

ACL-04

**42nd
Annual Meeting
of the
Association for
Computational Linguistics**

Proceedings of the Conference

21–26 July , 2004
Barcelona, Spain

Preface to the Student Research Workshop Proceedings

We are pleased to introduce the papers accepted for presentation at the Student Research Workshop of the *42nd Annual Meeting of the Association for Computational Linguistics* in Barcelona, Spain. We received 43 papers – a near-record number – of which we were able to accept 12. We thank all students who submitted and we hope the reviews provided useful feedback and directions for future research both for those presenting and for those whose papers we were not able to accept.

Many people contributed to the organization of the ACL-04 Student Research Workshop. First of all, we thank the many reviewers for the time they spent reading and reviewing the papers. A special thank you to those reviewers who helped us out with extra reviews, often on short notice, in the face of a number of submissions that surpassed our expectations.

The members of the ACL-04 Student Research Workshop Program Committee were:

Laura Alonso, *Universidad de Barcelona, Spain*
Jordi Atserias i Batalla, *Universidad Politecnica de Catalunya, Spain*
Amit Bagga, *Avaya Labs Research, USA*
John Beavers, *Stanford University, USA*
Francis Bond, *NTT, Japan*
Christina Bosco, *Turin University, Italy*
Gosse Bouma, *Rijksuniversiteit Groningen, Netherlands*
Thorsten Brants, *Google Inc, USA*
Eric de la Clergerie, *INRIA, France*
Eric Fosler-Lussier, *Ohio State University, USA*
Heidi Fox, *Brown University, USA*
Oren Glickman, *Bar Ilan University, Israel*
Michelle Gregory, *University of Buffalo, USA*
Jon Herring, *University of Brighton, UK*
Chu-Ren Huang, *Academia Sinica, Taiwan*
Alistair Knott, *University of Otago, New Zealand*
Milen Kouylekov, *ITC-irst, Trento, Italy*
Ashwani Kumar, *MIT, USA*
Alberto Lavelli, *ITC-irst, Trento, Italy*

Rob Malouf, *San Diego State University, USA*
Erwin Marsi, *University of Tilburg, Netherlands*
Fernando Martínez Santiago, *Universidad de Jaén, Spain*
Joakim Nivre, *Växjö University, Sweden*
Jahna Otterbacher, *University of Michigan, USA*
Sebastian Padó, *Saarland University, Germany*
Bill Raymond, *Ohio State University, USA*
Hannah Rohde, *University of California–San Diego, USA*
Anoop Sarkar, *Simon Fraser University, Canada*
Jennifer Spenader, *Stockholm University, Sweden*
Paul Tepper, *Northwestern University, USA*
Kristina Toutanova, *Stanford University, USA*
Olga Uryupina, *Saarland University, Germany*
Gertjan van Noord, *Rijksuniversiteit Groningen, Netherlands*
Menno van Zaanen, *University of Tilburg, Netherlands*
Begoña Villada, *Rijksuniversiteit Groningen, Netherlands*
Stephen Wan, *Macquarie University, Australia*
Mirjam Wester, *University of Edinburgh, UK*
Maria Wolters, *Rhetorical Systems, UK*
Milena Yankova, *Bulgaria Academy of Sciences*

Twenty researchers have agreed to serve as respondents to the papers presented during the Student Research Workshop. We are thankful for their thoughtful comments which are of prime importance to this next generation of computational linguists.

Special thanks to Donia Scott for her input, to Claire Gardent for her helpful suggestions, and to our tireless faculty advisor Justine Cassell for her advice, support, and obtaining funding. Above all, we wish to thank the American National Science Foundation for providing financial support that has allowed the student presenters to travel to Barcelona.

ACL-04 Student Research Workshop Co-chairs:
Leonoor van der Beek, *University of Groningen, Netherlands*
Dmitriy Genzel, *Brown University, USA*
Daniel Midgley, *University of Western Australia, Australia*

