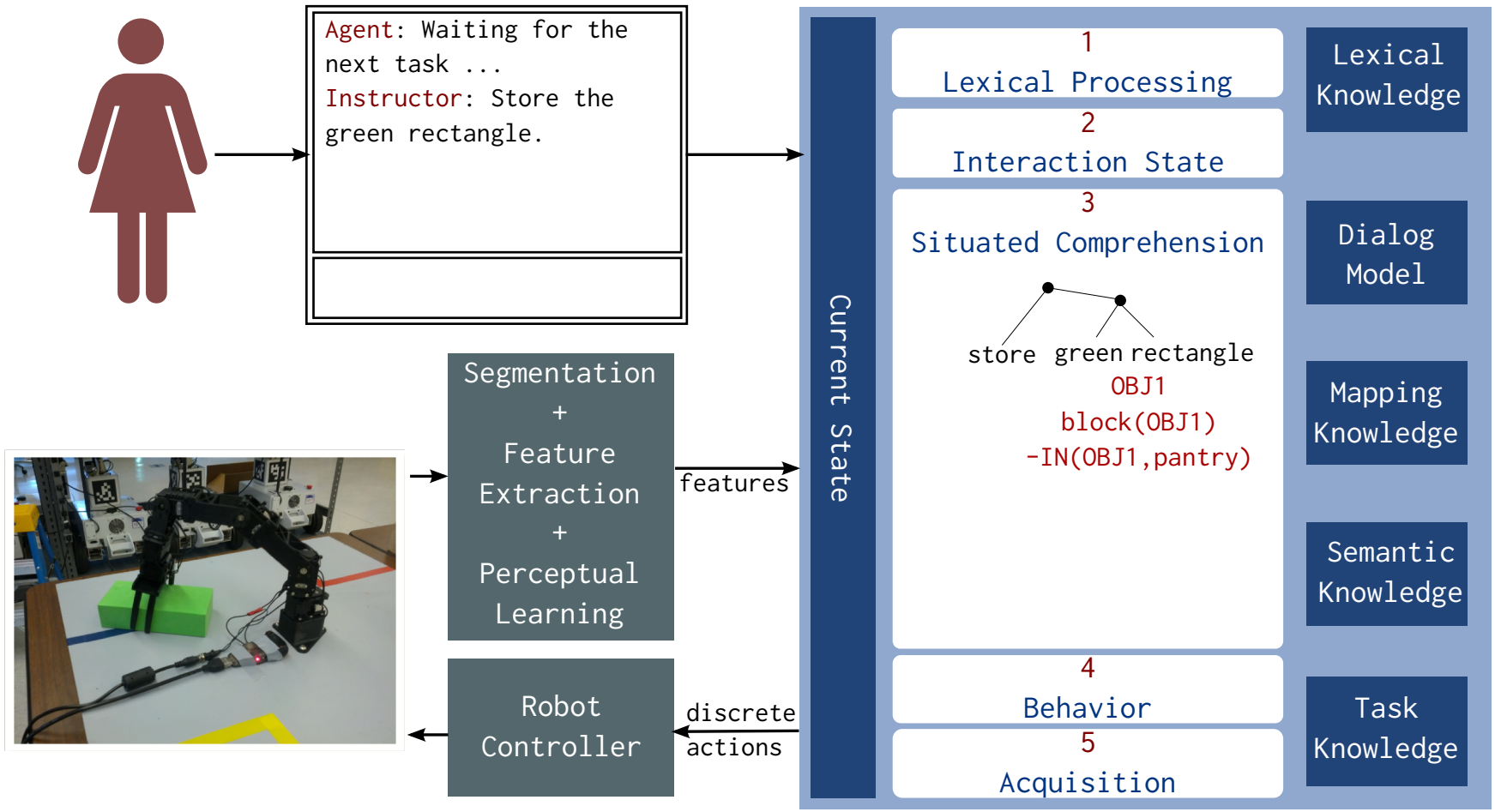
Interaction Cycle

Situated Comprehension



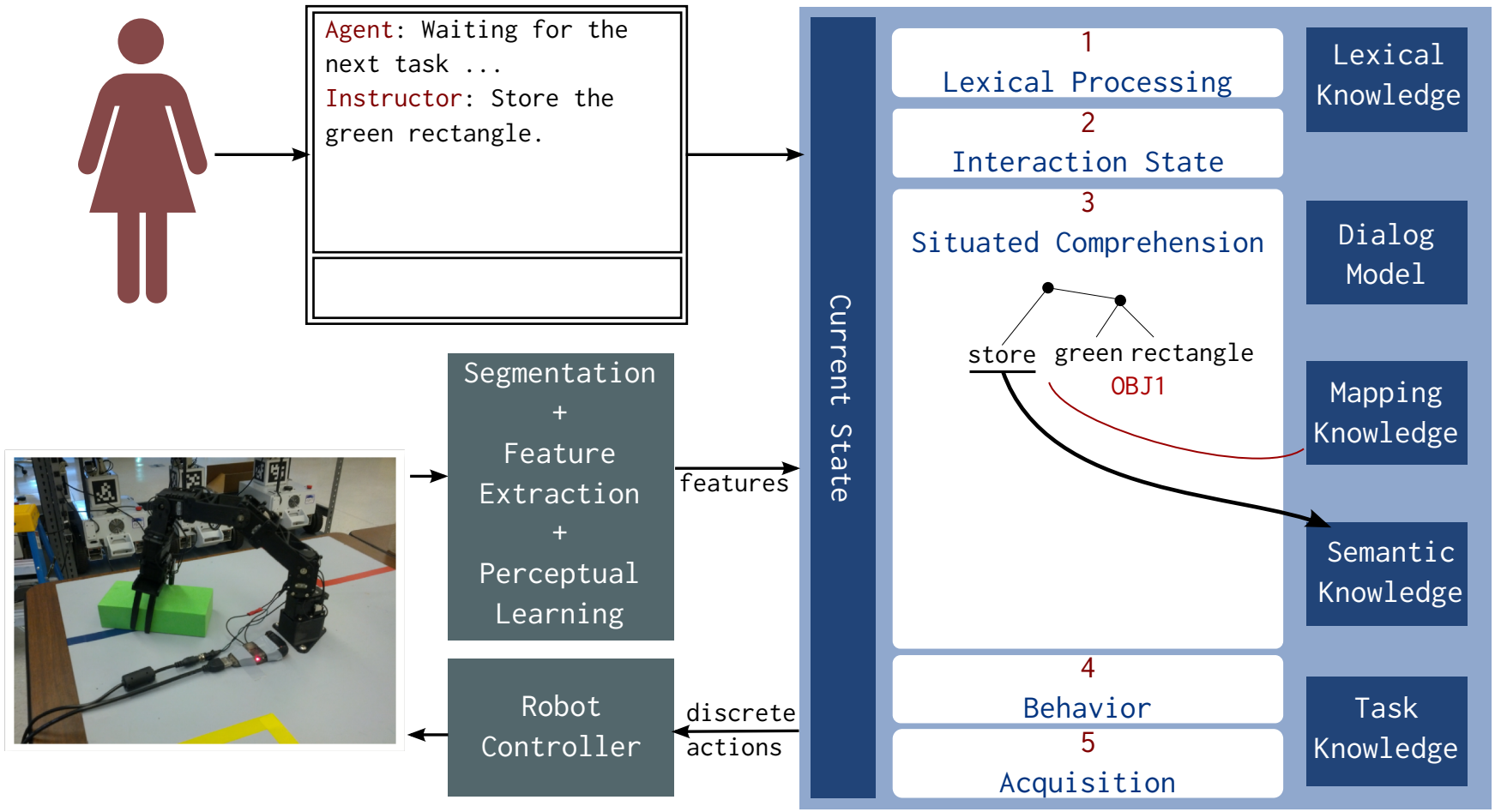
Interaction Cycle

Situated Comprehension



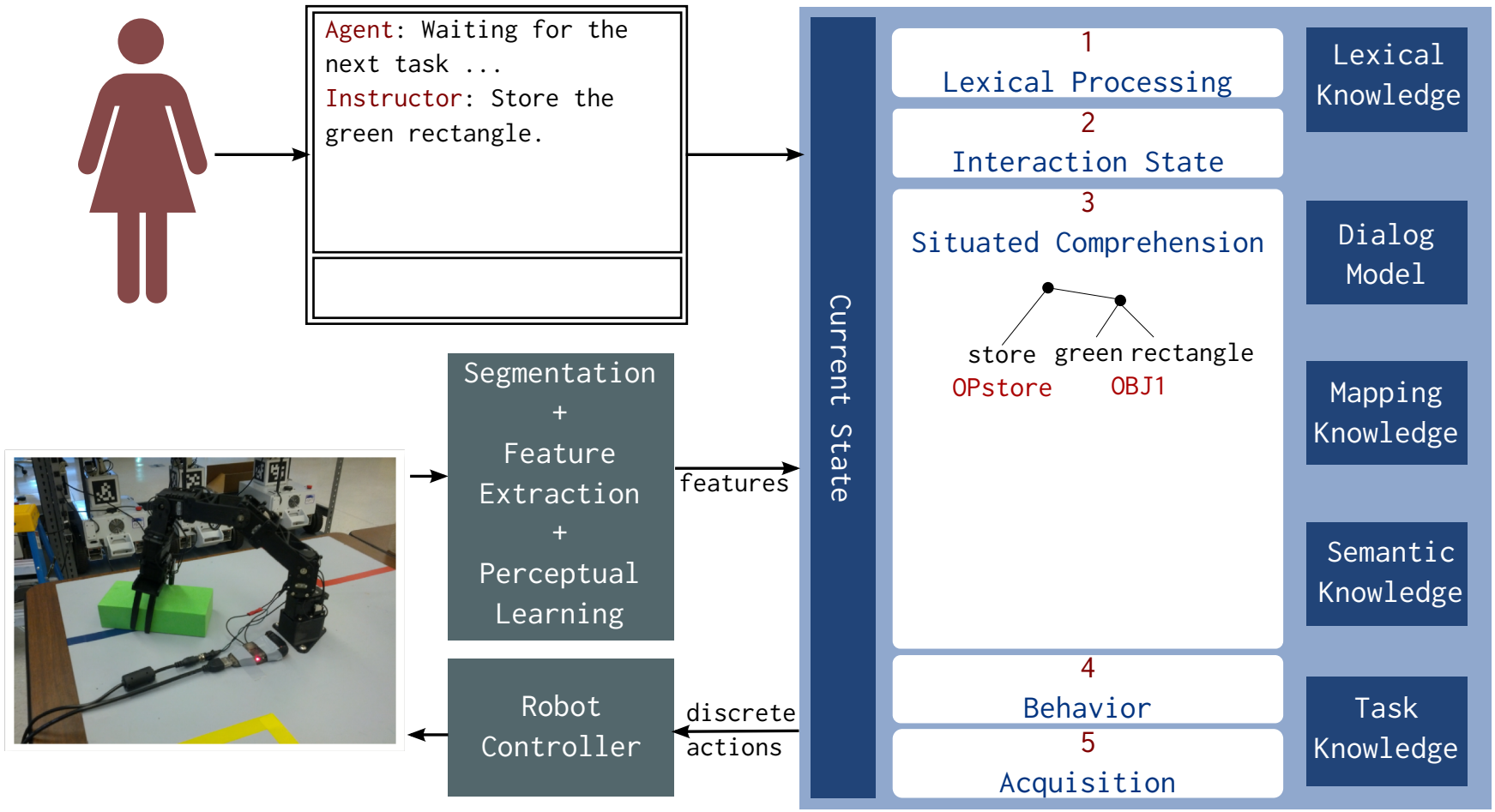
Interaction Cycle

Situated Comprehension



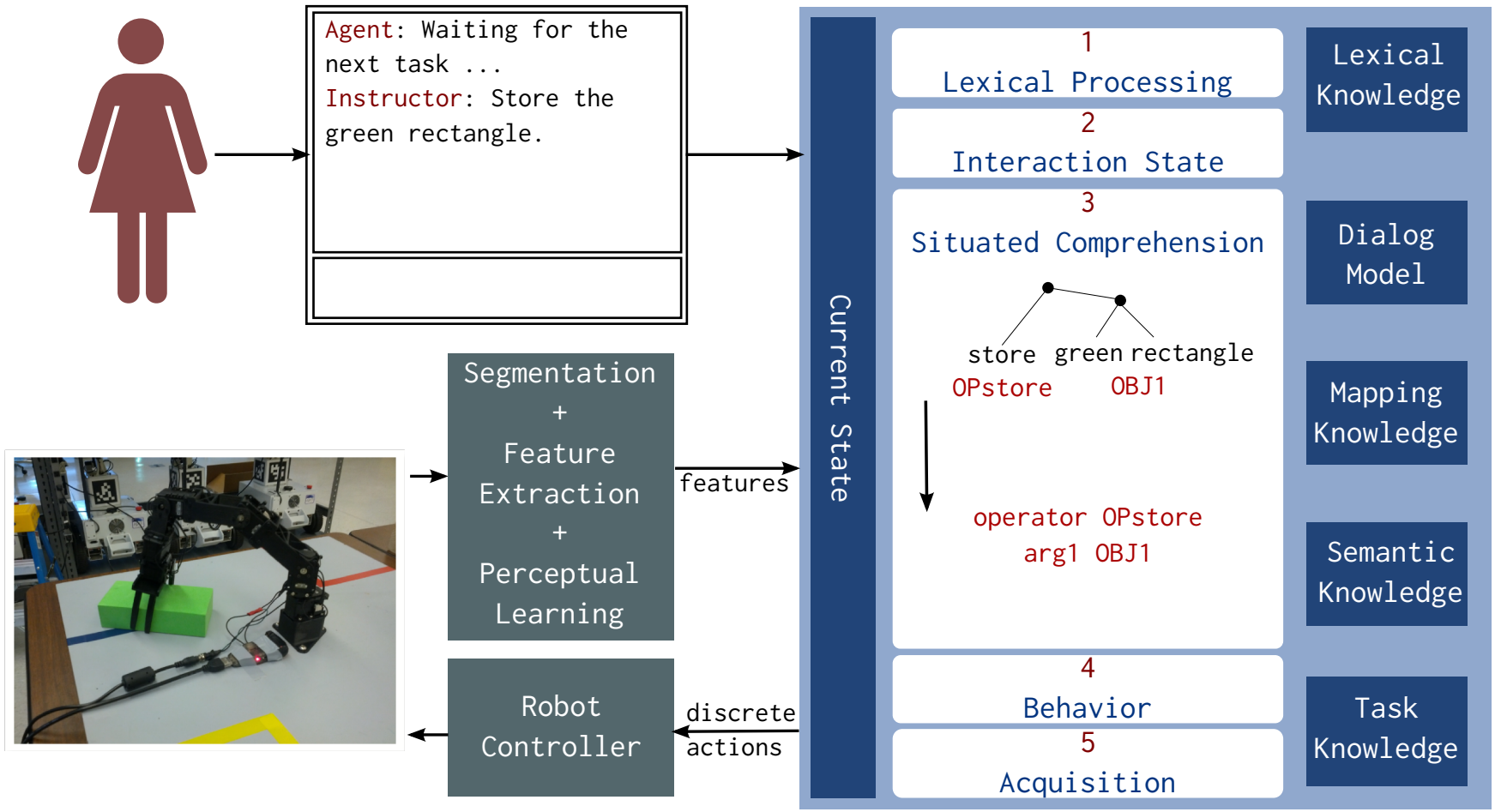
Interaction Cycle

Situated Comprehension



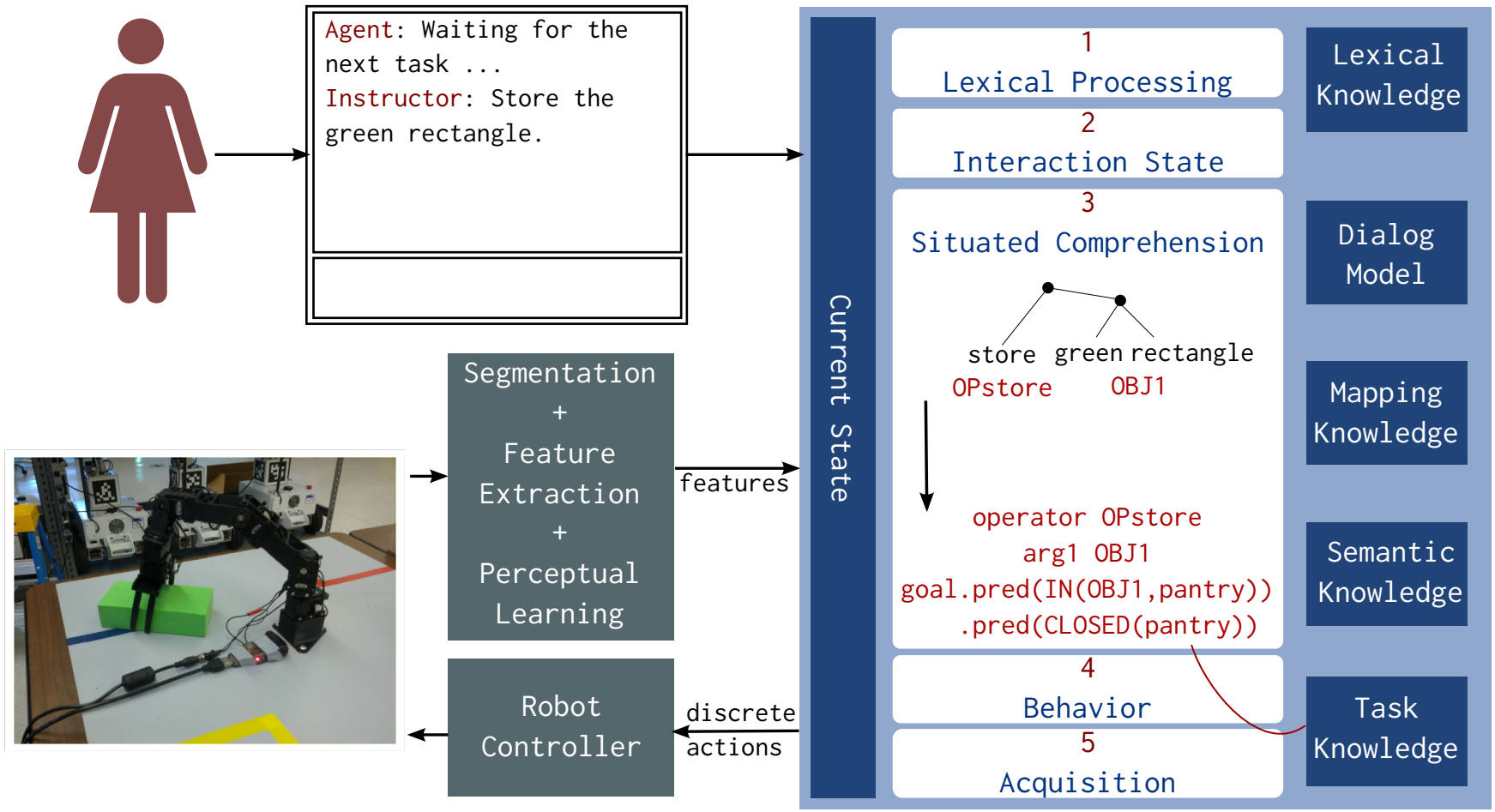
Interaction Cycle

Situated Comprehension



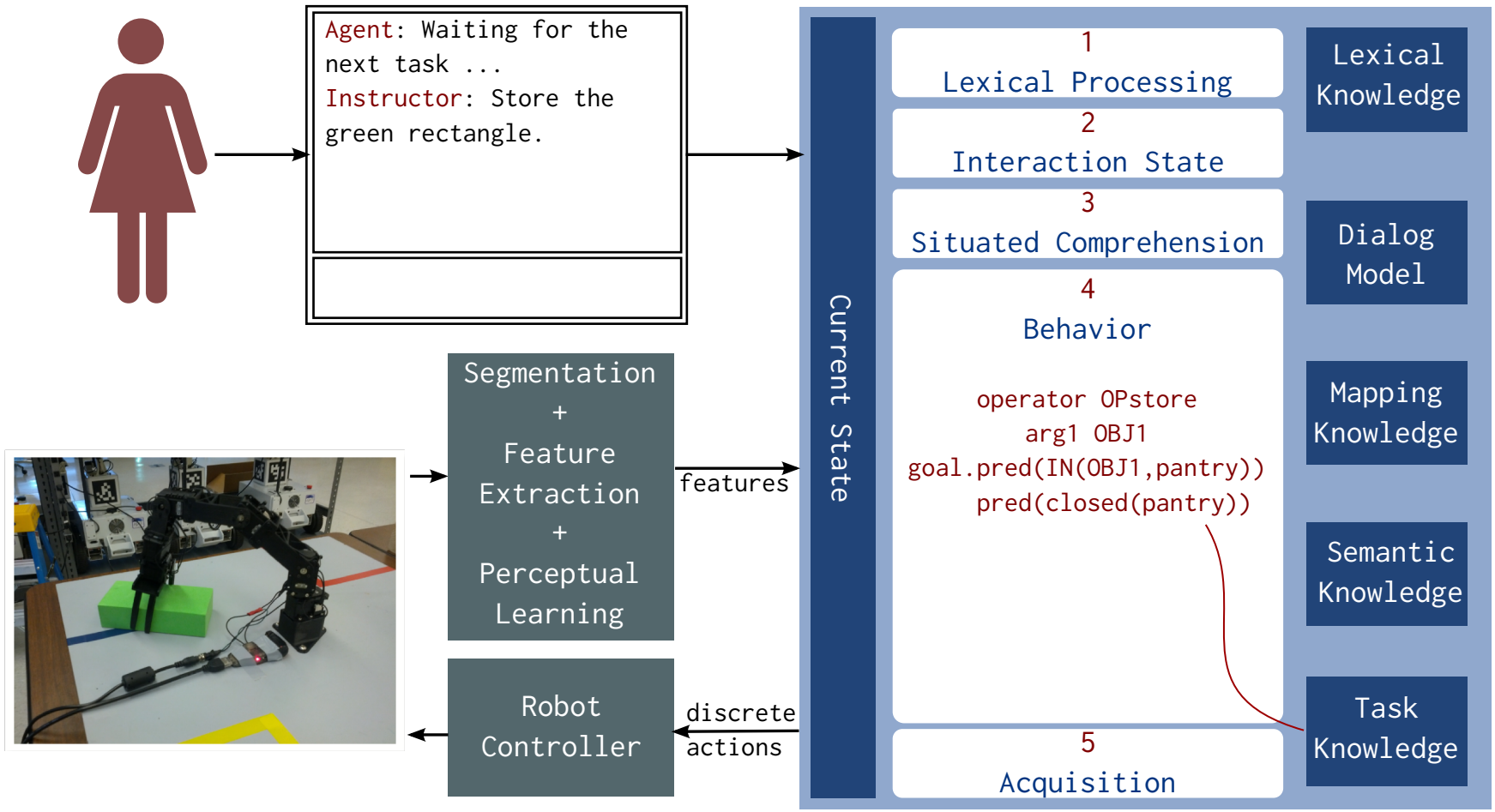
