A Flexible Mixed-Initiative Speech Interface for Restaurant Information

Stephanie Seneff† Chao Wang†, and Grace Chung‡*

†MIT Computer Science and Artificial Intelligence Laboratory
Stata Center, 32 Vassar Street, Cambridge, MA 02139
{seneff,wangc}@sls.csail.mit.edu

‡Corporation for National Research Initiatives
1895 Preston White Drive, Suite 100, Reston, VA 22209
gchung@cnri.reston.va.us

Abstract

This live demonstration will involve spoken conversational interaction over the telephone with a system that provides information about a set of nearly 1000 restaurants in the Greater Boston area. Several aspects of our recent research in spoken dialogue system development will be illustrated, including (1) the extensive use of simulated dialogues both for system development and for generating training data for the recognition and understanding components, (2) the ability to tailor a generic dialogue manager to a specific domain through external declarative knowledge, (3) the ability to adjust the vocabulary on the fly to reflect specific entities retrieved from a database, (4) flexible summarization capabilities for lists of tuples, where the summary's content takes into account the distribution of attributes over values, and (5) a capability to back off to alternative suggestions when the database contains no items matching a user's specific requests.

1 Introduction

For over fifteen years, researchers in the Spoken Language Systems Group at MIT have been developing human language technologies for *mixed initiative* conversational systems, which are distinguished from the emerging deployed commercial systems in that the interaction is natural and flexible, modelled after the style of human-human dialogue (Zue and Glass, 2000). The development of the Galaxy Communicator architecture (Seneff et al., 1998) has accelerated the pace at which experts can configure complex dialogue systems in a wide range of different domains. As the underlying technology components have matured, our research focus has evolved to include issues related to portability, flexibility, and dynamic vocabulary. We believe that the ability for naive system developers, even end users, to reconfigure existing systems to manage their personal needs will be crucial for the successful future deployment of these technologies.

The research at MIT was supported in part by an industrial consortium supporting the MIT Oxygen Alliance.

2 Illustrated Technical Issues

This live demonstration will illustrate a new telephone-access spoken dialogue system, which can answer a wide range of questions concerning a database of about 1000 restaurants in the Boston area. As with all of our conversational dialogue systems, this new system is based on the Galaxy Communicator architecture. However, there are several recent technology developments that have been incorporated into this new domain to set it apart from our earlier development efforts. These include (1) user simulations, (2) generic dialogue modeling, (3) dynamic vocabulary, and (4) flexible response generation. We address each of these in turn below.

(1) **User Simulations:** One of the critical barriers to widespread deployment of spoken dialogue systems is the chicken-and-egg problem of real user data. User utterances drawn from spoken dialogues with the system are critical for performance improvement, particularly with respect to the language model (LM) for the recognizer. However, such data are difficult to obtain, since users are reluctant to interact with the system during its early development stages. Our solution was to develop a *user simulator* which could role play the user in simulated dialogues with the system. Our language generation framework (Baptist and Seneff, 2000) supports multiple paraphrases of the same meaning representation into surface strings, which allows us to collect a large number of variants on ways to express queries about particular topics relevant to the domain. These simulated data are currently the main contribution to LM training. They also guide the development of the natural language grammar.

(2) **Generic Dialogue Modeling:** The dialogue manager is usually a very labor-intensive component of dialogue system development, particularly for mixed-initiative dialogue, where the system must be able to respond appro-

USER:	Are there any Chinese restaurants in Harvard square?
SYSTEM:	I found no Chinese restaurants in Harvard Square. However, there are in total 34 Chinese restaurants. None of them are high end. Many of them are on Beacon Street, Hudson Street, Beach Street, and Tyler Street. The particular neighborhoods are Chinatown, The South End, and Fenway. Most of them are located in Boston, and Brookline. The nearest one is New Asia in Somerville. It is less than 2 miles away from Harvard Square.

Figure 1: An example dialogue interaction between the user and the system, where the system offers several alternative suggestions given that the database fails to deliver any restaurants matching the user's original request.

priately to unexpected as well as expected moves, and also to recover gracefully from recognition error. For the new restaurant domain, we have made use of a *generic* dialogue manager, simultaneously under development, which is evolving to support many of the functions that are needed to respond appropriately to a wide range of user queries. All system turns are controlled by a set of rules in a simple scripting language, which execute specified *operations* when tested conditions are met, as described in (Seneff and Polifroni, 2000). The dialogue management task is decomposed into a set of about 70 such operations, each of which handles a specific aspect of query processing. Typically a dozen or more are executed at each turn. All domain-dependencies have been relegated to external declarative tables, such that the dialogue manager would be effective for any domain involving database queries against entities with attributes.

(3) **Dynamic Vocabulary:** This new restaurant domain represents, for the first time, a significant milestone in our research agenda, with the introduction of a capability to automatically update the working vocabulary with new proper names on-the-fly, on the basis of information retrieved from the database. It builds on research described in (Schalkwyk et al., 2003; Seneff et al., 2003). At the beginning of a new dialogue with a user, the system has no knowledge of any restaurants by name, but, as the dialogue unfolds, new names are constantly added, reflecting database retrievals. This allows the user to refer by name to any restaurants that have been previously mentioned by the system. Furthermore, the system is capable of understanding a restaurant name newly introduced in a single user turn, as long as the turn also contains other context to restrict the space of alternatives. For example, if the user says, "What is the phone number of Bertucci's in Cambridge," the system first recognizes "Bertucci's" as an unknown word, then dynamically updates the vocabulary for all restaurants in Cambridge, and reprocesses the waveform, resulting, in many cases, in a correct understanding of the name.

(4) **Flexible Response Generation:** Response generation in dialogue systems concerns both what to say and how to say it. It is important for spoken dialogue systems to be able to succinctly describe lists of items retrieved from the database, tailoring the level of detail to the size of the retrieved list as well as the distribution of attribute values. To achieve this goal, we have recently enhanced our generation system with the capability to speak effectively about database tuples under diverse conditions, through minimal external control of parameter settings.

In addition to determining the set of items from the database that meet the constraints of the query, the dialogue manager is also tasked with deciding how to back off when no data match the user's requests. We have devoted considerable effort to developing a generic framework for offering suggestions, as illustrated by the dialogue exchange in Figure 1.

3 Future Work

The restaurant domain is still in its infancy: we are only now beginning to collect real user data and use them to improve all aspects of the system's performance. However, we are encouraged that users are often able to interact effectively with the system to obtain useful information. We believe a critical milestone has been reached to greatly accelerate the pace at which new database-query domains can be instantiated, but this remains to be seen through future continued development of novel domains.

References

L. Baptist and S. Seneff. 2000. Genesis-II: A versatile system for language generation in conversational system applications. In *Proc. ICSLP*, pages 271–274.

J. Schalkwyk, L. Hetherington, and E. Story. 2003. Speech recognition with dynamic grammars using finite state transducers. In *Proc. EUROSPEECH*, pages 1969–1972.

S. Seneff and J. Polifroni. 2000. Dialogue management in the Mercury flight reservation system. In *Proc. ANLP-NAACL 2000, Satellite Workshop*, pages 1–6, Seattle, Washington.

S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid, and V. Zue. 1998. Galaxy-II: A reference architecture for conversational system development. In *Proceedings, ICSLP*, pages 931–934.

S. Seneff, G. Chung, and C. Wang. 2003. Empowering end users to personalize dialogue systems through spoken interaction. In *Proc. EUROSPEECH*, pages 749–752.

V. Zue and J. Glass. 2000. Conversational interfaces: Advances and challenges. *Proc. IEEE*, 88(11):1166–1180.