

Overview of the Fifth DARPA Speech and Natural Language Workshop

Mitchell P. Marcus, General Chair, Editor

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104

1. THE DARPA SPEECH AND NL WORKSHOPS

This volume presents revised versions of papers presented at the Fifth DARPA Speech and Natural Language Workshop, held in February 1992 at the Arden House Conference Center in Harriman, NY. This series of workshops provides a forum where researchers can exchange information about very recent technical progress in an informal, highly interactive setting. Topics of interest include speech recognition, speech understanding, text understanding, and machine translation.

The majority of the participants receive funding under DARPA's Spoken Language or Written Language programs, which are the focus of the workshops. Other participants include researchers not funded by DARPA who voluntarily participate in these programs or in related benchmark evaluations; government researchers and consumers of these research results; and, on a rotating basis, selected visitors from both inside and outside the US. The participants in this ongoing series of workshops form a tightly-coupled research community in which results and research breakthroughs are typically evaluated, disseminated, and exploited with very short latency.

Aspects of these meetings include: technical presentations of both new research results and the ongoing evolutionary development of large software systems, often months before this research is reported elsewhere; the presentation of results of a variety of carefully standardized system evaluations; discussion of the future direction of the various DARPA programs in light of recent progress; and much discussion of ongoing work between individual researchers.

To facilitate the timeliness of the work presented at Arden House, abstracts were due less than two months before the workshop, with accepted papers due less than two weeks before the workshop. A looseleaf proceedings was distributed at the workshop itself, with these revised versions submitted a month after.

2. THE 1992 WORKSHOP

For a sense of the workshop as a whole, the reader is encouraged to first read the overview which precedes the collected papers from each of the sessions. These overviews not only summarize the key points in each set of papers, but also serve to give a perspective on the larger research arena in which this work is done.

Technical highlights of the 1992 Workshop include (a) the continuing steep slope of improvement of spoken language systems working within the Air Travel Information System (ATIS) domain, (b) the clear payoff of very large coordinated inter-site data collection efforts, (c) beginning steps towards very large vocabulary speech recognition systems, and (d) the continuing explosion of research results in statistical natural language processing.

Over the past five years, error rates for benchmark speech recognition tasks have dropped by a factor of two to three every year. This year was no exception, with the best system word error rates for ATIS demonstration systems down to about 6% on reserved test material taken from spontaneously recorded natural speech, some with disfluencies.

The centrality of data to the efforts of this research community was underscored by the efforts of the many sites of the Multi-Site Data Collection Working group (also known as MADCOW). This group was able to collect over 10,000 utterances with transcriptions within the ATIS domain as of the February workshop. The training and test sets for the ATIS benchmarks came from this corpus.

Going well beyond the limits of the ATIS task, this year's workshop saw preliminary results from the first of a new generation of very large vocabulary speech recognition systems, trained and tested on read material taken from the Wall Street Journal. As a first, preliminary effort, the vocabulary for this initial test was limited to 5000 words. Again, the data was collected by a coordinated effort involving multiple sites.

The paradigm shift in natural language processing towards empirical, corpus based methods was nowhere clearer than at this workshop. Eleven papers were presented in one marathon four-hour session on statistical language modelling and statistical natural language processing. Other sessions on the lexicon and on information retrieval were also dominated by papers in this area. The percentage of natural language papers in this new paradigm has increased from 20% in the first workshop to 30% in the second and third to 50% last year, to above 90% this year. This workshop also saw the first entire session in the workshop series devoted to research which is the beginning of a new fusion of techniques from both natural language processing and knowledge representation with classic information retrieval technologies.

Several important events at the workshop itself are either not reflected in these proceedings by either papers or represented only by session summaries. The workshop began and ended with comments from government representatives. Eric Mettala opened the workshop by presenting DARPA's view of the future role of Speech and NLP technologies, and a picture of future funding patterns within DARPA. The most telling observation: all future systems now under consideration by the U.S. Army have a requirement for speech recognition. The workshop concluded with a panel of government personnel, each commenting on the direction of the workshop. Chaired and organized by Charles Wayne, the speakers were YT Chien, NSF; John Prange, NSA; Susan Chipman, ONR; Tim Anderson, Armstrong Laboratory; and Jordan Cohen, IDA. A session chaired by Mark Liberman discussed whether the time has arrived for efforts to produce a common lexicon and/or grammar for spoken and written language systems. Both conservative and fairly ambitious proposals were presented, with fairly strong consensus that at least initial steps toward full scale efforts should be taken in the near future.

A very successful demo and video session, chaired by Victor Zue, presented 10 demos, almost all of which were presented live. The first row of the conference hall was filled with workstations, each of which performed in turn. The real time or near-real time response of many of the systems, combined with high understanding rates was most impressive.

Many of these systems were also available for one-on-one demos throughout the meeting. System performance has now increased to a level that this speaker was able to get nearly perfect performance from at least one system in essentially real time for careful speech.

3. ACKNOWLEDGEMENTS

This workshop was the result of the hard work of many people. The Workshop Planning Committee, Madeline Bates, Julia Hirschberg, Lynette Hirschman, David Pallett, Patti Price, Salim Roukos, Richard Stern and Clifford Weinstein and myself, was responsible for reviewing and selecting papers for presentation, as well as shaping other events in the program, and for providing crucial input on many policy issues. Patti Price, as previous general chair, gave much good advice. Victor Zue, as chair for the demonstration and video session, assisted by Mike Phillips, did a wonderful job of organizing and arranging technical facilities for that session, a very difficult task. Rich Stern, as chair of the Workshop Coordinating Committee, provided constant guidance and advice, and put in much effort to insure that the needs of all parts of the community were met.

Responsible for the smooth and flawless running of the workshop itself was Vicky Palay, who served as workshop coordinator, registrar, compiler of conference notebooks within a three day span. Vicky handled the thousand-and-one details of such a workshop with constant grace and flawless competence. In addition to Victor and Vicky, almost the entire MIT group (Phillips, Seneff, Glass, Goodine, Polifroni and Goddeau) helped with the transfer of two vans worth of equipment for the demos. David Goodine was primarily responsible for setting up the public use Mac/printer.

Finally, the role of Charles Wayne in shaping the direction both of the research behind this workshop and of the nature of this workshop itself deserves strong acknowledgement. This workshop, in the view of many participants, has become one of a handful of key forums for the presentation of first rank work in speech recognition, natural language processing, and perhaps the strongest setting for the presentation of work in statistical natural language processing.