Overview of the Fourth DARPA Speech and Natural Language Workshop

Patti Price, General Chair, Editor

SRI International, Menlo Park, CA 94025

The goals of the DARPA programs in Spoken Language and Written Language are advanced, and progress assessed by holding regular workshops. These workshops provide a mechanism for demonstrating innovative designs and strategies, and for documenting advances in the state of the art. More important, however, than providing a snapshot of the program, the workshops also provide a forum for the participants to share and develop ideas, to discuss technical issues, to outline long-term goals, to develop strategies for shared resources, and to develop standards for evaluation. The workshop is attended by a mixture of DARPA contractors, government representatives, and non-DARPA researchers in the US and abroad, representing universities and industry.

This, the fourth, workshop marked a striking advance toward practical applications by representing the first full-scale evaluation in the ATIS (Air Travel Information System) domain, with results that would not have been predicted a year or two ago. The most challenging aspect of the ATIS domain, for both speech recognition and natural language understanding, is the use of spontaneous speech in a normal office setting from participants engaged in a travel planning task. Spontaneous speech is a significant technical challenge because, compared to read speech or typed text, the focus of previous work, it tends to be more casual, more variable, and contains pauses, false starts and other phenomena not previously observed. Furthermore, the vocabulary is not specified. Thus, the task is more challenging, but also more realistic. The results, as summarized in Session 2, are extremely encouraging, and it can be argued that a paradigm shift is resulting from the marriage of the two component technologies.

The Fourth Workshop consisted of twelve technical sessions, including sessions on European speech and natural language research, benchmark reports, demos/videos, corpora and evaluation, machine translation, speech, natural language, and systems and prosody. Included in this volume are papers representing the presentations at the workshop, and one-page site reports for the various DARPA speech and natural language projects.

Barry Boehm, in his opening remarks, presented an overview of the DARPA/ISTO program, which consists of efforts in Artificial Intelligence (AI), High Performance Computing, Networks, Software, and combinations of the above. The Spoken and Written Language programs are part of the AI effort, which also includes Vision and Symbolic AI. In the language of the Total Quality approach to management, Boehm described the "product" of DARPA/ISTO as paradigm shifts with the goals of creating required capabilities and of providing a means to meet next-generation needs. According to Boehm, the most striking recent paradigm shift in the Spoken and Written Language programs has been the "pioneering application of scientific methods to AI", i.e., explicit measures and testing and revising of hypotheses under controlled conditions.

Charles Wayne, in his opening remarks, outlined various DARPA programs related to the Spoken Language and Written Language Programs, both of which are managed by Wayne, and which are the focus of this workshop. The related programs include: Barbara Yoon's Neural Net program, which includes a speech part; Tom Crystal's Tipster program; and the brand new Linguistic Data Consortium. Wayne then emphasized Boehm's comments about the importance of the scientific method and performance evaluation, and the major role these have played in encouraging and measuring progress. A chart of progress in speech recognition as a function of time is included in Wayne's site report in Section 13. The remainder of Wayne's remarks concerned the new Linguistics Data Consortium (LDC), which at this writing is still awaiting final approval. The LDC has the goal of distributing existing corpora and collecting new corpora to meet the data needs of the community. Current plans involve billions of words of text and thousands of hours of speech. A planning committee for this program is chaired by Mark Liberman, of the University of Pennsylvania, who should be contacted for further information.

The panelists in Session 1, "Speech and Natural Language Efforts in the US and Abroad," included 7 representatives from Europe. There is a significant speech and NL effort in Europe, and much of it is similar to the DARPA effort. There are also clear differences in focus, most notably a stronger focus on multi-lingual work in Europe compared to the US. See Section 1 for an overview of these programs.

As summarized by Dave Pallett in Section 2, the DARPA Resource Management and ATIS Benchmark Test Poster Session contained fourteen posters. It is the results presented in this session, and the papers included here, that document and encourage measurable progress, and which are the core of the scientific method referred to by Boehm and by Wayne.

Machine translation is a new direction in the DARPA program, and a session was devoted to this topic for the first time in this series of workshops. Since this effort has just begun, the papers in this session, compared to others at the workshop, tend to be weighted more toward approaches than to results. However, as described in Jaime Carbonell's introduction to Session 3, an important challenge to this machine translation program is the

development of "appropriate, task-sensitive and comprehensive evaluation criteria," so that results and progress can be measured.

Immediately before Session 3, two "extra-sessional" talks were presented: Yorick Wilks outlined a new program, the Consortium for Lexical Research, and George Miller proposed a technique for lexical disambiguation. These papers are included in the "Additional Papers" section.

Sessions 4, 8 and 9 were devoted to speech recognition topics. Session 4, chaired by Richard Lyon, considered field tests of telephone application compared to laboratory results, and front-end techniques (microphone arrays and representations of acoustic information). Session 8, chaired by Kai-Fu Lee, focussed on new techniques and recent advances in acoustic modeling, including neural nets, stochastic segment modeling, and methods for dealing with the variability found in very large vocabularies. Session 9, chaired by Francis Kubala, considered a broad range of speech recognition topics, including acoustic modeling, statistical language modeling, search techniques and adaptation of acoustic and language models to new data.

