ACL-04

42nd Annual Meeting of the Association for Computational Linguistics

Proceedings of the Conference

21–26 July , 2004
Barcelona, Spain

Preface from the Conference Chair

I am delighted to welcome you to the 42nd meeting of the Association for Computational Linguistics: ACL'04 in Barcelona.

In keeping with the tradition of ACL conferences, we present you with a scientific programme of cutting-edge research in our field, in the form of papers, demonstrations and posters, supported by tutorials and workshops and complemented by an ACL-endorsed associated conference. The programme continues to show a healthy mixture of high standards of selectivity coupled with wide participation from our scientific community: ACL'04 received 619 submissions to the technical sessions and workshops (not including 260 to EMNLP) of which 221 were accepted, providing an overall acceptance rate of 36%.

Over the three days of the main conference, there will be 137 presentations, distributed over three parallel sessions, in addition to a fourth session on Day 2 when the Student Research Workshop will take place. We are continuing the format for Demonstrations and Interactive Posters introduced so successfully by our colleagues in Sapporo last year: in the first session of each day there will be short formal presentations of each system to be demonstrated, and poster to be (interactively) displayed, later that day. As in the last two years, the ACL executive committee will be presenting an ACL Lifetime Achievement Award to a distinguished researcher for his or her pioneering work in Computational Linguistics. Previous recipients have been Aravind Joshi and Makato Nagao. The winner of this year's award will be announced by the ACL Executive Committee at a special plenary session on Day 2, followed by a lecture by the recipient of the award. Also as in previous years, the Programme Chairs of the Main Session will be announcing, at the end of the Conference, the recipient of the Best Paper Award; this year the Chairs of the Student Research Workshop will also present an award, for the best Student Paper.

This scientific programme will be accompanied by a special programme of social events: you are all invited to join us at the Welcome Reception on the evening of the 21st July. I hope we will choose to also join us at the Conference Banquet on the 23rd in the medieval Saló de Contractacions at the nearby Llotja, which will be a feast of local Catalan cuisine and wines, and will be the venue for the Presidental Address by Johanna Moore.

Any ACL conference is only as good as the people who take time out of their very busy schedule to help put it all together. As General Chair, I would like to extend my sincere gratitude to all programme chairs for their dedication and care in producing this year's programme: Walter Daelemans and Marilyn Walker (main session); Philippe Blache and Horacio Rodríques (Demos and Posters); Srinivas Bangalore (Workshops); Inderjeet Mani (Tutorials); and Leonoor van der Beek, Dmitriy Genzel, Daniel Midgley and Justine Cassell (Student Research Workshop). Many thanks also to Owen Rambow and Sergi Balari for their sterling job of coordinating the production of several hundreds of papers over many proceedings, to Deborah Dahl and Antònia Martí for their work in securing sponsorships for the conference, to Felisa Verdejo and José Mariño Acebal for their work on publicity, to Joaquim Llisterri for coordinating the website and to Montse Caprubí for her administrative support.

This will be the second ACL conference to be held in Spain — the other being in Madrid in 1995, which was the first ACL meeting to take place outside North America. We liked it so much the first time that we couldn't resist the opportunity to do it again! I am extremely grateful to our colleagues from Barcelona, under the steer of Toni Badia, for offering to host this year's conference, for securing for us the very generous offer of an excellent venue at the Forum of Cultures, and for their invaluable work in providing the necessary infrastructure and support for a successful conference.

Finally, without the generosity of our sponsors — in particular the Forum of Cultures — this conference would not have been possible. On behalf of all members of the Association of Computational Linguistics, I would like to take this opportunity to express our gratitude and appreciation to them.

Donia Scott
ITRI, University of Brighton
General Chair

Preface from the Program Co-chairs

It is with great pleasure that we introduce the proceedings of ACL-04, the 42nd Annual Conference of the Association for Computational Linguistics. There is no sign of the field losing its vitality. The group of researchers considering themselves computational linguists is growing every year, numerous national and transnational projects are being defined and funded, and consequently, research results are presented in ever larger quantities. These results cover the complete research spectrum, from new annotated corpora and lexical resources over new statistical and parsing techniques to innovative applications.

