NAACL HLT 2009

Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics

Short Papers

May 31 – June 5, 2009
Boulder, Colorado

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

Sponsors:

- Rosetta Stone
- CNGL
- Microsoft Research
- Google
- AT&T
- Language Weaver
- J.D. Power
- IBM Research
- The Linguistic Data Consortium
- The Human Language Technology Center of Excellence at the Johns Hopkins University
- The Computational Language and Education Research Center at the University of Colorado at Boulder

©2009 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: 978-1-932432-42-8

Table of Contents

Cohesive Constraints in A Beam Search Phrase-based Decoder
Nguyen Bach, Stephan Vogel and Colin Cherry 1

Revisiting Optimal Decoding for Machine Translation IBM Model 4
Sebastian Riedel and James Clarke 5

Efficient Extraction of Oracle-best Translations from Hypergraphs
Zhifei Li and Sanjeev Khudanpur 9

Semantic Roles for SMT: A Hybrid Two-Pass Model
Dekai Wu and Pascale Fung 13

Comparison of Extended Lexicon Models in Search and Rescoring for SMT
Saša Hasan and Hermann Ney 17

A Simplex Armijo Downhill Algorithm for Optimizing Statistical Machine Translation Decoding Parameters
Bing Zhao and Shengyuan Chen 21

Translation Corpus Source and Size in Bilingual Retrieval
Paul McNamee, James Mayfield and Charles Nicholas 25

Large-scale Computation of Distributional Similarities for Queries
Enrique Alfonseca, Keith Hall and Silvana Hartmann 29

Text Categorization from Category Name via Lexical Reference
Libby Barak, Ido Dagan and Eyal Shnarch 33

Identifying Types of Claims in Online Customer Reviews
Shilpa Arora, Mahesh Joshi and Carolyn P. Rosé 37

Towards Automatic Image Region Annotation - Image Region Textual Coreference Resolution
Emilia Apostolova and Dina Demner-Fushman 41

TESLA: A Tool for Annotating Geospatial Language Corpora
Nate Blaylock, Bradley Swain and James Allen 45

Modeling Dialogue Structure with Adjacency Pair Analysis and Hidden Markov Models
Kristy Elizabeth Boyer, Robert Phillips, Eun Young Ha, Michael Wallis, Mladen Vouk and James Lester 49

Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue Systems
Kenji Sagae, Gwen Christian, David DeVault and David Traum 53

Spherical Discriminant Analysis in Semi-supervised Speaker Clustering
Hao Tang, Stephen Chu and Thomas Huang 57

Learning Bayesian Networks for Semantic Frame Composition in a Spoken Dialog System
Marie-Jean Meurs, Fabrice Lefèvre and Renato De Mori 61

Evaluation of a System for Noun Concepts Acquisition from Utterances about Images (SINCA) Using Daily Conversation Data
Yuzu Uchida and Kenji Araki 65

Web and Corpus Methods for Malay Count Classifier Prediction
Jeremy Nicholson and Timothy Baldwin 69

Minimum Bayes Risk Combination of Translation Hypotheses from Alternative Morphological Decompositions
Adrià de Gispert, Sami Virpioja, Mikko Kurimo and William Byrne 73

Generating Synthetic Children's Acoustic Models from Adult Models
Andreas Hagen, Bryan Pellom and Kadri Hacioglu 77

Detecting Pitch Accents at the Word, Syllable and Vowel Level
Andrew Rosenberg and Julia Hirschberg 81

Shallow Semantic Parsing for Spoken Language Understanding
Bonaventura Coppola, Alessandro Moschitti and Giuseppe Riccardi 85

Automatic Agenda Graph Construction from Human-Human Dialogs using Clustering Method
Cheongjae Lee, Sangkeun Jung, Kyungduk Kim and Gary Geunbae Lee 89

A Simple Sentence-Level Extraction Algorithm for Comparable Data
Christoph Tillmann and Jian-ming Xu 93

Learning Combination Features with L1 Regularization
Daisuke Okanohara and Jun'ichi Tsujii 97

Multi-scale Personalization for Voice Search Applications
Daniel Bolaños, Geoffrey Zweig and Patrick Nguyen 101

The Importance of Sub-Utterance Prosody in Predicting Level of Certainty
Heather Pon-Barry and Stuart Shieber 105

Using Integer Linear Programming for Detecting Speech Disfluencies
Kallirroi Georgila 109

Contrastive Summarization: An Experiment with Consumer Reviews
Kevin Lerman and Ryan McDonald 113

Topic Identification Using Wikipedia Graph Centrality
Kino Coursey and Rada Mihalcea 117

Extracting Bilingual Dictionary from Comparable Corpora with Dependency Heterogeneity
Kun Yu and Jun'ichi Tsujii 121

Domain Adaptation with Artificial Data for Semantic Parsing of Speech
Lonneke van der Plas, James Henderson and Paola Merlo 125

Extending Pronunciation Lexicons via Non-phonemic Respellings
Lucian Galescu 129

A Speech Understanding Framework that Uses Multiple Language Models and Multiple Understanding Models
Masaki Katsumaru, Mikio Nakano, Kazunori Komatani, Kotaro Funakoshi, Tetsuya Ogata and Hiroshi G. Okuno 133

