The DIAMOND project

Geertzen, J.; Girard, Y.; Morante, R.; van der Sluis, Ielka; van Dam, H.; Suijkerbuijk, B.; van der Werf, R.; Bunt, H.

Published in:
 Presented at the 8th Workshop on the Semantics and Pragmatics of Dialogue, Catalog'04

Publication date:
2004

The DIAMOND Project

Jeroen Geertzen, Yann Girard, Roser Morante, Ielka van der Sluis, Hans van Dam,
Barbara Suijkerbuijk, Rintse van der Werf and Harry Bunt

Tilburg University, Tilburg, The Netherlands

1 Introduction

In this abstract we describe the project "DIalogue Act MOdelling of Natural Dialogue" (DIAMOND) that is carried out at Tilburg University.

The project aims to develop formal and computational models of natural human-human and human-computer dialogue. Natural dialogue is multimodal, and indeed the DIAMOND project is concerned with multimodal interaction, in particular with combinations of speech, text and graphics. The project has a strong focus on what is felt to be at the heart of a (human or computer) dialogue agent: the design and implementation of a sophisticated *dialogue manager* component that decides what to do next in a dialogue, based on a representation of the agent's current information state and making use of a repertoire of *dialogue acts* in a rational and cooperative way to produce natural dialogue behaviour. The use of general principles for applying dialogue acts should provide the DIAMOND dialogue manager the flexibility to be useful in a wide range of domains. Several PhD projects are concerned with specific aspects of the project.

The tasks of the DIAMOND project are grouped into three categories: tasks related to the construction of a dialogue corpus; tasks related to elaboration of the theory of dialogue management, using these corpus data; and tasks related to the design and implementation of the DIAMOND dialogue manager.

2 Motivations

Many of the existing dialogue managers today are based on the exploitation of domain- and task-specific properties, using simple frame-based techniques or finite-state models, such as a parameterized task model that supports a slot-filling dialogue. This typically leads to a rigid form of dialogue, that may be acceptable for very simple tasks, but that is often strange to users and does not exploit the possibilities of natural multimodal dialogue. Another problem is that such dialogue managers typically work only for the domain for which they have been developed, which again signals the lack of general mechanisms and principles that have a solid theoretical basis.

3 Theoretical foundation

The DIAMOND project takes it theoretical basis mainly in formalized notions of common ground and communicative action, as modeled in Dynamic Interpretation Theory (DIT) [Bunt, 2000]. This theoretical framework has emerged from the study of spoken human-human and human-machine information dialogues, with the aim of uncovering fundamental principles that can be used in designing a theoretically sound model of dialogue management. One of the aims of the project is to further develop the theoretical framework of DIT both through the study of the DIAMOND dialogue corpus and through the implementation of prototype systems. Machine learning techniques, in particular memory-based learning as developed partly at Tilburg University, will be applied to develop algorithms for the automatic recognition of dialogue act types from utterance features.

DIT interprets multimodal dialogue utterances in terms of context-changing operators, called *dialogue acts*. Dialogue acts are functional units used by speakers to change the context, where context is understood as the totality of conditions that may influence the understanding and generation of communicative behavior. A 'context' in DIT includes

1

both an agent's information about the domain of discourse and the task that the participants are trying to perform, as well as various kinds of information about the communicative situation and about the communicative and cognitive process that are involved. More specifically, DIT conceptually distinguishes five types of information in an agent's context: linguistic (in particular: the dialogue history), semantic, social, cognitive, and perceptual. Cognitive and linguistic context are particularly important in DIT for the modelling of *dialogue control*, i.e. the use of dialogue acts to create and maintain the conditions for successful communication.

4 The DIAMOND corpus

The DIAMOND corpus initially consists of annotated transcriptions of WoZ-dialogue with a multimodal interactive user manual for a fax machine, developed in the MATIS project[1]. The corpus will be used for several purposes:

- for the study of dialogue phenomena, to strengthen the empirical basis of the DIT framework.
- to train algorithms to recognise dialogue acts from utterance features.
- to evaluate the prototype dialogue manager and dialogue system.

In the construction of the corpus, the following activities are performed:

- **Transcription**. The audio recordings from the WoZ experiments are transcribed orthographically using the PRAAT speech analysis tool[2]. The transcription is made according to the strict conventions of a transcription manual. To maintain a direct link between the transcription and the audio- and video recordings, the transcriptions will contain time stamps at utterance level.
- **Annotation** The utterances in the transcribed dialogues are annotated with dialogue act type and semantic content type information, again following strict guidelines described in a annotation guide. So far, the focus is

on coreference and task-related information in addition to dialogue act information, i.e. on those information types that are most relevant for local context updating.

The results of the transcription and annotation are stored in an XML-based representation, together with the results of parsing and semantic analysis steps. We are building a DTD for these representations that reflects the relationship between methodological choices and the background theories used in the DIAMOND project, which turns out to be a useful way to secure the consistency of our studies. This is useful both for working on the theoretical framework and for designing the DIAMOND prototype system. The XML-based representation also allows a direct link between our definitions of dialogue acts and the establishment of a repository of dialogue act definitions as currently undertaken by the ACL SIGSEM Working Group on the Representation of Multimodal Semantic Information, instigated by the international standards organization ISO[3].

5 Dialogue manager and system prototyping

The DIAMOND dialogue manager will be embedded in a dialogue system that relies on a distributed message-based publish/subscribe framework, which is currently under development. Since the focus of the project is on dialogue management, other modules (e.g. speech-recognizer, PoS tagger, semantic analyser,..) will make use of software partly developed at Tilburg University in other projects, and of (open source) third party software that has been made available by other universities.

References

H. Bunt. Dialogue pragmatics and context specification. In H. Bunt and W. Black, editors, *Abduction, Belief and Context in Dialogue*, pages 81–150. John Benjamins, Amsterdam, 2000.

---

[1]Multimodal Access to Transaction and Information Systems, see http://lands.let.kun.nl/TSpublic/MATIS/

[2]See http://www.fon.hum.uva.nl/praat/

[3]See http://let.uvt.nl/research/ti/sigsem/wg and http://www.tc37sc4.org