Interaction Cycle

Situated Comprehension



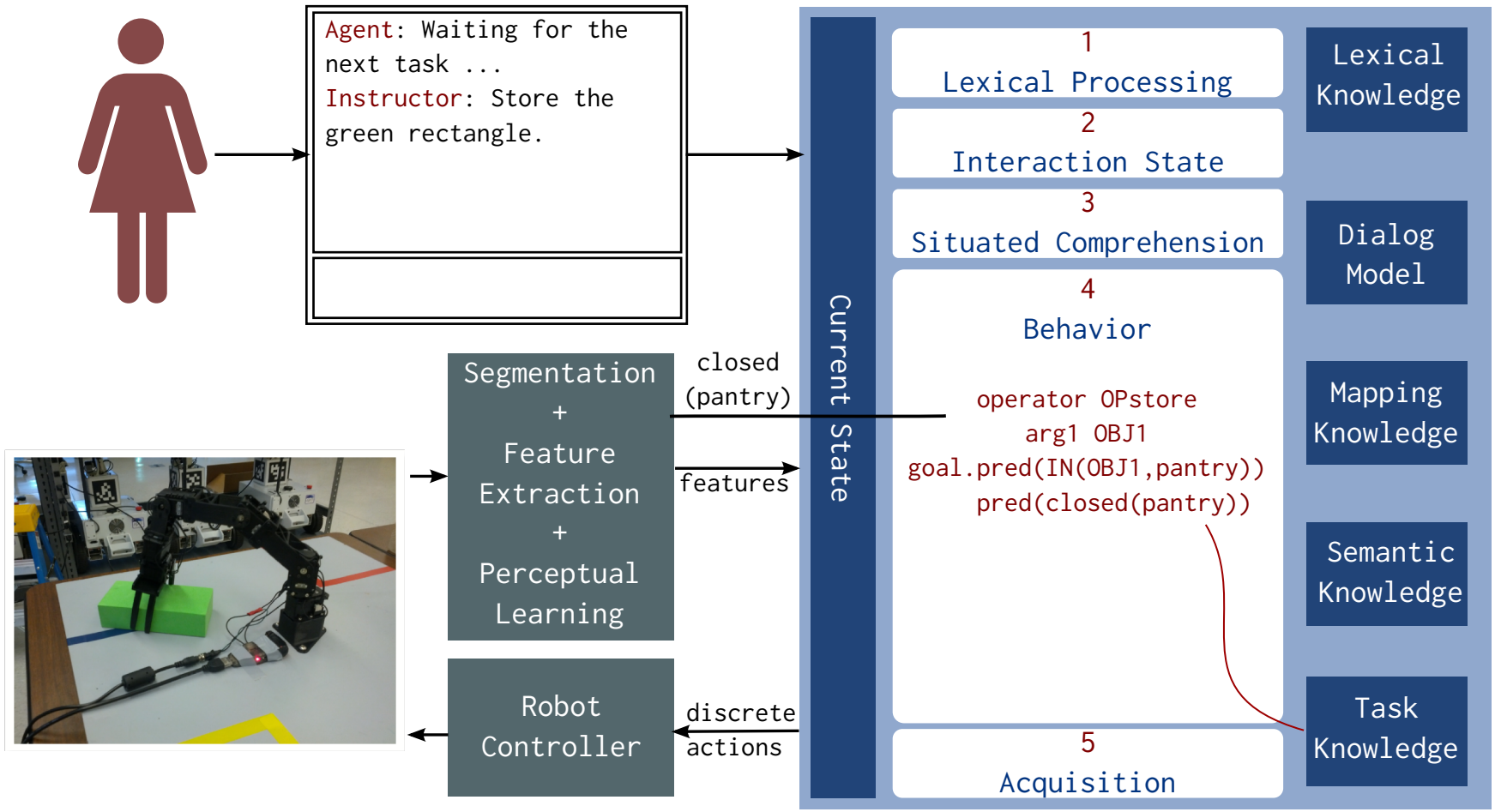
Interaction Cycle

Phase IV: Behavior



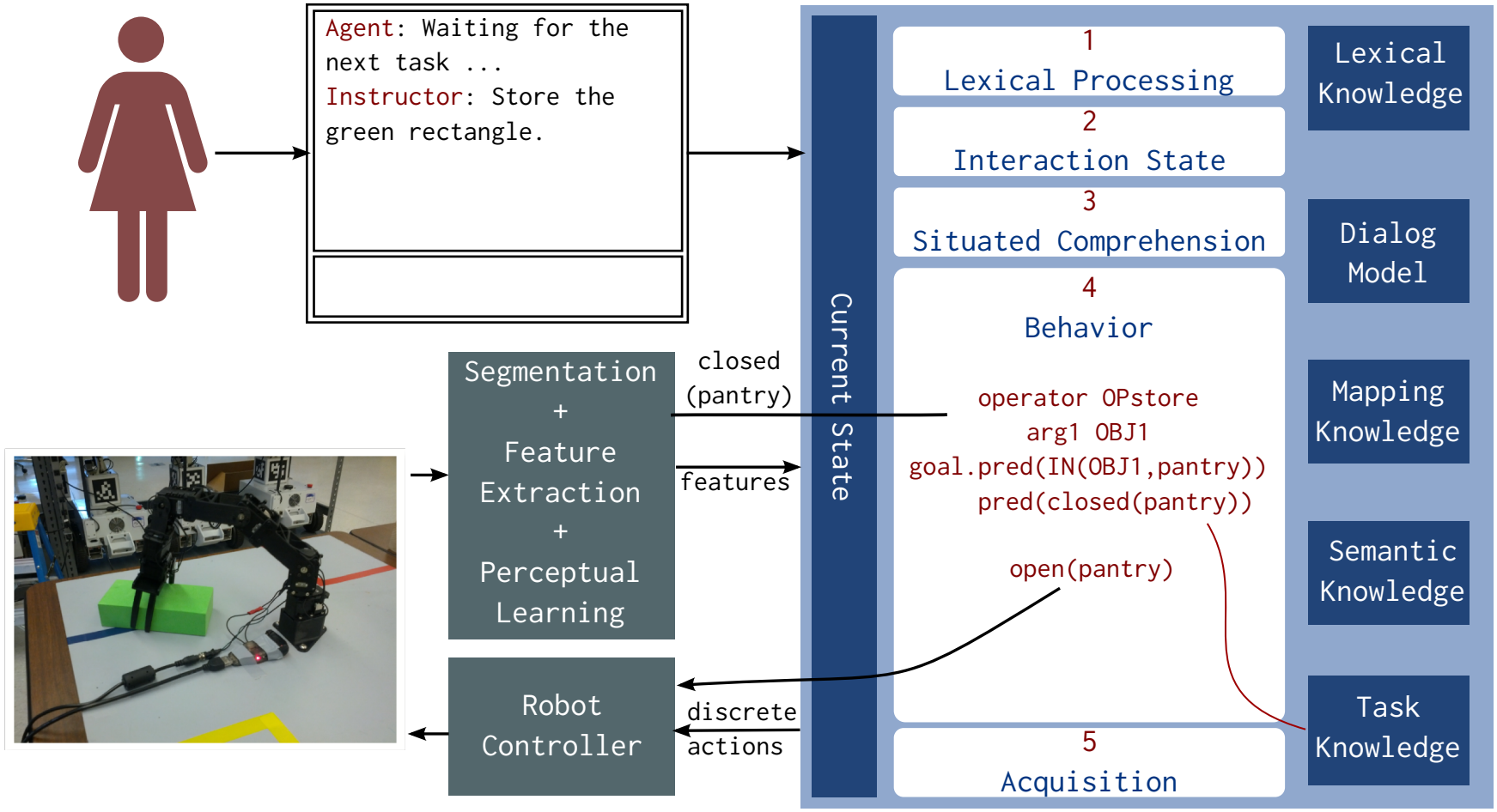
Interaction Cycle

Phase IV: Behavior



Interaction Cycle

Phase IV: Behavior



Reference Resolution

use of Referring Expressions

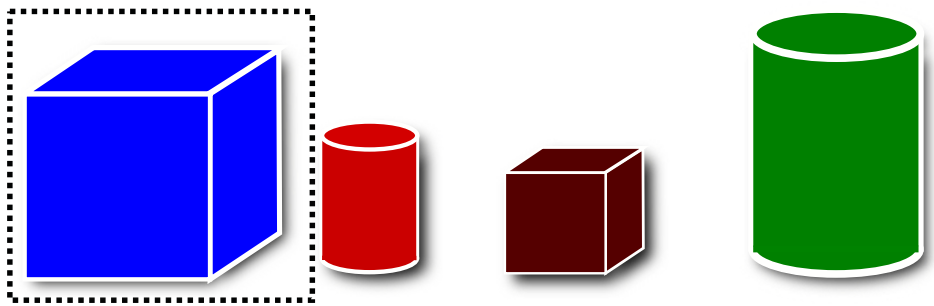
- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Reference Resolution

use of Referring Expressions

