

Overview of the Third DARPA Speech and Natural Language Workshop

Richard M. Stern, General Chair

School of Computer Science
Department of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

The Third DARPA Speech and Natural Language Workshop was designed to provide a forum for the presentation of research results and the discussion of technical issues related to a variety of DARPA research programs in spoken language systems, speech recognition, natural language processing, and related areas. The workshop also provided an opportunity to discuss the long-term goals of these programs, to develop strategies that will facilitate the sharing of databases and useful technology by all sites in the DARPA community, and to develop standards by which progress can be evaluated.

One of the major focal points of the three-day workshop was an extended discussion of the new DARPA spoken-language common task, the Air Travel Information System (ATIS). The topics of these sessions, which were organized and chaired by Dave Pallett, included descriptions of the rationale and infrastructure of the task, as well as reports from sites that had just completed the first round of common evaluations of spoken-language systems in the ATIS domain. The other special session, organized by Mitch Marcus, concerned the automatic acquisition of linguistic structure. This session highlighted significant progress being made in the exploitation of large corpora and in the fusion of symbolic and probabilistic methods. The remainder of the technical program consisted of reports on recent progress and new technical developments in the areas of speech recognition (including results from the latest benchmark evaluations on the Resource Management task), natural language processing, and spoken language system. David Nagel delivered a keynote address on spoken language and the personal computer.

The technical program was augmented by an evening session of system demonstrations coordinated by Victor Zue, with videotaped presentations by AT&T Bell Laboratories, Columbia, and Texas Instruments, and live demos by Dragon Systems, Entropic Systems, MIT, and MIT Lincoln Laboratory. Both the videotapes and the live demos were highly effective: the live systems were the focus of a great deal of attention and scrutiny throughout the meeting, and the videotapes provided a good picture of the current status of other systems that were not so easily transported. Many of us were extremely impressed by the performance of the systems demonstrated live at Hidden Valley. The promotional film *Advances in Spoken-*

Language Research 1970-1990: A DARPA Perspective was also screened to an approving audience. The CMU speech group hosted the attendees at an open house following the workshop which included demonstrations of their spoken language and speech recognition systems.

This volume includes the papers of the technical program, plus site reports from the various DARPA contractors, and a white paper by Cliff Weinstein on speech processing in military computer-based systems.

* * * * *

In recent years the DARPA speech and natural language workshops have evolved from a small informal discussion of current progress and plans to a much larger and more formal meeting which has become a primary forum for the exchange of major research results. The success of this series of workshops, and of the underlying DARPA-sponsored program of research in spoken language, has been confirmed by the recent widespread surge of interest in spoken-language systems outside the DARPA community, and by the extremely high standards of excellence of the papers presented at the meeting itself.

Many people were responsible for the success of this workshop. The inputs of the Organizing Committee, Lyn Bates, Lynette Hirschman, Mitch Marcus, Kathy McKeown, Doug Paul, Patti Price, Salim Roukos, and Victor Zue, were crucial not only in the selection of papers for the technical program, but also for many meta-level decisions which greatly affected the overall shape of the workshop. The advice of the two previous workshop chairs, Lynette Hirschman and Victor Zue, was extremely valuable. Dave Pallett's assistance in organizing the technical program is gratefully acknowledged, as well. I also thank Judy Bandola, Romina Fincher, and Shilpa Agarwal for their help in coordinating the flow of paper and information, John Hampshire and Christopher Young for their help in translating document formats, as well as Bob Brennan, Eric Thayer, and Bob Weide for their technical assistance at the meeting itself. Finally, Charles Wayne deserves major credit for his oversight of the spoken language program and for his vision of what the speech and natural language workshops could become.