Grounding Natural Language References to Unvisited and Hypothetical Locations

T. Williams, R. Cantrell, G. Briggs, P. Schermerhorn and M. Scheutz
Human-Robot Interaction Laboratory, Tufts University, 200 Boston Avenue, Medford, MA

Abstract

While much research exists on resolving spatial natural language references to known locations, little work deals with handling references to unknown locations. In this paper we introduce and evaluate algorithms integrated into a cognitive architecture which allow an agent to learn about its environment while resolving references to both known and unknown locations. We also describe how multiple components in the architecture jointly facilitate these capabilities.

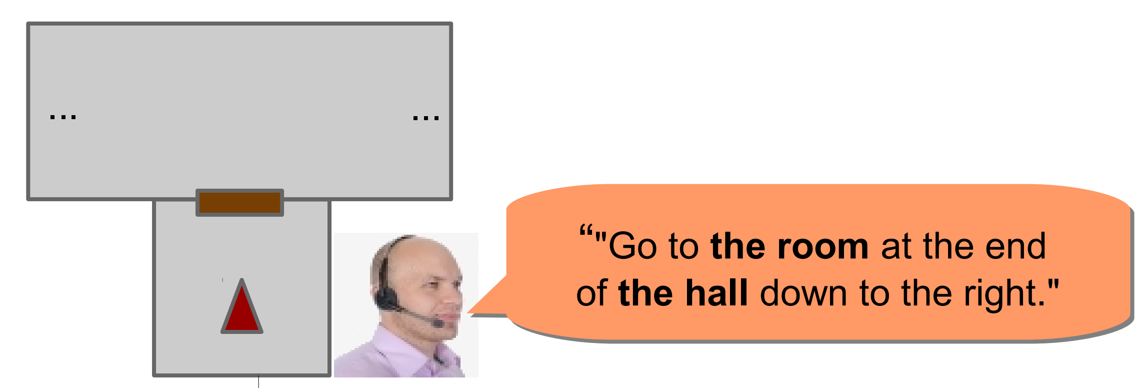
Motivation

Most approaches to spatial reference resolution have focused on resolving references to known locations. Only Matuszek et al. (2012) also resolve references to unknown locations, but they do so in such a way that a robot cannot store any information acquired while traveling to the unknown location.

All previous approaches to spatial reference resolution have used a static environmental map which cannot be changed once reference resolution begins, and have only dealt with natural language *commands,* and not interrogative or declarative utterances.

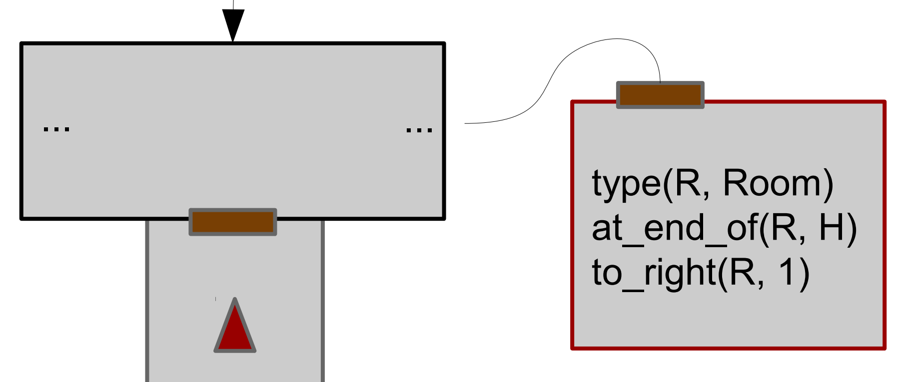
We present algorithms for spatial reference resolution integrated into a cognitive robotic architecture that significantly improve previous proposals by:
(1) systematically adding unknown places to the map, which allows robots to meaningfully communicate about unknown places without having to first discover their exact location,
(2) updating the map as the agent discovers unknown environments, which allows robots to have natural language interactions about new environmental features discovered while navigating to an unknown place, and
(3) generating action sequences only when they are actually needed to visit the referenced location (information is stored in a location-independent form, which affords the robot the capability of learning a map entirely through dialogue).

These algorithms are used by the SPEX (the SPatial EXpert), a component of the ADE implementation of the DIARC architecture which is responsible for map-making and spatial reference resolution.



