

9807-11

Multilingual Conversational System Research

James Glass and Stephanie Seneff

Project Overview:

The long-term goals of this research project are to foster collaboration between MIT and NTT speech and language researchers and to develop language-independent approaches to speech understanding and generation. We will initiate this effort by developing the necessary human language technologies that will enable us to port our conversational interfaces from English to Japanese. The Jupiter weather information system will be used as the basis of this porting process. This work will involve the close collaboration with NTT researchers both in Japan and at MIT.

Third Year Research Plan

By the end of 1999 we developed a laboratory prototype for a Jupiter system which can handle Japanese input and output. This system, which we call Mokusei, has been used to collect preliminary data from native Japanese speakers. We are currently devoting a significant effort to incorporating a more flexible language generation framework into Mokusei. This framework, which is under active development, can produce more natural sounding language generation output for Japanese. We expect this modification to be complete by June, and will begin a second round of data collection from NTT employees at this stage.

Our third year goals are as follows:

1. Our top priority for the next year will be a wider-scale data collection from native Japanese speakers. Since data collection and system development are iterative processes, we expect at least two rounds of data collection. In the first stage we would like to have NTT employees use the Mokusei system. Once these data have been used to test and improve the recognition and understanding components, we will then initiate a second round of data collection with a wider distribution of users.

2. The data collection process will allow us to develop a more robust natural language capability using our TINA NL component. We expect that this will involve both expanding the coverage of the complete NL analysis which is currently performed, as well as augmenting it with a robust parsing mechanism modeled after the approach used in our English systems. We will also be able to use the collected corpus to better train the probabilities in the NL component.

3. The collected data will also be invaluable in improving the performance of our speech recognizer. The areas we plan to explore are improved acoustic-phonetic and phonological modeling through the integration of additional Japanese data with English data, and improved language modeling, possibly by using our NL component to help generate a statistical language model. Another area we plan to address is the addition of word and sentence-level confidence scoring for more robust understanding.

4. In addition to processing English sources of weather forecast information, we will also explore Japanese language content processing, in order to improve the quality and scope of the information the system can deliver concerning weather in Japan. For example, we will explore the feasibility of parsing weather reports available in Japanese from Web sites maintained in Japan, and incorporating the results into our weather database. The result will be improved weather information for both our English-based and Japanese-based systems.

5. Finally, depending on the expertise and interest of our NTT visitor, it may be possible to develop a Japanese version of our corpus-based concatenative speech synthesizer, called Envoice, for the Mokusei domain. The Envoice framework, which is integrated with our language generation component for meaning-to-speech synthesis, has been successfully used for very natural sounding speech generation for several of our English domains. We would be interested in exploring its use for Japanese to understand how well the employed techniques generalize to other languages.

Project-end Goals:

By the end of the third year, we hope that we will be able to have a complete Mokusei system available to a wider population of Japanese speakers. This will of course depend in part on the logistical feasibility of providing a publicly available telephone connection, and monitoring the system operation. However, our experience has shown that such a data collection framework is invaluable for collecting speech corpora of human-computer interactions, and will enable long term research in human language technology in this area.