

2006



COLING • ACL

COLING • ACL 2006

EMNLP 2006

2006 Conference on Empirical Methods
in Natural Language Processing

Proceedings of the Conference

Chairs:

Dan Jurafsky and Eric Gaussier

22-23 July 2006
Sydney, Australia

Production and Manufacturing by
BPA Digital
11 Evans St
Burwood VIC 3125
AUSTRALIA

A meeting of SIGDAT, a Special Interest Group of the ACL,
held in conjunction with COLING/ACL 2006

©2006 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 1-932432-73-6

Table of Contents

Preface	ix
Organizers	xi
Conference Program Outline	xiii
Conference Program in Detail	xiv
<i>Unsupervised Discovery of a Statistical Verb Lexicon</i> Trond Grenager and Christopher D. Manning	1
<i>An Empirical Approach to the Interpretation of Superlatives</i> Johan Bos and Malvina Nissim	9
<i>Paraphrase Recognition via Dissimilarity Significance Classification</i> Long Qiu, Min-Yen Kan and Tat-Seng Chua	18
<i>Detecting Parser Errors Using Web-based Semantic Filters</i> Alexander Yates, Stefan Schoenmackers and Oren Etzioni	27
<i>Distributional measures of concept-distance: A task-oriented evaluation</i> Saif Mohammad and Graeme Hirst	35
<i>SPMT: Statistical Machine Translation with Syntactified Target Language Phrases</i> Daniel Marcu, Wei Wang, Abdessamad Echihabi and Kevin Knight	44
<i>Phrasetable Smoothing for Statistical Machine Translation</i> George Foster, Roland Kuhn and Howard Johnson	53
<i>The impact of parse quality on syntactically-informed statistical machine translation</i> Chris Quirk and Simon Corston-Oliver	62
<i>Statistical Machine Reordering</i> Marta R. Costa-jussà and José A. R. Fonollosa	70
<i>Re-evaluating Machine Translation Results with Paraphrase Support</i> Liang Zhou, Chin-Yew Lin and Eduard Hovy	77
<i>Exploiting Discourse Structure for Spoken Dialogue Performance Analysis</i> Mihai Rotaru and Diane J. Litman	85
<i>Learning Information Status of Discourse Entities</i> Malvina Nissim	94
<i>Automatic classification of citation function</i> Simone Teufel, Advaith Siddharthan and Dan Tidhar	103
<i>Is it Really that Difficult to Parse German?</i> Sandra Kübler, Erhard W. Hinrichs and Wolfgang Maier	111
<i>Domain Adaptation with Structural Correspondence Learning</i> John Blitzer, Ryan McDonald and Fernando Pereira	120