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Pick up the blue cube.

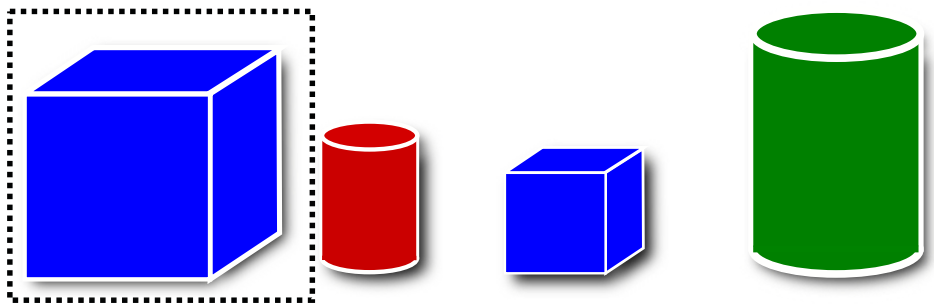


Reference Resolution

use of Referring Expressions

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Pick up the large, blue cube.

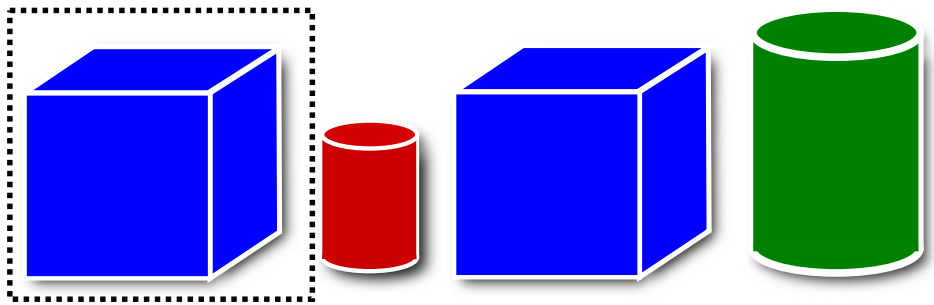


Reference Resolution

use of Referring Expressions

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

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