Table of Contents

<i>Determining the Specificity of Terms using Compositional and Contextual Information</i> Pum-Mo Ryu	1
<i>Minimizing the Length of Non-Mixed Initiative Dialogs</i> R. Bryce Inouye	7
<i>Towards a Semantic Classification of Spanish Verbs Based on Subcategorisation Information</i> Eva Esteve Ferrer	13
<i>Improving the Accuracy of Subcategorizations Acquired from Corpora</i> Naoki Yoshinaga	19
<i>Robust VPE Detection using Automatically Parsed Text</i> Leif Arda Nielsen	25
<i>A Machine Learning Approach to German Pronoun Resolution</i> Beata Kouchnir	31
<i>Searching for Topics in a Large Collection of Texts</i> Martin Holub, Jiří Semecký, and Jiří Diviš	37
<i>Temporal Context: Applications and Implications for Computational Linguistics</i> Robert Leibscher	43
<i>Automatic Acquisition of English Topic Signatures Based on a Second Language</i> Xinglong Wang	49
<i>iSTART: Paraphrase Recognition</i> Chutima Boonthum	55
<i>Beyond N in N-gram Tagging</i> Robbert Prins	61
<i>A Framework for Unsupervised Natural Language Morphology Induction</i> Christian Monson	67

Table of Contents

<i>Optimization in Multimodal Interpretation</i> Joyce Y. Chai, Pengyu Hong, Michelle X. Xhou and Zahar Prasov	1
<i>Constructivist Development of Grounded Construction Grammar</i> Luc Steels	9
<i>A Distributional Model of Semantic Context Effects in Lexical Processing</i> Scott McDonald and Chris Brew	17
<i>Analysis of Mixed Natural and Symbolic Input in Mathematical Dialogs</i> Magdalena Wolska and Ivana Kruijff-KorBayová	25
<i>A TAG-based noisy-channel model of speech repairs</i> Mark Johnson and Eugene Charniak	33
<i>Attention Shifting for Parsing Speech</i> Keith B. Hall and Mark Johnson	40
<i>Discriminative Language Modeling with Conditional Random Fields and the Perceptron Algorithm</i> Brian Roark, Murat Saraclar, Michael Collins and Mark Johnson	47
<i>Statistical Modeling for Unit Selection in Speech Synthesis</i> Mehryar Mohri, Cyril Allauzen and Michael Riley	55
<i>Developing a Flexible Spoken Dialog System Using Simulation</i> Grace Chung	63
<i>Data-Driven Strategies for an Automated Dialogue System</i> Hilda Hardy, Tomek Strzalkowski, Min Wu, Cristian Ursu, Nick Webb, Alan Biermann, R. Bryce Inouye and Ashley McKenzie	71
<i>Trainable Sentence Planning for Complex Information Presentations in Spoken Dialog Systems</i> Amanda Stent, Rashmi Prasad and Marilyn Walker	79
<i>User Expertise Modeling and Adaptivity in a Speech-Based E-Mail System</i> Kristiina Jokinen and Kari Kanto	87
<i>Discriminative Training of a Neural Network Statistical Parser</i> James Henderson	95
<i>Parsing the WSJ Using CCG and Log-Linear Models</i> Stephen Clark and James R. Curran	103
<i>Incremental Parsing with the Perceptron Algorithm</i> Michael Collins and Brian Roark	111
<i>Convolution Kernels with Feature Selection for Natural Language Processing Tasks</i> Jun Suzuki, Hideki Isozaki and Eisaku Maeda	119
<i>Improving Pronoun Resolution by Incorporating Coreferential Information of Candidates</i> Xiaofeng Yang, Jian Su, Guodong Zhou and Chew-Lim Tan	127
<i>A Mention-Synchronous Coreference Resolution Algorithm Based On the Bell Tree</i> Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla and Salim Roukos	135
<i>Learning to Resolve Bridging References</i> Massimo Poesio, Rahul Mehta, Axel Maroudas and Janet Hitzeman	143