<i>Incremental Integer Linear Programming for Non-projective Dependency Parsing</i> Sebastian Riedel and James Clarke	129
<i>Semantic Role Labeling of NomBank: A Maximum Entropy Approach</i> Zheng Ping Jiang and Hwee Tou Ng	138
<i>Identification of Event Mentions and their Semantic Class</i> Steven Bethard and James H. Martin	146
<i>Extremely Lexicalized Models for Accurate and Fast HPSG Parsing</i> Takashi Ninomiya, Takuya Matsuzaki, Yoshimasa Tsuruoka, Yusuke Miyao and Jun'ichi Tsujii	155
<i>Multilingual Deep Lexical Acquisition for HPSGs via Supertagging</i> Phil Blunsom and Timothy Baldwin	164
<i>Lexical Reference: a Semantic Matching Subtask</i> Oren Glickman, Eyal Shnarch and Ido Dagan	172
<i>Semantic Role Labeling via Instance-Based Learning</i> Chi-san Althon Lin and Tony C. Smith	180
<i>Inducing Temporal Graphs</i> Philip Bramsen, Pawan Deshpande, Yoong Keok Lee and Regina Barzilay	189
<i>A Weakly Supervised Learning Approach for Spoken Language Understanding</i> Wei-Lin Wu, Ru-Zhan Lu, Jian-Yong Duan, Hui Liu, Feng Gao and Yu-Quan Chen	199
<i>Humor: Prosody Analysis and Automatic Recognition for F*R*I*E*N*D*S*</i> Amruta Purandare and Diane Litman	208
<i>Distributed Language Modeling for N-best List Re-ranking</i> Ying Zhang, Almut Silja Hildebrand and Stephan Vogel	216
<i>Efficient Search for Inversion Transduction Grammar</i> Hao Zhang and Daniel Gildea	224
<i>A Discriminative Model for Tree-to-Tree Translation</i> Brooke Cowan, Ivona Kučerová and Michael Collins	232
<i>Modeling Impression in Probabilistic Transliteration into Chinese</i> LiLi Xu, Atsushi Fujii and Tetsuya Ishikawa	242
<i>Unsupervised Named Entity Transliteration Using Temporal and Phonetic Correlation</i> Tao Tao, Su-Youn Yoon, Andrew Fister, Richard Sproat and ChengXiang Zhai	250
<i>Capturing Out-of-Vocabulary Words in Arabic Text</i> Abdusalam F.A. Nwesri, S.M.M. Tahaghoghi and Falk Scholer	258
<i>Using linguistically motivated features for paragraph boundary identification</i> Katja Filippova and Michael Strube	267
<i>BESTCUT: A Graph Algorithm for Coreference Resolution</i> Cristina Nicolae and Gabriel Nicolae	275

<i>Automatic Construction of Predicate-argument Structure Patterns for Biomedical Information Extraction</i>	
Akane Yakushiji, Yusuke Miyao, Tomoko Ohta, Yuka Tateisi and Jun'ichi Tsujii	284
<i>Protein folding and chart parsing</i>	
Julia Hockenmaier, Aravind K. Joshi and Ken A. Dill	293
<i>Learning Phrasal Categories</i>	
William P. Headden III, Eugene Charniak and Mark Johnson	301
<i>Priming Effects in Combinatory Categorical Grammar</i>	
David Reitter, Julia Hockenmaier and Frank Keller	308
<i>Better Informed Training of Latent Syntactic Features</i>	
Markus Dreyer and Jason Eisner	317
<i>Get out the vote: Determining support or opposition from Congressional floor-debate transcripts</i>	
Matt Thomas, Bo Pang and Lillian Lee	327
<i>Partially Supervised Coreference Resolution for Opinion Summarization through Structured Rule Learning</i>	
Veselin Stoyanov and Claire Cardie	336
<i>Sentiment Retrieval using Generative Models</i>	
Koji Eguchi and Victor Lavrenko	345
<i>Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis</i>	
Hiroshi Kanayama and Tetsuya Nasukawa	355
<i>A Skip-Chain Conditional Random Field for Ranking Meeting Utterances by Importance</i>	
Michel Galley	364
<i>Style & Topic Language Model Adaptation Using HMM-LDA</i>	
Bo-June (Paul) Hsu and James Glass	373
<i>Text data acquisition for domain-specific language models</i>	
Abhinav Sethy, Panayiotis G. Georgiou and Shrikanth Narayanan	382
<i>Corrective Models for Speech Recognition of Inflected Languages</i>	
Izhak Shafran and Keith Hall	390
<i>Lexicon Acquisition for Dialectal Arabic Using Transductive Learning</i>	
Kevin Duh and Katrin Kirchhoff	399
<i>Arabic OCR Error Correction Using Character Segment Correction, Language Modeling, and Shallow Morphology</i>	
Walid Magdy and Kareem Darwish	408
<i>Partially Supervised Sense Disambiguation by Learning Sense Number from Tagged and Untagged Corpora</i>	
Zheng-Yu Niu, Dong-Hong Ji and Chew Lim Tan	415
<i>Automatically Assessing Review Helpfulness</i>	
Soo-Min Kim, Patrick Pantel, Tim Chklovski and Marco Pennacchiotti	423

