

# Theoretically Grounded Conversational Interfaces to Digital Information

Yookyung Kim, Amy Perfors, Stanley Peters, and Cynthia Thompson

July 18, 2000

## Abstract

This paper introduces a general methodology for the development and study of conversational systems. Closer coupling of linguistic and psycholinguistic theories about dialogue with the development of dialogue systems allows each to inform the other more fully. Three activities support our goals. We develop theories through the empirical investigation of dialogue in conversational settings. We use this to inform our theoretical research on dialogue and our design of advanced conversational interface systems. We then experiment with these systems to compare their behavior to human-to-human interactions. This paper discusses a study and analysis of conversations about library information, the development of a system based on this analysis, lessons learned, and future goals.

## 1 Introduction

Conversations are a ubiquitous, universal form of human communication. It is difficult for humans to jointly carry out a task without at some point discussing the task and their progress on it, either by spoken or written communication. It is desirable for computers to also be capable of this natural sort of communication. Research in this area has recently become more feasible with advances in the underlying technology, and as evidenced by the recent creation of SIGdial<sup>1</sup>, the Special Interest Group on discourse processing, more and more researchers are taking advantage of this opportunity.

The research on conversation discussed in this paper has two main goals. The first is to advance the theory of dialogue, especially dialogue in the service of cooperative activities. Second, we aim to use this theory in designing conversational interfaces that allow human users to interact with resources or devices through the mediation of computers. These interfaces are meant to be conversational in two senses. One is the literal sense of allowing spoken<sup>2</sup> interaction between user and device. The other is that users should experience interactions as natural, cooperative, flexible, and informal.

Our research is influenced by the work of (Grosz, 1978), who was among the first to point out the importance of understanding the task being carried out in a dialogue, in order to understand the dialogue. Later work by (Chu-Carroll & Brown, 1997) showed the importance of distinguishing between the *task*-level and the *dialogue*-level in conversations. Using this as a basis leads us to view dialogue as a cooperative activity (Clark, 1996) between two participants.

We also draw on past approaches to modeling dialogue. An early approach was to use dialogue grammars (e.g., (Sinclair & Coulthard, 1975; Reichman, 1981)) to define the constraints on acceptable dialogues. While simple and theoretically grounded, this approach leads to inflexible, domain-specific grammars and systems. Plan-based approaches (e.g., (Allen & Perrault, 1980; Carberry, 1990)) show a greater ability to handle complex dialogues, but the resulting systems are often slow because of the difficulty of determining user's underlying plans. Finally, more recent collaborative approaches (e.g., (Clark & Wilkes-Gibbs, 1986;

---

<sup>1</sup>See [www.iet.com/Projects/sigdial](http://www.iet.com/Projects/sigdial)

<sup>2</sup>although we eventually plan to incorporate multi-modal interaction, we focus here on speech and natural language

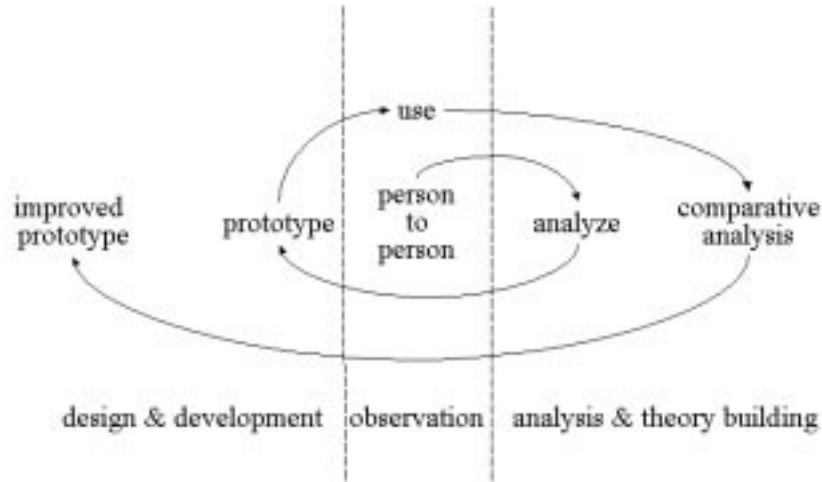


Figure 1: Dialogue Research Methodology

Grosz & Kraus, 1996)) have captured more of the underlying motivations and mechanisms of dialogues. While we favor the last approach, there is a natural tension between our broader goals, and limited resources, both in the underlying technology and in time. For example, while attempting to make progress on discourse theory, we take advantage of existing speech recognition tools, and thus face some limitations imposed by them, as discussed later.