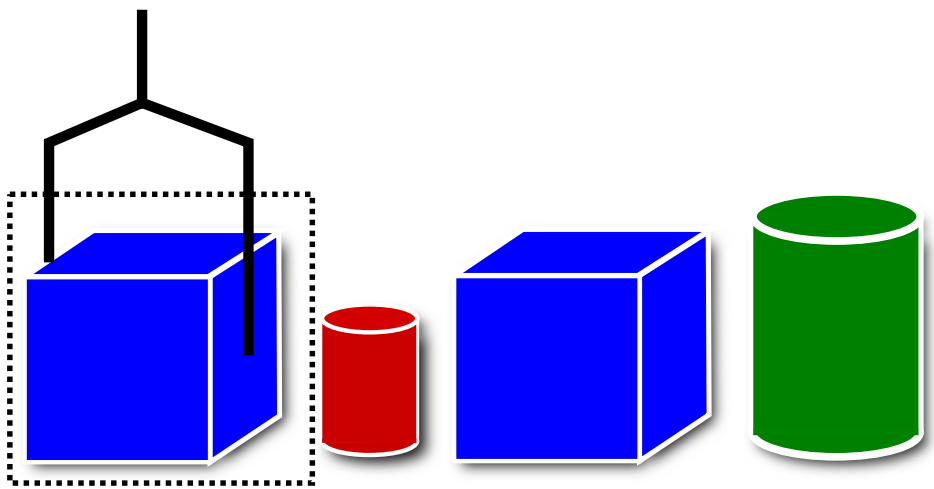


Reference Resolution

use of Referring Expressions

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Put it down.



Reference Resolution

Linguistic Theory

- Saliency based cognitive-status of referent objects
  - discourse, action, surprise
- Use of referring expression (RE) is dependent on saliency
  - more informative RE for less salient objects and vice-versa
- Givenness hierarchy (?)