<i>Learning Noun Phrase Anaphoricity to Improve Conference Resolution: Issues in Representation and Optimization</i> Vincent Ng	151
<i>A Joint Source-Channel Model for Machine Transliteration</i> Haizhou Li, Min Zhang and Jian Su	159
<i>Collocation Translation Acquisition Using Monolingual Corpora</i> Yajuan Lü and Ming Zhou	167
<i>Statistical Machine Translation with Word- and Sentence-Aligned Parallel Corpora</i> Chris Callison-Burch, David Talbot and Miles Osborne	175
<i>Finding Ideographic Representations of Japanese Names Written in Latin Script via Language Identification and Corpus Validation</i> Yan Qu and Gregory Grefenstette	183
<i>Extracting Regulatory Gene Expression Networks From Pubmed</i> Jasmin Saric, Lars J. Jensen, Peer Bork, Rossitza Ouzounova and Isabel Rojas	191
<i>Linguistic Profiling for Authorship Recognition and Verification</i> Hans van Halteren	199
<i>An Empirical Study of Information Synthesis Task</i> Enrique Amigo, Julio Gonzalo, Victor Peinado, Anselmo Peñas and Felisa Verdejo	207
<i>Mining Metalinguistic Activity in Corpora to Create Lexical Resources Using Information Extraction Techniques: the MOP System</i> Carlos Rodriguez Penagos	215
<i>Optimizing Typed Feature Structure Grammar Parsing through Non-Statistical Indexing</i> Cosmin Munteanu and Gerald Penn	223
<i>Head-Driven Parsing for Word Lattices</i> Christopher Collins, Bob Carpenter and Gerald Penn	231
<i>Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying</i> Gerald Penn	239
<i>Minimal Recursion Semantics as Dominance Constraints: Translation, Evaluation, and Analysis</i> Ruth Fuchss, Alexander Koller, Joachim Niehren and Stefan Thater	247
<i>Learning with Unlabeled Data for Text Categorization Using a Bootstrapping and a Feature Projection Technique</i> Youngjoong Ko and Jungyun Seo	255
<i>The Sentimental Factor: Improving Review Classification Via Human-Provided Information</i> Philip Beineke, Trevor Hastie and Shivakumar Vaithyanathan	263
<i>A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts</i> Bo Pang and Lillian Lee	271
<i>Finding Predominant Word Senses in Untagged Text</i> Diana McCarthy, Rob Koeling, Julie Weeds and John Carroll	279
<i>Unsupervised Sense Disambiguation Using Bilingual Probabilistic Models</i> Indrajit Bhattacharya, Lise Getoor and Yoshua Bengio	287
<i>Chinese Verb Sense Discrimination Using an EM Clustering Model with Rich Linguistic Features</i> Jinying Chen and Martha Palmer	295
<i>Relieving the data Acquisition Bottleneck in Word Sense Disambiguation</i>	

Mona Diab	303
<i>Enriching the Output of a Parser Using Memory-based Learning</i> Valentin Jijkoun and Maarten de Rijke	311
<i>Long-Distance Dependency Resolution in Automatically Acquired Wide-Coverage PCFG-Based LFG Approximations</i> Aoife Cahill, Michael Burke, Ruth O'Donovan, Josef Van Genabith and Andy Way	319
<i>Deep Dependencies from Context-Free Statistical Parsers: Correcting the Surface Dependency Approximation</i> Roger Levy and Christopher Manning	327
<i>A Study on Convolution Kernels for Shallow Statistic Parsing</i> Alessandro Moschitti	335
<i>Combining Acoustic and Pragmatic Features to Predict Recognition Performance in Spoken Dialogue Systems</i> Malte Gabsdil and Oliver Lemon	343
<i>Predicting Student Emotions in Computer-Human Tutoring Dialogues</i> Diane J. Litman and Kate Forbes-Riley	351
<i>Building Verb Predicates: A Computational View</i> Fernando Gomez	359
<i>Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II Treebank</i> Ruth O'Donovan, Michael Burke, Aoife Cahill, Josef Van Genabith and Andy Way	367
<i>Inducing Frame Semantic Verb Classes from WordNet and LDOCE</i> Rebecca Green, Bonnie J. Dorr and Philip Resnik	375
<i>Paragraph-, Word-, and Coherence-based Approaches to Sentence Ranking: A Comparison of Algorithm and Human Performance</i> Florian Wolf and Edward Gibson	383
<i>Evaluating Centering-Based Metrics of Coherence</i> Nikiforos Karamanis, Massimo Poesio, Chris Mellish and Jon Oberlander	391
<i>Computing Locally Coherent Discourses</i> Ernst Althaus, Nikiforos Karamanis and Alexander Koller	399
<i>Generating Referring Expressions in Open Domains</i> Advaith Siddharthan and Ann Copestake	407
<i>Discovering Relations among Named Entities from Large Corpora</i> Takaaki Hasegawa, Satoshi Sekine and Ralph Grishman	415
<i>Dependency Tree Kernels for Relation Extraction</i> Aron Culotta and Jeffrey Sorensen	423
<i>Classifying Semantic Relations in Bioscience Texts</i> Barbara Rosario and Marti Hearst	430
<i>Collective Information Extraction with Relational Markov Networks</i> Razvan Bunescu and Raymond Mooney	438
<i>Error Mining for Wide-Coverage Grammar Engineering</i> Gertjan van Noord	446
<i>Alternative approaches for Generating Bodies of Grammar Rules</i> Gabriel Infante-Lopez and Maarten de Rijke	454