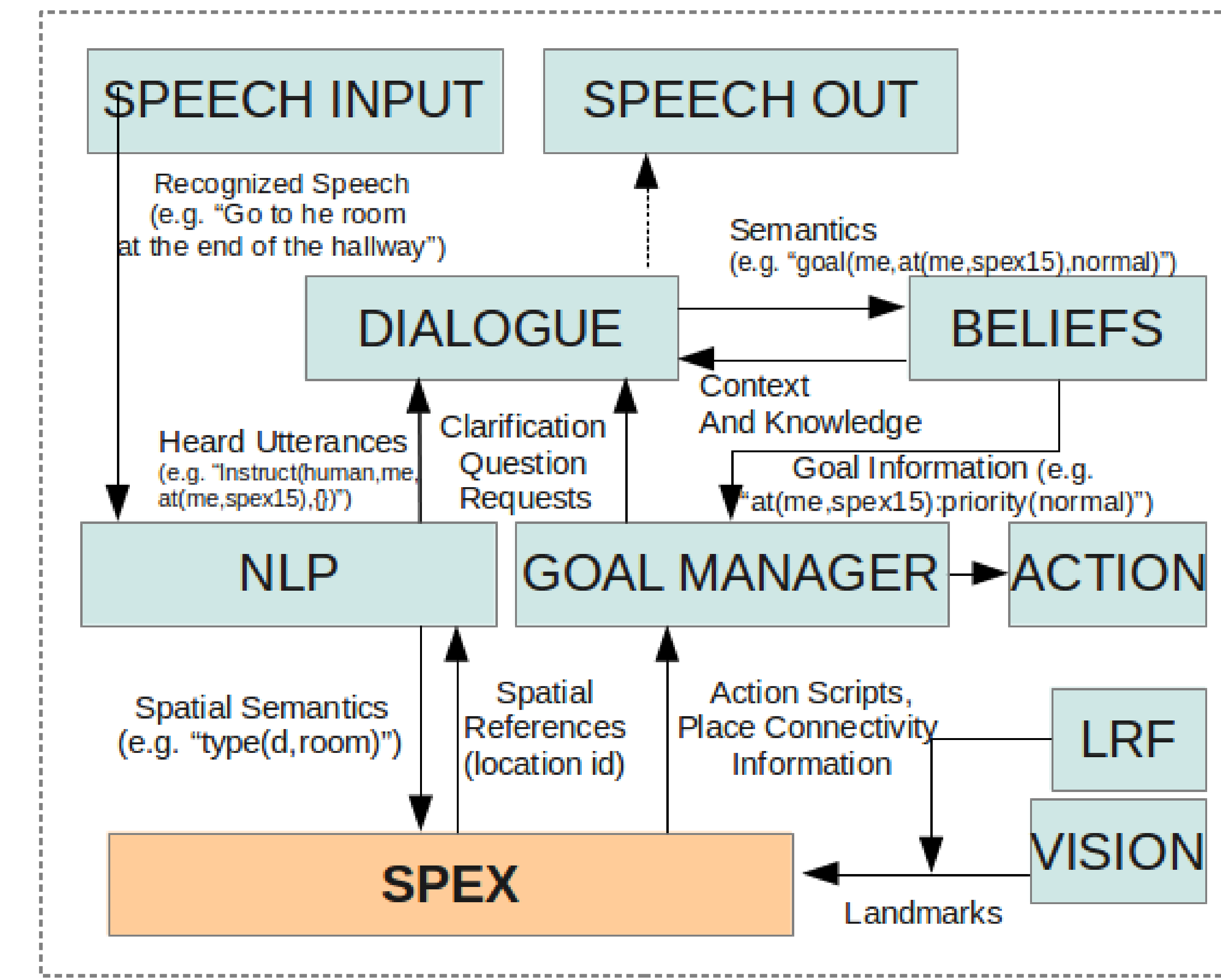
SPEX finds a suitable identifier for "the hall" but not for "the room" as it knows of no rooms at the end of the nearby hallway. SPEX thus creates a representation for a new room and notes that it is connected at a topological evel to the nearby hallway.

(P is a list of Predicates, I is a Map(String $\mapsto$ Identifier)List)
for all $p \in P$ **do**
  if p is a type predicate $\wedge I(p \rightarrow arg_0) == \emptyset$ **then**
    generate new places as necessary for type $type(p)$
    $n \leftarrow$ identifier of the new place referred to by p
    $I(p \rightarrow arg_0) \leftarrow I(p \rightarrow arg_0) \cup \{n\}$
  else if p is a descriptive predicate **then**
    $properties(I(p \rightarrow arg_0)[0]) \leftarrow properties(I(p \rightarrow arg_0)[0]) \cup \{p\}$
  else if p is a relational predicate **then**
    make p true in M
  end if
end for



At a metric level, SPEX can only plan as far ahead as he other side of the doorway, but can use the fact that the target location is at the *end* of the hall and is *to the right* of the current location to create an action script which should bring the robot to the target location.

