

Multilingual Conversational Speech Research

9807-11

Progress Report: July 1, 2000—December 31, 2000

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.

Progress Through December 2000

In the past six months, we have continued improving all system capabilities, and have continued our data collection efforts. The following sections describe our activities in more detail.

Data collection and transcription

To date we have collected 1,900 read sentences, over 2,500 sentences from expert users, and 8,400 sentences from novice users. All of these data have been transcribed. In order to maintain consistency at word boundaries, we performed a morphological analysis to do automatic segmentation. Phonetic transcriptions of entire sentences were first manually segmented into bunsetsu sequences, which could then be converted into words. With this method, variations in pronunciations such as 'kyaroraina' and 'karoraina' (for Carolina) are reduced to a common word. The pronunciation variation information used here is shared with the baseform file for recognition. The morphological analyzer was built with the MIT Tina natural language component, and the grammar for morphological analysis is a part of the grammar for sentence parsing so that the consistency in word definitions is maintained. The morphological analyzer can analyze about 96% of bunsetsu in naive and expert sentences.

Speech Recognition

The acoustic model and language model have been trained with the transcribed data. The current recognizer has an active vocabulary of 1,061 words with a trigram test set perplexity of 12.9. On average there are 3.0 morae per word. On in-vocabulary test data containing no artifacts the word error rate is 9.4% with a sentence error rate

of 34.9% (average of 5.8 words/sentence). On the complete test set the word and sentence error rates increase to 20.5% and 48.1% respectively. These results are similar to the performance we obtain for our English weather system. We expect these results to improve significantly as we collect much larger amounts of data.

Language Understanding

The language understanding grammar has been rewritten during this period, and now covers 75% of the data collected from novice users. This is much better than we have achieved previously. Furthermore, the parser is now 10 times faster, due to better re-organization of grammar rules. We plan to continue improving the parser and incorporate robust parsing in the future. Currently, the concept error rate (analogous to word error rate on important concepts) is 12%, while the overall understanding error (correct concepts for the entire sentence) is 44%.

Other Activities

We have explored methods for increasing the amount of Japanese weather information available to the system. We have made contact with the Japan branch of Weather News Inc, who distributes weather information in Japan.

We have initiated work in corpus-based speech synthesis for Mokusei by preparing an NTT weather corpus for use with our synthesizer. We are currently exploring methods for Japanese intonation generation.

Finally, we have continued to augment our language generation capability for weather forecasts, as novel constructs (e.g., winter terminology) arise.

By the end of this period we had installed all of the software onto standalone computers, in preparation for deploying the Mokusei system at NTT for continuous data collection in Japan.

Research Plan for the Next Six Months

Over the next six months we plan to deploy our Japan-based Mokusei system and begin collecting and transcribing data. We expect these data to be extremely valuable in improving the performance of all aspects of the system. Finally, we plan to continue our efforts on more natural sounding output for the Mokusei system, which were described in the previous progress report.