The number and quality of submissions witnesses the healthy state of our field. We received 348 submissions for the main sessions, of which 88 were accepted, resulting in an acceptance rate of 25%, once more an extremely competitive selection. Among the accepted papers, 57% originates from North America, Canada and Mexico, 11% from Asia and the Pacific Rim area, and 32% from Europe. As a whole, the accepted papers provide an interesting overview of high-quality research in the field of Computational Linguistics.

We would like to thank all the authors of submitted papers for choosing ACL-2004 for the presentation of their research results. For the selection we are indebted to the programme committee with 11 area chairs from Europe, the U.S. and Asia, and the army of more than 200 reviewers that they recruited. With at least 3 reviews per submitted paper, the programme committee had to manage well over 1000 reviews (an interesting volume of more than 800 pages of critical literature when printed). We are deeply grateful to the area chairs and the reviewers for their conscientious judgement and for the time and energy they contributed. Our special thanks go to the 11 area chairs who spent many hours reaching a consensus on every submitted paper, and who attended a one day Program Committee meeting in Brighton in which papers were discussed and acceptances were decided. Many thanks to Donia Scott for making available the ITRI premises for this meeting, and to Kees van Deemter for hosting it.

The program is structured into plenary sessions, three parallel paper sessions, demo/poster sessions, and the student workshop. In addition to the submitted papers, the main program includes two invited presentations by distinguished researchers in neighboring fields. We are honored and very grateful that Anne Cutler (Max Planck Institute for Psycholinguistics, The Netherlands) and Jack Mostow (Carnegie Mellon University, Robotics Institute) have accepted our invitation.

As in the previous two years, the Association for Computational Linguistics will present an ACL Lifetime Achievement Award to a distinguished researcher for his or her pioneering work in Computational Linguistics. Previous recipients of this prestigious prize, awarded by the ACL Executive Committee, are Aravind Joshi and Makoto Nagao. The winner of this year's award will be announced at a special plenary session, followed by a lecture of the awardee. Also, as in previous years, a Best Paper award will be announced in a special plenary session at the end of the meeting.

We are very grateful for the advice and support of the General Conference Chair, Donia Scott, and the Local Organizing Chair, Toni Badia and the local organization committee. For the preparation of this volume and for the production of the CD-ROM version of the proceedings we would like to thank the Publications Co-Chairs, Owen Rambow and Sergio Balari. Furthermore, we could always rely on the experience, guidance, and support of the ACL executive committee, and the PC-chairs of ACL-03, Erhard Hinrichs and Dan Roth. Thanks are due also to the University of Sheffield and the University of Antwerp for financial and administrative support. Special mention is deserved by Guy De Pauw of the University of Antwerp who installed and maintained the electronic conference management system.

We hope that you enjoy the conference!

Walter Daelemans and Marilyn Walker
Program Committee Co-Chairs
June 2004

Table of Contents

Optimization in Multimodal Interpretation
Joyce Y. Chai, Pengyu Hong, Michelle X. Xhou and Zahar Prasov 1

Constructivist Development of Grounded Construction Grammar
Luc Steels 9

A Distributional Model of Semantic Context Effects in Lexical Processing
Scott McDonald and Chris Brew 17

Analysis of Mixed Natural and Symbolic Input in Mathematical Dialogs
Magdalena Wolska and Ivana Kruijff-KorBayová 25

A TAG-based noisy-channel model of speech repairs
Mark Johnson and Eugene Charniak 33

Attention Shifting for Parsing Speech
Keith B. Hall and Mark Johnson 40

Discriminative Language Modeling with Conditional Random Fields and the Perceptron Algorithm
Brian Roark, Murat Saraclar, Michael Collins and Mark Johnson 47

Statistical Modeling for Unit Selection in Speech Synthesis
Mehryar Mohri, Cyril Allauzen and Michael Riley 55

