Summary of Invited Speech

Mitch Marcus
Chairman
Department of Computer and Information Science
University of Pennsylvania
mitch@cis.upenn.edu

PART I

Over the past several years, there has been very significant and continuing progress in the development of accurate parsers for unconstrained text; much of this breakthrough has depended crucially on the use of statistical methods which estimate model parameters from a "tree bank" of hand-parsed sentences. This use of statistical models was clearly inspired by the use of statistical methods for speech recognition, where self-organizing systems based on statistics are now beginning to achieve commercial success, and greatly outperform systems which attempt to explicitly encode linguistic knowledge. Many of the techniques used in statistical parsing derive rather directly from methods used for speech recognition; this is particularly true of methods for dealing with sparse data.

This progress in parsing accuracy continues; within the last month, several new parsers have been reported, two developed within the speaker's group at the University of Pennsylvania, which show distinct improvement over the best parsing results of even a year ago. Most surprisingly, each of of these parsers incorporates a very different model, yet they perform similarly.

This talk will argue that this progress in parsing, unlike the earlier progress in speech recognition, depends crucially on the combination of explicit linguistic representation and statistical estimation. I will claim that the reason these recent systems perform so similarly is that they explicitly encode very similar levels of linguistic representation. Furthermore, even in the use of smoothing techniques to deal with sparse data, paying close attention to underlying representational issues is often crucial; systems that utilize designer-encoded linguistic knowledge about representations currently significantly outperform representation-poor "self-organizing" smoothing methods.

PART II

Recently, the executive committe of the Association for Computational Linguistics (ACL) reached a new vision of the structure and functioning of this organization; this talk will outline both the current activities of the ACL and this new vision. As part of this new vision, the ACL now seeks a structure that both creates a truly international ACL and that reflects the fact that there are strong regional interests and special circumstances. To implement this in part, the ACL both will adopt a schedule of future annual meetings and applied natural language processing conferences that gives much wider geographic distribution to these meetings and will propose to its membership changes in the composition of the executive committee itself which better reflect the geographical distribution of activity in computational linguistics worldwide.