Taking into Account the Differences between Actively and Passively Acquired Data: The Case of Active Learning with Support Vector Machines for Imbalanced Datasets
Michael Bloodgood and Vijay Shanker 137

Faster MT Decoding Through Pervasive Laziness
Michael Pust and Kevin Knight 141

Evaluating the Syntactic Transformations in Gold Standard Corpora for Statistical Sentence Compression
Naman K. Gupta, Sourish Chaudhuri and Carolyn P. Rosé 145

Incremental Adaptation of Speech-to-Speech Translation
Nguyen Bach, Roger Hsiao, Matthias Eck, Paisarn Charoenpornsawat, Stephan Vogel, Tanja Schultz, Ian Lane, Alex Waibel and Alan Black 149

Name Perplexity
Octavian Popescu 153

Answer Credibility: A Language Modeling Approach to Answer Validation
Protima Banerjee and Hyoil Han 157

Exploiting Named Entity Classes in CCG Surface Realization
Rajakrishnan Rajkumar, Michael White and Dominic Espinosa 161

Search Engine Adaptation by Feedback Control Adjustment for Time-sensitive Query
Ruiqiang Zhang, Yi Chang, Zhaohui Zheng, Donald Metzler and Jian-yun Nie 165

A Local Tree Alignment-based Soft Pattern Matching Approach for Information Extraction
Seokhwan Kim, Minwoo Jeong and Gary Geunbae Lee 169

Classifying Factored Genres with Part-of-Speech Histograms
Sergey Feldman, Marius Marin, Julie Medero and Mari Ostendorf 173

Towards Effective Sentence Simplification for Automatic Processing of Biomedical Text
Siddhartha Jonnalagadda, Luis Tari, Jörg Hakenberg, Chitta Baral and Graciela Gonzalez 177

Improving SCL Model for Sentiment-Transfer Learning
Songbo Tan and Xueqi Cheng 181

MICA: A Probabilistic Dependency Parser Based on Tree Insertion Grammars (Application Note)
Srinivas Bangalore, Pierre Boullier, Alexis Nasr, Owen Rambow and Benoît Sagot 185

Lexical and Syntactic Adaptation and Their Impact in Deployed Spoken Dialog Systems
Svetlana Stoyanchev and Amanda Stent 189

Analysing Recognition Errors in Unlimited-Vocabulary Speech Recognition
Teemu Hirsimäki and Mikko Kurimo 193

The independence of dimensions in multidimensional dialogue act annotation
Volha Petukhova and Harry Bunt 197

Improving Coreference Resolution by Using Conversational Metadata
Xiaoqiang Luo, Radu Florian and Todd Ward 201

Using N-gram based Features for Machine Translation System Combination
Yong Zhao and Xiaodong He 205

Language Specific Issue and Feature Exploration in Chinese Event Extraction
Zheng Chen and Heng Ji 209

Improving A Simple Bigram HMM Part-of-Speech Tagger by Latent Annotation and Self-Training
Zhongqiang Huang, Vladimir Eidelman and Mary Harper 213

Statistical Post-Editing of a Rule-Based Machine Translation System
Antonio-L. Lagarda, Vicent Alabau, Francisco Casacuberta, Roberto Silva and Enrique Díaz-de-Liaño 217

On the Importance of Pivot Language Selection for Statistical Machine Translation
Michael Paul, Hirofumi Yamamoto, Eiichiro Sumita and Satoshi Nakamura 221

Tree Linearization in English: Improving Language Model Based Approaches
Katja Filippova and Michael Strube 225

Determining the position of adverbial phrases in English
Huayan Zhong and Amanda Stent 229

Estimating and Exploiting the Entropy of Sense Distributions
Peng Jin, Diana McCarthy, Rob Koeling and John Carroll 233

Semantic Classification with WordNet Kernels
Diarmuid Ó Séaghdha 237

Sentence Boundary Detection and the Problem with the U.S.
Dan Gillick 241

Quadratic Features and Deep Architectures for Chunking
Joseph Turian, James Bergstra and Yoshua Bengio 245

Active Zipfian Sampling for Statistical Parser Training
Onur Çobanoğlu 249

Combining Constituent Parsers
Victoria Fossum and Kevin Knight 253

Recognising the Predicate-argument Structure of Tagalog
Meladel Mistica and Timothy Baldwin 257

Reverse Revision and Linear Tree Combination for Dependency Parsing
Giuseppe Attardi and Felice Dell'Orletta 261

Anchored Speech Recognition for Question Answering
Sibel Yaman, Gokan Tur, Dimitra Vergyri, Dilek Hakkani-Tur, Mary Harper and Wen Wang . 265

Score Distribution Based Term Specific Thresholding for Spoken Term Detection
Dogan Can and Murat Saraclar 269

Automatic Chinese Abbreviation Generation Using Conditional Random Field
Dong Yang, Yi-Cheng Pan and Sadaoki Furui 273

Fast decoding for open vocabulary spoken term detection
Bhuvana Ramabhadran, Abhinav Sethy, Jonathan Mamou, Brian Kingsbury and Upendra Chaudhari 277

