Spatial Reference Resolution for an Embodied Dialogue Agent

Timothy Weale
The Ohio State University
395 Dreese Labs, 2015 Neil Ave.
Columbus, OH 43210
weale@cse.ohio-state.edu

In dialogue systems, reference resolution is an essential part of the semantic interpretation of an utterance. The resolution task is a variable-assignment problem – given the user's utterance, we want to connect unknown discourse entities to our system's internal knowledge base. To illustrate this problem, consider the following example sentences:

(1)
A: Mary likes John.
B: I heard that **he** gave **her** a rose.

In order to understand the meaning of these sentences, we must solve two resolution problems. The first sentence introduces two entities to our domain, *Mary* and *John* and can be semantically represented as **likes(Mary, John)**. In the second sentence, we encounter the two pronouns *he* and *her*. The task is to then resolve the variables *he* to *John* and *her* to *Mary*, providing the appropriate semantics of the sentence. [**gave(John, rose, Mary)**]

Dialogue systems like COMMUNICATOR, TRIPS and TRAINS all include some manner of reference resolution (Byron & Allen 1998). These traditional systems only consider elements of the discourse history and planning state as referents for an ambiguous expression. Extra-linguistic information, such as gesturing or spatial orientation, is ignored. Our work looks to extend this previous work to include spatial context in reference resolution. This is a natural extension of previous work, and holds potential in embedded applications (cars, houses) and robotics.

Spatial Reference Resolution

Context of an ambiguous expression is not limited to just language. We often refer to an item by its spatial location. A host might tell his guest that "the spoons are in the right drawer." In resolving this reference, the guest must figure out how the description "right" relates to the spatial context and aids in picking out the intended referent.

Using spatial context for reference resolution requires both knowledge of how spatial relationships are organized and also how they are encoded linguistically. Linguistic investigation into spatial reference (Levinson 2003) has yielded three related strategies of encoding spatial reference. To illustrate, consider the following sentences, each referring to a particular door in the world:

(2)
A: Which door should I use?
B_1: Use **the north door**.
B_2: Use **the back door**.
B_3: Use **the left door**.

Traditional resolution algorithms will not be able to resolve these references because the information required to interpret the reference does not come solely from the linguistic context.

In each of the response sentences in Example 2, the speaker has utilized a different, distinct strategy for describing the door. Sentence B_1 utilizes the **absolute** reference strategy. In this scheme, reference is based on an imposed external arrangement (*North/South, Above/Below*). In B_2, **intrinsic** reference is used, whereby the properties of a grounding item (front/back of the building) are used for reference. Finally, **relative** reference is utilized, whereby a particular point-of-view anchors the spatial context.

Each strategy places different requirements on reference resolution, but the basic strategy for interpretation remains fairly consistent. First, one must use the spatial description to generate a general search space strategy (right/left, north/south). From that, find all appropriate objects that can be used as a center point for the search (a "grounding mechanism") in order to define the search space. Finally, search for an item matching other predicates in the description (e.g. "door", "drawer") beginning at the grounding item and continuing in the direction indicated by the the spatial phrase.

Computational research in this direction has been limited. There have been investigations into combining spatial information, language and robotics (Moratz & Tenbrink 2006). However, these systems only understand simple commands like "go to the left box." Spatial reference in this work consists only of robot-centered, relative references and ignore other reference strategies. Additionally, it is not a dialogue agent, as it lacks discourse history and feedback beyond the option to move (success) or not (failure).

Proposed Plan of Research

We are developing a software agent capable of interpreting reference based on a description of the object's spatial properties. As part of this, we have collected a dialogue corpus of spatial references, allowing us to get insight into the language theory and ensure that our work is based on attested examples.

Corpus Development

Seed Corpus Last summer, we collected a corpus to serve as a seed for this thesis work (Weale & Byron 2007). The corpus consists of two participants arranging a group of boxes into a target configuration. The task elicits the underlying frame of reference strategies by forcing participants to use the spatial properties of the target boxes for reference, as the boxes were identical in all other respects.