Developing a Flexible Spoken Dialog System Using Simulation
Grace Chung 63

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 71

Trainable Sentence Planning for Complex Information Presentations in Spoken Dialog Systems
Amanda Stent, Rashmi Prassad and Marilyn Walker 79

User Expertise Modeling and Adaptivity in a Speech-Based E-Mail System
Kristiina Jokinen and Kari Kanto 87

Discriminative Training of a Neural Network Statistical Parser
James Henderson 95

Parsing the WSJ Using CCG and Log-Linear Models
Stephen Clark and James R. Curran 103

Incremental Parsing with the Perceptron Algorithm
Michael Collins and Brian Roark 111

Convolution Kernels with Feature Selection for Natural Language Processing Tasks
Jun Suzuki, Hideki Isozaki and Eisaku Maeda 119

Improving Pronoun Resolution by Incorporating Coreferential Information of Candidates
Xiaofeng Yang, Jian Su, Guodong Zhou and Chew-Lim Tan 127

A Mention-Synchronous Coreference Resolution Algorithm Based On the Bell Tree
Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla and Salim Roukos 135

Learning to Resolve Bridging References
Massimo Poesio, Rahul Mehta, Axel Maroudas and Janet Hitzeman 143

Learning Noun Phrase Anaphoricity to Improve Conference Resolution: Issues in Representation and Optimization
Vincent Ng 151

A Joint Source-Channel Model for Machine Transliteration
Haizhou Li, Min Zhang and Jian Su 159

Collocation Translation Acquisition Using Monolingual Corpora
Yajuan Lü and Ming Zhou 167

Statistical Machine Translation with Word- and Sentence-Aligned Parallel Corpora
Chris Callison-Burch, David Talbot and Miles Osborne 175

Finding Ideographic Representations of Japanese Names Written in Latin Script via Language Identification and Corpus Validation
Yan Qu and Gregory Grefenstette 183

Extracting Regulatory Gene Expression Networks From Pubmed
Jasmin Saric, Lars J. Jensen, Peer Bork, Rossitza Ouzounova and Isabel Rojas 191

Linguistic Profiling for Authorship Recognition and Verification
Hans van Halteren 199

An Empirical Study of Information Synthesis Task
Enrique Amigo, Julio Gonzalo, Victor Peinado, Anselmo Peñas and Felisa Verdejo 207

Mining Metalinguistic Activity in Corpora to Create Lexical Resources Using Information Extraction Techniques: the MOP System
Carlos Rodriguez Penagos 215

Optimizing Typed Feature Structure Grammar Parsing through Non-Statistical Indexing
Cosmin Munteanu and Gerald Penn 223

Head-Driven Parsing for Word Lattices
Christopher Collins, Bob Carpenter and Gerald Penn 231

Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying
Gerald Penn 239

Minimal Recursion Semantics as Dominance Constraints: Translation, Evaluation, and Analysis
Ruth Fuchss, Alexander Koller, Joachim Niehren and Stefan Thater 247

Learning with Unlabeled Data for Text Categorization Using a Bootstrapping and a Feature Projection Technique
Youngjoong Ko and Jungyun Seo 255

The Sentimental Factor: Improving Review Classification Via Human-Provided Information
Philip Beineke, Trevor Hastie and Shivakumar Vaithyanathan 263

A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts
Bo Pang and Lillian Lee 271

Finding Predominant Word Senses in Untagged Text
Diana McCarthy, Rob Koeling, Julie Weeds and John Carroll 279

Unsupervised Sense Disambiguation Using Bilingual Probabilistic Models
Indrajit Bhattacharya, Lise Getoor and Yoshua Bengio 287

Chinese Verb Sense Discrimination Using an EM Clustering Model with Rich Linguistic Features
Jinying Chen and Martha Palmer 295

Relieving the data Acquisition Bottleneck in Word Sense Disambiguation

Mona Diab 303

Enriching the Output of a Parser Using Memory-based Learning
Valentin Jijkoun and Maarten de Rijke 311

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 319

Deep Dependencies from Context-Free Statistical Parsers: Correcting the Surface Dependency Approximation
Roger Levy and Christopher Manning 327