Sessions 5, 7, and 11 were devoted to natural language understanding topics. Session 5, chaired by James Allen, concentrated on parsing issues, including robustness, flexibility, efficiency and coverage. A key challenge, pointed out by the discussion in this session, is that of finding methods that combine the robustness of template-based approaches (which seems to be especially necessary with spontaneous speech) and the coverage of complex phenomena possible with more syntactically-based models. Session 7, chaired by Salim Roukos, focussed on methods that involve the use of probabilities in context-free grammars. This session represents an important trend, or, paradigm shift: more than half of the natural language papers at the workshop concerned some use of probabilistic models. This compares to about a third of the NL papers in the Second and Third Workshops, and to about 20% at the First Workshop. The growth in number corresponds to a growth in diversity of uses of probabilities in NL research and NL systems. Session 11, chaired by Mitch Marcus, also represented a paradigm shift: all the papers were corpus-based (as opposed to relying on the intuitions of experts), and most have at least one statistical subcomponent.

Session 6, chaired by Mari Ostendorf, included demonstrations and videotapes of speech and natural language technology. Demonstrations and videotapes are becoming increasingly important promotional tools for showing off the technology in a greater number of potential applications. Accompanying papers were optional for presenters in this session, because of the difficulty of translating the multi-media presentations to written form. The session summary describes the presentations; the people named in the summary can be contacted for further information.

Session 10, the Corpora and Evaluation Session, chaired by Cliff Weinstein, consisted of two parts: one focussed on issues in NL and SLS evaluation (including a summary of initial MUC-3 evaluations), and the other concerned Corpora and Performance Evaluation Committee reports and discussion. The surprising result was reported that a group of fourteen grammarians came to agreement on a "skeletal parse" very similar to those produced in the UPenn Treebank project. This represents an important milestone in the development of techniques for NL evaluation. The major issue in the second half of the session appeared to be the creation of a corpus for benchmarking large vocabulary speech recognition technology. The key cause for discussion related to similarities and differences between spontaneous and read speech. The discussion revealed a need for more data on the factors in a system that affect the speech style of users, how these styles differ from read speech, and how the differences may affect performance of systems trained on (cheaper-to-collect) read speech.

Ned Neuberg chaired Session 12, the last technical session of the workshop, which was devoted to topics that involve both speech and natural language: SLS systems issues, and prosody. As advances are made in component technologies, as various integration techniques evolve, as the cultural differences between the speech and natural language communities melt into a new culture, sessions such as this will likely dominate future workshops.

In his closing remarks Charles Wayne announced that the Fifth DARPA Workshop will be chaired by Mitch Marcus. It will take place at Arden House, in Harriman, New York, in February of 1992. Wayne then relayed remarks from Barry Boehm expressing pleasure in the workshop's lively interactions and our "healthy balance of cooperation and competition." Boehm was pleased with progress on relevant and practical problems, with quantitative results, and with the balance of domain knowledge and concern with generality (components of the application of the scientific method).

Thanks are due to the many people who contributed to the success of this workshop. The workshop committee played a major role in the technical direction and the logistics of the workshop, thanks to: James Allen, Lyn Bates, Kai-Fu Lee, Mitch Marcus, Mari Ostendorf, Dave Pallett, Fernando Pereira, and especially to the two previous chairs Richard Stern and Victor Zue for much useful help and advice. The session chairs (James Allen, Jaime Carbonell, Francis Kubala, Mark Liberman, Kai-Fu Lee, Dick Lyon, Mitch Marcus, Ned Neuberg, Mari Ostendorf, Dave Pallett, Salim Roukos, and Cliff Weinstein) are to be thanked for running the sessions, moderating the discussions and for provided valuable summaries and session introductions. I thank Chris Barker, Horacio Franco, Psi Mankoski, Louise Mason and Bev Harlan for their help in the organization and running of the workshop. Special thanks to Liz Shriberg for peerless assistance for many months preceding the workshop, and for her competent assistance in running the workshop. Special thanks also to Romina Fincher for her efficient handling of registration and the conference notebook, and to Inara Gravitis for her professional organization of the workshop proceedings. Thanks, too, to all participants for the lively discussion, and to Asilomar, to Photo and Sound for the audio/visual equipment, and to the weather. Finally, thanks go to Charles Wayne and to Barry Boehm for their direction of the program and for making possible the workshop and the research program behind it.

I also thank the participants who entered the cover design contest. Many interesting photos were submitted of people in various poses and states, and of beautiful landscapes at and near Asilomar. John Garafolo's photo of a craggy wind-bent cypress tree with truncated branches and restarts wins honorable mention, and is a nice metaphor for spontaneous speech. The winning cover design, however, is an original water color of the beach at Asilomar, by Chris Barker. I could say that it best represents the conference because it best captures the next paradigm shift: how to elegantly integrate the continuously varying (e.g., water, probabilities) with more discrete elements (e.g., rocks, words). But I have to admit that I chose it because I really liked it. Thanks, Chris.

4