Our view of dialogue as communication serving a cooperative activity while also being regulated by linguistic principles motivates a particular approach to designing and developing computer systems that carry out one side of a dialogue. This approach has three crucial aspects: (i) empirical observation, (ii) analysis and theory building, (iii) design and development. Actual observation of naturally occurring person-to-person interactions is an essential step for solidly grounding the design and development of artifacts intended to perform the role of one or more participants. Analysis of how people talk with one another in the process of performing a certain cooperative activity constrains the design of a dialogue system to aid in that activity. Developing a computational prototype can thus be supported by empirical fact and dialogue theory as well as by methods and tools for computing.

These types of research activity recur over and over again in a spiraling fashion (see Figure 1), for the prototype can be deployed and its use by actual people observed for the purpose of further analysis. Comparing results with the human-to-human case feeds into the design of improvements in the prototype, and deployment of a new version sets the stage for further iterations of the process. This paper addresses the first three stages in the spiral. Use of analysis of human-to-human conversation to design computational tools is not new; however, relatively few systems have been actually implemented and analyzed with this methodology.

Our first efforts have focussed on two application domains. The first is a library reference system with spoken and typed input and output. This system will augment the existing visual display and keyboard input of the library catalog system with additional capabilities and enhanced ease of use. Users will interact with the system using natural speech in order to find the answers to their questions or items or databases that they are looking for. The second domain is an adaptive destination advisor with a spoken interface

Category	Words	Utterances	Turns
book finding	31	8	6
periodical/article finding	47	13	9
stack location	68	19	14
library finding	101	19	17
search/system query	118	17	17
subject search	900 odd	87	69
miscellany	35	6	4

Table 1: Dialogue Categories

for use in an automobile. This paper will primarily discuss the library system.

Our approach is based on the analysis of real-life conversations as examples of task-oriented dialogues. We made audio tapes of conversations between librarians and people seeking information at the reference desk in order to study the linguistic and conversational strategies used to solve specific tasks in this domain. We then transcribed these recordings and annotated them with tags indicating the communicative functions of each utterance segment. Based on the annotated dialogues, we constructed a discourse grammar for this class of dialogues and built grammars including typical word sequences and constructions for this domain. These pieces were incorporated into a complete prototype for spoken and typed interaction with a library interface system. The next section discusses the collection and analysis of the conversations. This is followed by a description of our development process and implemented system. We then discuss our next steps and related research, followed by our conclusions.

## 2 Collection & Analysis

Our first step was to collect human-to-human conversations in a natural environment. Because of our collaboration with the library, we chose the reference desk as our environment. 55 hours (20 tapes) of conversation between librarians and patrons were recorded and 8 of the tapes transcribed, for a total of 259 dialogues.<sup>3</sup>

The next step was to analyze the conversations linguistically and in terms of library related features. We first categorized dialogues into several groups, listed in Table 1, according to the type of task of the patron, assuming that different tasks would exploit different strategies not only in solving the task but also in conversational structure. This categorization would confirm or nullify this assumption, but also simplify further analysis.

Keeping in mind that the object of this empirical work was to guide the design of the prototype conversational interface, we focused on certain kinds of statistics. These included length of conversations: number of words, utterances, and turns; control of conversations; the amount of inherent structure in the conversations; and task-solving strategy. Table 1 gives some counts for typical conversations in each category. Categories with lots of structure allowed pre-stored responses, so were good candidates for early development, because we only had to consider conversational features without worrying about how to access databases to provide correct answers for queries. Since they also involve some clarification sub-dialogues and adjacency pairs such as question and answer, they were interesting enough to implement.