A Study on Convolution Kernels for Shallow Statistic Parsing
Alessandro Moschitti 335

Combining Acoustic and Pragmatic Features to Predict Recognition Performance in Spoken Dialogue Systems
Malte Gabsdil and Oliver Lemon 343

Predicting Student Emotions in Computer-Human Tutoring Dialogues
Diane J. Litman and Kate Forbes-Riley 351

Building Verb Predicates: A Computational View
Fernando Gomez 359

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 367

Inducing Frame Semantic Verb Classes from WordNet and LDOCE
Rebecca Green, Bonnie J. Dorr and Philip Resnik 375

Paragraph-, Word-, and Coherence-based Approaches to Sentence Ranking: A Comparison of Algorithm and Human Performance
Florian Wolf and Edward Gibson 383

Evaluating Centering-Based Metrics of Coherence
Nikiforos Karamanis, Massimo Poesio, Chris Mellish and Jon Oberlander 391

Computing Locally Coherent Discourses
Ernst Althaus, Nikiforos Karamanis and Alexander Koller 399

Generating Referring Expressions in Open Domains
Advaith Siddharthan and Ann Copestake 407

Discovering Relations among Named Entities from Large Corpora
Takaaki Hasegawa, Satoshi Sekine and Ralph Grishman 415

Dependency Tree Kernels for Relation Extraction
Aron Culotta and Jeffrey Sorensen 423

Classifying Semantic Relations in Bioscience Texts
Barbara Rosario and Marti Hearst 430

Collective Information Extraction with Relational Markov Networks
Razvan Bunescu and Raymond Mooney 438

Error Mining for Wide-Coverage Grammar Engineering
Gertjan van Noord 446

Alternative approaches for Generating Bodies of Grammar Rules
Gabriel Infante-Lopez and Maarten de Rijke 454

Adaptive Chinese Word Segmentation
Jianfeng Gao, Andi Wu, Cheng-Ning Huang, Hong qiao Li, Xinsong Xia and Hauwei Qin 462

Experiments in parallel-text based grammar induction
Jonas Kuhn 470

Corpus-Based Induction of Syntactic Structure: Models of Dependency and Constituency
Dan Klein and Christopher Manning 478

Annealing Techniques For Unsupervised Statistical Language Learning
Noah A. Smith and Jason Eisner 486

Multi-Engine Machine Translation with Voted Language Model
Tadashi Nomoto 494

Aligning words using matrix factorisation
Cyril Goutte, Kenji Yamada and Eric Gaussier 502

FSA: An Efficient and Flexible C++ Toolkit for Finite State Automata Using On-Demand Computation
Stephan Kanthak and Hermann Ney 510

Improving IBM Word Alignment Model 1
Robert C. Moore 518

A Geometric View on Bilingual Lexicon Extraction from Comparable Corpora
Eric Gaussier, J.M. Renders, I. Matveeva, C. Goutte and H. Dejean 526

Creating Multilingual Translation Lexicons with Regional Variations Using Web Corpora
Pu-Jen Cheng, Wen-Hsiang Lu, Jei-Wen Teng and Lee-Feng Chien 534

Probabilistic Parsing Strategies
Mark-Jan Nederhof and Giorgio Satta 542

An Alternative Method of Training Probabilistic LR Parsers
Mark-Jan Nederhof and Giorgio Satta 550

Wrapping of Trees
James Rogers 558

Splitting Complex Temporal Questions for Question Answering Systems
E. Saquete, P. Martínez-Barco, R. Muñoz and J.L. Vicedo 566

Question Answering Using Constraint Satisfaction: QA-By-Dossier-With-Contraints
John Prager, Jennifer Chu-Carroll and Krzysztof Czuba 574

Applying Machine Learning to Chinese Temporal Relation Resolution
Wenjie Li, Kam-Fai Wong, Guihong Cao and Chunfa Yuan 582