Tightly coupling Speech Recognition and Search
Taniya Mishra and Srinivas Bangalore 281

Conference Program Overview

Monday, June 1, 2009

9:00–10:10 Plenary Session – Invited Talk by Antonio Torralba: *Understanding Visual Scenes*

10:40–11:20 Session 1A: Semantics
Session 1B: Multilingual Processing / Morphology and Phonology
Session 1C: Syntax and Parsing
Student Research Workshop Session 1

2:00–3:30 Short Paper Presentations:
Session 2A: Machine Translation
Session 2B: Information Retrieval / Information Extraction / Sentiment
Session 2C: Dialog / Speech / Semantics
Student Research Workshop Session 2

4:00–5:40 Session 3A: Machine Translation
Session 3B: Semantics
Session 3C: Information Retrieval
Student Research Workshop Session 3

6:30–9:30 Poster and Demo Session
Student Research Workshop Poster Session

Tuesday, June 2, 2009

9:00-10:10 Plenary Session: Paper Award Presentations

10:10–11:40 Session 4A: Machine Translation
Session 4B: Sentiment Analysis / Information Extraction
Session 4C: Machine Learning / Morphology and Phonology

2:00–3:30 Short Paper Presentations:
Session 5A: Machine Translation / Generation / Semantics
Session 5B: Machine Learning / Syntax
Session 5C: SPECIAL SESSION – Speech Indexing and Retrieval

4:00–5:15 Session 6A: Syntax and Parsing
Session 6B: Discourse and Summarization
Session 6C: Spoken Language Systems

Wednesday, June 3, 2009

9:00–10:10	Plenary Session – Invited Talk by Dan Jurafsky: *Ketchup, Espresso, and Chocolate Chip Cookies: Travels in the Language of Food*
10:40–12:20	Session 7A: Machine Translation Session 7B: Speech Recognition and Language Modeling Session 7C: Sentiment Analysis
12:40-1:40	Panel Discussion: *Emerging Application Areas in Computational Linguistics*
1:40–2:30	NAACL Business Meeting
2:30–3:45	Session 8A: Large-scale NLP Session 8B: Syntax and Parsing Session 8C: Discourse and Summarization
4:15–5:30	Session 9A: Machine Learning Session 9B: Dialog Systems Session 9C: Syntax and Parsing

Conference Program

Monday, June 1, 2009

Plenary Session

9:00–10:10 Welcome and Invited Talk: *Understanding Visual Scenes*
Antonio Torralba

10:10–10:40 **Break**

Session 1A: Semantics

Note: all full papers are located in the Main volume of the proceedings

10:40–11:05 *Subjectivity Recognition on Word Senses via Semi-supervised Mincuts*
Fangzhong Su and Katja Markert

11:05–11:30 *Integrating Knowledge for Subjectivity Sense Labeling*
Yaw Gyamfi, Janyce Wiebe, Rada Mihalcea and Cem Akkaya

11:30–11:55 *A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches*
Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Pasca and Aitor Soroa

11:55–12:20 *A Fully Unsupervised Word Sense Disambiguation Method Using Dependency Knowledge*
Ping Chen, Wei Ding, Chris Bowes and David Brown

Session 1B: Multilingual Processing / Morphology and Phonology

10:40–11:05 *Learning Phoneme Mappings for Transliteration without Parallel Data*
Sujith Ravi and Kevin Knight

11:05–11:30 *A Corpus-Based Approach for the Prediction of Language Impairment in Monolingual English and Spanish-English Bilingual Children*
Keyur Gabani, Melissa Sherman, Thamar Solorio, Yang Liu, Lisa Bedore and Elizabeth Peña

11:30–11:55 *A Discriminative Latent Variable Chinese Segmenter with Hybrid Word/Character Information*
Xu Sun, Yaozhong Zhang, Takuya Matsuzaki, Yoshimasa Tsuruoka and Jun'ichi Tsujii

11:55–12:20 *Improved Reconstruction of Protolanguage Word Forms*
Alexandre Bouchard-Côté, Thomas L. Griffiths and Dan Klein

Monday, June 1, 2009 (continued)

Session 1C: Syntax and Parsing

10:40–11:05 *Shared Logistic Normal Distributions for Soft Parameter Tying in Unsupervised Grammar Induction*
Shay Cohen and Noah A. Smith

11:05–11:30 *Adding More Languages Improves Unsupervised Multilingual Part-of-Speech Tagging: a Bayesian Non-Parametric Approach*
Benjamin Snyder, Tahira Naseem, Jacob Eisenstein and Regina Barzilay

11:30–11:55 *Efficiently Parsable Extensions to Tree-Local Multicomponent TAG*
Rebecca Nesson and Stuart Shieber

11:55–12:20 *Improving Unsupervised Dependency Parsing with Richer Contexts and Smoothing*
William P. Headden III, Mark Johnson and David McClosky

Student Research Workshop Session 1:

Note: all student research workshop papers are located in the Companion volume of the proceedings

10:40–11:10 *Classifier Combination Techniques Applied to Coreference Resolution*
Smita Vemulapalli, Xiaoqiang Luo, John F. Pitrelli and Imed Zitouni

