Aspects of Context for Understanding Multi-modal Communication

Elise H. Turner, Roy M. Turner, John Phelps, Mark Neal, Charles Grunden and Jason Mailman

Department of Computer Science, University of Maine, Orono, ME 04469

Abstract. Context is important for AI applications that interact with users. This is true both for natural language interfaces as well as for multi-modal interfaces. In this paper, we consider aspects of context that are important in a multi-modal interface combining natural language and graphical input to describe locations. We have identified several aspects of contexts in our preliminary study. We describe them here and discuss plans for future work.

1 Introduction

In this paper, we will discuss contextual aspects which we have identified for understanding user input in speech and graphics. We consider context to be anything that is required to understand an utterance beyond syntax and semantics. We call the components of context *contextual aspects* or simply *aspects*. A distinct aspect was recognized for one of several reasons. First, all of the knowledge contained in the aspect might be connected by a theme. Second, different aspects would be relevant to, or function differently when, handling different phenomena. Third, some parts of the context were divided into separate aspects because they are managed differently. Similarly, some aspects were separated from others due to the duration of the information in the aspect (i.e., is the information relevant during a single session or across many sessions?). Fourth, some aspects have been separated out, for now, because they are well-studied elsewhere. This is the case for the discourse aspect.

Our work is to be applied to Sketch-and-Talk, a multi-modal interface to geographical information systems that is being created by Max Egenhofer and his colleagues in the Department of Spatial Information Science and Engineering at the University of Maine. The system will construct database queries from spoken natural language and graphical input from the user. Because the implementation of the initial system has not yet been completed, we have begun our work by studying ten videotaped examples of members of our research group describing locations or spatial information. This preliminary work has led us to identify several contextual aspects that affect the interpretation of multi-modal interaction. These aspects are presented below. More detail can be found elsewhere [1,2].

P. Bouquet et al. (Eds.): CONTEXT'99, LNAI 1688, pp. 523–526, 1999.

2 Contextual Aspects for Multi-Modal Interactions

Discourse Aspect. The discourse aspect is known in natural language processing as the *discourse context*. It contains all of the entities that are mentioned in the discourse. This context is broken into several subparts, or *discourse segments*. Discourse segments are made up of contiguous utterances that are related to the same topic. Many techniques already exist for creating the discourse context and moving between its segments (e.g., [3,4,5]), and any of these could be adopted for our system.

Graphics Aspect. The graphics aspect includes all of the entities that have been drawn and their spatial relations. For our work with Sketch-and-Talk, we will use the entity and relation representations used by that project [6].

We have found that, like discourse, the graphics context should be divided into *graphics spaces*. We have seen indications that users consider the graphics context to be subdivided. Users speak of "the area around *some entity*". They also deviate from their established order of drawing to draw certain related objects. For example, a user who has been drawing entities from left to right may deviate from this pattern to draw all of the outbuildings surrounding a house. Users also draw detailed views of particular regions of the location and move between the overview and detailed views during a session. Entities in a graphics space are often all related to a single entity or function. For example "where we fished" may constitute a graphics space. Also, users can easily refer to a graphics space with a single reference, for example, by pointing or referring to the most significant entity.

Clearly, the graphics spaces and discourse segments will be closely related because users are expected to talk as they draw. For now, we keep the discourse and graphics aspects separate to take advantage of the work that has been done to develop representations and management algorithms for the discourse aspect. In future work, we plan to explore the relationship between these two aspects. This includes determining if they are truly separate. Future work will also include discovering exactly what constitutes a graphics space and how a speaker/drawer moves between them.

Task Aspect. This aspect provides information related to the task that the user is pursuing. For our application, the representation of this aspect will include likely goals of the user as well as procedures for achieving those goals. In addition, we saw evidence of a *social interaction task aspect*, in which users put aside the task of describing a location to interact with or entertain the observers, and a *drawing correction task aspect*. The task context influences the flow of the communication [5], as well as helping to identify important entities and concepts. The information represented for this aspect will vary, depending on the task.

Location Aspect. In Sketch-and-Talk, the kind of location that is the target of the query also constitutes an important aspect of the context. We expect the application to have world knowledge about the location that it can access to build or respond to database queries. The location aspect brings this information into the context. Other types of world knowledge will be needed to understand the speech, and, at times, the graphics. However, we create a separate aspect for