Steps towards Incremental Semantics for Spoken Dialog Systems

Gregory S. Aist* **Scott Stoness** **James Allen**
Computer Science Department
University of Rochester
{gaist,stoness,james}@cs.rochester.edu

1 Introduction

Traditionally, spoken dialog systems have interpreted a user's speech one complete utterance at a time, and operated on one level of processing at a time as well. It is clear, however, that people understand language incrementally: they can backchannel, interrupt, and begin taking actions while the speaker's utterance unfolds. It has also become clear that human language understanding involves the rapid integration of multiple sources of information including syntax, semantics, and pragmatics. On the system side, as a reviewer pointed out, deployment scenarios such as driving or flight simulation would present cases where incremental understanding would nicely enable features such as beginning to act on partial commands, or changing the goal of an action as more information arrives. We have recently shown that computational methods for incremental understanding have various advantages over their nonincremental counterparts, including better parsing (Stoness et al. 2005).

In this brief note, we describe work on representations of natural language semantics for incremental understanding. First we describe how rampant use of free variables helps represent not only partial sentences, but utterances as they arrive word-by-word. Second, we show how the commonalities in the meaning of phrases such as those used to indicate fine-grained adjustments ("a bit more to the right") can be captured in semantic representations. Third, we show how allowing actions ("moving to central park") to be states can cleanly and concisely represent otherwise messy facts such as interrupted actions. These are three initial steps towards incremental semantics for spoken dialog systems.

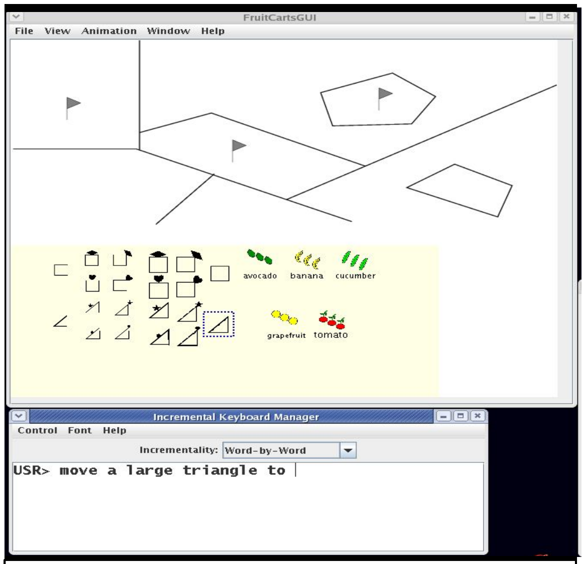
2 Semantics for Partial Sentences and for Incremental Processing

In many cases in spoken dialog systems, utterances do not consist of complete sentences, but rather of fragments of various sorts. We have taken an aggressive approach to allowing free variables in semantic representations, which allows for representation not only of partial sentences but also for representation of incoming utterances while the words are still arriving. Let us consider the example sentence *move a large triangle to central park* (Figure 1, from a testbed domain) and how we can construct its semantics incrementally as words arrive from the speech recognizer. For brevity we have shown several words arriving together; other parts of the system handle the segmentation issue.

MOVE a: *move(X, Y)*
LARGE TRIANGLE to: *move(triangle1, Y)*
CENTRAL PARK: *move(triangle1, centralpark1)*

This simple change to the sort of semantic representation we are willing to take action on allows the dialog system to not only process the incoming utterance incrementally, but to take action incrementally as well. So, if desired, the system could respond to the representation *move(system, triangle1, Y)* by estimating the most likely destination for the triangle and initiating the move – if its estimate turns out to be wrong, a revised move can be started when further information is available.

*To whom correspondence should be addressed.


FruitCartsGUI
File View Animation Window Help
avocado
banana
cucumber
grapefruit
tomato
Incremental Keyboard Manager
Control Font Help
Incrementality: Word-by-Word

```
USR> move a large triangle to
```

Figure 1. At this point in the incoming sentence we have *move(triangle1, Y)* which might be enough for the system to act. The middle flag is in Central Park.

3 DHL Semantics for States

Consider the two dialog excerpts shown below:

A.1. Move a triangle to the flag
A.2. A bit more to the right

B.1. Rotate the triangle clockwise thirty degrees
B.2. A bit more to the right

There is a great deal of semantic similarity in the goals specified in A2 and B2 that is not captured in a representation such as *move(triangle1, east 1cm*) and *rotate(triangle1, clockwise 5 degrees*). In order to capture this sort of generality, we have constructed a semantic framework that uses a triple (distance, heading, location) to represent values of various attributes such as position, angle, and color. This representation yields the same underlying goal structure for A2 and B2, namely *(a bit, right, X)* where X is a free variable so that the location remains underspecified. In a sense, representing the semantics of goals in this way defers the decision about how best to interpret this command in context to other levels of processing where information such as the most recently used verb can be brought to bear. It is also a more efficient representation than explicitly representing all possible ways in which *a bit more to the right* could serve as the goal for various actions.