One of the most important findings from this empirical study was that we cannot simply wrap a speech interface around the existing library search interface and obtain a truly useful conversational interface. The main reason for this is that we found that librarians try a variety of methods for assisting users. Librarians

<sup>3</sup>We first collected data from the humanities and social science library. The findings reported here are based on that data. We later recorded reference transactions in the engineering library to see whether there was any significant difference in conversational structure or problem solving strategy depending on the field of patrons, and there was none.

attempt to teach patrons these methods for finding information in the future, while also providing them with the specific results of applying these methods to the current query. This helps patrons evaluate both the results and the appropriateness of their original query. We consider this particular point to be a merit of interaction with a librarian. Therefore, we made an effort to include this in our conversational interface design, since conversation is the best, if not only, method to teach patrons these search techniques.

We found that conversational initiative changes hands even within one conversation. When a patron has a specific query, the task is accomplished with only a few turns, with the librarian yielding control to the patron. On the other hand, when negotiation is necessary, for instance to find out what the patron wants or to revise the original query, the initiative changes often between the patron and the librarian. As discussed in the next section, this flexibility was not fully achieved in the first prototype, but will be an important goal in future implementations.

In summary, we tried to find an overall pattern of communicative acts in each category, to guide the design of a discourse grammar which recognizes and predicts discourse intentions.

### 3 Prototype Development

We designed and implemented the prototype with two main goals in mind. First, we wanted to create a conversational interface that incorporated the findings from our empirical analyses of the library dialogues. Second, we wanted to create a prototype that could be quickly implemented and deployed in an actual library context. This would make it possible to collect data about the effectiveness the system, and modify it accordingly at an early stage in the project. This constraint meant that the prototype we ultimately developed, since it was only intended as the first iteration of a much farther-reaching goal, was not a full-scale implementation of all levels of the original proposal.

Designing the dialogues of which our computer interface was capable of involved the following:

1. Generating (from the collected dialogues) overall patterns of interaction that are natural for human users of the system while remaining within current algorithmic capabilities.
2. Specifying what vital information the system needed to consider, and at what point in a dialogue this information should be communicated.
3. Spelling out precisely what vocabulary and grammatical patterns should be recognized by the interface system at each stage in the dialogue.

These steps were made possible by the collection and analysis of the data discussed in the prior section. With the exception of the “subject search” and “miscellany” categories, which we did not include in the prototype, most of the dialogues in the library domain followed regular, almost scripted patterns. This made it possible to characterize natural-seeming patterns of interaction, as well as which information was vital at every stage of the dialogue. Building these patterns into the prototype allowed natural speech while still remaining within current recognition capabilities. Furthermore, the small range of dialogue patterns was paralleled by a relatively minimal vocabulary. This factor made it possible to spell out precisely the vocabulary and grammatical patterns likely at each point, and made implementation of a speech recognition system far more feasible than otherwise.

Since the thrust of our research was not in generating better speech recognition capability, we used Entropic’s off-the-shelf recognition tools and programming interfaces in the implementation. We were searching for software that could handle continuous speech among a large number of users and over a (relatively) large vocabulary. At the time, Entropic’s software best maximized all three variables. As is now becoming standard, recognition of multiple users was facilitated by restricting the number of recognizable utterances, through the specification of recognition grammars of legal utterances.

The main limitation we encountered came from our initial goal of recognizing the majority of the user’s utterances. This meant we needed to incorporate book author and title names, which would add tens of

thousands of words to the vocabulary. Adding some names caused accuracy to plummet (down to 48.7% from over 60% in one trial with only 150 book titles); therefore, we eliminated the ability to recognize these words, forcing the user to type them in at the appropriate point in the dialogue. This was not ideal given the goal of a natural interaction, but it seemed the best compromise in view of the opposing goal of getting a prototype up and working quickly.

