A Dialog Architecture for Military Story Capture

**Ronnie W. Smith, Brian Manning, Jon Rogers,
Brian Adams, Mujibur Abdul**
Department of Computer Science
East Carolina University
Greenville, NC, USA 27858
email: rws@cs.ecu.edu

Amaury Alvarez
General Dynamics - Advanced Information Systems
P.O. Box 26002
Greensboro, NC, USA 27420-6002
email: Amaury.Alvarez@gd-ais.com

Abstract

This paper describes a prototype multimodal spoken natural language dialog system for capturing a commander's expectations about planned military actions. Dialog context is used for handling issues concerning military echelons, multiple databases, dynamic name creation, and relative time references.

1 Introduction

With the proliferation of sensor technology for military applications, commanders now have access to more information but also have greater difficulty in gaining access to the proper information to provide them with appropriate situational awareness about what is actually happening. The DARPA-sponsored Command Post of the Future (CPOF) project is an attempt to develop technology to enable commanders to more quickly obtain accurate awareness of the actual military situation so that commanders can be more proactive rather than reactive. Our research team has focused on the notion of a story (Gershon and Page, 2001) as a representation for the expectations of a military action. This paper focuses on our work on a prototype multimodal spoken natural language dialog system that allows military commanders to specify the expectations for a story. The machine representation for these expectations will be a *story*

```
 1. C: What's next?
 2. U: Move delta to the four pack in
       two hours.
 3. C: What's next?
 4. U: Move alpha to checkpoint alpha
       one starting one hour after that.
       (CLICK to designate on map)
 5. C: What is the destination time?
 6. U: Start plus one hour.
 7. C: What's next?
 8. U: Initiate when alpha at checkpoint.
       <DRAG GESTURE moves charlie
       company towards peepsite building)
 9. C: What is destination time?
10. U: Start plus three zero minutes.
```

Figure 1: Sample Dialog Scenario

graph. The story graph can then be used as input to other software that filters battle reports to ensure that commanders maintain appropriate situational awareness about the important events of an ongoing military action.

2 System Requirements

The discussion of system requirements takes place in the context of the brief interaction with our current prototype system given in figure 1.[1] This scenario is based on an urban military action from the Vietnam War.

---

[1] C denotes the computer and U the user.

2.1 Contextual Reference of Military Echelon

The name *delta* in utterance 2 refers to delta company of the 1st marines division, 5th battalion. Military forces are organized into hierarchies. A company is composed of platoons which is composed of squads which is composed of soldiers. Thus, at any given time, context information about the various values at each level of the hierarchy must be assumed or ascertained as needed. Therefore, if the next user utterance had been, "First platoon will lead.", the established context should imply that this is first platoon of delta company. If ambiguity arises, clarification is needed.

2.2 Knowledge of Data Sources

Besides the user, the dialog controller must be able to take advantage of system knowledge bases as needed. In utterance 2, a system knowledge base must be consulted to determine both the type of entity and the location of the entity four pack (it is a cluster of buildings) in order to properly represent the commander's expectation in the story. Our organization of knowledge bases is described in section 4.

2.3 Multimodal Integration

The utility of gestural input in conjunction with speech, particularly for spatial references, has been clearly demonstrated (Oviatt, 1996). Utterances 4 and 8 illustrate contexts where multimodal integration is needed. Dynamic association of names with locations (such as checkpoint alpha one in utterance 4) and specifications of force movements are just two of the many possibilities for multimodal input. Another under development includes specification of geographic regions for various purposes (such as direction of enemy movement, designation of flanks, etc). An analysis of a videotape of an exercise performed by Brigadier General Keith Holcomb, U.S. Marine Corps, Ret., as part of a DARPA CPOF exercise, revealed a total of at least 7 different semantic classes for gestures. The capability for multimodal input for any map based task is essential.