<i>Joint Extraction of Entities and Relations for Opinion Recognition</i> Yejin Choi, Eric Breck and Claire Cardie	431
<i>Feature Subsumption for Opinion Analysis</i> Ellen Riloff, Siddharth Patwardhan and Janyce Wiebe	440
<i>Relevance Feedback Models for Recommendation</i> Masao Utiyama and Mikio Yamamoto	449
<i>Random Indexing using Statistical Weight Functions</i> James Gorman and James R. Curran	457
<i>A Hybrid Markov/Semi-Markov Conditional Random Field for Sequence Segmentation</i> Galen Andrew	465
<i>Boosting Unsupervised Relation Extraction by Using NER</i> Ronen Feldman and Benjamin Rosenfeld	473
<i>Short Text Authorship Attribution via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation</i> Conrad Sanderson and Simon Guenter	482
<i>Entity Annotation based on Inverse Index Operations</i> Ganesh Ramakrishnan, Sreeram Balakrishnan and Sachindra Joshi	492
<i>Unsupervised Information Extraction Approach Using Graph Mutual Reinforcement</i> Hany Hassan, Ahmed Hassan and Ossama Emam	501
<i>Empirical Study on the Performance Stability of Named Entity Recognition Model across Domains</i> Hong Lei Guo, Li Zhang and Zhong Su	509
<i>Statistical Ranking in Tactical Generation</i> Erik Velldal and Stephan Oepen	517
<i>Sentence ordering with manifold-based classification in multi-document summarization</i> Paul D Ji and Stephen Pulman	526
<i>Quality Assessment of Large Scale Knowledge Resources</i> Montse Cuadros and German Rigau	534
<i>Graph-based Word Clustering using a Web Search Engine</i> Yutaka Matsuo, Takeshi Sakaki, Kôki Uchiyama and Mitsuru Ishizuka	542
<i>Context-Dependent Term Relations for Information Retrieval</i> Jing Bai, Jian-Yun Nie and Guihong Cao	551
<i>Loss Minimization in Parse Reranking</i> Ivan Titov and James Henderson	560
<i>Unsupervised Relation Disambiguation with Order Identification Capabilities</i> Jinxu Chen, Donghong Ji, ChewLim Tan and Zhengyu Niu	568
<i>Competitive generative models with structure learning for NLP classification tasks</i> Kristina Toutanova	576

<i>Two graph-based algorithms for state-of-the-art WSD</i> Eneko Agirre, David Martínez, Oier López de Lacalle and Aitor Soroa	585
<i>Broad-Coverage Sense Disambiguation and Information Extraction with a Supersense Sequence Tagger</i> Massimiliano Ciaramita and Yasemin Altun	594
<i>Learning Field Compatibilities to Extract Database Records from Unstructured Text</i> Michael Wick, Aron Culotta and Andrew McCallum	603
<i>Discriminative Methods for Transliteration</i> Dmitry Zelenko and Chinatsu Aone	612
<i>Solving the Problem of Cascading Errors: Approximate Bayesian Inference for Linguistic Annotation Pipelines</i> Jenny Rose Finkel, Christopher D. Manning and Andrew Y. Ng	618
Author Index	627

Preface

We are delighted to introduce the proceedings of the 11th Conference on Empirical Methods in Natural Language Processing, organized under the auspices of SIGDAT, the ACL Special Interest Group for linguistic data and corpus-based approaches to NLP.

This was a wonderfully fruitful year for EMNLP; we received 234 submissions, drawn from every area of language processing. Of these we were able to accept 73 papers (an acceptance rate of 31%), making for an unusually broad and exciting program. 43 of the papers were presented as talks, and 30 as posters.

The papers were selected by a program committee of 13 area chairs from Asia, Australia, Europe, and North America, ably assisted by a superb panel of 258 reviewers, also from all over the world. We are deeply indebted to the area chairs and the reviewers for their tireless and generous work.