  `in-focus` (*it*) > `activated` (*this, that rectangle*) > `uniquely-identifiable`(*the green rectangle*) > `type-identifiable` (*a green rectangle*)

Reference Resolution

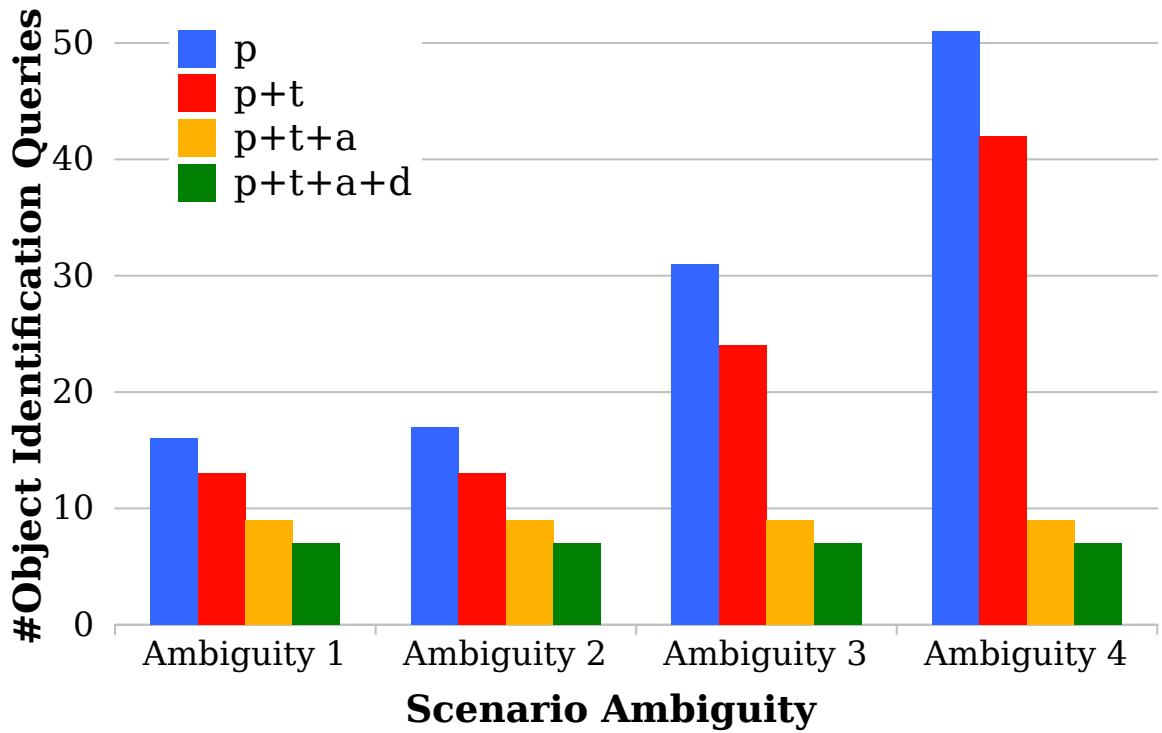
in Indexical Model

- Assign cognitive status to all objects
  - perceptible objects: $O_{percept}$
  - recently addressed objects: O_{active}
  - objects in the dialog state: O_{focus}
- Identify the referent set based on surface form of RE
  - *this cylinder* $\rightarrow O_{cand} = O_{active}$
- Apply filters
  - perceptual filter, spatial filter, task filter
- Obtain partial-ordering
  - $O_{focus} > O_{active} > O_{percept}$
- If unsuccessful, ask object-identification queries

Reference Resolution

Evaluation

scenarios: number of distractors
models: *p*, *p+t*, *p+t+a*, *p+t+a+d*
corpus: instructional dialogs, 12 personal pronouns (*it*), 4 demonstrative pronouns (*this*), 3 demonstrative phrases (*that cylinder*), and 14 noun phrases (*the red cylinder*)



Overview

- Comprehensiveness
  - representation is sufficient for a variety of verbs/actions
  - C1, C2
- Flexibility
  - learning model allows for different kinds of supervision
  - L1
- Domain complexity
  - learning and comprehension models are useful in larger, distance-limited sensing domains
  - D1, D2, D3

Proposed Work
Comprehensiveness [C1]

- Implement diverse goals (verbs)
- Predicates
  - spatial [✓]: `IN([obj1], pantry)`, `RIGHT-OF([obj1],[obj2])`
  - state [✓]: `OPEN(pantry)`, `ON(stove)`
  - know-value: `WEIGHT([obj1],[w1])`
- Compositions:
  - conjunctions [✓]
  - disjunctions
  - negations
- Policy and Goal
  - achievement [✓]: *move*, *store*, *shift*
  - maintenance: *keep the table clean*
  - performance: *patrol*
- Extend learning model, representation

Proposed Work
Comprehensiveness [C2]

- Verb execution with different type of arguments
  - different goal, similar action-space/policy
    - *store NP*
    - NP → perishable object, goal: `IN([obj],refrigerator)`
    - NP → non-perishable object, goal: `IN([obj], pantry)`
    - NP → to tool object, goal: `IN([obj], drawer)`
  - similar goal, different action-space/policy
    - *cook NP*
    - NP → *eggs*, action-space: `{move(ON,[obj],stove), turn-on(stove), wait(5), turn-off(stove)}`
    - NP → *chicken*, action-space: `{thaw([obj]), move (IN,[obj],oven), set-temprature(450), wait(30), turn-off(oven)}`
- Extend learning and comprehension models

Proposed Work
Flexibility [L1]

- Policy learning
  - complete [✓]

    `store([obj1])` → `open(pantry)`, `move([obj1], pantry)`, `close(pantry)`
  - partial

    infer implicit actions

    `store(<obj1>)` → `move([obj1], pantry)`
  - unguided

    infer actions from goal description
- Extend learning model

Proposed Work
Domain Complexity [D1]

- Current learning model learns from a small set of primitives
- Expand the set for larger, complex domains
- Larger hierarchies of action execution

Proposed Work
Domain Complexity [D2]

- Current comprehension model is grounded in perceptions.
- Complex, partially-observable domains with distance limited sensing.
- Displaced
  - *move the red cylinder to the kitchen*
  - *kitchen* is not observable
  - rely on prior experience with the domain, episodic memory, semantic memory
- Inferential indexing
  - *get the red cylinder.*
  - *red cylinders* are not observable
  - domain knowledge: red cylinders are in the pantry in the kitchen.

Proposed Work
Domain Complexity [D3]

- Current acquisition model learns from execution traces
- Execution in complex domains is time consuming
- *Store the red cylinder*
  - *pick it up*

    go to the kitchen,
    open the pantry,
    put it in the pantry,
    close the pantry.
  - learn via knowledge of the domain, and symbolic simulation

Timeline

- April 2013 - August 2013
  - Implement diverse goals (C1)
  - Expand the set of primitive actions (D1)
- August 2013 - December 2013
  - Verb execution polysemy (C2)
  - Extend active learning (partial/un-guided learning) (L1)
- January 2014 - April 2014
  - Comprehension with displaced indexing (D2)
  - Learning with displaced indexing (D3)
- May 2014 - August 2014
  - analysis, writing, and defense

Thanks.

Questions?

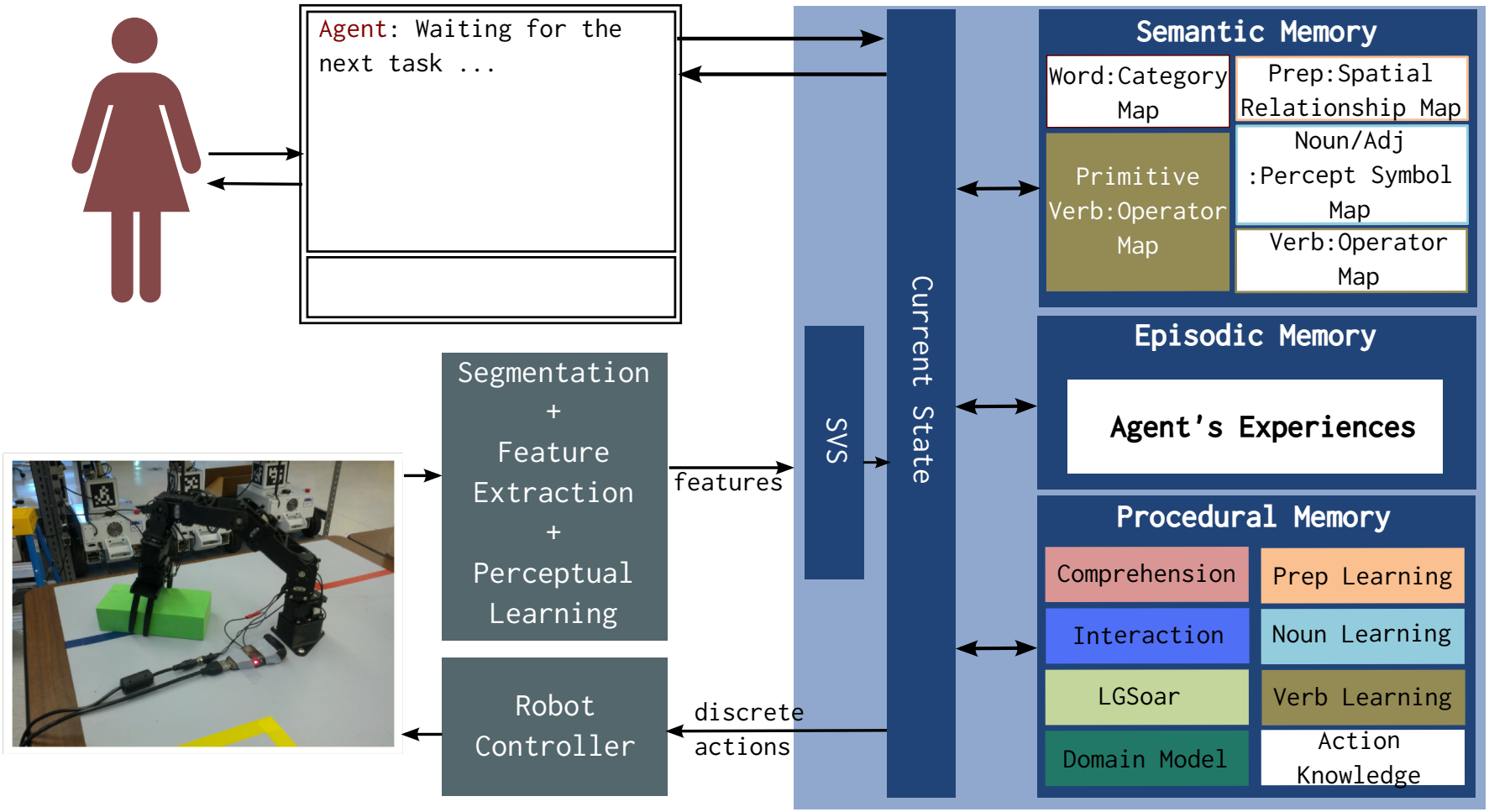
Summary

situation →	immediate	hypothetical	past
capability ↓			
representation	C1		
comprehension	C2, D2	D3	×
learning	C1,C2,L1	D3	×

