

# Technical Program Schedule

## First Day: Thursday, July 22, 2004

8:30-9:00 Opening Session  
9:00-10:00 *Invited Talk 1: Jack Mostow*

10:00-10:30 Break

### Session 1: Multimodal and Situated Language Processing

10:30-10:55 *Optimization in Multimodal Interpretation*  
Joyce Y. Chai, Pengyu Hong, Michelle X. Xhou and Zahar Prasov

10:55-11:20 *Constructivist Development of Grounded Construction Grammar*  
Luc Steels

11:20-11:45 *A Distributional Model of Semantic Context Effects in Lexical Processing*  
Scott McDonald and Chris Brew

11:45-12:10 *Analysis of Mixed Natural and Symbolic Input in Mathematical Dialogs*  
Magdalena Wolska and Ivana Kruijff-KorBayová

12:10-13:40 Lunch

### Session 2: Spoken Language Processing

13:40-14:05 *A TAG-based noisy-channel model of speech repairs*  
Mark Johnson and Eugene Charniak

14:05-14:30 *Attention Shifting for Parsing Speech*  
Keith B. Hall and Mark Johnson

14:30-14:55 *Discriminative Language Modeling with Conditional Random Fields and the Perceptron Algorithm*  
Brian Roark, Murat Saraclar, Michael Collins and Mark Johnson

14:55-15:20 *Statistical Modeling for Unit Selection in Speech Synthesis*  
Mehryar Mohri, Cyril Allauzen and Michael Riley

15:20-15:50 Coffee Break

### Session 3: Dialogue Systems

15:50-16:15 *Developing a Flexible Spoken Dialog System Using Simulation*  
Grace Chung

16:15-16:40 *Data-Driven Strategies for an Automated Dialogue System*  
Hilda Hardy, Tomek Strzalkowski, Min Wu, Cristian Ursu, Nick Webb, Alan Biermann, R. Bryce Inouye and Ashley McKenzie

16:40-17:05 *Trainable Sentence Planning for Complex Information Presentations in Spoken Dialog Systems*  
Amanda Stent, Rashmi Prasad and Marilyn Walker

17:05-17:30 *User Expertise Modeling and Adaptivity in a Speech-Based E-Mail System*  
Kristiina Jokinen and Kari Kanto

#### **Session 4: Statistical Parsing**

- 10:30-10:55 *Discriminative Training of a Neural Network Statistical Parser*  
James Henderson
- 10:55-11:20 *Parsing the WSJ Using CCG and Log-Linear Models*  
Stephen Clark and James R. Curran
- 11:20-11:45 *Incremental Parsing with the Perceptron Algorithm*  
Michael Collins and Brian Roark
- 11:45-12:10 *Convolution Kernels with Feature Selection for Natural Language Processing Tasks*  
Jun Suzuki, Hideki Isozaki and Eisaku Maeda

#### **Session 5: Coreference and Anaphora**

- 13:40-14:05 *Improving Pronoun Resolution by Incorporating Coreferential Information of Candidates*  
Xiaofeng Yang, Jian Su, Guodong Zhou and Chew-Lim Tan
- 14:05-14:30 *A Mention-Synchronous Coreference Resolution Algorithm Based On the Bell Tree*  
Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla and Salim Roukos
- 14:30-14:55 *Learning to Resolve Bridging References*  
Massimo Poesio, Rahul Mehta, Axel Maroudas and Janet Hitzeman
- 14:55-15:20 *Learning Noun Phrase Anaphoricity to Improve Conference Resolution: Issues in Representation and Optimization*  
Vincent Ng

#### **Session 6: Machine Translation**

- 15:50-16:15 *A Joint Source-Channel Model for Machine Transliteration*  
Haizhou Li, Min Zhang and Jian Su
- 16:15-16:40 *Collocation Translation Acquisition Using Monolingual Corpora*  
Yajuan Lü and Ming Zhou
- 16:40-17:05 *Statistical Machine Translation with Word- and Sentence-Aligned Parallel Corpora*  
Chris Callison-Burch, David Talbot and Miles Osborne
- 17:05-17:30 *Finding Ideographic Representations of Japanese Names Written in Latin Script via Language Identification and Corpus Validation*  
Yan Qu and Gregory Grefenstette

