SESSION 12: SLS AND PROSODY

Edward P. Neuburg

IDA
Center for Communications Research
Thanet Road
Princeton, N.J. 08540

This last session of the Meeting was really two sub-sessions on two quite different topics. The session opened with two thought-provoking papers from MIT on what might be called a philosophy of building a spoken language system.

The first paper, by Seneff, Hirschman and Zue, concerned the ATIS domain. The authors state what seems, after it is pointed out, an obvious truth, but which too few system builders seem to understand: when attempting a project like ATIS, the time to worry about the kinds of interaction that will take place in a dialogue is at the very beginning of the project. The MIT ATIS system has not yet been fully designed, but the paper discusses, and gives some examples of, the kinds information exchange it is going to have to handle.

The second paper, by Hirschman, Seneff, Goodine and Phillips, was of the same flavor, but in a more practical vein. It describes some actual Natural Language improvements to the MIT Voyager system, having to do with merging of acoustic and grammatical evidence during the tree-search part of the recognition algorithm. Some simple (new) ideas have already provided a 33% improvement in recognition score.

In the discussion period, it was clear that questioners had no quarrel with the aims expressed by the talkers; all questions had to do with the architecture, the actual implementation, of the algorithms discussed in the papers.

The final three papers were on prosody. Suggestions that prosody might be used in automatic speech understanding go back at least to the ARPA SUR project of the early 1970s; however no recognition system has actually ever used prosody. In the early days, when compute-time was a primary issue, the notion was that prosodic information could be used to order competing theories, and thus speed the search. Today, with computational resources faster and cheaper, the emphasis is on use of prosody for disambiguation.

The papers in this sub-session reported work that is very early in the process of folding prosody into the recognition process; in fact, all three might be described as feasibility tests. None reported work in which prosodic information is automatically extracted from an unknown incoming utterance.

The first paper, by Price, Ostendorf, Shattuck-Hufnagel and Fong, and the second by Wang and Hirschberg, are somewhat in the nature of thought experiments. Price et al. are interested in whether or not prosody actually can be used to disambiguate; they show that human listeners can indeed use it to some extent. Wang and Hirschberg show that in the ATIS domain, which is syntactically simple, it is possible to predict fairly well, given the text of a sentence, where intonation boundaries will occur.

The third paper, by Wightman, Veilleux and Ostendorf, describes what comes closest to a practical experiment. Here, given incoming sentences where the words are known in advance, their algorithm measures certain phoneme durations, and uses them in a simple disambiguation task, with very encouraging results.

Questions focused on three topics. The first was the statistical gathering of prosodic evidence. There was general agreement that there is not now any set of statistics on prosody from a large corpus. Further, since prosodic units are much longer than phonemes, for example, it will take a lot of text to get reliable estimates. It was thought that some of the corpora now being collected will be large enough. It is perhaps of interest that there was no discussion of just what measurements we need to find the distribution of.

There was a question about whether 90% accuracy in prosody (reported in one of the papers) was good or bad. The fact that the question was asked, and that in the discussion there was no solid opinion on either side, is very revealing of our state of knowledge of prosody and its usefulness in automatic understanding of speech.

The third topic was the general usefulness of prosody. Opinions expressed were that prosody is not just good for speeding up search; that when prosody can be successfully extracted from speech it will be a useful addition to the probabilistic recognition framework; and that in fact there are many situations in which prosody will be the only way to get at the true meaning of a spoken sentence.