Evaluation Approach

- Representation
  - comprehensive: can represent diverse kinds of *action* verbs
  - polysemous: verb with different policies
  - learn with interaction
- Learning evaluation
  - fast generalization: learn from a few number of example executions
  - incremental: learning from `incomplete' example executions
  - performance: correctness of execution
  - robustness: non-deterministic actions, partially observable domains
- Comprehension evaluation
  - correctness
  - usefulness of diverse contexts
    - reference resolution, verb execution ambiguity
  - usefulness of experience
    - partially observable environments

System Overview



Requirements

- Requirements
  - R1 diverse
    - represent different kinds of knowledge
  - R2 comprehensive
    - represent a variety of verbs and actions
  - A1 active
    - distributed onus of learning
    - agent contributes to learning by asking questions
  - A2 assimilative
    - information in interactions is dispersed in time
    - agent assimilates information
  - A3 general
    - exploit all information in instruction
    - fast generalization

Challenges (representation)

Earlier work has focussed on uni-directional mapping for simple actions;
verb $\rightarrow$ action.
move $\rightarrow$ `move-one-step` in simulated environment (??)
move in real-world domain corresponds to a policy

- Related Work
  - Verb: VerbNet (?)
    - encodes semantic knowledge + syntactical constraints
    - thematic roles, selectional restrictions, frames, semantic predicates (cause, manner etc.)
  - Action: goals and behavior (?)
    - types of predicates, compositions of predicates, temporal behavior
  - Action: behavior ?
    - operator proposal, availability, termination

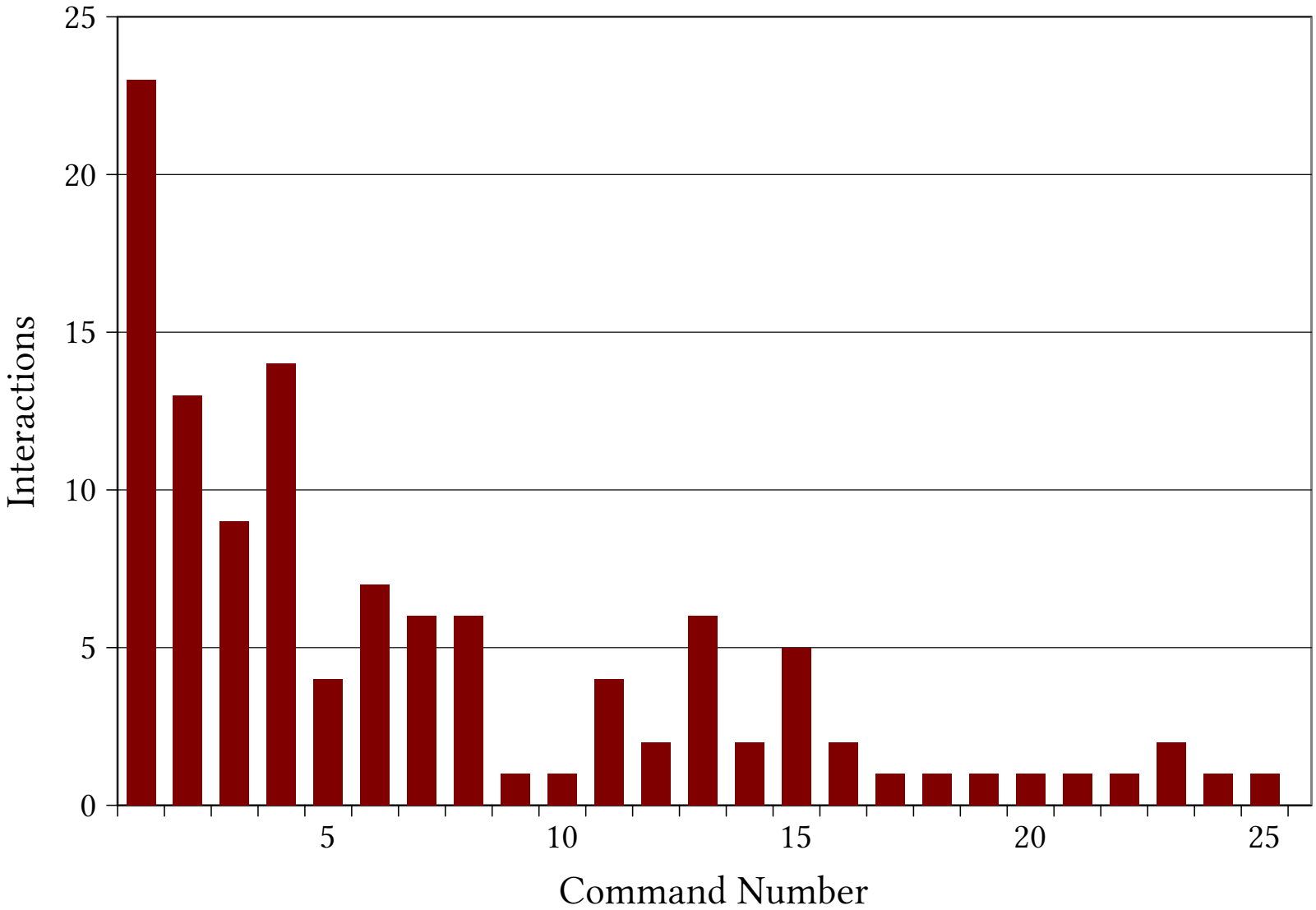
Challenges (learning)

Learn various components of action representation (and syntactic constraints) through interaction.

- Action representation components
  - parameters: inductive, heuristics
  - termination conditions: declarative, inductive
  - availability conditions: analytical (known termination, model), inductive
  - action space: inductive, heuristics
  - policy: analytical (known termination, model), inductive
  - model: inductive, derivative (simulate policy)
- Active, assimilative learning
  - When to ask a query?
  - What query to ask?
  - How to integrate the information in prior knowledge?