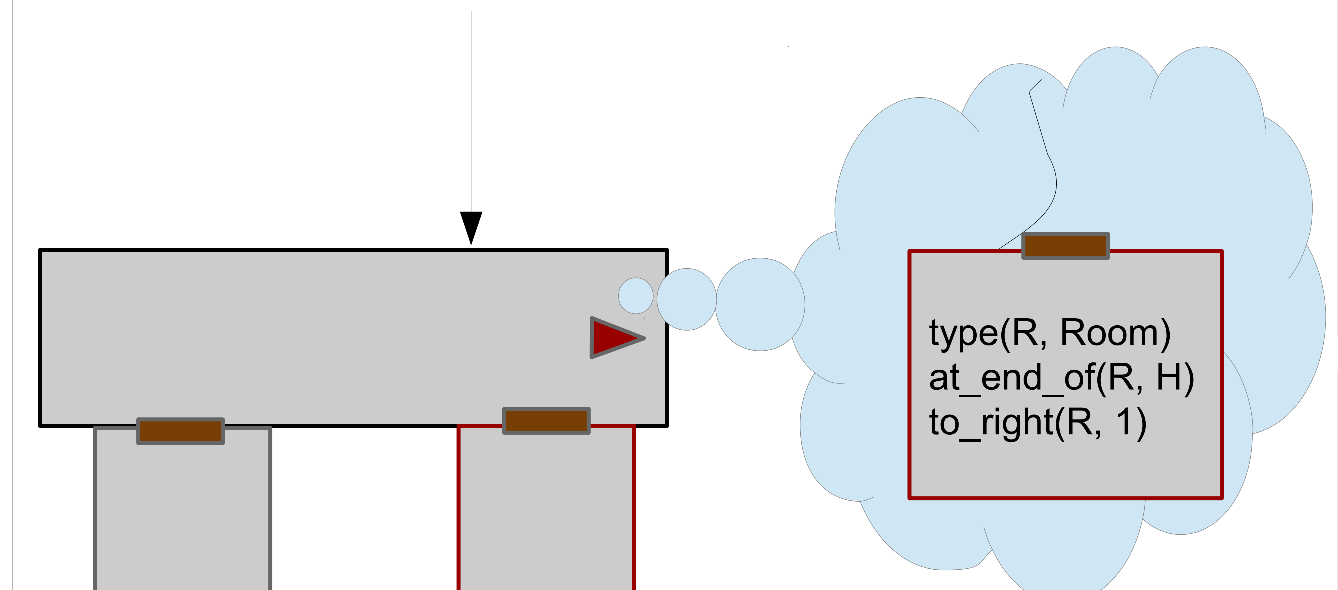
(S is the id of the source, D is the id of the destination)
A is a new action script
if S, D are small scale places with the same parent **then**
  if $locationKnown(D)$ **then**
    add instruction to A to move to $position(D)$
  else
    $useCluesToPlanMotion(S, D, A)$
  end if
else if S, D are small scale places with connected parents **then**
  if $locationKnown(D)$ **then**
    add instructions to A to approach and move through the door which leads to D
  else
    $useCluesToPlanMotion(S, D, A)$
  end if
else if $\neg locationKnown(D)$ **then**
  I_0, I_1 are new lists of identifiers
  $I_1 \leftarrow reverse($route from D to the closest point to it whose location *is* known$)$
  $D \leftarrow first(I_1)$
  $I_0 \leftarrow planRoute(S, D)$
  for all $i \in concatenate(I_0, I_1)$ **do**
    add instruction to A to move to i
  end for
end if
return A




```
[moveTo, self, exitposition]
[exitRoom, self]
[moveTo, self, entryposition]
[moveTo self, currentroom]
[turnRel, self, ang]
[traverse, self]
[informSpexEnd, AtEndOfHall]
[moveTo, self, destination]
```

Upon reaching the end of the hallway, SPEX reconciles its representations of the room it observed at the hallway's end and the room it was originally told about.

This allows SPEX to successfully reach the target location.



Discussion

Our approach has a number of advantages. Storing the information gleaned from natural language and through exploration in a location-independent format allows the robot to (1) travel to previously described locations, (2) describe how two unknown locations are positioned relative to each other, (3) pause an action sequence and then later resume it from another location, and (4) return to a known location after visiting an unknown one.
Finally, augmenting the robot's world model based only on descriptions allows a robot to learn a map purely through dialogue if it is able to extract sufficiently accurate semantic representations; previous approaches would not be able to learn a map of their environment
from dialogue alone.

Future Work

Ongoing and future work includes reference resolution when there is uncertainty as to a location's properties, referential ambiguity resolution, determining when to ask for help rather than explore, and modeling other agents' spatial beliefs.

References

Cantrell, R.; Scheutz, M.; Schermerhorn, P.; and Wu, X. 2010. Robust spoken instruction understanding for HRI. In Proceedings of the 2010 Human-Robot Interaction Conference.
Chen, D. L., and Mooney, R. J. 2011. Learning to interpret natural language navigation instructions from observations. In Proceedings of the 25th AAAI Conference on Artificial Intelligence.
Hemachandra, S.; Kollar, T.; Roy, N.; and Teller, S. 2011. Following and interpreting narrated guided tours. In Proceedings of the IEEE International Conference on Robotics and Automation.
Kuipers, B. 2000. The spatial semantic hierarchy. Artificial Intelligence 119:191–233.
Matuszek, C.; Herbst, E.; Zettlemoyer, L.; and Fox, D. 2012. Learning to parse natural language commands to a robot control system. In Proc. of the 13th Intl Symposium on Experimental Robotics (ISER).
Scheutz, M.; Schermerhorn, P.; Kramer, J.; and Anderson, D. 2007. First steps toward natural human-like HRI. Autonomous Robots 22(4):411–423.

Acknowledgments

This work was in part funded by NSF grant 111323 and in part funded by ONR grant #N00014-11-1-0493