SPOKEN LANGUAGE SYSTEMS

PI: John Makhoul

BBN STC, 10 Moulton St., Cambridge, MA 02138
makhoul@bbn.com

The objective of this project is to develop a spoken language system capable of understanding and responding to spoken English commands and queries for interactive human-machine applications, such as battle management, command and control, and training of personnel on complex tasks. The system will also include a capability to adapt to new speakers and a capability to detect when a user says a new word, and allows the user to add the word to the system.

Work in this area requires the integration of three technologies: large-vocabulary continuous speech recognition, natural language understanding, and system integration. In our work at BBN, we have integrated our BYBLOS continuous speech recognition technology with a new unification-based natural language understanding component, resulting in an initial complete spoken language system, called HARC (Hear And Respond to Continuous speech).

Our most recent contribution is the development of a new strategy for integrating speech and natural language components, called "N-best". This method takes a spoken utterance and produces the N highest scoring sentences that match the input utterance withing some threshold, based on a statistical language model. The natural language component then searches these N sentences for the highest scoring sentence for which the system can produce a semantic interpretation. The meaning representations are passed to a discourse component that resolves reference ambiguities and chooses the best meaning. Finally, the chosen meaning representation is passed to a response component which carries out the user's request.

Initial experiments have shown that for applications of interest, the correct sentence is usually one of the top five and almost always within the top twenty (i.e., N=20). One important feature of this N-best integration strategy is that it provides a very clean interface between speech and natural language and, therefore, allows for greater sharing of resources among researchers in spoken language systems.

The natural language knowledge sources in HARC use a Unification formalism for describing the syntax and semantics of English and a higher-order intensional logic for representing the meaning of an utterance. The system uses unification to enforce syntactic as well as semantic constraints, and provides for the incremental application of syntax and semantics. Advantages of this approach are that unproductive search paths are cut off more quickly, and any improvements in unification parsing (through better algorithms, special hardware, etc.) apply automatically to semantics as well as syntax.

In this project, we have been very instrumental in the design and collection of spoken language data for the purpose of objective system evaluation. We previously helped specify the DARPA Resource Management Corpus that is now in common use for speech recognition evaluation, and we provided a Word-Pair Grammar to be used with the corpus. We have recently developed and made available to the DARPA community, with full documentation, a personnel database for use in spoken language evaluations, and a relational database language (ERL) in Common LISP to interface to the database. We have also provided software to aid in the collection of an appropriate corpus by Texas Instruments.

Most recently, we have started work on the automatic detection of out-of-vocabulary words. This is an important problem for any realistic system with a large vocabulary, since the user is unlikely to be able to remember which words are in the vocabulary. Initial results for the detection of open-class words have been very encouraging.