11:15–11:45 *Solving the "Who's Mark Johnson Puzzle": Information Extraction Based Cross Document Coreference*
Jian Huang, Sarah M. Taylor, Jonathan L. Smith, Konstantinos A. Fotiadis and C. Lee Giles

11:50–12:20 *Exploring Topic Continuation Follow-up Questions using Machine Learning*
Manuel Kirschner and Raffaella Bernardi

12:20–2:00 **Lunch Break**

Monday, June 1, 2009 (continued)

Session 2A: Short Paper Presentations: Machine Translation

2:00–2:15 *Cohesive Constraints in A Beam Search Phrase-based Decoder*
Nguyen Bach, Stephan Vogel and Colin Cherry

2:15–2:30 *Revisiting Optimal Decoding for Machine Translation IBM Model 4*
Sebastian Riedel and James Clarke

2:30–2:45 *Efficient Extraction of Oracle-best Translations from Hypergraphs*
Zhifei Li and Sanjeev Khudanpur

2:45–3:00 *Semantic Roles for SMT: A Hybrid Two-Pass Model*
Dekai Wu and Pascale Fung

3:00–3:15 *Comparison of Extended Lexicon Models in Search and Rescoring for SMT*
Saša Hasan and Hermann Ney

3:15–3:30 *A Simplex Armijo Downhill Algorithm for Optimizing Statistical Machine Translation Decoding Parameters*
Bing Zhao and Shengyuan Chen

Session 2B: Short Paper Presentations: Information Retrieval / Information Extraction / Sentiment

2:00–2:15 *Translation Corpus Source and Size in Bilingual Retrieval*
Paul McNamee, James Mayfield and Charles Nicholas

2:15–2:30 *Large-scale Computation of Distributional Similarities for Queries*
Enrique Alfonseca, Keith Hall and Silvana Hartmann

2:30–2:45 *Text Categorization from Category Name via Lexical Reference*
Libby Barak, Ido Dagan and Eyal Shnarch

2:45–3:00 *Identifying Types of Claims in Online Customer Reviews*
Shilpa Arora, Mahesh Joshi and Carolyn P. Rosé

3:00–3:15 *Towards Automatic Image Region Annotation - Image Region Textual Coreference Resolution*
Emilia Apostolova and Dina Demner-Fushman

Monday, June 1, 2009 (continued)

3:15–3:30 *TESLA: A Tool for Annotating Geospatial Language Corpora*
Nate Blaylock, Bradley Swain and James Allen

Session 2C: Short Paper Presentations: Dialog / Speech / Semantics

2:00–2:15 *Modeling Dialogue Structure with Adjacency Pair Analysis and Hidden Markov Models*
Kristy Elizabeth Boyer, Robert Phillips, Eun Young Ha, Michael Wallis, Mladen Vouk and James Lester

2:15–2:30 *Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue Systems*
Kenji Sagae, Gwen Christian, David DeVault and David Traum

2:30–2:45 *Spherical Discriminant Analysis in Semi-supervised Speaker Clustering*
Hao Tang, Stephen Chu and Thomas Huang

2:45–3:00 *Learning Bayesian Networks for Semantic Frame Composition in a Spoken Dialog System*
Marie-Jean Meurs, Fabrice Lefèvre and Renato De Mori

3:00–3:15 *Evaluation of a System for Noun Concepts Acquisition from Utterances about Images (SINCA) Using Daily Conversation Data*
Yuzu Uchida and Kenji Araki

3:15–3:30 *Web and Corpus Methods for Malay Count Classifier Prediction*
Jeremy Nicholson and Timothy Baldwin

Monday, June 1, 2009 (continued)

Student Research Workshop Session 2

Note: all student research workshop papers are located in the Companion volume of the proceedings

2:00–2:30 *Sentence Realisation from Bag of Words with Dependency Constraints*
Karthik Gali and Sriram Venkatapathy

2:35–3:05 *Using Language Modeling to Select Useful Annotation Data*
Dmitriy Dligach and Martha Palmer

3:30–4:00 **Break**

Session 3A: Machine Translation

4:00–4:25 *Context-Dependent Alignment Models for Statistical Machine Translation*
Jamie Brunning, Adrià de Gispert and William Byrne

4:25–4:50 *Graph-based Learning for Statistical Machine Translation*
Andrei Alexandrescu and Katrin Kirchhoff

4:50–5:15 *Intersecting Multilingual Data for Faster and Better Statistical Translations*
Yu Chen, Martin Kay and Andreas Eisele

5:15–5:40 No Presentation

Session 3B: Semantics

4:00–4:25 *Without a 'doubt'? Unsupervised Discovery of Downward-Entailing Operators*
Cristian Danescu-Niculescu-Mizil, Lillian Lee and Richard Ducott

4:25–4:50 *The Role of Implicit Argumentation in Nominal SRL*
Matthew Gerber, Joyce Chai and Adam Meyers

4:50–5:15 *Jointly Identifying Predicates, Arguments and Senses using Markov Logic*
Ivan Meza-Ruiz and Sebastian Riedel