#### **Session 7: Demonstrations & Interactive Posters: Short Presentations**

10:30-12:10

#### **Session 8: Information Extraction**

- 13:40-14:05 *Extracting Regulatory Gene Expression Networks From Pubmed*  
Jasmin Saric, Lars J. Jensen, Peer Bork, Rossitza Ouzounova and Isabel Rojas
- 14:05-14:30 *Linguistic Profiling for Authorship Recognition and Verification*  
Hans van Halteren
- 14:30-14:55 *An Empirical Study of Information Synthesis Task*  
Enrique Amigo, Julio Gonzalo, Victor Peinado, Anselmo Peñas and Felisa Verdejo
- 14:55-15:20 *Mining Metalinguistic Activity in Corpora to Create Lexical Resources Using Information Extraction Techniques: the MOP System*  
Carlos Rodriguez Penagos

### **Session 9: Syntax/Semantics/Parsing**

- 15:50-16:15 *Optimizing Typed Feature Structure Grammar Parsing through Non-Statistical Indexing*  
Cosmin Munteanu and Gerald Penn
- 16:15-16:40 *Head-Driven Parsing for Word Lattices*  
Christopher Collins, Bob Carpenter and Gerald Penn
- 16:40-17:05 *Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying*  
Gerald Penn
- 17:05-17:30 *Minimal Recursion Semantics as Dominance Constraints: Translation, Evaluation, and Analysis*  
Ruth Fuchss, Alexander Koller, Joachim Niehren and Stefan Thater

### **Second Day: Friday, July 23, 2004**

#### **Session 10: Text Categorization**

- 8:45-9:10 *Learning with Unlabeled Data for Text Categorization Using a Bootstrapping and a Feature Projection Technique*  
Youngjoong Ko and Jungyun Seo
- 9:10-9:35 *The Sentimental Factor: Improving Review Classification Via Human-Provided Information*  
Philip Beineke, Trevor Hastie and Shivakumar Vaithyanathan
- 9:35-10:00 *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*  
Bo Pang and Lillian Lee
- 10:00-10:30 Coffee Break

#### **Session 11: Word Sense Disambiguation**

- 10:30-10:55 *Finding Predominant Word Senses in Untagged Text*  
Diana McCarthy, Rob Koeling, Julie Weeds and John Carroll
- 10:55-11:20 *Unsupervised Sense Disambiguation Using Bilingual Probabilistic Models*  
Indrajit Bhattacharya, Lise Getoor and Yoshua Bengio
- 11:20-11:45 *Chinese Verb Sense Discrimination Using an EM Clustering Model with Rich Linguistic Features*  
Jinying Chen and Martha Palmer
- 11:45-12:10 *Relieving the data Acquisition Bottleneck in Word Sense Disambiguation*  
Mona Diab
- 12:10-13:40 Lunch

### **Session 12: Syntax/Semantics/Parsing**

- 13:40-14:05 *Enriching the Output of a Parser Using Memory-based Learning*  
Valentin Jijkoun and Maarten de Rijke
- 14:05-14:30 *Long-Distance Dependency Resolution in Automatically Acquired Wide-Coverage PCFG-Based LFG Approximations*  
Aoife Cahill, Michael Burke, Ruth O'Donovan, Josef Van Genabith and Andy Way
- 14:30-14:55 *Deep Dependencies from Context-Free Statistical Parsers: Correcting the Surface Dependency Approximation*  
Roger Levy and Christopher Manning
- 14:55-15:20 *A Study on Convolution Kernels for Shallow Statistic Parsing*  
Alessandro Moschitti
- 15:20-15:50 Coffee Break