Multi-Criteria-based Active Learning for Named Entity Recognition
Dan Shen, Jie Zhang, Jian Su, Guodong Zhou and Chew-Lim Tan 589

Weakly Supervised Learning for Cross-document Person Name Disambiguation Supported by Information Extraction
Cheng Niu, Wei Li and Rohini K. Srihari 597

Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics
Chin-Yew Lin and Franz Josef Och 605

A Unified Framework For Automatic Evaluation Using 4-Gram Co-occurrence Statistics
Radu Soricut and Eric Brill 613

Extending the BLEU MT Evaluation Method with Frequency Weightings
Bogdan Babych and Tony Hartley 621

Learning Word Sense With Feature Selection and Order Identification Capabilities
Zheng-Yu Niu, Dong-Hong Ji and Chew-Lim Tan 629

A Kernel PCA Method for Superior Word Sense Disambiguation
Dekai Wu, Weifeng Su and Marine Carpuat 637

Using Linguistic Principles to Recover Empty Categories
Richard Campbell 645

Statistical Machine Translation by Parsing
I. Dan Melamed 653

Generalized Multitext Grammars
I. Dan Melamed, Giorgio Satta and Benjamin Wellington 661

Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies
Michel Galley, Kathleen McKeown, Julia Hirschberg and Elizabeth Shriberg 669

Using Conditional Random Fields to Predict Pitch Accents in Conversational Speech
Michelle Gregory and Yasemin Altun 677

Acquiring the Meaning of Discourse Markers
Ben Hutchinson 684

FLSA: Extending Latent Semantic Analysis with Features for Dialogue Act Classification
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 Prassad 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-Contraints*
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

THIS IS A BLANK PAGE PLEASE IGNORE

Author Index

Allauzen, Cyril 55
Althaus, Ernst 399
Altun, Yasemin 677
Amigo, Enrique 207
Babych, Bogdan 621
Beineke, Philip 263
Bengio, Yoshua 287
Bhattacharya, Indrajit 287
Biermann, Alan 71
Bork, Peer 191
Brew, Chris 17
Brill, Eric 613
Bunescu, Razvan 438
Burke, Michael 319, 367
Cahill, Aoife 319, 367
Callison-Burch, Chris 175
Campbell, Richard 645
Cao, Guihong 582
Carpenter, Bob 231
Carpuat, Marine 637
Carroll, John 279
Chai, Joyce Y. 1
Charniak, Eugene 33
Chen, Jinying 295
Cheng, Pu-Jen 534
Chien, Lee-Feng 534
Chu-Carroll, Jennifer 574
Chung, Grace 63
Clark, Stephen 103
Collins, Christopher 231
Collins, Michael 47, 111
Copestake, Ann 407
Culotta, Aron 423
Curran, James R. 103
Czuba, Krzysztof 574
de Rijke, Maarten 311, 454
Dejean, H. 526
Di Eugenio, Barbara 692
Diab, Mona 303
Dorr, Bonnie J. 375
Eisner, Jason 486
Forbes-Riley, Kate 351
Fuchss, Ruth 247
Gabsdil, Malte 343
Galley, Michel 669
Gao, Jianfeng 462
Gaussier, Eric 502, 526
Getoor, Lise 287
Gibson, Edward 383
Gomez, Fernando 359
Gonzalo, Julio 207
Goutte, C. 526
Goutte, Cyril 502
Green, Rebecca 375
Grefenstette, Gregory 183
Gregory, Michelle 677
Grishman, Ralph 415
Hall, Keith B. 40
Hardy, Hilda 71
Hartley, Tony 621
Hasegawa, Takaaki 415
Hastie, Trevor 263
Hearst, Marti 430
Henderson, James 95
Hirschberg, Julia 669
Hitzeman, Janet 143
Hong, Pengyu 1
Huang, Cheng-Ning 462
Hutchinson, Ben 684
Infante-Lopez, Gabriel 454
Inouye, R. Bryce 71
Isozaki, Hideki 119
Ittycheriah, Abe 135
Jensen, Lars J. 191
Ji, Dong-Hong 629
Jijkoun, Valentin 311
Jing, Hongyan 135
Johnson, Mark 33, 40, 47
Jokinen, Kristiina 87
Kambhatla, Nanda 135
Kanthak, Stephan 510
Kanto, Kari 87
Karamanis, Nikiforos 391, 399
Klein, Dan 478
Ko, Youngjoong 255
Koeling, Rob 279
Koller, Alexander 247, 399
Kruijff-KorBayová, Ivana 25
Kuhn, Jonas 470
Lee, Lillian 271
Lemon, Oliver 343
Levy, Roger 327
Li, Haizhou 159
Li, Hong qiao 462
Li, Wei 597
Li, Wenjie 582
Lin, Chin-Yew 605
Litman, Diane J. 351
Lu, Wen-Hsiang 534
Lü, Yajuan 167
Luo, Xiaoqiang 135
Maeda, Eisaku 119
Manning, Christopher 327, 478
Maroudas, Axel 143