This corpus work utilized Quake II, which has been used before in dialogue research (Byron & Fosler-Lussier 2006). This engine allows us to fully capture the spatial context of our dialogue. We can make re-playable recordings for sharing, easily create experimental maps and modify the source code (Weale, Mellen, & Byron 2005), facilitating utterance and action synchronization.

The collected descriptions follow the three reference strategies, including phrases like "the north box" or "the front left box". Thus, this corpus provides a good variety of spatial descriptions under different contexts.

Supplemental Corpus We will extend the corpus to ensure broader coverage and better system generalization. Most importantly, the task environment will now incorporate multiple grounding objects. Additionally, the manipulation of the grounding object will provide support for spatial grouping references or intrinsic reference strategies.

The extension will be implemented this summer. The bulk of the environment creation will be completed by the end of June, and the dialogue collection will be completed by the end of August.

Agent Development

This thesis will focus on learning and implementing an algorithm for resolving references based on a spatial description. For example, given a room containing two tables with three boxes placed on each, the system must be able to resolve references like "the north box" from context.

The spatial reference will require three steps:

Semantic Analysis: From the input utterance, determine the most likely reference strategy and what relationship the description is attempting to describe. For example, "the north box" would become **north(?G, [?T: type=box])**, a search in the *north* direction from the center of an undetermined *ground* for an entity with type *box*.

Grounding Object Selection: Given the spatial information in conjunction with the initial semantic interpretation, generate referent searches based on available grounding objects from both discourse history and spatial context. If the room contains two tables, this may be **north([G=TABLE1(.9),=TABLE2(.1)], [?T:type=box])**.

Reference Object Selection: Given the grounding object and relationship, resolve the referent to an object within the search space. In our example, TABLE1 has BOX2 in its *north* area. Therefore, the most likely relationship found was **north(TABLE1,BOX2)** and the entire module will return **BOX2** as its output.

The corpus data will be used for feature engineering in order to define the context, based on visual saliency of objects, spatial proximity, dialogue recency and description information. We will then train a reference resolution strategy to predict the referent from the context (dialogue and spatial) by using machine learning techniques on these features. This will enable more expandability and flexibility than by simply hand-coding rules.

The output of our module will be the resolution of the description to its denotation in the knowledge-base. The mapping will be saved as part of the discourse context for use in future contexts. In the future, probabilistic weights may be added to the module outputs.

We have used the Quake II engine as our virtual environment. In it, we can simulate embodied technology without the overhead of a robotics lab, allowing for faster and more refined research into spatial language strategies. However, we can broaden the general applicability of our research by simulating noise problems like visual uncertainty, incomplete knowledge and path planning.

A formal dissertation proposal will be completed by the end of May as part of the author's Ph.D. program. This proposal will detail the timetable of agent development and the interpretation pipeline, which will be based on corpus analysis and current literature in the field.

During the remainder of the summer and the following school year, the agent system will be under active development, based on the seed corpus, feedback from the proposal and supplemental corpus analysis. The anticipated completion date of this project is spring 2008.

References

Byron, D. K., and Allen, J. F. 1998. Resolving Demonstrative Anaphora in the TRAINS93 Corpus. In *New Approaches to Discourse Anaphora: Proceedings of the Second Colloquium on Discourse Anaphora and Anaphor Resolution (DAARC2)*.

Byron, D. K., and Fosler-Lussier, E. 2006. The OSU Quake 2004 corpus of two-party situated problem-solving dialogs. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC'06)*.

Levinson, S. C. 2003. *Space in Language and Cognition - Explorations in Cognitive Diversity*. Cambridge University Press.

Moratz, R., and Tenbrink, T. 2006. Spatial Reference in Linguistic Human-Robot Interaction: Iterative, Empirically Supported Development of a Model of Projective Relations. *Spatial Cognition and Computation* 6(1):63–106.

Weale, T., and Byron, D. K. 2007. The OSU CONCERT 2006 corpus of two-party situated problem-solving dialogs. Technical Report In Preparation, The Ohio State University.

Weale, T.; Mellen, B.; and Byron, D. K. 2005. Modifying the Quake II Engine for 'Civilian' Use. Technical Report OSU-CISRC-11/05-TR74, The Ohio State University.