5:15–5:40 *Structured Generative Models for Unsupervised Named-Entity Clustering*
Micha Elsner, Eugene Charniak and Mark Johnson

Monday, June 1, 2009 (continued)

Session 3C: Information Retrieval

4:00–4:25 *Hierarchical Dirichlet Trees for Information Retrieval*
Gholamreza Haffari and Yee Whye Teh

4:25–4:50 *Phrase-Based Query Degradation Modeling for Vocabulary-Independent Ranked Utterance Retrieval*
J. Scott Olsson and Douglas W. Oard

4:50–5:15 *Japanese Query Alteration Based on Lexical Semantic Similarity*
Masato Hagiwara and Hisami Suzuki

5:15–5:40 *Context-based Message Expansion for Disentanglement of Interleaved Text Conversations*
Lidan Wang and Douglas Oard

Student Research Workshop Session 3

Note: all student research workshop papers are located in the Companion volume of the proceedings

4:00–4:30 *Pronunciation Modeling in Spelling Correction for Writers of English as a Foreign Language*
Adriane Boyd

4:35–5:05 *Building a Semantic Lexicon of English Nouns via Bootstrapping*
Ting Qian, Benjamin Van Durme and Lenhart Schubert

5:10–5:40 *Multiple Word Alignment with Profile Hidden Markov Models*
Aditya Bhargava and Grzegorz Kondrak

6:30–9:30 **Poster and Demo Session**

Note: all demo abstracts are located in the Companion volume of the proceedings

Minimum Bayes Risk Combination of Translation Hypotheses from Alternative Morphological Decompositions
Adrià de Gispert, Sami Virpioja, Mikko Kurimo and William Byrne

Generating Synthetic Children's Acoustic Models from Adult Models
Andreas Hagen, Bryan Pellom and Kadri Hacioglu

Monday, June 1, 2009 (continued)

Detecting Pitch Accents at the Word, Syllable and Vowel Level
Andrew Rosenberg and Julia Hirschberg

Shallow Semantic Parsing for Spoken Language Understanding
Bonaventura Coppola, Alessandro Moschitti and Giuseppe Riccardi

Automatic Agenda Graph Construction from Human-Human Dialogs using Clustering Method
Cheongjae Lee, Sangkeun Jung, Kyungduk Kim and Gary Geunbae Lee

A Simple Sentence-Level Extraction Algorithm for Comparable Data
Christoph Tillmann and Jian-ming Xu

Learning Combination Features with L1 Regularization
Daisuke Okanohara and Jun'ichi Tsujii

Multi-scale Personalization for Voice Search Applications
Daniel Bolaños, Geoffrey Zweig and Patrick Nguyen

The Importance of Sub-Utterance Prosody in Predicting Level of Certainty
Heather Pon-Barry and Stuart Shieber

Using Integer Linear Programming for Detecting Speech Disfluencies
Kallirroi Georgila

Contrastive Summarization: An Experiment with Consumer Reviews
Kevin Lerman and Ryan McDonald

Topic Identification Using Wikipedia Graph Centrality
Kino Coursey and Rada Mihalcea

Extracting Bilingual Dictionary from Comparable Corpora with Dependency Heterogeneity
Kun Yu and Jun'ichi Tsujii

Domain Adaptation with Artificial Data for Semantic Parsing of Speech
Lonneke van der Plas, James Henderson and Paola Merlo

Monday, June 1, 2009 (continued)

Extending Pronunciation Lexicons via Non-phonemic Respellings
Lucian Galescu

A Speech Understanding Framework that Uses Multiple Language Models and Multiple Understanding Models
Masaki Katsumaru, Mikio Nakano, Kazunori Komatani, Kotaro Funakoshi, Tetsuya Ogata and Hiroshi G. Okuno

Taking into Account the Differences between Actively and Passively Acquired Data: The Case of Active Learning with Support Vector Machines for Imbalanced Datasets
Michael Bloodgood and Vijay Shanker

Faster MT Decoding Through Pervasive Laziness
Michael Pust and Kevin Knight

Evaluating the Syntactic Transformations in Gold Standard Corpora for Statistical Sentence Compression
Naman K. Gupta, Sourish Chaudhuri and Carolyn P. Rosé

Incremental Adaptation of Speech-to-Speech Translation
Nguyen Bach, Roger Hsiao, Matthias Eck, Paisarn Charoenpornsawat, Stephan Vogel, Tanja Schultz, Ian Lane, Alex Waibel and Alan Black

Name Perplexity
Octavian Popescu

Answer Credibility: A Language Modeling Approach to Answer Validation
Protima Banerjee and Hyoil Han

Exploiting Named Entity Classes in CCG Surface Realization
Rajakrishnan Rajkumar, Michael White and Dominic Espinosa

Search Engine Adaptation by Feedback Control Adjustment for Time-sensitive Query
Ruiqiang Zhang, Yi Chang, Zhaohui Zheng, Donald Metzler and Jian-yun Nie

A Local Tree Alignment-based Soft Pattern Matching Approach for Information Extraction
Seokhwan Kim, Minwoo Jeong and Gary Geunbae Lee

Classifying Factored Genres with Part-of-Speech Histograms
Sergey Feldman, Marius Marin, Julie Medero and Mari Ostendorf