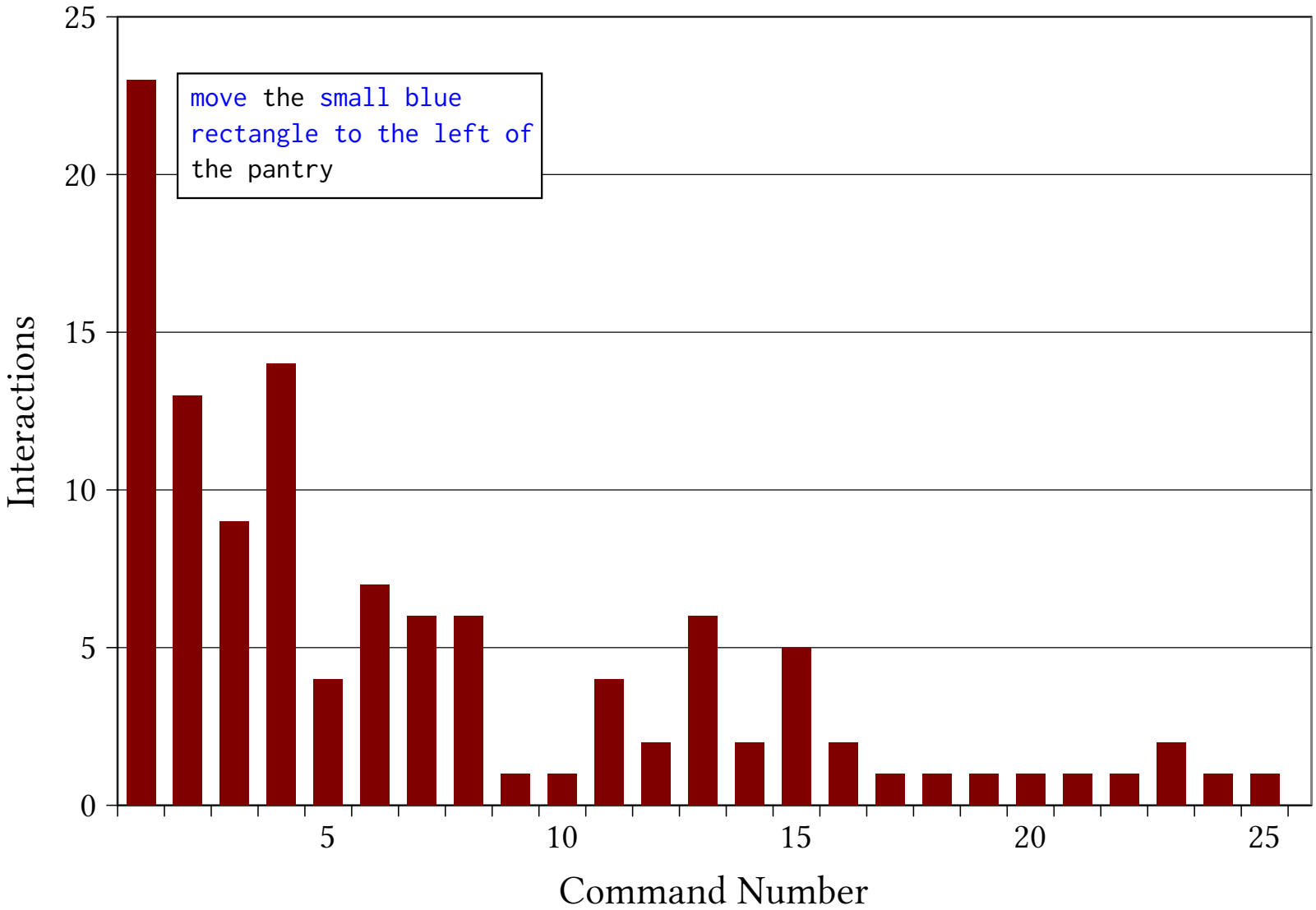
System Evaluation

3 verbs: *move*, *store*, *discard*, 9 nouns/adjectives, 3 prepositions



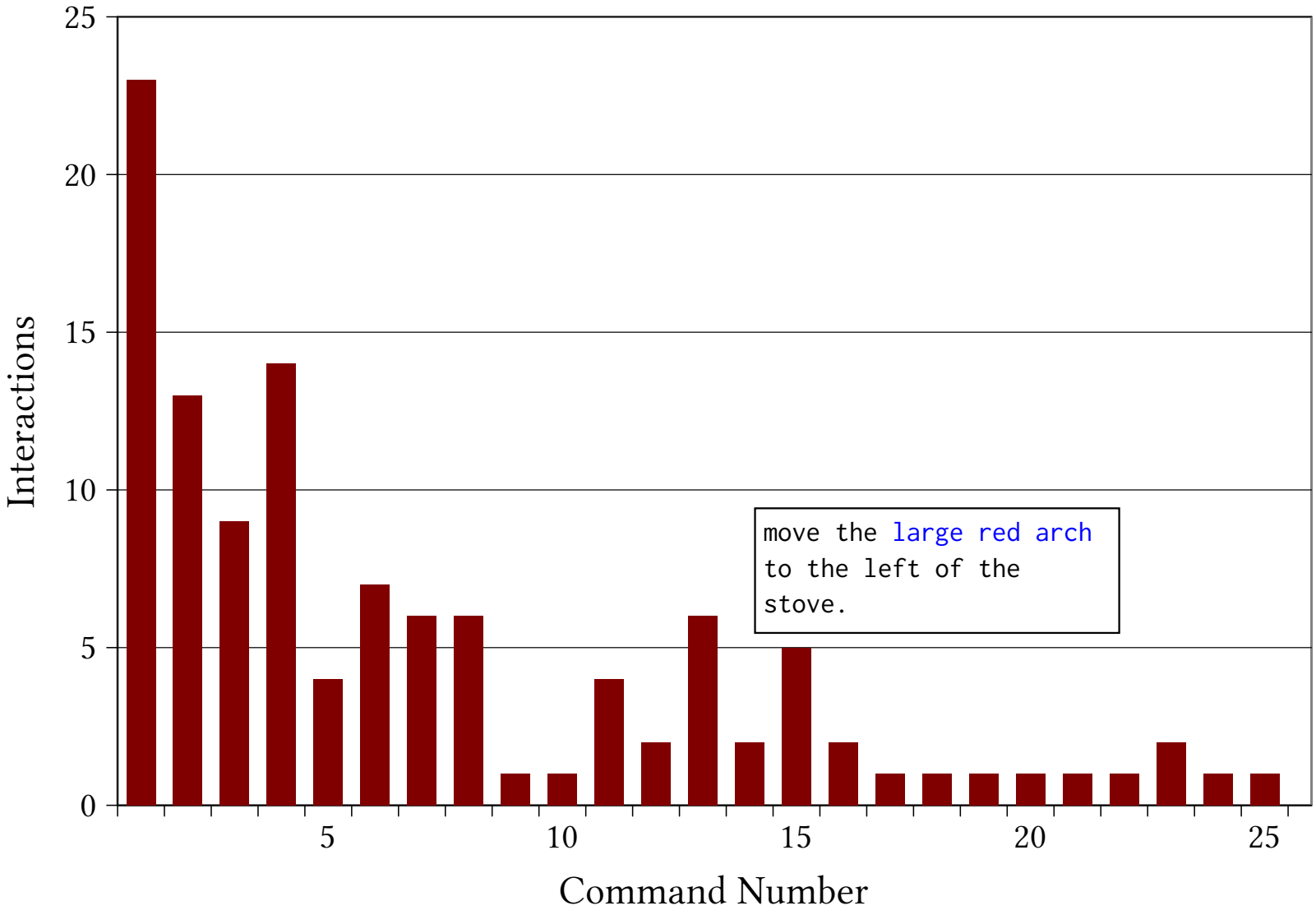
System Evaluation

3 verbs: *move*, *store*, *discard*, 9 nouns/adjectives, 3 prepositions



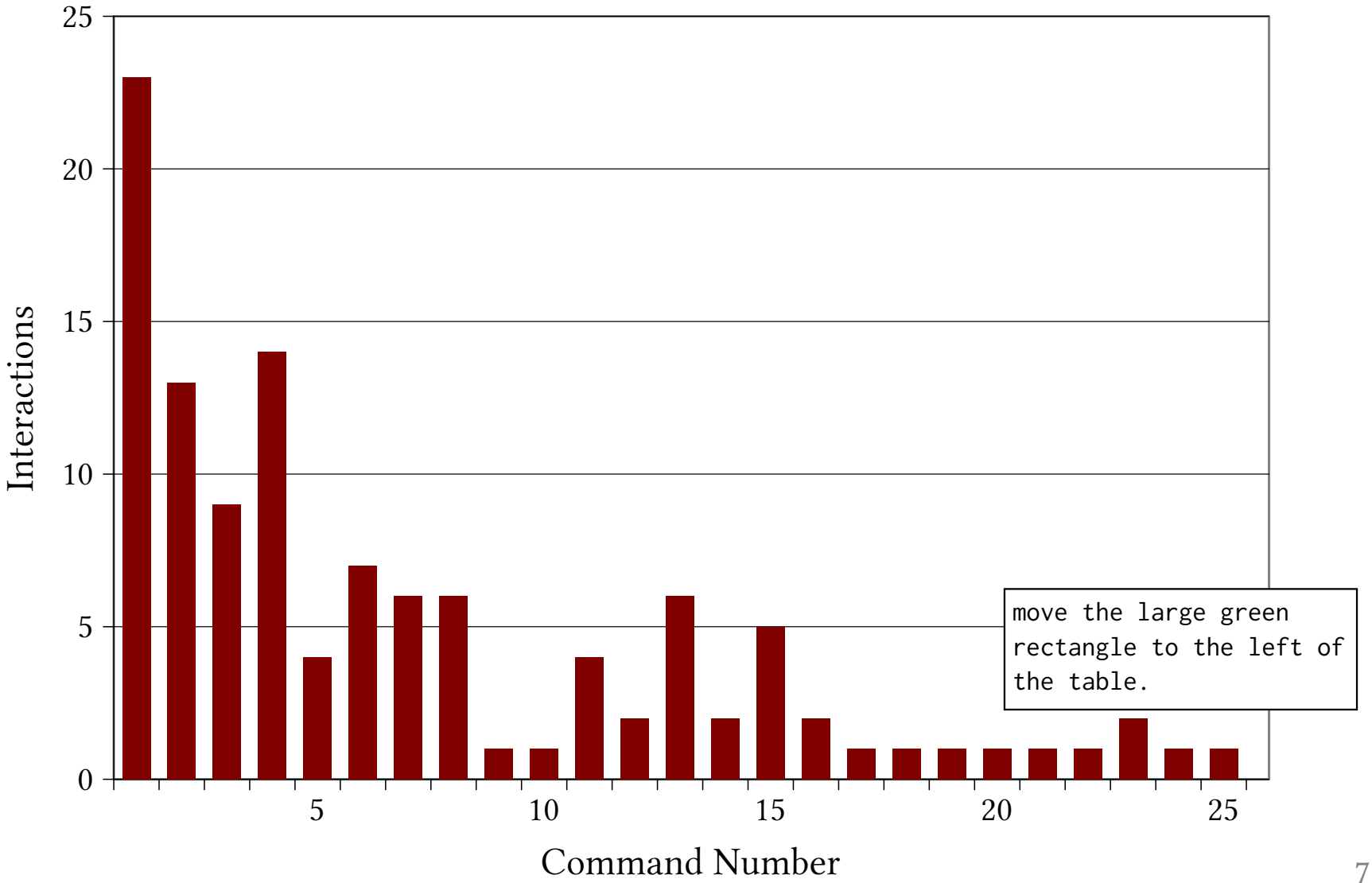
System Evaluation

3 verbs: *move*, *store*, *discard*, 9 nouns/adjectives, 3 prepositions



System Evaluation

3 verbs: *move*, *store*, *discard*, 9 nouns/adjectives, 3 prepositions



Proposal

Goal: A model to translate linguistic utterance to internal representation used for reasoning, planning, actions, and learning using shared perceptions, shared knowledge.
Implement: Grounding functions $G^K_{action}(i) : I_{AC} \rightarrow A$, $G^K_{goal}(i); I_{GD} \rightarrow S$ ...