A related limitation arose because although the dialogues we analyzed were fairly circumscribed in comparison with natural conversation, they were still enormously complicated. The analysis helped us divide the user's possible utterances into several classes. Since recognition accuracy decreases as size of the recognition grammar increases, this was helpful in designing smaller grammars, one for each class of utterance. This led naturally to a finite state model of dialogue management in which we implemented a distinct recognition grammar for each stage in a dialogue. This limitation, however, was mostly hidden from the user. The final network had over 850,000 word nodes and just under 1,000,000 transitions.

The states in the dialogue model were constructed according to the empirical evidence obtained earlier, based on natural states arising in typical library interactions. Transitions were at first based on key-word matching of words in a user's utterances. This method became cumbersome as the grammars were expanded, so we augmented the networks with "tags" that had no associated recognition input, but which added these tags to the output recognition strings, allowing for more flexible transitions between states.

For every state we decided: i) what the system would display on the screen; ii) what the system would say; iii) what user utterances would be recognized; iv) what the system would suggest the user should say, if they asked for help; v) what other states the system would go to, for each class of allowed user inputs. The possible user inputs were determined from the dialogue analysis stage.

This basic construction resulted in a prototype that could engage in limited but natural-sounding interaction with human clients.

## 4 Next Steps

We learned many lessons in our analysis and implementation, and we hope to apply them to our future endeavors in this area. Our prototype was never connected to the library databases for lack of resources (also library resources), so we were unable to evaluate its effectiveness as we originally planned. However, we are currently working on another project, discussed further below, that should ease the process of deploying an updated version of the system. At that point, we will deploy the new system, gather user data, and perform a comparative analysis with the original conversation data. This will contribute to further stages of our design methodology in Figure 1.

We are also currently investigating methods for facilitating the development of conversational interfaces. Many of the systems in the literature, including our prototype, are specific to one domain. Our proposed solution is to develop a tool for the construction of dialogue systems. Similar in working to a compiler, the tool would take as input a script specifying dialogue states, allowed utterances, and conditions for moving from state to state. The mechanisms for managing state transitions, coded in a generic form, would then be merged with the script to produce a complete system, including speech recognition grammars. This work is still in progress, but can be compared to the work at SRI (Stent, Dowding, Gawron, Bratt, & Moore, 1999) for producing grammars for both speech recognition and parsing, though we concentrate on entire dialogue systems and a scripting language that is easy for a novice programmer to write.

Finally, we are also pursuing the personalization of dialogues to individual users, while also building prototypes in the domain of destination advice. This work should also lend insight into the development of general, flexible, and cooperative interfaces.

## 5 Related Research

There is currently much activity and excitement in the area of conversational interfaces. This section can not hope to mention all of this work, so we instead strive to include the most closely related research.

First, many others have collected and analyzed human-to-human conversations, but only a few have used this to motivate system design. One exception is the Human Communication Research Center project to record participants in the Map Task (Anderson, Bader, Bard, Boyle, Doherty, Garrod, Isard, Kowtko, McAllister, Miller, Sotillo, Thompson, & Weinert, 1991). The corpus was used primarily for aiding the development of speech recognition and syntactic parsing tools, and to develop a lexicon, grammar, and parsing scheme for the corpus. While being noteworthy in using a corpus of real data for guiding some of the implementation, the project did not move on to the next stage of developing a complete conversational interface. Also, (Zhou, Freedman, Glass, Michael, Rovick, & Evens, 1999) describe a dialogue-based tutoring system which incorporates strategies based on analyses of human transcripts.

The research for the TRAINS (Allen, Schubert, Ferguson, Heeman, Hwang, Kato, Light, Martin, Miller, Poesio, & Traum, 1995) project began with the collection and analysis of two humans collaborating to plan train delivery schedules and routes (Traum & Hinkelman, 1992). They use some results of analysis of the dialogues (tagging, etc.) to influence other components of the system. Together with other groups, an annotation scheme for communicative acts in dialogue, DAMSL (Dialogue Act Markup in Several Layers)<sup>4</sup>, has been developed and used by many other researchers. The TRAINS group does not, however, focus on the interaction between analysis of dialogues to influence future system development, nor comparative analysis of the system to human-to-human dialogues.