2.4 Relative Time References

While the schedule for military actions often has precise timing, the timing is frequently relative rather than absolute. Furthermore, the uncertainties of military actions (often referred to as the "fog of war") virtually ensures that no precise scheduling will ever be achieved during the actual action. Consequently, any system that attempts to assist commanders with their situational awareness must be able to reason about the relative time of events.

In utterance 4 of figure 1, the time in context comes from utterance 2, the time at which delta company gains position in the four pack. In utterance 8 the time reference, when alpha at checkpoint refers to the time when alpha company gains position at checkpoint alpha one.

Any system dealing with coordination of complex sequences of events must be capable of handling pronominal as well as more complex linguistic references to other times and events.

3 Activity-Based Dialog Model

The interaction is based on the notion of an *activity*. Various types of activities in which a military force can be engaged include movement, position establishment, and reconnaissance. Associated with an activity type are parameters, some of which are mandatory and some are optional. The dialog system will attempt to continue interaction about the current activity until values for all mandatory parameters are supplied. This approach is an instantiation of the Missing Axiom Theory of dialog that we have used in the past ((Smith et al., 1995) and (Biermann et al., 1997)).

In the prototype system we have focused on movement activities. There are a total of five required parameters: force, start time, start location, end time, and end location. We will illustrate how the model functions from the scenario in figure 1.

Utterance 2: Move delta to the four pack in two hours.

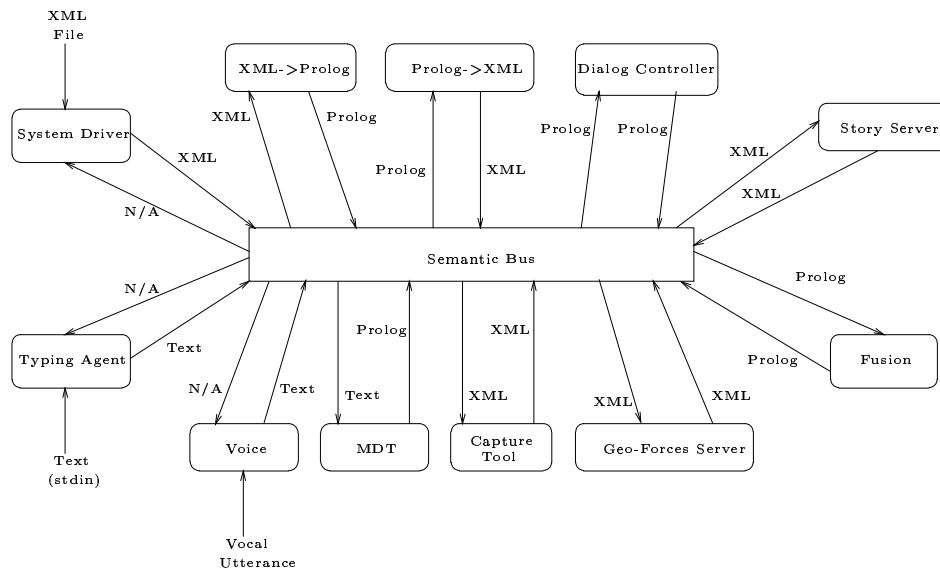Initial context specifies start time and start location for delta company. Since initial context

Figure 2: System Architecture

is discussing only one battalion, there is only one company named delta; thus, it is a unique reference. The end location is `four pack` and the end time is two hours after the start. Since all mandatory parameters are specified, the system simply issues a generic query. Future work will include dealing with issues concerning confirmation and verification.

Utterance 4: `Move alpha to checkpoint alpha one (MOUSE CLICK) starting one hour after that.`

Context provides initial location of alpha company. Start time is specified linguistically while the end location is specified by mouse click, but end time is not, which leads to utterance 5, `What is the destination time?` In the response `Start plus one hour`, current time refers to when alpha begins its movement. After processing each utterance, the dialog controller must update its context model to correctly reflect current time.