Proposal

Goal: A model to translate linguistic utterance to internal representation used for reasoning, planning, actions, and learning using shared perceptions, shared knowledge.
Implement: Grounding functions $G^K_{action}(i) : I_{AC} \rightarrow A$, $G^K_{goal}(i); I_{GD} \rightarrow S$ ...

- Assumptions
  - constrained by interactions required for learning
  - intentions assigned through heuristics
  - *imperative sentences* $\rightarrow$ action commands
  - *goal descriptions* $\rightarrow$ state description

  ...

Proposal

Goal: A model to translate linguistic utterance to internal representation used for reasoning, planning, actions, and learning using shared perceptions, shared knowledge.
Implement: Grounding functions $G^K_{action}(i) : I_{AC} \rightarrow A$, $G^K_{goal}(i); I_{GD} \rightarrow S$ ...

- Assumptions
  - constrained by interactions required for learning
  - intentions assigned through heuristics
  - *imperative sentences* $\rightarrow$ action commands
  - *goal descriptions* $\rightarrow$ state description

    ...
- Requirements
  - Referential
    - a theory of how linguistic symbols are connected to extra-linguistic knowledge
  - Expandable
    - capability grows with learning new verbs and actions
  - Active
    - agent asks questions to assist comprehension

Challenges: Incomplete Information

- Under-specified commands
  - *store the blue cylinder* `IN(O1,pantry)`.

Challenges: Incomplete Information

- Under-specified commands
  - *store the blue cylinder* `IN(O1,pantry)`.
- Verb execution ambiguity: action-space, policy, goal may change with argument types
  - *cook rice* → `{stove; move(ON,<o>,stove), turn-on(stove), turn-off(stove)}`
  - *cook chicken* → `{oven; move(IN,<o>,oven), turn-on(oven), turn-off(oven)}`

Challenges: Incomplete Information

- Under-specified commands
  - *store the blue cylinder* `IN(O1,pantry)`.
- Verb execution ambiguity: action-space, policy, goal may change with argument types
  - *cook rice* → `{stove; move(ON,<o>,stove), turn-on(stove), turn-off(stove)}`
  - *cook chicken* → `{oven; move(IN,<o>,oven), turn-on(oven), turn-off(oven)}`
- Argument agreement: verb, goal, spatial-relationships
  - *move the blue cylinder to the pantry*

    ```
    ACT: operator OP1, argument1 O42, argument2 L42, goal.predicate
    (IN42(O42,L42) CLOSED42(L42)).
    ```

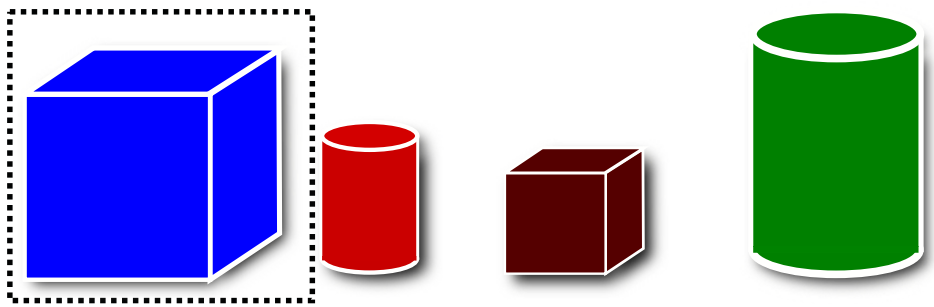
Challenges: Linguistic Ambiguity

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Challenges: Linguistic Ambiguity

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Pick up the blue cube.



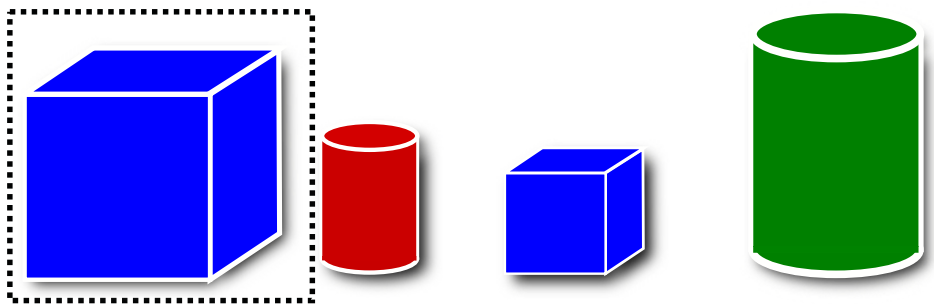
- Preposition-phrase attachment

  place the red large object to the right of the green cylinder in the pantry
  place (the red large object) (to the right of the green cylinder in the pantry)
  or
  place (the red large object to the right of the green cylinder) (in the pantry)

Challenges: Linguistic Ambiguity

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Pick up the large, blue cube.



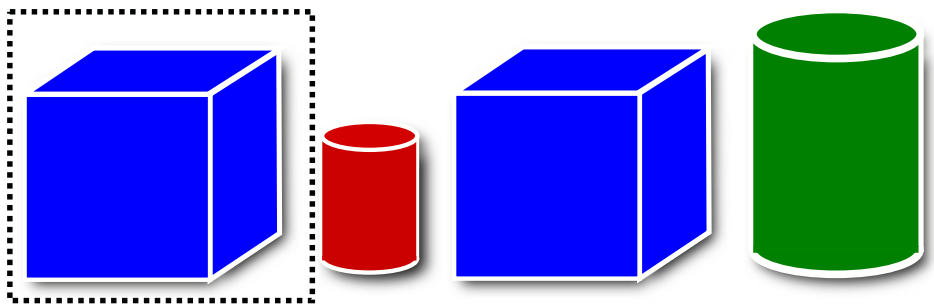
- Preposition-phrase attachment

  place the red large object to the right of the green cylinder in the pantry
  place (the red large object) (to the right of the green cylinder in the pantry)
  or
  place (the red large object to the right of the green cylinder) (in the pantry)

Challenges: Linguistic Ambiguity

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

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



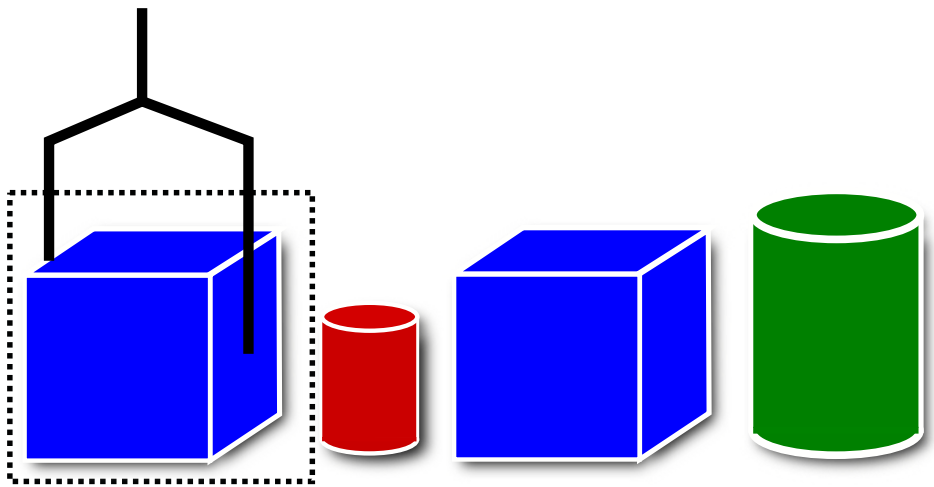
- Preposition-phrase attachment

  place the red large object to the right of the green cylinder in the pantry
  place (the red large object) (to the right of the green cylinder in the pantry)
  or
  place (the red large object to the right of the green cylinder) (in the pantry)

Challenges: Linguistic Ambiguity

- Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Put it down.



- Preposition-phrase attachment

  place the red large object to the right of the green cylinder in the pantry
  place (the red large object) (to the right of the green cylinder in the pantry)
  or
  place (the red large object to the right of the green cylinder) (in the pantry)

Challenges: Perceptual

- Unobservable referents
  move the cylinder to the kitchen where *kitchen* is not perceivable
  to go to the market, take a left on Plymouth road
- Past events
  the location you last visited, is kitchen
  the place next to the dining hall, is kitchen
  you picked up a block, and then put it in the pantry

Approach

- Immediacy of Situation (?)
  - how does the instructor utterance related to experience?
  - current, hypothetical, past
- Sources of referents
  - perceptions: objects (immediate indexing)
  - semantic memory: spatial relationships, action definitions, domain semantic knowledge (immediate, displaced indexing)
  - episodic memory: past events (displaced indexing)
- Diverse contexts
  - linguistic and extra-linguistic
  - knowledge for identifying and constraining the set of referents
  - two dimensions of context (?)
  - informational
    - identifying the set of plausible referents
    - syntactical, perceptual, spatial, verb-action mapping, action
  - temporal
    - recency based ambiguity resolution
    - dialog (discourse model ?), attentional

Evaluation Approach

- Identify a linguistic ambiguity/perceptual challenge
- Generate a comprehensive corpora
- Generate a set of varying scenarios
- Different designs: initial state of knowledge, contexts exploited
- Empirical metrics: correct behavior, number of interactions, precision/recall, time (decision cycles)
- Qualitative analysis: requirements