Towards Effective Sentence Simplification for Automatic Processing of Biomedical Text
Siddhartha Jonnalagadda, Luis Tari, Jörg Hakenberg, Chitta Baral and Graciela Gonzalez

Improving SCL Model for Sentiment-Transfer Learning
Songbo Tan and Xueqi Cheng

MICA: A Probabilistic Dependency Parser Based on Tree Insertion Grammars (Application Note)
Srinivas Bangalore, Pierre Boullier, Alexis Nasr, Owen Rambow and Benoît Sagot

Lexical and Syntactic Adaptation and Their Impact in Deployed Spoken Dialog Systems
Svetlana Stoyanchev and Amanda Stent

Analysing Recognition Errors in Unlimited-Vocabulary Speech Recognition
Teemu Hirsimäki and Mikko Kurimo

The independence of dimensions in multidimensional dialogue act annotation
Volha Petukhova and Harry Bunt

Improving Coreference Resolution by Using Conversational Metadata
Xiaoqiang Luo, Radu Florian and Todd Ward

Using N-gram based Features for Machine Translation System Combination
Yong Zhao and Xiaodong He

Language Specific Issue and Feature Exploration in Chinese Event Extraction
Zheng Chen and Heng Ji

Improving A Simple Bigram HMM Part-of-Speech Tagger by Latent Annotation and Self-Training
Zhongqiang Huang, Vladimir Eidelman and Mary Harper

6:30–9:30 **Student Research Workshop Poster Session**

Note: all student research workshop papers are located in the Companion volume of the proceedings

Also: All papers presented in the morning and afternoon sessions of the student research workshop will also be shown as posters.

Monday, June 1, 2009 (continued)

Using Emotion to Gain Rapport in a Spoken Dialog System
Jaime Acosta

Interactive Annotation Learning with Indirect Feature Voting
Shilpa Arora and Eric Nyberg

Loss-Sensitive Discriminative Training of Machine Transliteration Models
Kedar Bellare, Koby Crammer and Dayne Freitag

Syntactic Tree-based Relation Extraction Using a Generalization of Collins and Duffy Convolution Tree Kernel
Mahdy Khayyamian, Seyed Abolghasem Mirroshandel and Hassan Abolhassani

Towards Building a Competitive Opinion Summarization System: Challenges and Keys
Elena Lloret, Alexandra Balahur, Manuel Palomar and Andres Montoyo

Domain-Independent Shallow Sentence Ordering
Thade Nahnsen

Towards Unsupervised Recognition of Dialogue Acts
Nicole Novielli and Carlo Strapparava

Modeling Letter-to-Phoneme Conversion as a Phrase Based Statistical Machine Translation Problem with Minimum Error Rate Training
Taraka Rama, Anil Kumar Singh and Sudheer Kolachina

Disambiguation of Preposition Sense Using Linguistically Motivated Features
Stephen Tratz and Dirk Hovy

Tuesday, June 2, 2009

Plenary Session

9:00–9:10 Paper Awards

9:10–9:40 *Unsupervised Morphological Segmentation with Log-Linear Models*
Hoifung Poon, Colin Cherry and Kristina Toutanova

9:40–10:10 *11,001 New Features for Statistical Machine Translation*
David Chiang, Kevin Knight and Wei Wang

10:10-10:40 **Break**

Session 4A: Machine Translation

10:10–10:35 *Efficient Parsing for Transducer Grammars*
John DeNero, Mohit Bansal, Adam Pauls and Dan Klein

10:35–10:50 *Preference Grammars: Softening Syntactic Constraints to Improve Statistical Machine Translation*
Ashish Venugopal, Andreas Zollmann, Noah Smith and Stephan Vogel

10:50–11:15 *Using a Dependency Parser to Improve SMT for Subject-Object-Verb Languages*
Peng Xu, Jaeho Kang, Michael Ringgaard and Franz Och

11:15–11:40 *Learning Bilingual Linguistic Reordering Model for Statistical Machine Translation*
Han-Bin Chen, Jian-Cheng Wu and Jason S. Chang

Tuesday, June 2, 2009 (continued)

Session 4B: Sentiment Analysis / Information Extraction

10:10–10:35 *May All Your Wishes Come True: A Study of Wishes and How to Recognize Them*
Andrew Goldberg, Nathanael Fillmore, David Andrzejewski, Zhiting Xu, Bryan Gibson and Xiaojin Zhu

10:35–10:50 *Predicting Risk from Financial Reports with Regression*
Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi and Noah A. Smith

10:50–11:15 *Domain Adaptation with Latent Semantic Association for Named Entity Recognition*
Honglei Guo, Huijia Zhu, Zhili Guo, Xiaoxun Zhang, Xian Wu and Zhong Su

11:15–11:40 *Semi-Automatic Entity Set Refinement*
Vishnu Vyas and Patrick Pantel

Session 4C: Machine Learning / Morphology and Phonology

10:10–10:35 *Unsupervised Constraint Driven Learning For Transliteration Discovery*
Ming-Wei Chang, Dan Goldwasser, Dan Roth and Yuancheng Tu

