SESSION 1: SPOKEN LANGUAGE SYSTEMS

Alexander I. Rudnicky

School of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213 USA

By themselves, speech recognition and natural language processing have limited applications. the reason is that each only accomplishes a part of what humans are capable of when they use language. Spoken language systems represent the merger of these two technologies and provide an integrated functionality that more closely approximates humans capabilities in speech communication.

The addition of natural language processing enhances speech recognition by allowing people to speak naturally without the need to remember specific command words or to keep within a specific grammar. In principle, as long as we are able to describe something verbally, a spoken language system should be able to understand us.

Without the ability to interpret natural language, speech recognition is suited only for a subset of tasks (though certainly not trivial ones), such as data entry, simple commands or dictation. Similarly, without speech recognition natural language is restricted to the interpretation of written language, a stylized form of human communication. Spoken language systems thus represent an attempt to automate speech communication. While limited in terms of the target behavior, they still represents an advance over the capabilities of the individual technologies.

The development of spoken language systems has many facets. Certainly it pushes the development of the basic technologies, speech recognition and natural language processing, since successful understanding ultimately depends on high quality processing at these levels. At the same time, spoken language generates the need for advances in other domains.

The requirements of fluent communication require engineering of algorithms to produce real-time understanding of speech; since human speech is a highly interactive medium real-time response is necessary for fluency. It also raises the need for incorporating additional aspects of analysis into the process. Prosody is a case in point. The development of spoken language systems also requires the implementation of running systems that can be used to perform non-trivial tasks. This in turn creates the need for studying the usability of spoken language systems, since some issues critical to performance reveal themselves only in the course of actual use. In turn these issues impact the development of processing strategies at the level of the individual technologies. The use of spoken language systems under more realistic conditions also serves as a stimulus for the study of robustness, a system's ability to handle variations in environment, microphone characteristics as well as the vagaries of human speech.

The development of spoken language systems requires support activities, such as the development of training and test corpora and the development of evaluation techniques. The spoken language community has developed a large corpus of speech data that attempts to approximate the speech that would occur under real conditions. t the same time, elaborate evaluation techniques have been developed that allow us to compare and diagnose the characteristics of spoken language systems.

The papers in this session at first glance appear to fall into two groups. One group concerns itself with evaluation and data collection while the other concerns itself with language processing techniques. In fact the two sets of papers share an important theme in common, that of *real data*. In reading these papers one is struck by the extent to which the use of real data has shaped what we do and how we do it.

The paper by Pallett *et al* describes the evaluation procedures currently in use in the spoken language technology program. In comparing the current evaluation procedures with those in use a few years ago, one is struck by the extent to which the program has progressed from the use of carefully controlled data to the use of more natural (and more difficult) speech. While the continuous speech recognition evaluation uses read speech for its main evaluations, a new "stress test" has been introduced which exposes systems to a variety of unpredictable material, a condition that begins to approximate what recognition system might actually be exposed to under realistic conditions. The spoken

language evaluations have seen a similar progression from single sentences to edited scenarios to an attempt to use complete scenarios for evaluation.

Data collection, as described in the paper by Hirshman and the MADCOW committee has seen a similar progression, from read sentences generated from an artificial grammar, to collection through the use of wizard systems, to the use of real systems to collect data fro both training and testing. The paper by Thompson and Bard represents perhaps the logical conclusion of this process, the collection of speech from natural human-human interactions. While the Edinburgh corpus is meant for analysis rather than for speech system development, it nevertheless represents the kind of data that spoken language systems will ultimately be asked to process. In the discussion of this paper it was pointed out that human-computer communication might turn out to be quite different from human-human communication and that things learned from this corpus might not be transferrable to that situation. Some interest was also expressed in the phenomenon of overlap and its role in communication.

The three following papers, from Paramax, BBN and SRI describe current improvements to the natural language components of spoken language systems. All three papers attempt to deal with the problem of how to adapt a syntactic-based parser to the realities of language as spoken by humans and further transcribed (perhaps erroneously) by speech recognition systems. The paper by Linebarger *et al* describes a robustness heuristic (based on the ability to skip non-keywords) that allows the Paramax parser to interpret otherwise unprocessed inputs. The paper by Stallard and Bobrow also addresses the problem of salvaging otherwise unparsable inputs by the use of semantic structure when the use of syntactic structure fails to produce an interpretation. It was pointed out in the discussion that the semantic post-processing might not be portable. The paper by Dowding et al presents a parsing strategy that uses the mutual constraints of syntax an semantics to generate parses, together with heuristics that allow the system to produce an interpretation even if no satisfactory initial parse is found. Interestingly, for the ATIS task, none of the syntax based parsers can currently outperform a frame-based parser.

One of the key arguments that have been made in favor of syntax-based parsing is that the knowledge gained in one domain will transfer to other domains and will save the work of having to build a parser for the new domain. The paper by Linebarger *et al* describes how the Paramax robustness heuristics can be easily ported between domains. Portability can refer not only to transfer between domains but also to transfer between languages, a potentially more difficult task. The paper by Glass *et al* describes experiences in porting the MIT Voyager system from English to Japanese. While this paper is a good example of how a syntactic parser can be successfully ported to a new domain, it is nevertheless significant that the parsing strategy had to be altered in order to accommodate the structure of the Japanese language. In the discussion, it was pointed out that discourse-level processing might not, in principle, be portable, though the elementary processing need for the Voyager domain turned out to be portable.