Utterance 8: `Initiate when alpha at checkpoint (DRAG ACTION).`

Force, start location, and end location are all specified via mouse gesture. Start time is specified linguistically. Again, end time is not specified and must be requested.

In general, we believe that using the idea of activity for dialog context together with pa-

rameter specification as the engine for choosing what to say will provide a robust interaction model as we expand its capabilities.

4 Architecture

A block diagram of the system is shown in figure 2. The arrow labels specify the type of data transmitted. Prolog refers to attribute-value (AV) lists in Prolog syntax. A brief description of each module is included below.

- *Semantic Bus* - handles interprocess communication between agents.

- *System Driver* - provides a graphical user interface (GUI) from which other system components can be initiated.

- *Dialog Controller* - manages the interaction with the user through contextual interpretation of the user's multimodal inputs.

- *Capture Tool* - provides a GUI for viewing the map and story graph and specifying gestural inputs.

- *Fusion* - handles multimodal integration of input via a unification-based approach (e.g. QuickSet (Johnston et al., 1997))

- *Voice* - handles speech recognition using ViaVoice from IBM.

- *Typing Agent* - allows for typed inputs when speech recognition is not functioning.

- *MDT* - handles parsing via minimum distance translation.

- *Geo-Forces Server* - provides a database for all geographic entities and military forces.

- *Story Server* - provides a database for the story graph.

- *XML -> Prolog* and *Prolog->XML* - converts messages between XML format and Prolog attribute-value (AV) list format.

5   Related Work

The two most relevant systems previously developed are QuickSet ((Johnston et al., 1997)) and CommandTalk ((Stent et al., 1999)). QuickSet is a multimodal interface for designing military simulations. The emphasis of the work is on proper integration of the various input modalities. QuickSet uses a unification based approach over typed feature structures for determining the appropriate interpretation. Due to the command-driven nature of the application, a great deal of functionality can be achieved without a complex model of the ongoing dialog.

CommandTalk is primarily a spoken natural language dialog system for the same application. Its gestural capabilities are limited to specification of points on the map analogous to utterance 4 of our scenario in figure 1. Notable features of CommandTalk include its bidirectional grammar and its dialog manager which consists of a collection of finite state machines (FSM). The main purpose of these FSM's is to coordinate initiative and handle parameter specification. Neither system is designed to be able to maintain a representation of an actual planned military action for purposes of helping commanders maintain appropriate situational awareness via use of intelligent information filtering and reporting.

6   Status

A prototype system has been implemented for location activities. There is currently limited linguistic capability but it can process inputs such as those specified in the sample scenario. Work continues on expanding its capabilities, including specification of geographic regions, battle resources such as helicopters and ships, and development of a more robust dialog model.

7   Acknowledgements

References

A.W. Biermann, C.I. Guinn, M.S. Fulkerson, G. Keim, Z. Liang, D. Melamed, and K. Rajagopalan. 1997. Goal-oriented multimedia dialogue with variable initiative. In Z.W. Ras and A. Skowron, editors, *Foundations of Intelligent Systems*, pages 1–16. Springer-Verlag, Berlin.

Nahum Gershon and Ward Page. 2001. What storytelling can do for information visualization. *Communications of the ACM*, pages 31–37, August.

M. Johnston, P.R. Cohen, D. McGee, S.L. Oviatt, J.A. Pittman, and I. Smith. 1997. Unification-based multimodal integration. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 281–288.

S.L. Oviatt. 1996. Multimodal interfaces for dynamic interactive maps. In *Proceedings of Conference on Human Factors in Computing Systems: CHI' 96*, pages 95–102, New York. ACM Press.

R.W. Smith, D.R. Hipp, and A.W. Biermann. 1995. An architecture for voice dialog systems based on Prolog-style theorem-proving. *Computational Linguistics*, 21:281–320.

A. Stent, Dowding J., Gawron J. M., Bratt E.O., and R. Moore. 1999. The commandtalk spoken dialogue system. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 183–190.