Additional thanks to to the Publications Chair, Eric Ringger, who put this volume together, to the Local Arrangements Chair, James Curran, to the COLING/ACL Organizing Committee, especially Claire Cardie and Suzanne Stevenson, for constant advice, and to David Yarowsky and Ken Church of SIGDAT for fielding many questions. Special thanks go to the student volunteers at Stanford (Dan Cer, Pi-Chuan Chang, Surabhi Gupta, William Morgan, Yun-Hsuan Sung, and Huihsin Tseng).

We wish you all an enjoyable and thought-provoking conference.

Dan Jurafsky and Eric Gaussier
EMNLP Co-Chairs
June 2006

Organizers

Program Chairs:

Dan Jurafsky (Stanford University)
Eric Gaussier (Xerox Research Centre Europe)

Area Chairs:

Regina Barzilay (MIT)
Grace Chung
Jason Eisner (Johns Hopkins University)
Sanda M Harabagiu (University of Texas at Dallas)
James Henderson (University of Edinburgh)
Philipp Koehn (University of Edinburgh)
Paola Merlo (University of Geneva)
Rada Mihalcea (University of North Texas)
Grace Ngai (Hong Kong Polytechnic University)
Yan Qu (Clairvoyance Corporation)
Owen Rambow (Columbia University)
Dekai Wu (The Hong Kong University of Science and Technology)
Francois Yvon (ENST)

Local Arrangements Chair:

James Curran (University of Sydney)

Publications Chair:

Eric Ringger (Brigham Young University)

Reviewers:

Eneko Agirre, Yaser Al-Onaizan, Yasemin Altun, Sophia Ananiadou, Hidir Aras, Jason Baldridge, Srinivas Bangalore, Roberto Basili, Frederic Bechet, Oliver Bender, Nicola Bertoldi, Dan Bikel, Johan Bos, Costas Boulis, Chris Brew, Bill Byrne, Chris Callison-Burch, Nicola Cancedda, Yunbo Cao, Giuseppe Carenini, Michael Carl, Marine Carpuat, Xavier Carreras, Mauro Cettolo, Joyce Chai, Hsin-Hsi Chen, John Chen, Stanley Chen, Colin Cherry, Timothy Chklovski, Kenneth Church, Massimiliano Ciaramita, Stephen Clark, Nigel Collier, Michael Collins, Michael Connor, Koby Crammer, Josep Maria Crego, Andras Csomai, Silviu Cucerzan, Walter Daelemans, Hal Daume III, Adria de Gispert, Rodolfo Delmonte, Matthias Denecke, Maarten de Rijke, Mona Diab, Gaël Dias, Phil Edmonds, Andreas Eisele, Noemie Elhadad, Katrin Erk, Afsaneh Fazly, Marcello Federico, Radu Florian, George Foster, Alex Fraser, Pascale Fung, Jianfeng Gao, Claire Gardent, Dan Gildea, Jesus Gimenez, Roxana Girju, Natalie Glance, Sharon Goldwater, Gregory Grefenstette, Trond Grenager, Iryna Gurevych, Joakim Gustafson, Jan Hajic, Keith Hall, Sanda Harabagiu, Sasa Hasan, Vasileios Hatzivassiloglou, James Henderson, Iris Hendrickx, Andrew Hickl, Graeme Hirst, Julia Hockenmaier, Thomas Hofmann, Veronique Hoste, David Hull, Matthew Hurst, Rebecca Hwa, Martin Jansche, Michèle Jardino, Zhang Jie, Kristiina Jokinen, Min-Yen Kan, Nikiforos Karamanis, Jussi Karlgren, Adam Kilgarriff, Katrin Kirchhoff, Dan Klein, Alex Klemntiev, Kevin Knight, Philipp Koehn, Moshe Koppel, Roland Kuhn, Shankar Kumar, Olivia Kwong, Patrik Lambert, Irene Langkilde-Geary, Philippe Langlais, Guy Lapalme, Yoong Keok Lee, Oliver Lemon, Roger Levy, Maggie Li, Chin-Yew Lin, Dekang Lin, Bing Liu, Hugo Liu, Berenike Loos, Qin Lu, Klaus Macherey, Bernardo Magnini, Steven Maiorano, Igor Malioutov, Gideon Mann, Christopher Manning, Daniel Marcu, Lluís Marquez, Jim Martin, David Martinez, Yuji Matsumoto, Takuya Matsuzaki, Andrew McCallum, Diana McCarthy, Kathy McKeown, I. Dan Melamed, Arul Menezes, Helen Meng, Eleni Miltsakaki, Sebastian Moeller, Jacques Moeschler, Dan Moldovan, Christof Monz, Robert C. Moore, Rumen Moralyiski, Dragos Stefan Munteanu, Gabriel Murray, Mikio Nakano, Alexis Nasr, Vivi Nastase, Roberto Navigli, Mark-Jan Nederhof, Ani Nenkova, Andrew Ng, Hwee Tou Ng, Vincent Ng, Grace Ngai, Nicolas Nicolov, Malvina Nissim, Cheng Niu, Joakim Nivre, Franz Josef Och, Kemal Oflazer, Miles Osborne, Sebastian Pado, Tim Paek, Becky Passonneau, Jon Patrick, Fuchun Peng, Gerald Penn, Fernando Pereira, Fabio Pianesi, Joe Polifroni, Jay Ponte, Andrei Popescu-Belis, Robert Porzel, Oana Postolache, Pascal Poupard, John Prager, John Prange, Rashmi Prasad, James Pustejovsky, Chris Quirk, Steve Renals, German Rigau, Ellen Riloff, Brian Roark, Dan Roth, Alex Rudnicky, Marta Ruiz Costa-jussá, Fatiha Sadat, Kenji Sagae, Horacio Saggion, Franco Salvetti, Mark Sammons, Manabu Sassano, Charles Schafer, Holger Schwenk, Libin Shen, Wade Shen, Yihai Shen, Lei Shi, Advait Siddharthan, Candy Sidner, Michel Simard, Noam Slonim, Kevin Small, David Smith, Noah Smith, Rion Snow, Caroline Sporleder, Mark Steedman, Amanda Stent, Suzanne Stevenson, Veselin Stoyanov, Carlo Strapparava, Jian Su, Zhifang Sui, Mihai Surdeanu, Charles Sutton, Chew Lim Tan, Franck Thollard, Christoph Tillmann, Kristina Toutanova, Roy Tromble, Huishin Tseng, Jun'ichi Tsujii, Yoshimasa Tsuruoka, Dan Tufis, Peter Turney, Nicola Ueffing, Takehito Utsuro, Kees Van Deemter, Antal van den Bosch, Sebastian Varges, Ashish Venugopal, David Vilar, Jean-Yves Vion-Dury, Stephan Vogel, Chao Wang, Haifeng Wang, Taro Watanabe, Andy Way, Bonnie Webber, Michael White, Janyce Wiebe, Jason Williams, Theresa Wilson, Dekai Wu, Peng Xu, XiaoFeng Yang, Deniz Yuret, Richard Zens, Tong Zhang, Jun Zhao, GuoDong Zhou, Liang Zhou, Ming Zhou, Michael Zock, Andreas Zollmann, Hans-Peter Zorn, Ingrid Zukerman

Conference Program Outline

Saturday, 22 July 2006

- 7:45–9:00 Registration
- 8:15–8:25 Welcome from the Organizers
- 8:25–10:30 Sessions 1a and 1b
- 10:30–11:00 Morning Coffee Break
- 11:00–12:15 Sessions 2a and 2b
- 12:15–1:45 Lunch
- 1:45–2:35 Sessions 3a and 3b
- 2:35–3:30 Invited Talk
- 3:30–4:00 Afternoon Coffee Break
- 4:00–6:30 Long Poster Session 1 and Welcome Reception

