

SESSION 1: SPEECH AND NATURAL LANGUAGE EFFORTS IN THE U. S. AND ABROAD

Mark Y. Liberman, University of Pennsylvania

Patti Price, SRI International

We see two purposes for this first session: increased communications among research communities in some danger of drifting apart, and a comparison of alternative goals and organizational structures for such communities. Obviously, a single hour-long session is no more than a symbolic gesture in this direction, even if the time had not been truncated further by schedule overruns pressing against an inflexible dinner hour, but we feel that the symbol was nevertheless a worthwhile and important one.

Recently, programs of research in speech and natural language have been increasing in number and size all around the world. At the same time, workshops like those sponsored by DARPA have become increasingly important, as research communities develop around the thrust of each funding agency's program. Inevitably, increasing cohesion within these communities raises the possibility of fragmentation among them, especially since the sheer number and complexity of new efforts make it hard to stay informed about everything.

We have the impression that many researchers in the DARPA community have an increasing number of blank spots for overseas research projects, even for major efforts like ESPRIT SUNDIAL, or EDRI in Japan, or the German ASL project. So far, there does not seem to be much divergence in the underlying technologies, except that the communications channel for technical details is narrower and slower across the oceans than it is within each of the three major communities. However, there is an increasing divergence in goals.

For instance, the European efforts see spoken dialogue systems as involving natural language generation and speech synthesis, as well as speech recognition and natural language understanding, while the DARPA community has generally seen the problem as "speech in, something else out," thus there is little American effort on generation, and less on speech synthesis. Another example is a lively European interest, mentioned by several of the panelists in this session, in "multimodal dialogue; systems," a fascinating concept whose future seems likely to be more concretely elaborated than its present. An important difference in focus is that all the European efforts are multilingual in essence and by necessity, while most American work is on English only.

In addition, the European, Japanese and American communities have developed different styles and approaches in organizing large research projects, at least in the speech and natural language area. For instance, the Europeans have emphasized cooperation among laboratories in developing common modules that fit together into a single overall system, in contrast to the main DARPA pattern of encouraging researchers to engage in a quan-

titatively-scored competition on a well-defined common task. The European projects also tend to stress university-industry cooperation in projects aimed at particular commercial applications, rather than pushing competitive technology development as motivated by military goals, with possible civilian commercialization left to market forces.

Overall, these divergences seem quite healthy. It would be a bad thing if all researchers around the world were working on exactly the same problems in exactly the same way for exactly the same reasons. However, it is possible for these differences in goals and modes of organization to create organizational and cultural barriers that make the transfer of ideas increasingly slower and more difficult across community boundaries. It is likely that a freer trade in ideas leads to faster technical advances and to ultimate benefits for everyone. We appreciate the participation of our European colleagues in this workshop, and we strongly encourage continued invitations to prominent European and Japanese researchers.

The panelists included: Louis Boves, Nijmegen University, Netherlands; Rolf Carlson, Royal Institute of Technology, Stockholm, Sweden; Jeremy Peckham, Logica Cambridge LTD, Cambridge, England; Keith Ponting, Royal Signals & Radar Establishment, Malvern, England; Christel Sorin, CNET, Lannion, France; Wolfgang Wahlster, German Research Center for AI, Saarbruecken, Germany; and Susan Warwick-Armstrong, University of Geneva, Switzerland. Unfortunately, Sadaoki Furui of NTT in Japan was unable to attend due to travel restrictions, and we were unable to benefit from an overview of the vast amount of speech and natural language research and development being undertaken in Japan. Due to time constraints, the planned discussion period did not occur; we apologize to the panelists. We greatly appreciate the participation in this workshop of all our foreign guests, and thank them for their participation in many informal discussions, which, though not documented in this proceedings, form an important component of the workshop.

Peckham described similarities and differences between the ESPRIT program and the DARPA program, concentrating on the Sundial project, of which he is the project director. He stressed several of the general points made above, while observing that the underlying technologies remain very close indeed. His paper in this volume provides an overview of the SUNDIAL project, which includes an interactive flight reservation application in French and in English. This application has notable similarities to ATIS, while providing a complementary approach (more focus

on dialogue, on the interactive system, and on expert, as opposed to naive, users).

Wolfgang Wahlster, of the DFKI in Saarbruecken, discussed speech and natural language research in Germany. All European countries have their own locally-funded research programs, in addition to EEC-wide efforts such as ESPRIT, but Germany has a particularly large amount of such work. For instance, the German Ministry for Research and Technology is putting 15 million DM per year into "Verbmobil", a project to develop speech-to-speech translation in the context of multimodal interaction. This is only one of several specific projects or basic research programs of three to six years each, which cover a range of topics including: syntax and semantics, multi-modal access to expert systems, bi-directional NL models (generation and understanding), models of uncooperative dialogue, prosody, and the integration of speech recognition and natural language understanding.

Susan Warwick-Armstrong, of ISSCO at the University of Geneva, discussed European efforts in machine translation, and especially the various Eurotra projects. She stressed the economic, cultural and political centrality in Europe of multi-lingual projects in general, and of translation in particular, and the level of on-going commitment to making progress in this area. Applications of the research, in addition to machine translation, include multilingual abstracting and indexing, document generation, computer aided instruction and training. A corresponding paper is included in this volume.

Finally, Christel Sorin, of CNET in Lannion, described speech research in France. She stressed the French interest in high-quality text-to-speech synthesis as an important component of speech technology, and also their interest in multimodal dialogue systems. She also pointed out that some years ago French researchers collaborated to produce a rather large speech database for research, but that little use has been made of it yet, compared to the crucial role of speech databases in common task definition for the DARPA community. She stressed that this is an area where they intend to follow the American example in the future, observing that they had been ahead in forming the idea of gather-

ing such data, but behind in making efficient use of the data once gathered.

In addition to the communications mentioned above, this volume contains additional related papers: a summary of the ESPRIT project Polyglot, by Louis Boves; an overview of research and development at KTH in Stockholm, by Rolf Carlson; and a general overview of ESPRIT.

As outlined in the paper by Boves, the ESPRIT project POLY-GLOT aims to develop multi-lingual speech-to-text and text-to-speech in a number of prototype applications, including dictation, office automation and teaching aids. The project includes components focussed on (1) speaker-adaptive isolated word recognition for very large vocabularies, (2) continuous speech recognition for mid-sized vocabularies (1000-5000 words), and (3) text-to-speech conversion

The paper by Rolf Carlson describes the research and development effort at the Department of Speech Communication and Music Acoustics of the Royal Institute of Technology, focussing on the speech effort. These activities include the study of individual speaker characteristics, speaking styles, text-to-speech synthesis, knowledge-based recognition for large vocabularies, voice source characteristics (including modeling of speech production for recognition), and artificial neural networks for speech recognition. Applications efforts have been aimed at air traffic control, the process industry, devices for the disabled, and mobile telephony.

With the help of Patrick Van Hove, of the ESPRIT program, we have compiled a complete list of the various ESPRIT projects, including contact points for each. Surveying the list should make it clear that there is a good deal of activity in Europe that is very similar in aims and in methods to the DARPA program. It is also clear that there are differences in focus, most notably a stronger focus on multi-lingual work in Europe compared to the US. We hope this compilation will serve as a reference for US researchers and as a means for initiating future trans-Atlantic communication. We hope that European and Japanese researchers will continue to be invited to these workshops, and that more American researchers will be invited to comparable gatherings overseas.