### **Session 13: Dialogue Systems**

- 15:50-16:15 *Combining Acoustic and Pragmatic Features to Predict Recognition Performance in Spoken Dialogue Systems*  
Malte Gabsdil and Oliver Lemon
- 16:15-16:40 *Predicting Student Emotions in Computer-Human Tutoring Dialogues*  
Diane J. Litman and Kate Forbes-Riley
- 16:40 - 17:40 Lifetime Achievement Award

### **Session 14: Lexical Semantics**

- 8:45-9:10 *Building Verb Predicates: A Computational View*  
Fernando Gomez
- 9:10-9:35 *Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II Treebank*  
Ruth O'Donovan, Michael Burke, Aoife Cahill, Josef Van Genabith and Andy Way
- 9:35-10:00 *Inducing Frame Semantic Verb Classes from WordNet and LDOCE*  
Rebecca Green, Bonnie J. Dorr and Philip Resnik

### **Session 15: Generation**

- 10:30-10:55 *Paragraph-, Word-, and Coherence-based Approaches to Sentence Ranking: A Comparison of Algorithm and Human Performance*  
Florian Wolf and Edward Gibson
- 10:55-11:20 *Evaluating Centering-Based Metrics of Coherence*  
Nikiforos Karamanis, Massimo Poesio, Chris Mellish and Jon Oberlander
- 11:20-11:45 *Computing Locally Coherent Discourses*  
Ernst Althaus, Nikiforos Karamanis and Alexander Koller
- 11:45-12:10 *Generating Referring Expressions in Open Domains*  
Advaith Siddharthan and Ann Copestake

### **Session 16: Information Extraction/Information Retrieval**

- 13:40-14:05 *Discovering Relations among Named Entities from Large Corpora*  
Takaaki Hasegawa, Satoshi Sekine and Ralph Grishman
- 14:05-14:30 *Dependency Tree Kernels for Relation Extraction*  
Aron Culotta and Jeffrey Sorensen
- 14:30-14:55 *Classifying Semantic Relations in Bioscience Texts*  
Barbara Rosario and Marti Hearst
- 14:55-15:20 *Collective Information Extraction with Relational Markov Networks*  
Razvan Bunescu and Raymond Mooney

### **Session 17: Syntax/Semantics/Parsing (Grammar Construction)**

- 15:50-16:15 *Error Mining for Wide-Coverage Grammar Engineering*  
Gertjan van Noord
- 16:15-16:40 *Alternative approaches for Generating Bodies of Grammar Rules*  
Gabriel Infante-Lopez and Maarten de Rijke

### **Session 18: Demonstrations & Interactive Posters: Short Presentations**

8:45-10:00

### **Session 19: Statistical Parsing**

- 10:30-10:55 *Adaptive Chinese Word Segmentation*  
Jianfeng Gao, Andi Wu, Cheng-Ning Huang, Hong qiao Li, Xinsong Xia and Hauwei Qin
- 10:55-11:20 *Experiments in parallel-text based grammar induction*  
Jonas Kuhn
- 11:20-11:45 *Corpus-Based Induction of Syntactic Structure: Models of Dependency and Constituency*  
Dan Klein and Christopher Manning
- 11:45-12:10 *Annealing Techniques For Unsupervised Statistical Language Learning*  
Noah A. Smith and Jason Eisner

### **Session 20: Machine Translation**

- 13:40-14:05 *Multi-Engine Machine Translation with Voted Language Model*  
Tadashi Nomoto
- 14:05-14:30 *Aligning words using matrix factorisation*  
Cyril Goutte, Kenji Yamada and Eric Gaussier
- 14:30-14:55 *FSA: An Efficient and Flexible C++ Toolkit for Finite State Automata Using On-Demand Computation*  
Stephan Kanthak and Hermann Ney
- 14:55-15:20 *Improving IBM Word Alignment Model 1*  
Robert C. Moore

### **Session 21: Multilingual Lexicons**

- 15:50-16:15 *A Geometric View on Bilingual Lexicon Extraction from Comparable Corpora*  
Eric Gaussier, J.M. Renders, I. Matveeva, C. Goutte and H. Dejean
- 16:15-16:40 *Creating Multilingual Translation Lexicons with Regional Variations Using Web Corpora*  
Pu-Jen Cheng, Wen-Hsiang Lu, Jei-Wen Teng and Lee-Feng Chien