Sunday, 23 July 2006

- 8:25–10:30 Sessions 4a and 4b
- 10:30–11:00 Morning Coffee Break
- 11:00–12:30 Short Poster Session 2
- 12:30–1:45 Lunch
- 1:45–3:25 Sessions 5a and 5b
- 3:30–4:00 Afternoon Coffee Break
- 4:00–5:15 Sessions 6a and 6b

Conference Program In Detail

Saturday, 22 July 2006

Session 1a: Computational Semantics

- 8:25–8:50 *Unsupervised Discovery of a Statistical Verb Lexicon*
Trond Grenager and Christopher D. Manning
- 8:50–9:15 *An Empirical Approach to the Interpretation of Superlatives*
Johan Bos and Malvina Nissim
- 9:15–9:40 *Paraphrase Recognition via Dissimilarity Significance Classification*
Long Qiu, Min-Yen Kan and Tat-Seng Chua
- 9:40–10:05 *Detecting Parser Errors Using Web-based Semantic Filters*
Alexander Yates, Stefan Schoenmackers and Oren Etzioni
- 10:05–10:30 *Distributional measures of concept-distance: A task-oriented evaluation*
Saif Mohammad and Graeme Hirst

Session 1b: MT

- 8:25–8:50 *SPMT: Statistical Machine Translation with Syntactified Target Language Phrases*
Daniel Marcu, Wei Wang, Abdessamad Echihabi and Kevin Knight
- 8:50–9:15 *Phrasetable Smoothing for Statistical Machine Translation*
George Foster, Roland Kuhn and Howard Johnson
- 9:15–9:40 *The impact of parse quality on syntactically-informed statistical machine translation*
Chris Quirk and Simon Corston-Oliver
- 9:40–10:05 *Statistical Machine Reordering*
Marta R. Costa-jussà and José A. R. Fonollosa
- 10:05–10:30 *Re-evaluating Machine Translation Results with Paraphrase Support*
Liang Zhou, Chin-Yew Lin and Eduard Hovy

Saturday, 22 July 2006 (continued)

Session 2a: Discourse

11:00–11:25 *Exploiting Discourse Structure for Spoken Dialogue Performance Analysis*
Mihai Rotaru and Diane J. Litman

11:25–11:50 *Learning Information Status of Discourse Entities*
Malvina Nissim

11:50–12:15 *Automatic classification of citation function*
Simone Teufel, Advaith Siddharthan and Dan Tidhar

Session 2b: Parsing

11:00–11:25 *Is it Really that Difficult to Parse German?*
Sandra Kübler, Erhard W. Hinrichs and Wolfgang Maier

11:25–11:50 *Domain Adaptation with Structural Correspondence Learning*
John Blitzer, Ryan McDonald and Fernando Pereira

11:50–12:15 *Incremental Integer Linear Programming for Non-projective Dependency Parsing*
Sebastian Riedel and James Clarke

Session 3a: Computational Semantics

1:45–2:10 *Semantic Role Labeling of NomBank: A Maximum Entropy Approach*
Zheng Ping Jiang and Hwee Tou Ng

2:10–2:35 *Identification of Event Mentions and their Semantic Class*
Steven Bethard and James H. Martin

Session 3b: Parsing

1:45–2:10 *Extremely Lexicalized Models for Accurate and Fast HPSG Parsing*
Takashi Ninomiya, Takuya Matsuzaki, Yoshimasa Tsuruoka, Yusuke Miyao and Jun'ichi Tsujii

2:10–2:35 *Multilingual Deep Lexical Acquisition for HPSGs via Supertagging*
Phil Blunsom and Timothy Baldwin

Saturday, 22 July 2006 (continued)