Viewing dialogue as a cooperative activity is not a novel approach. Some foundational theoretical work was done by (Clark & Schaefer, 1989) and later extended by (Novick & Hansen, 1995) and others. Several of the groups mentioned above also take this perspective. More recently, (Rich & Sidner, 1998) discuss a tool for building collaborative interfaces to applications. Although they are not interested in natural language understanding, they are interested in theories of discourse and collaboration. Their tool provides an hierarchically structured interaction history, containing the user's and agent's goals and intentions. They also allow collaboration where both the user and agent can interact with the application of interest, whereas we view the agent as accessing information from an application or other source to assist the user.

Finally, an alternative model for collecting dialogues to use as a basis of system design is the so-called "Wizard of Oz" technique. A user is made to think they are in conversation with a computational interface, but the system is actually incomplete in some way, typically with a human actually performing the natural language processing and generating appropriate responses. In this way, people's natural tendencies to lower their expectations when interacting with machines can be taken into account during system design.

## 6 Conclusion

We found the methodology of first studying empirically the natural phenomenon we seek to replicate computationally to be quite helpful in the design of even our first prototype system. Our experience was that even for relatively simple tasks such as helping a person find information, a separate task module or "agent" is important for good system design. Our initial prototype suffered from an inadequate separation between constraints arising from the task to be performed and linguistic constraints on dialogue. We are rectifying this by building both a separate task engine and a dialogue "scripting" engine. The latter will also be helpful in allowing linguists to perform work that does not require programming knowledge.

Dialogues oriented to the task of finding information, though not all alike, are remarkably stylized and restricted in vocabulary (apart from the open-ended categories of author names and publication titles). This fact offers real hope of constructing truly usable conversational systems to help people find information in various domains, including libraries.

---

<sup>4</sup>See [www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html](http://www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html) for details

## References

- Allen, J., Schubert, L., Ferguson, G., Heeman, P., Hwang, C., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M., & Traum, D. (1995). The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 7, 7–48.
- Allen, J. F., & Perrault, C. R. (1980). Analyzing intention in utterances. *Artificial Intelligence*, 15(3), 143–178.
- Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H., & Weinert, R. (1991). The HCRC map task corpus. *Language and Speech*, 34(4), 351–366.
- Carberry, S. (1990). *Plan recognition in natural language dialogue*. MIT Press, Cambridge, MA.
- Chu-Carroll, J., & Brown, M. (1997). Tracking initiative in collaborative dialogue interactions. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics* Montreal, Canada.
- Clark, H., & Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.
- Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39.
- Clark, H. (1996). *Using Language*. Cambridge University Press.
- Grosz, B. (1978). Discourse analysis. In Walker, D. (Ed.), *Understanding Spoken Language*. Elsevier North-Holland, New York, NY.
- Grosz, B., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence*, 86(2), 269–357.
- Novick, D., & Hansen, B. (1995). Mutuality strategies for reference in task-oriented dialogue. In *Proceedings of the 9th Twente Workshop on Language Technology* Enschede, Netherlands.
- Reichman, R. (1981). *Plain-speaking: A theory and grammar of spontaneous discourse*. Ph.D. thesis, Harvard University, Cambridge, MA.
- Rich, C., & Sidner, C. (1998). COLLAGEN: A collaboration manager for software interface agents. *User Modeling and User-Adapted Interaction*, 8, 315–350.
- Sinclair, J., & Coulthard, R. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press, London, England.
- Stent, A., Dowding, J., Gawron, J., Bratt, E., & Moore, R. (1999). The CommandTalk spoken dialogue system. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics* College Park, MD.
- Traum, D., & Hinkelman, E. (1992). Conversation acts in task oriented spoken dialogue. *Computational Intelligence*, 8(3), 575–599.
- Zhou, Y., Freedman, R., Glass, M., Michael, J., Rovick, A., & Evens, M. (1999). Delivering hints in a dialogue-based intelligent tutoring system. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 128–134 Orlando, FL.