Martínez-Barco, P. 566
Matveeva, I. 526
McCarthy, Diana 279
McDonald, Scott 17
McKenzie, Ashley 71
McKeown, Kathleen 669
Mehta, Rahul 143
Melamed, I. Dan 653, 661
Mellish, Chris 391
Mohri, Mehryar 55
Mooney, Raymond 438
Moore, Robert C. 518
Moschitti, Alessandro 335
Muñoz, R. 566
Munteanu, Cosmin 223
Nederhof, Mark-Jan 542, 550
Ney, Hermann 510
Ng, Vincent 151
Niehren, Joachim 247
Niu, Cheng 597
Niu, Zheng-Yu 629
Nomoto, Tadashi 494
Oberlander, Jon 391
Och, Franz Josef 605
O'Donovan, Ruth 319, 367
Osborne, Miles 175
Ouzounova, Rossitza 191
Palmer, Martha 295
Pang, Bo 271
Peinado, Victor 207
Penagos, Carlos Rodriguez 215
Peñas, Anselmo 207
Penn, Gerald 223, 231, 239
Poesio, Massimo 143, 391
Prager, John 574
Prasov, Zahar 1
Prassad, Rashmi 79
Qin, Hauwei 462
Qu, Yan 183
Renders, J.M. 526
Resnik, Philip 375
Riley, Michael 55
Roark, Brian 47, 111
Rogers, James 558
Rojas, Isabel 191
Rosario, Barbara 430
Roukos, Salim 135
Saquete, E. 566
Saraclar, Murat 47
Saric, Jasmin 191
Satta, Giorgio 542, 550, 661
Sekine, Satoshi 415
Seo, Jungyun 255
Serafin, Riccardo 692
Shen, Dan 589
Shriberg, Elizabeth 669
Siddharthan, Advaith 407
Smith, Noah A. 486
Sorensen, Jeffrey 423
Soricut, Radu 613
Srihari, Rohini K. 597
Steels, Luc 9
Stent, Amanda 79
Strzalkowski, Tomek 71
Su, Jian 127, 159, 589
Su, Weifeng 637
Suzuki, Jun 119
Talbot, David 175
Tan, Chew-Lim 127, 589, 629
Teng, Jei-Wen 534
Thater, Stefan 247
Ursu, Cristian 71
Vaithyanathan, Shivakumar 263
Van Genabith, Josef 319, 367
van Halteren, Hans 199
van Noord, Gertjan 446
Verdejo, Felisa 207
Vicedo, J.L. 566
Walker, Marilyn 79
Way, Andy 319, 367
Webb, Nick 71
Weeds, Julie 279
Wellington, Benjamin 661
Wolf, Florian 383
Wolska, Magdalena 25
Wong, Kam-Fai 582
Wu, Andi 462
Wu, Dekai 637
Wu, Min 71
Xhou, Michelle X. 1
Xia, Xinsong 462
Yamada, Kenji 502
Yang, Xiaofeng 127
Yuan, Chunfa 582
Zhang, Jie 589
Zhang, Min 159
Zhou, Guodong 127, 589
Zhou, Ming 167