**Long Poster Session 1: Discourse, Dialogue, MT, Computational Semantics, Parsing
(4:00–6:30)**

Lexical Reference: a Semantic Matching Subtask

Oren Glickman, Eyal Shnarch and Ido Dagan

Semantic Role Labeling via Instance-Based Learning

Chi-san Althon Lin and Tony C. Smith

Inducing Temporal Graphs

Philip Bramsen, Pawan Deshpande, Yoong Keok Lee and Regina Barzilay

A Weakly Supervised Learning Approach for Spoken Language Understanding

Wei-Lin Wu, Ru-Zhan Lu, Jian-Yong Duan, Hui Liu, Feng Gao and Yu-Quan Chen

*Humor: Prosody Analysis and Automatic Recognition for F*R*I*E*N*D*S**

Amruta Purandare and Diane Litman

Distributed Language Modeling for N-best List Re-ranking

Ying Zhang, Almut Silja Hildebrand and Stephan Vogel

Efficient Search for Inversion Transduction Grammar

Hao Zhang and Daniel Gildea

A Discriminative Model for Tree-to-Tree Translation

Brooke Cowan, Ivona Kučerová and Michael Collins

Modeling Impression in Probabilistic Transliteration into Chinese

LiLi Xu, Atsushi Fujii and Tetsuya Ishikawa

Saturday, 22 July 2006 (continued)

**Long Poster Session 1: Discourse, Dialogue, MT, Computational Semantics, Parsing
(4:00–6:30) (continued)**

Unsupervised Named Entity Transliteration Using Temporal and Phonetic Correlation

Tao Tao, Su-Youn Yoon, Andrew Fister, Richard Sproat and ChengXiang Zhai

Capturing Out-of-Vocabulary Words in Arabic Text

Abdusalam F.A. Nwesri, S.M.M. Tahaghoghi and Falk Scholer

Using linguistically motivated features for paragraph boundary identification

Katja Filippova and Michael Strube

BESTCUT: A Graph Algorithm for Coreference Resolution

Cristina Nicolae and Gabriel Nicolae

Automatic Construction of Predicate-argument Structure Patterns for Biomedical Information Extraction

Akane Yakushiji, Yusuke Miyao, Tomoko Ohta, Yuka Tateisi and Jun'ichi Tsujii

Protein folding and chart parsing

Julia Hockenmaier, Aravind K. Joshi and Ken A. Dill

Learning Phrasal Categories

William P. Headden III, Eugene Charniak and Mark Johnson

Priming Effects in Combinatory Categorical Grammar

David Reitter, Julia Hockenmaier and Frank Keller

Better Informed Training of Latent Syntactic Features

Markus Dreyer and Jason Eisner

Sunday, 23 July 2006

Session 4a: Sentiment

- 8:25–8:50 *Get out the vote: Determining support or opposition from Congressional floor-debate transcripts*
Matt Thomas, Bo Pang and Lillian Lee
- 8:50–9:15 *Partially Supervised Coreference Resolution for Opinion Summarization through Structured Rule Learning*
Veselin Stoyanov and Claire Cardie
- 9:15–9:40 *Sentiment Retrieval using Generative Models*
Koji Eguchi and Victor Lavrenko
- 9:40–10:05 *Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis*
Hiroshi Kanayama and Tetsuya Nasukawa
- 10:05–10:30 *A Skip-Chain Conditional Random Field for Ranking Meeting Utterances by Importance*
Michel Galley

Session 4b: Language Modeling

- 8:25–8:50 *Style & Topic Language Model Adaptation Using HMM-LDA*
Bo-June (Paul) Hsu and James Glass
- 8:50–9:15 *Text data acquisition for domain-specific language models*
Abhinav Sethy, Panayiotis G. Georgiou and Shrikanth Narayanan
- 9:15–9:40 *Corrective Models for Speech Recognition of Inflected Languages*
Izhak Shafran and Keith Hall
- 9:40–10:05 *Lexicon Acquisition for Dialectal Arabic Using Transductive Learning*
Kevin Duh and Katrin Kirchhoff
- 10:05–10:30 *Arabic OCR Error Correction Using Character Segment Correction, Language Modeling, and Shallow Morphology*
Walid Magdy and Kareem Darwish

