# Prototyping Dialogues for Information Access

**C.J. Rupp**
Computerlinguistik
Universität des Saarlandes
D-66041 Saarbrücken
cj@coli.uni-sb.de

## 1 Introduction

This project note describes work conducted in the larger COLLATE project, concerned with fundamental issues in the development of information systems. The focus of this part of the project is the conjunction of Question Answering and Dialogue Modelling in several different ways.

- Comparing the notion of question answering in information retrieval with the use of questions as an organisational construct in an information state model for dialogue.
- Extending natural interaction with information systems beyond the question answering task.
- Examining the type of dialogues that arise between users and complex information systems.

Since the use of dialogue to access information services is not yet an accepted technology, suitable dialogue data is in short supply. To counter this, we prototype a partial functionality both as a technical infrastructure into which extended components can be slotted and as a research tool for data collection.

## 2 Questions and Information Access

Research in Question Answering has been stimulated by the competitive framework of TREC which has spawned a variety of question answering systems. However, most existing systems exhibit a functionality that lies somewhere between that of a web search engine and a human response to a direct question. This is usually a one shot activity, similar to answering quiz questions for humans, a highly stylised and minimal form of dialogue.

A more natural dialogue interaction in response to questions allows for clarification. An answer is not a fact in isolation but part of a system of justified beliefs to which the answerer has committed himself. Therefore, he must be prepared to offer a justification and it is, similarly, natural for an answer to trigger further related questions.

In addition, an information access dialogue may address aspects of how information is obtained. This is a kind of meta-level subdialogue, concerning the process of fulfilling the task. Thus, the system plays two roles, as an expert on information resources, as well as the agent who holds the information that is the answer to a question. There is no natural parallel to this dual role in human dialogue.

## 3 Implicit Questions

Recent development in dialogue modelling has also been concerned with the relationship between questions and answers as the dynamic driving dialogue interactions. However, such *questions under discussion* (Ginzburg, 1994; Ginzburg, 1996), are more likely to be implicit questions that are raised and answered in the course of a dialogue. Explicit questions enter this representation domain, but in so doing they tend to raise further implicit questions, e.g. as a prototype method for arriving at an answer.

Much of this work has been concentrated within the ISU (Information State Update) approach to dialogue modelling, based on an information state, as a single representation of all the relevant context information for the current point in the discourse, and a system of update rules which map one state to another. The ISU approach provides us with a general, high level tool for prototyping dialogue systems. TRINDIKIT (Larsson et al., 2002) combines a framework for the definition of information states and update rules with the ability to integrate an ISU dialogue model and external resources, such as parsers, speech recognition, and information resources.

## 4 Available Resources

Dialogue is, characteristically, a complete application for language technology. In principle, resources are required at all levels of analysis, in addition to the application domain. In practice, compromises are possible, provided that the linguistic domain can be adequately restricted. Prototyping also allows the development of different components at different speeds. In the context of COLLATE we have a number of resources available:

- A corpus of business reports from a newspaper, hand annotated with relational templates,
- A partial parser, SPROUT (Becker et al., 2002), based on finite state technology,
- A grammar of German, defined for named entity extraction in SPROUT over a business domain.

To obtain realistic data for information access dialogues, we set about combining these resources in a prototype dialogue system for closed domain question answering.

## 5 Our Prototype System

To construct an application from the available resources, a number of additional steps were required:

- The relational templates in the corpus annotation were compiled into a database.
- A robust parser for question constructions was defined, using technology developed for parsing spoken language (Pinkal et al., 2000; Worm, 2000). Fragmentary analyses from the SPROUT parser were taken as input.
- The SPROUT parser, robust parser and business database were encapsulated as modules for TRINDIKIT.

For the remaining modules where we had no specific resources we followed standard practice in TRINDIKIT prototypes, e.g. defining template generation for system outputs.

The dialogue model divides the core task into three phases:

1. obtaining information to consult the database,
2. interaction with information resources,
3. the appropriate presentation of results.

The first and third phases are conducted as dialogue, including clarification subdialogues. The order in which information is exchanged in each phase can be left relatively free, because of the ISU model. In the presentation phase, follow-up questions that can be handled with information already present in the information state are regarded as clarifications, but a question requiring new information from the database is treated as a new query. However, follow-up questions are interpreted in the existing context, thus considerably shortening the new initial phase.

## 6 Usage

The purpose of constructing this prototype was to get a foothold on the phenomena of realistic information access dialogues, using a restricted system as an environment for eliciting dialogue data. There are two parallel goals here: to conduct usability experiments to see how natural interactions in such an environment can be, and to obtain linguistic data to improve the linguistic components. Since these results feed into subsequent versions of the system, our strategy is effectively one of bootstrapping by means of prototyping. To some extent, this replicates on a small scale the development process effectively employed in Verb*mobil* (Wahlster, 2000).

## Acknowledgement

This work is part of the COLLATE project funded by the German Federal Ministry of Education and Research (BMBF) under grant 01INA01B.

## References

Markus Becker, Witold Drozdzynski, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. 2002. Sprout - shallow processing with typed feature structures and unification. In *Proceedings of the International Conference on NLP (ICON 2002). December 18-21*, Mumbai, India.

Jonathan Ginzburg. 1994. An update semantics for dialogue. In Harry Bunt, Reinhard Muskens, and Gerrit Rentier, editors, *1st International Workshop on Computational Semantics*, Tilburg, The Netherlands. ITK.

Jonathan Ginzburg. 1996. Dynamics and the semantics of dialogue. In Jerry Seligman and Dag Westerstaahl, editors, *Language, Logic and Computation, Volume 1*, CSLI Lecture Notes. CSLI, Stanford.

Staffan Larsson, Alexander Berman, Leif Grönqvist, and Fredrik Kronlid. 2002. Trindikit 3.0 manual. Siridus Report 6.4, Göteborg University, Department of Linguistics, Göteborg, Dezember.

Manfred Pinkal, C. J. Rupp, and Karsten Worm. 2000. Robust semantic processing of spoken language. In W. Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, Artificial Intelligence, pages 322–336. Springer, Berlin.

Wolfgang Wahlster, editor. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin.

Karsten L. Worm. 2000. *Robust Semantic Processing for Spoken Language*. Ph.D. thesis, Universität des Saarlandes, Saarbrücken, Germany, June.