

Session 11: Natural Language II

Deborah A. Dahl, Chair
Unisys Defense Systems
Center for Advanced Information Technology
Paoli, PA 19301

Three of the four papers in this session described work aimed at reducing the amount of labor-intensive "hand-crafting" of natural language systems. The fourth paper described the current need for machine translation and machine-assisted translation and advocated an interlingual approach involving a coordinated effort by several sites.

It is extremely encouraging to see that new ideas are being explored for reducing the effort required to develop new applications of natural language understanding. These ideas will clearly be needed to enable natural language technology to be applied to real problems. The first three papers described three different approaches to this problem – using statistical automatic training techniques (BBN), developing "generic" text processing capabilities (GE), and the use of analogical reasoning to hypothesize new word senses (Berkeley). The work described in these papers was generally exploratory in nature. It will be exciting to see how these approaches continue to develop and become integrated into full-scale systems.

The first paper, by Ayuso, et. al. from BBN, described three experiments in statistical approaches – part of speech tagging, probabilistic parsing, and acquisition of lexical syntax. The most intriguing aspect of this work was the potential for synergy among the various tools described. For example, one suggestion was that the probabilistic parsing could be used to control the ambiguity inherent in dealing with unknown words.

The second paper, by Jacobs, et. al., from GE, described an approach to handling large amounts of unrestricted text which involves developing generic text processing

capabilities. This paper reports on some of the tools which underlie this approach, including the development of a 10,000 word lexicon and various text preprocessing tools. These tools are used to produce a text tagged with word senses. This result is interesting because it shows that it is possible to produce at least part of a semantic analysis (word sense tagging) for arbitrary text.

Robert Wilensky's paper on lexical acquisition was the third paper in this session. This work uses lexical sub-regularities to extend the meaning of words used in new senses. This paper lists a number of lexical subregularities and outlines a preliminary procedure for lexical acquisition. This work is in its early stages and it should be very interesting to see how it develops. This capability should be quite useful for making natural language systems more independent of their developers.

The final paper in the session (Wilks, et. al.) was a proposal for a new effort in machine translation, based on an interlingua approach. It makes a strong case for the need for such work, since currently available machine translation systems are based on very old technology, although the need for translation assistance is so great that they are still being used.

Several interesting points were made in the discussion of the Wilks paper. One criticism was made that the proposal offered no striking new ideas. The response to this was that the project would pull together existing skills. In addition, it should also be noted that there are many current ideas in areas such as parsing that have not yet been exploited in machine translation systems. Other discussions concerned the relative merits of transfer vs. interlingua approaches, and statistical vs. knowledge-based techniques.