A Dialogue System for Language Learning

Chao Wang and Stephanie Seneff
MIT Computer Science and Artificial Intelligence Laboratory
Stata Center
Cambridge, MA 02139, USA
(wangc,seneff)@csail.mit.edu

Abstract

We demonstrate an English/Chinese bilingual weather information system that is configured as both a conversational partner and a tutor for foreign language (L2) students to practice conversational skills.

1 Introduction

The main components to language learning are reading, writing, listening and speaking. While it is possible for diligent students to gain adequate proficiency in the first three areas, the goal of improving conversational skills cannot be achieved by simply working hard. This is mainly due to the lack of a proper environment and adequate opportunity to practice speaking.

Over the past 15 years, members of the Spoken Language Systems group at MIT have been developing multilingual spoken conversational systems typically supporting access to on-line information sources. We have long believed that such systems could benefit foreign language learning, because conversational interaction is a critical part of acquiring fluency, and the computer provides an entertaining and non-threatening environment.

2 Technical Challenges

Dialogue systems for language tutoring pose additional challenges compared to systems used for information access. Above all, speech recognition of non-native speech is a difficult problem. Frequent recognition and understanding failures can easily frustrate the student and make the dialogue interaction unappealing. In the absence of non-native speech data to adapt the acoustic models, we adopted two strategies to address this problem. The first strategy is to customize the speech recognizer for focused lesson plans, and re-generate the recognizer language model for different lessons. In this way, the recognition system will have a constrained search space, and hence a better chance to understand the input correctly. Our second strategy is to provide the ability to generate high-quality translations for speech inputs produced in the native language, to provide L2 students with immediate assistance when they have difficulty expressing themselves in the new language. With this assistance, their conversation with the computer can progress even before they have sufficient proficiency in the foreign language.

Our translation framework is based on an interlingual approach, which is shared with the dialogue system development. Our goal is to obtain extremely high quality translation, which can be achieved if the domain is sufficiently restricted. Given an input sentence, a parse tree is derived and critical elements in the parse tree are extracted. The resulting semantic frame can be used to generate key-value information for the dialogue system, and to generate a sentence in the original language (paraphrasing) or in a different language (translation). The generation is controlled by a set of rules, which can be fine-tuned by experts to achieve high quality.

In order for our approach to succeed, the grammar used for parsing must satisfy the following requirements: first, it must be able to capture detailed linguistic constructs of the input sentence, to ensure an accurate translation; and second, it must be easily adaptable for different lessen plans. Our solution is to construct a generic core grammar that is linguistically rich and easily adaptable for different domains within the dialogue interaction scenario. The grammar is induced semi-automatically, making heavy use of existing resources such as several grammars developed for various application domains, as well as speech corpora collected using those systems. The core grammar contains top-level rules specifying detailed syntactic structure, with part-of-speech nodes to be expanded into words or word classes specific to an application domain. Parse rules for general semantic concepts such as dates and times are organized into sub-grammars that are easily embedded into any applications. Phrases for sub-grammars are also extracted from existing speech transcripts, which can be re-used to train a statistical language model for the new domain (Seneff et al., 2003).

User asks in Chinese:	bo1 shi4 dun4 ming2 tian1 hui4 xia4 yu3 ma5?
	(Is it going to rain in Boston tomorrow?)
System replies in Chinese:	tian1 qi4 yu4 bao4 ming2 tian1 bo1 shi4 dun4 mei2 you3 yu3.
	(The forecast calls for no rain in Boston tomorrow.)
User asks in English:	What is the temperature?
System translates into Chinese:	qi4 wen1 shi4 duo1 shao5?
User imitates in Chinese:	qi4 wen1 shi4 duo1 shao5?
System replies in Chinese:	bo1 shi4 dun4 ming2 tian1, zui4 gao1 qi4 wen1 er4 she4 shi4 du4,
	ming2 tian1 ye4 jian1, zui4 di1 qi4 wen1 ling2 xia4 wu3 she4 shi4 du4.
User requests translation:	Could you translate that?
System translates:	In Boston tomorrow, high 2 Celsius, tomorrow night, low -5 Celsius.

Figure 1: An example dialogue interaction between the user and the system. Chinese inputs are represented using tone-marked pinyin. English translation of Chinese sentence is provided in parentheses.

3 System Overview

In our initial effort, we have chosen to focus on the scenario of an American learning Mandarin Chinese. Built upon our multilingual Jupiter weather information system (Zue et al., 2000; Wang et al., 2000), we have configured a system which is capable of interacting with a user on weather related questions. When the student asks a question in Chinese, the system will provide an answer in Chinese. When the student asks a question in English, the system will try to provide a Chinese translation of the input sentence. The student could then imitate the Chinese translation and obtain an answer back. The student could also explicitly request the system to repeat or to translate a piece of information. The system is able to keep track of the proper dialogue context, despite the digressions in the dialogue interaction. The system also echoes a paraphrase of the recognized user input in the same language, before responding to the user query, which is helpful to keep the user aware of any recognition errors. A novel use of our translation capability is to automatically generate Chinese queries from a large existing corpus of English utterances in the weather domain, to provide language model training for the Chinese recognizer.

Another feature of our system is the capability to support seamless switching between the two input languages. We believe that it is much more natural than the alternative strategy, which is to use some kind of meta command (verbal or via the telephone keypad) to explicitly request a switching of input language (and hence, the recognizer). Language switching is achieved by joining the acoustic models and search spaces of English and Chinese recognizers and simultaneously decoding both the words and the language (Hazen et al., 2001).

Figure 1 illustrates an example dialogue between the user and the system. The system also echoes a paraphrasing of each recognized user input in the same language, which is omitted from the figure for simplicity.

4 Future Work

In order to make dialogue systems a viable tool for improving conversational skills, we must make our underlying technologies highly portable, so that systems specializing in different lesson plans can be rapidly configured. We have made some progress towards making the grammar and dialogue manager more generic. However, much work remains to be done in natural language generation and high-quality speech synthesis.

We plan to expand the conversational domains and collect data from real L2 students interacting with our systems. These data can be used to improve and evaluate system performance, especially in the area of automatic recognition of non-native spontaneous speech. Furthermore, these data are valuable for developing algorithms to automatically assess pronunciation and other aspects of speech production.

References

T. J. Hazen, I. L. Hetherington, and A. Park. 2001. Fst-based recognition techniques for multi-lingual and multi-domain spontaneous speech. In *Proc. EUROSPEECH*, pages 1591–1594, Aalborg, Denmark.

S. Seneff, C. Wang, and T. J. Hazen. 2003. Automatic induction of n-gram language models from a natural language grammar. In *Proc. EUROSPEECH*, pages 641–644, Geneva, Switzerland.

C. Wang, D. S. Cyphers, X. Mou, J. Polifroni, S. Seneff, J. Yi, and V. Zue. 2000. MUXING: A telephone-access mandarin conversational system. In *Proc. ICSLP*, pages 715–718, Beijing, China.

V. Zue, S. Seneff, J. Glass, J. Polifroni, C. Pao, T. J. Hazen, and L. Hetherington. 2000. JUPITER: A telephone-based conversational interface for weather information. *IEEE Transactions on Speech and Audio Processing*, 8(1):85–96.