10:35–10:50 *On the Syllabification of Phonemes*
Susan Bartlett, Grzegorz Kondrak and Colin Cherry

10:50–11:15 *Improving nonparameteric Bayesian inference: experiments on unsupervised word segmentation with adaptor grammars*
Mark Johnson and Sharon Goldwater

11:15–11:40 No Presentation

12:20–2:00 **Lunch Break**

Tuesday, June 2, 2009 (continued)

Session 5A: Short Paper Presentations: Machine Translation / Generation / Semantics

2:00–2:15 *Statistical Post-Editing of a Rule-Based Machine Translation System*
Antonio-L. Lagarda, Vicent Alabau, Francisco Casacuberta, Roberto Silva and Enrique Díaz-de-Liaño

2:15–2:30 *On the Importance of Pivot Language Selection for Statistical Machine Translation*
Michael Paul, Hirofumi Yamamoto, Eiichiro Sumita and Satoshi Nakamura

2:30–2:45 *Tree Linearization in English: Improving Language Model Based Approaches*
Katja Filippova and Michael Strube

2:45–3:00 *Determining the position of adverbial phrases in English*
Huayan Zhong and Amanda Stent

3:00–3:15 *Estimating and Exploiting the Entropy of Sense Distributions*
Peng Jin, Diana McCarthy, Rob Koeling and John Carroll

3:15–3:30 *Semantic Classification with WordNet Kernels*
Diarmuid Ó Séaghdha

Session 5B: Short Paper Presentations: Machine Learning / Syntax

2:00–2:15 *Sentence Boundary Detection and the Problem with the U.S.*
Dan Gillick

2:15–2:30 *Quadratic Features and Deep Architectures for Chunking*
Joseph Turian, James Bergstra and Yoshua Bengio

2:30–2:45 *Active Zipfian Sampling for Statistical Parser Training*
Onur Çobanoğlu

2:45–3:00 *Combining Constituent Parsers*
Victoria Fossum and Kevin Knight

3:00–3:15 *Recognising the Predicate-argument Structure of Tagalog*
Meladel Mistica and Timothy Baldwin

Tuesday, June 2, 2009 (continued)

3:15–3:30 *Reverse Revision and Linear Tree Combination for Dependency Parsing*
Giuseppe Attardi and Felice Dell'Orletta

Session 5C: Short Paper Presentations: SPECIAL SESSION – Speech Indexing and Retrieval

2:00–2:15 *Introduction to the Special Session on Speech Indexing and Retrieval*

2:15–2:30 *Anchored Speech Recognition for Question Answering*
Sibel Yaman, Gokan Tur, Dimitra Vergyri, Dilek Hakkani-Tur, Mary Harper and Wen Wang

2:30–2:45 *Score Distribution Based Term Specific Thresholding for Spoken Term Detection*
Dogan Can and Murat Saraclar

2:45–3:00 *Automatic Chinese Abbreviation Generation Using Conditional Random Field*
Dong Yang, Yi-Cheng Pan and Sadaoki Furui

3:00–3:15 *Fast decoding for open vocabulary spoken term detection*
Bhuvana Ramabhadran, Abhinav Sethy, Jonathan Mamou, Brian Kingsbury and Upendra Chaudhari

3:15–3:30 *Tightly coupling Speech Recognition and Search*
Taniya Mishra and Srinivas Bangalore

3:30–4:00 **Break**

Session 6A: Syntax and Parsing

4:00–4:25 *Joint Parsing and Named Entity Recognition*
Jenny Rose Finkel and Christopher D. Manning

4:25–4:50 *Minimal-length linearizations for mildly context-sensitive dependency trees*
Y. Albert Park and Roger Levy

4:50–5:15 *Positive Results for Parsing with a Bounded Stack using a Model-Based Right-Corner Transform*
William Schuler

Tuesday, June 2, 2009 (continued)

Session 6B: Discourse and Summarization

4:00–4:25 *Hierarchical Text Segmentation from Multi-Scale Lexical Cohesion*
Jacob Eisenstein

4:25–4:50 *Exploring Content Models for Multi-Document Summarization*
Aria Haghighi and Lucy Vanderwende

4:50–5:15 *Global Models of Document Structure using Latent Permutations*
Harr Chen, S.R.K. Branavan, Regina Barzilay and David R. Karger

Session 6C: Spoken Language Systems

4:00–4:25 *Assessing and Improving the Performance of Speech Recognition for Incremental Systems*
Timo Baumann, Michaela Atterer and David Schlangen

4:25–4:50 *Geo-Centric Language Models for Local Business Voice Search*
Amanda Stent, Ilija Zeljkovic, Diamantino Caseiro and Jay Wilpon

4:50–5:15 *Improving the Arabic Pronunciation Dictionary for Phone and Word Recognition with Linguistically-Based Pronunciation Rules*
Fadi Biadsy, Nizar Habash and Julia Hirschberg

Wednesday, June 3, 2009

Plenary Session

9:00–10:10 Invited Talk: *Ketchup, Espresso, and Chocolate Chip Cookies: Travels in the Language of Food*
Dan Jurafsky

10:10–10:40 **Break**

Wednesday, June 3, 2009 (continued)