Sunday, 23 July 2006 (continued)

**Short Poster Session 2: Sentiment, WSD, Machine Learning Models and Methods,
Term and Entity Extraction (11:00–12:30)**

Partially Supervised Sense Disambiguation by Learning Sense Number from Tagged and Untagged Corpora

Zheng-Yu Niu, Dong-Hong Ji and Chew Lim Tan

Automatically Assessing Review Helpfulness

Soo-Min Kim, Patrick Pantel, Tim Chklovski and Marco Pennacchiotti

Joint Extraction of Entities and Relations for Opinion Recognition

Yejin Choi, Eric Breck and Claire Cardie

Feature Subsumption for Opinion Analysis

Ellen Riloff, Siddharth Patwardhan and Janyce Wiebe

Relevance Feedback Models for Recommendation

Masao Utiyama and Mikio Yamamoto

Random Indexing using Statistical Weight Functions

James Gorman and James R. Curran

A Hybrid Markov/Semi-Markov Conditional Random Field for Sequence Segmentation

Galen Andrew

Boosting Unsupervised Relation Extraction by Using NER

Ronen Feldman and Benjamin Rosenfeld

Short Text Authorship Attribution via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation

Conrad Sanderson and Simon Guenter

Entity Annotation based on Inverse Index Operations

Ganesh Ramakrishnan, Sreeram Balakrishnan and Sachindra Joshi

Unsupervised Information Extraction Approach Using Graph Mutual Reinforcement

Hany Hassan, Ahmed Hassan and Ossama Emam

Empirical Study on the Performance Stability of Named Entity Recognition Model across Domains

Hong Lei Guo, Li Zhang and Zhong Su

Sunday, 23 July 2006 (continued)

Session 5a: Generation, Summarization, and Lexical Semantics

- 1:45–2:10 *Statistical Ranking in Tactical Generation*
Erik Velldal and Stephan Oepen
- 2:10–2:35 *Sentence ordering with manifold-based classification in multi-document summarization*
Paul D Ji and Stephen Pulman
- 2:35–3:00 *Quality Assessment of Large Scale Knowledge Resources*
Montse Cuadros and German Rigau
- 3:00–3:25 *Graph-based Word Clustering using a Web Search Engine*
Yutaka Matsuo, Takeshi Sakaki, Kôki Uchiyama and Mitsuru Ishizuka

Session 5b: Machine Learning Models and Methods

- 1:45–2:10 *Context-Dependent Term Relations for Information Retrieval*
Jing Bai, Jian-Yun Nie and Guihong Cao
- 2:10–2:35 *Loss Minimization in Parse Reranking*
Ivan Titov and James Henderson
- 2:35–3:00 *Unsupervised Relation Disambiguation with Order Identification Capabilities*
Jinxu Chen, Donghong Ji, ChewLim Tan and Zhengyu Niu
- 3:00–3:25 *Competitive generative models with structure learning for NLP classification tasks*
Kristina Toutanova

Sunday, 23 July 2006 (continued)

Session 6a: Word Senses

4:00–4:25 *Two graph-based algorithms for state-of-the-art WSD*
Eneko Agirre, David Martínez, Oier López de Lacalle and Aitor Soroa

4:25–4:50 *Broad-Coverage Sense Disambiguation and Information Extraction with a Supersense Sequence Tagger*
Massimiliano Ciaramita and Yasemin Altun

Session 6b: Machine Learning Models and Methods

4:00–4:25 *Learning Field Compatibilities to Extract Database Records from Unstructured Text*
Michael Wick, Aron Culotta and Andrew McCallum

4:25–4:50 *Discriminative Methods for Transliteration*
Dmitry Zelenko and Chinatsu Aone

4:50–5:15 *Solving the Problem of Cascading Errors: Approximate Bayesian Inference for Linguistic Annotation Pipelines*
Jenny Rose Finkel, Christopher D. Manning and Andrew Y. Ng