4 Actions as States in Interval Events

Consider the dialog excerpts shown below:

C.1. Move a triangle to the flag

D.1. Move a triangle to the flag
D.2. Stop

We might consider representing the action done in response to C1 as follows: *move(triangle1, flag1)*. However, that results in a problem when C2 is heard: how do we show that C1 was interrupted? Part of the answer to this concern is to represent events as intervals rather than points; for example, begin(move(triangle1, flag1)) for the start of an action and end(move(triangle1, flag1)) for the end of an action. That allows us to represent the time interval between when the action for D1 starts and when it would have stopped if not for D2.

But where is an object when it's started moving, but hasn't yet reached its goal? Suspend disbelief for a moment, and consider what happens if we allow actions to be used as states. We then get a representation as follows:

C.1. previous location: X
current location: move(flag1)
C.1. previous location: move(flag1)
current location: flag1
and for D we get:
D.1. previous location: X
current location: move(flag1)
D.2. previous location: move(flag1)
current location: (x1, y1)

This representation allows us to nicely represent overlapped and interrupted actions. There is also a very nice naturally occurring analogue. Suppose you call someone on their mobile phone, and have this conversation:

E1. Hello?
E2. Hello. Where are you?
Several reasonable E3s are possible:
E3'. I'm in my car.
E3''. I'm on Second Avenue.
E3'''. I'm on my way home.

Our representation for actions mirrors exactly the type of language that naturally occurs in sentences such as E3'''.

5 Related Work and Future Directions

Incremental semantics is related to theories such as Discourse Representation Theory (Kamp 1981) in that early parts of sentences are available (and thus at least partially interpreted) for use in later parts of sentences. A number of other researchers have worked in the area of incremental interpretation of natural language; space precludes a full review here but see for example Haddock (1988), Milward and Cooper (1994) and Fischer, Geistert, and Gorz (1995). Interesting work by Vermeulen (1994) proposes three constraints for the semantics of texts which we find informative for the semantics of dialogs as well:

INCREMENTALITY: "we can interpret texts *as we hear them*" (emphasis in original), and thus we should insist that in semantic interpretation "everything *can* be done incrementally" even if some things are in fact delayed until more information is available (244). We agree.

BREAK IN PRINCIPLE: "Any segment of a text can be interpreted. (In general its meaning will be partial.)" Here too, we agree, and we would say that not only can a segment of an utterance in dialog be interpreted, but that many segments yield interpretations with enough material to possibly warrant actions on the part of the hearer.

PURE COMPOSITIONALITY: "the meaning of a text depends on nothing but the meaning of its parts" (245). Here we part ways: in dialog (as perhaps opposed to text) there is more going on than simply the speech – in particular, there are the actions in the physical (or virtual) world that affect the meaning of what is said. (Not just in a pragmantic sense, either; a word with multiple senses might be thoroughly disambiguated by the presence or absence of certain objects in the visual world.) Here then is one interesting direction for future research: how information from the visual world can be brought to bear on the problem and process of incremental semantic understanding in spoken dialog systems.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0328810. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

I. Fischer, B. Geistert, and G. Gorz. 1995. Chart-based incremental semantics construction with anaphora resolution using DRT. Proceedings of the Fourth International Workshop on Parsing Technologies.

N. Haddock. 1988. Incremental semantics and interactive syntactic processing. Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

H. Kamp. 1981. A theory of truth and semantic representation. pp. 277-322 in *Formal methods in the study of language*, edited by J. Groenendijk et al., Mathematisch Centrum, Amsterdam, 1981.

D. Milward and R. Cooper. 1994. Incremental interpretation: Applications, theory, and relationship to dynamic semantics. COLING 1994.

S.C. Stoness, J. Allen, G. Aist, and M. Swift. 2005. Using real-world reference to improve spoken language understanding. AAAI Workshop on Spoken Language Understanding, Pittsburgh, Pennsylvania, July. pp. 38-45.

C. F. M. Vermeulen. 1994. Incremental semantics for propositional texts. *Notre Dame Journal of Formal Logic* **35**(2):243-271.