Session 7A: Machine Translation

10:40–11:05 *Using a maximum entropy model to build segmentation lattices for MT*
Chris Dyer

11:05–11:30 *Active Learning for Statistical Phrase-based Machine Translation*
Gholamreza Haffari, Maxim Roy and Anoop Sarkar

11:30–11:55 *Semi-Supervised Lexicon Mining from Parenthetical Expressions in Monolingual Web Pages*
Xianchao Wu, Naoaki Okazaki and Jun'ichi Tsujii

11:55–12:20 *Hierarchical Phrase-Based Translation with Weighted Finite State Transducers*
Gonzalo Iglesias, Adrià de Gispert, Eduardo R. Banga and William Byrne

Session 7B: Speech Recognition and Language Modeling

10:40–11:05 *Improved pronunciation features for construct-driven assessment of non-native spontaneous speech*
Lei Chen, Klaus Zechner and Xiaoming Xi

11:05–11:30 *Performance Prediction for Exponential Language Models*
Stanley Chen

11:30–11:55 *Tied-Mixture Language Modeling in Continuous Space*
Ruhi Sarikaya, Mohamed Afify and Brian Kingsbury

11:55–12:20 *Shrinking Exponential Language Models*
Stanley Chen

Wednesday, June 3, 2009 (continued)

Session 7C: Sentiment Analysis

10:40–11:05 *Predicting Response to Political Blog Posts with Topic Models*
Tae Yano, William W. Cohen and Noah A. Smith

11:05–11:30 *An Iterative Reinforcement Approach for Fine-Grained Opinion Mining*
Weifu Du and Songbo Tan

11:30–11:55 *For a few dollars less: Identifying review pages sans human labels*
Luciano Barbosa, Ravi Kumar, Bo Pang and Andrew Tomkins

11:55–12:20 *More than Words: Syntactic Packaging and Implicit Sentiment*
Stephan Greene and Philip Resnik

12:20–1:40 **Lunch Break**

12:40-1:40 Panel Discussion: *Emerging Application Areas in Computational Linguistics*
Chaired by Bill Dolan, Microsoft
Panelists: Jill Burstein, Educational Testing Service; Joel Tetreault, Educational Testing Service; Patrick Pantel, Yahoo; Andy Hickl, Language Computer Corporation + Swingly

1:40–2:30 NAACL Business Meeting

Session 8A: Large-scale NLP

2:30–2:55 *Streaming for large scale NLP: Language Modeling*
Amit Goyal, Hal Daume III and Suresh Venkatasubramanian

2:55–3:20 *The Effect of Corpus Size on Case Frame Acquisition for Discourse Analysis*
Ryohei Sasano, Daisuke Kawahara and Sadao Kurohashi

3:20–3:45 *Semantic-based Estimation of Term Informativeness*
Kirill Kireyev

Wednesday, June 3, 2009 (continued)

Session 8B: Syntax and Parsing

2:30–2:55 *Optimal Reduction of Rule Length in Linear Context-Free Rewriting Systems*
Carlos Gómez-Rodríguez, Marco Kuhlmann, Giorgio Satta and David Weir

2:55–3:20 *Inducing Compact but Accurate Tree-Substitution Grammars*
Trevor Cohn, Sharon Goldwater and Phil Blunsom

3:20–3:45 *Hierarchical Search for Parsing*
Adam Pauls and Dan Klein

Session 8C: Discourse and Summarization

2:30–2:55 *An effective Discourse Parser that uses Rich Linguistic Information*
Rajen Subba and Barbara Di Eugenio

2:55–3:20 *Graph-Cut-Based Anaphoricity Determination for Coreference Resolution*
Vincent Ng

3:20–3:45 *Using Citations to Generate surveys of Scientific Paradigms*
Saif Mohammad, Bonnie Dorr, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishan, Vahed Qazvinian, Dragomir Radev and David Zajic

3:45–4:15 **Break**

Session 9A: Machine Learning

4:15–4:40 *Non-Parametric Bayesian Areal Linguistics*
Hal Daume III

4:40–5:05 *Hierarchical Bayesian Domain Adaptation*
Jenny Rose Finkel and Christopher D. Manning

5:05–5:30 *Online EM for Unsupervised Models*
Percy Liang and Dan Klein

Wednesday, June 3, 2009 (continued)

Session 9B: Dialog Systems

4:15–4:40 *Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts*
Feifan Liu, Deana Pennell, Fei Liu and Yang Liu

4:40–5:05 *A Finite-State Turn-Taking Model for Spoken Dialog Systems*
Antoine Raux and Maxine Eskenazi

5:05–5:30 *Extracting Social Meaning: Identifying Interactional Style in Spoken Conversation*
Dan Jurafsky, Rajesh Ranganath and Dan McFarland

Session 9C: Syntax and Parsing

4:15–4:40 *Linear Complexity Context-Free Parsing Pipelines via Chart Constraints*
Brian Roark and Kristy Hollingshead

4:40–5:05 *Improved Syntactic Models for Parsing Speech with Repairs*
Tim Miller

5:05–5:30 *A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs*
Klinton Bicknell and Roger Levy