<i>Adaptive Chinese Word Segmentation</i> Jianfeng Gao, Andi Wu, Cheng-Ning Huang, Hong qiao Li, Xinsong Xia and Hauwei Qin	462
<i>Experiments in parallel-text based grammar induction</i> Jonas Kuhn	470
<i>Corpus-Based Induction of Syntactic Structure: Models of Dependency and Constituency</i> Dan Klein and Christopher Manning	478
<i>Annealing Techniques For Unsupervised Statistical Language Learning</i> Noah A. Smith and Jason Eisner	486
<i>Multi-Engine Machine Translation with Voted Language Model</i> Tadashi Nomoto	494
<i>Aligning words using matrix factorisation</i> Cyril Goutte, Kenji Yamada and Eric Gaussier	502
<i>FSA: An Efficient and Flexible C++ Toolkit for Finite State Automata Using On-Demand Computation</i> Stephan Kanthak and Hermann Ney	510
<i>Improving IBM Word Alignment Model 1</i> Robert C. Moore	518
<i>A Geometric View on Bilingual Lexicon Extraction from Comparable Corpora</i> Eric Gaussier, J.M. Renders, I. Matveeva, C. Goutte and H. Dejean	526
<i>Creating Multilingual Translation Lexicons with Regional Variations Using Web Corpora</i> Pu-Jen Cheng, Wen-Hsiang Lu, Jei-Wen Teng and Lee-Feng Chien	534
<i>Probabilistic Parsing Strategies</i> Mark-Jan Nederhof and Giorgio Satta	542
<i>An Alternative Method of Training Probabilistic LR Parsers</i> Mark-Jan Nederhof and Giorgio Satta	550
<i>Wrapping of Trees</i> James Rogers	558
<i>Splitting Complex Temporal Questions for Question Answering Systems</i> E. Saquete, P. Martínez-Barco, R. Muñoz and J.L. Vicedo	566
<i>Question Answering Using Constraint Satisfaction: QA-By-Dossier-With-Constraints</i> John Prager, Jennifer Chu-Carroll and Krzysztof Czuba	574
<i>Applying Machine Learning to Chinese Temporal Relation Resolution</i> Wenjie Li, Kam-Fai Wong, Guihong Cao and Chunfa Yuan	582
<i>Multi-Criteria-based Active Learning for Named Entity Recognition</i> Dan Shen, Jie Zhang, Jian Su, Guodong Zhou and Chew-Lim Tan	589
<i>Weakly Supervised Learning for Cross-document Person Name Disambiguation Supported by Information Ex- traction</i> Cheng Niu, Wei Li and Rohini K. Srihari	597
<i>Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics</i> Chin-Yew Lin and Franz Josef Och	605
<i>A Unified Framework For Automatic Evaluation Using 4-Gram Co-occurrence Statistics</i> Radu Soricut and Eric Brill	613

<i>Extending the BLEU MT Evaluation Method with Frequency Weightings</i> Bogdan Babych and Tony Hartley	621
<i>Learning Word Sense With Feature Selection and Order Identification Capabilities</i> Zheng-Yu Niu, Dong-Hong Ji and Chew-Lim Tan	629
<i>A Kernel PCA Method for Superior Word Sense Disambiguation</i> Dekai Wu, Weifeng Su and Marine Carpuat	637
<i>Using Linguistic Principles to Recover Empty Categories</i> Richard Campbell	645
<i>Statistical Machine Translation by Parsing</i> I. Dan Melamed	653
<i>Generalized Multitext Grammars</i> I. Dan Melamed, Giorgio Satta and Benjamin Wellington	661
<i>Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies</i> Michel Galley, Kathleen McKeown, Julia Hirschberg and Elizabeth Shriberg	669
<i>Using Conditional Random Fields to Predict Pitch Accents in Conversational Speech</i> Michelle Gregory and Yasemin Altun	677
<i>Acquiring the Meaning of Discourse Markers</i> Ben Hutchinson	684
<i>FLSA: Extending Latent Semantic Analysis with Features for Dialogue Act Classification</i> Riccardo Serafin and Barbara Di Eugenio	692

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