

# CENTER FOR COGNITIVE SCIENCE

University at Buffalo, State University of New York

**Wednesday, January 23, 2002**

280 Park Hall  
North Campus  
2:00 pm –3:30 pm

David Pierce, Ph.D.  
Department of Computer Science and Engineering  
University at Buffalo

## **Machine Learning Strategies for Corpus-Based Natural Language Processing**

Corpus-based natural language processing refers to the use of techniques from machine learning for training systems to understand natural language. These techniques generally require annotated training data as input. For example, building a parser requires pre-parsed sentences as input. This talk will consider strategies at a "meta-learning" level for using training data more efficiently and effectively. One such strategy, called active learning, tries to select training instances based on their predicted utility. Additionally, I will describe experiments demonstrating the effectiveness of strategies such as active learning for a simple natural language learning task, namely base noun phrase identification.

**Refreshments will be available  
Everyone is welcome to attend!**

For information please call the Cognitive Science Office at (716) 645-3794 or check  
<http://wings.buffalo.edu/cogsci/html/2002spring.htm>