**Third Day: Saturday, July 24, 2004**

**Session 22: Linguistic and Mathematical Models of Language**

- 9:05-9:30 *Probabilistic Parsing Strategies*  
Mark-Jan Nederhof and Giorgio Satta
- 9:30-9:55 *An Alternative Method of Training Probabilistic LR Parsers*  
Mark-Jan Nederhof and Giorgio Satta
- 9:55-10:20 *Wrapping of Trees*  
James Rogers
- 10:20-11:00 Coffee Break  
11:00-12:00 Business Meeting  
12:00-13:30 Lunch  
13:30-14:30 Invited Talk 2: Anne Cutler

**Session 23: Question Answering**

- 14:30-14:55 *Splitting Complex Temporal Questions for Question Answering Systems*  
E. Saquete, P. Martínez-Barco, R. Muñoz and J.L. Vicedo
- 14:55-15:20 *Question Answering Using Constraint Satisfaction: QA-By-Dossier-With-Constraints*  
John Prager, Jennifer Chu-Carroll and Krzysztof Czuba

**Session 24: Information Retrieval/Information Extraction**

- 15:50-16:15 *Applying Machine Learning to Chinese Temporal Relation Resolution*  
Wenjie Li, Kam-Fai Wong, Guihong Cao and Chunfa Yuan
- 16:15-16:40 *Multi-Criteria-based Active Learning for Named Entity Recognition*  
Dan Shen, Jie Zhang, Jian Su, Guodong Zhou and Chew-Lim Tan
- 16:40-17:05 *Weakly Supervised Learning for Cross-document Person Name Disambiguation Supported by Information Extraction*  
Cheng Niu, Wei Li and Rohini K. Srihari
- 17:05-17:30 Best Paper Award and Closing

### **Session 25: Evaluation of Machine Translation**

- 9:05-9:30 *Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics*  
Chin-Yew Lin and Franz Josef Och
- 9:30-9:55 *A Unified Framework For Automatic Evaluation Using 4-Gram Co-occurrence Statistics*  
Radu Soricut and Eric Brill
- 9:55-10:20 *Extending the BLEU MT Evaluation Method with Frequency Weightings*  
Bogdan Babych and Tony Hartley

### **Session 26: Word Sense Disambiguation**

- 14:30-14:55 *Learning Word Sense With Feature Selection and Order Identification Capabilities*  
Zheng-Yu Niu, Dong-Hong Ji and Chew-Lim Tan
- 14:55-15:20 *A Kernel PCA Method for Superior Word Sense Disambiguation*  
Dekai Wu, Weifeng Su and Marine Carpuat

### **Session 27: Syntax/Semantics/Parsing**

- 15:50-16:15 *Using Linguistic Principles to Recover Empty Categories*  
Richard Campbell
- 16:15-16:40 *Statistical Machine Translation by Parsing*  
I. Dan Melamed
- 16:40-17:05 *Generalized Multitext Grammars*  
I. Dan Melamed, Giorgio Satta and Benjamin Wellington

### **Session 28: Demonstrations & Interactive Posters: Short Presentations**

9:05-10:20

### **Session 29: Conversational Spoken Language Processing**

- 14:30-14:55 *Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies*  
Michel Galley, Kathleen McKeown, Julia Hirschberg and Elizabeth Shriberg
- 14:55-15:20 *Using Conditional Random Fields to Predict Pitch Accents in Conversational Speech*  
Michelle Gregory and Yasemin Altun

### **Session 30: Discourse and Dialogue**

- 15:50-16:15 *Acquiring the Meaning of Discourse Markers*  
Ben Hutchinson
- 16:15-16:40 *FLSA: Extending Latent Semantic Analysis with Features for Dialogue Act Classification*  
Riccardo Serafin and Barbara Di Eugenio