

9807-11

Multilingual Conversational System Research

Progress Report
January 1 — June 30, 1999

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 to Date

As outlined in our previous progress report, we have continued to develop component capabilities for the Japanese Jupiter system called Mokusei. In particular we have made substantial progress in the area of language generation, and have created a small corpus of sentences to begin developing a language understanding component.

In the area of language generation, we have completed an initial version of the generation files that govern translation of semantic frames representing weather reports into Japanese. We estimate that approximately 70% of the translations are correct, with the remainder being in many cases comprehensible but ill-formed in minor ways. The process required only a few alterations to the original semantic frame structure, for instance distinguishing between the preposition "in" in a locative vs temporal reference.

As a first step in creating a corpus of Japanese weather queries for processing inputs in the Mokusei domain, we created a set of 600 text sentences, which was obtained by a combined effort of translating English queries and soliciting example questions (typed) from colleagues at NTT.

We have developed a preliminary version of a grammar to parse the 600 queries in the sample data set. At this time, approximately half of the sentences are parsable. Some of the non-parsable queries are beyond the scope of Mokusei, but others are just unusual phrasings that will require more extensive rules. Thus far, the grammar is strictly context-free. We have taken the approach of retaining the English parse-tree categories in the upper layers of the parse tree, and maintaining an equivalent hierarchical organization of the structure as much as possible. The fact that it is left-branching rather than right-branching is mostly irrelevant for the conversion of the parse tree to a meaning representation. Thus we are able to utilize the same rules file to govern the translation of the parse tree into a semantic frame as was used for English.

Research Plan

Over the next six months, we will continue the plan we outlined in our previous report. Specifically, we will continue to develop rudimentary natural language and speech recognition capabilities for

Mokusei. These efforts will be aided by continuing data collection efforts, both in the form of read and wizard-based data collection. Once we have a complete system in place, we will begin collecting data from subjects talking to it, and will be able to refine the various component technologies.

We will continue to improve the translations of the system responses, aiming towards a more natural phrasing in Japanese. In the area of natural language we plan to add constraints to the rules file, in particular to implement a trace mechanism to decrease the branching factor near the beginning of the sentence. We will continue to expand rules to enhance coverage, and investigate more extensive use of robust parsing.

We will create an initial speech recognizer by developing a Japanese vocabulary which covers the cities and concepts understood by the Jupiter system. The initial acoustic models will be seeded from our English Jupiter models, and will be retrained as data become available. We will also use these data to ascertain whether a probabilistic phonological component can be useful for Mokusei. Language models will be created based on available data. We will consider several different types of class n-gram language models to see which provide the most constraint for the domain.

Read speech data collection will utilize tools already in place for English data collection. We will build templates based on the 600 queries that we have obtained thus far, by automatically substituting different cities randomly in pre-existing templates. We plan to collect data from two pools of subjects, all from NTT, but divided into those who are calling from the United States and those who are calling from Japan. We will be comparing these two pools in part with an interest in comparing the quality of the long-distance phone lines with that of the local phone lines. Lines for the MIT subjects.

For wizard-based data collection we plan to configure a version of the system that will be appropriate for data collection in a realistic setting, in order to obtain spontaneous speech to enhance our read-speech corpus. We envision that the system will operate with two audio servers, where the subject asks a question which gets directly piped to the wizard, who then rephrases it (either carefully enunciated in Japanese or translated into English). The system answers the wizard's rephrased question and speaks the reply in Japanese to the subject.

We expect to begin work on the speech recognizer and read-speech data collection over the summer months. Work on wizard-based data collection will most likely begin in the fall, once the natural language capability has been extended for richer coverage.