Main Conference Program

Monday, June 5

9:00–9:10   Opening Session

9:10–10:10   **Keynote Speaker I: Joshua Goodman**
               Email and Spam and Spim and Spat

10:10–10:40   Break

Machine Translation I

10:40–11:05   *Capitalizing Machine Translation*
               Wei Wang, Kevin Knight and Daniel Marcu

11:05–11:30   *Do we need phrases? Challenging the conventional wisdom in Statistical Machine Translation*
               Chris Quirk and Arul Menezes

11:30–11:55   *Improved Statistical Machine Translation Using Paraphrases*
               Chris Callison-Burch, Philipp Koehn and Miles Osborne

11:55–12:20   *Segment Choice Models: Feature-Rich Models for Global Distortion in Statistical Machine Translation*
               Roland Kuhn, Denis Yuen, Michel Simard, Patrick Paul, George Foster, Eric Joanis and Howard Johnson

Inference and Entailment

10:40–11:05   *Effectively Using Syntax for Recognizing False Entailment*
               Rion Snow, Lucy Vanderwende and Arul Menezes

11:05–11:30   *Learning to recognize features of valid textual entailments*
               Bill MacCartney, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer and Christopher D. Manning

11:30–11:55   *Acquisition of Verb Entailment from Text*
               Viktor Pekar

11:55–12:20   *Acquiring Inference Rules with Temporal Constraints by Using Japanese Coordinated Sentences and Noun-Verb Co-occurrences*
               Kentaro Torisawa

Named Entity Recognition

10:40–11:05   *Role of Local Context in Automatic Deidentification of Ungrammatical, Fragmented Text*
Tawanda Sibanda, Ozlem Uzuner and Ozlem Uzuner

11:05–11:30   *Exploiting Domain Structure for Named Entity Recognition*
Jing Jiang and ChengXiang Zhai

11:30–11:55   *Named Entity Transliteration and Discovery from Multilingual Comparable Corpora*
Alexandre Klementiev and Dan Roth

11:55–12:20   *Reducing Weight Undertraining in Structured Discriminative Learning*
Charles Sutton, Michael Sindelar and Andrew McCallum

12:20–1:50   Lunch

Short Papers: Machine Translation, Multi-Lingual Speech

1:50–2:05   *Spectral Clustering for Example Based Machine Translation*
Rashmi Gangadharaiah, Ralf Brown and Jaime Carbonell

2:05–2:20   *Bridging the Inflection Morphology Gap for Arabic Statistical Machine Translation*
Andreas Zollmann, Venugopal Ashish and Vogel Stephan

2:20–2:35   *Arabic Preprocessing Schemes for Statistical Machine Translation*
Nizar Habash and Fatiha Sadat

2:35–2:50   *Thai Grapheme-Based Speech Recognition*
Paisarn Charoenpornsawat, Sanjika Hewavitharana and Tanja Schultz

2:50–3:05   *Story Segmentation of Broadcast News in English, Mandarin and Arabic*
Andrew Rosenberg and Julia Hirschberg

3:05–3:20   *Word Pronunciation Disambiguation using the Web*
Eiichiro Sumita and Fumiaki Sugaya

Short Papers: Discourse/Dialogue

1:50–2:05   *Agreement/Disagreement Classification: Exploiting Unlabeled Data using Contrast Classifiers*
Sangyun Hahn, Richard Ladner and Mari Ostendorf

2:05–2:20   *Using Phrasal Patterns to Identify Discourse Relations*
Manami Saito, Kazuhide Yamamoto and Satoshi Sekine

2:20–2:35   *Evaluating Centering for Sentence Ordering in Two New Domains*
Nikiforos Karamanis

2:35–2:50   *Computational Modelling of Structural Priming in Dialogue*
David Reitter, Frank Keller and Johanna D. Moore

2:50–3:05   *Museli: A Multi-Source Evidence Integration Approach to Topic Segmentation of Spontaneous Dialogue*
Jaime Arguello and Carolyn Rose

3:05–3:20   *Automatic Recognition of Personality in Conversation*
Françis Mairesse and Marilyn Walker

Short Papers: Retrieval, Language Models

1:50–2:05   *Using the Web to Disambiguate Acronyms*
Eiichiro Sumita and Fumiaki Sugaya

2:05–2:20   *Lycos Retriever: An Information Fusion Engine*
Brian Ulicny

2:20–2:35   *BioEx: A Novel User-Interface that Accesses Images from Abstract Sentences*
Hong Yu and Minsuk Lee

2:35–2:50   *Selecting relevant text subsets from web-data for building topic specific language models*
Abhinav Sethy, Panayiotis Georgiou and Shrikanth Narayanan

2:50–3:05   *Factored Neural Language Models*
Andrei Alexandrescu and Katrin Kirchhoff

3:05–3:20   *Quantitative Methods for Classifying Writing Systems*
Gerald Penn and Travis Choma

3:20–3:50   Break

Monday, June 5 (continued)

Word Alignment

3:50–4:15   *A Maximum Entropy Approach to Combining Word Alignments*
Necip Fazil Ayan and Bonnie J. Dorr

4:15–4:40   *Alignment by Agreement*
Percy Liang, Ben Taskar and Dan Klein

4:40–5:05   *Word Alignment via Quadratic Assignment*
Simon Lacoste-Julien, Ben Taskar, Dan Klein and Michael I. Jordan

Semantics I

3:50–4:15   *An Empirical Study of the Behavior of Active Learning for Word Sense Disambiguation*
Jinying Chen, Andrew Schein, Lyle Ungar and Martha Palmer

4:15–4:40   *Unknown word sense detection as outlier detection*
Katrin Erk

4:40–5:05   *Understanding Temporal Expressions in Emails*
Benjamin Han, Donna Gates and Lori Levin

Parsing I

3:50–4:15   *Partial Training for a Lexicalized-Grammar Parser*
Stephen Clark and James Curran

4:15–4:40   *Effective Self-Training for Parsing*
David McClosky, Eugene Charniak and Mark Johnson

4:40–5:05   *Multilingual Dependency Parsing using Bayes Point Machines*
Simon Corston-Oliver, Anthony Aue, Kevin Duh and Eric Ringger

Tuesday, June 6

Parsing II

9:00–9:25   *Multilevel Coarse-to-Fine PCFG Parsing*
Eugene Charniak, Mark Johnson, Micha Elsner, Joseph Austerweil, David Ellis, Isaac Haxton, Catherine Hill, R. Shrivaths, Jeremy Moore, Michael Pozar and Theresa Vu

9:25–9:50   *A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis*
Daisuke Kawahara and Sadao Kurohashi

9:50–10:15   *Fully Parsing the Penn Treebank*
Ryan Gabbard, Seth Kulick and Mitchell Marcus

Discourse

9:00–9:25   *Exploiting Semantic Role Labeling, WordNet and Wikipedia for Coreference Resolution*
Simone Paolo Ponzetto and Michael Strube

9:25–9:50   *Identifying and Analyzing Judgment Opinions*
Soo-Min Kim and Eduard Hovy

9:50–10:15   *Learning to Detect Conversation Focus of Threaded Discussions*
Donghui Feng, Erin Shaw, Jihie Kim and Eduard Hovy

Spoken and Acoustic Aspects of Language

9:00–9:25   *Towards Automatic Scoring of Non-Native Spontaneous Speech*
Klaus Zechner and Isaac Bejar

9:25–9:50   *Unsupervised and Semi-supervised Learning of Tone and Pitch Accent*
Gina-Anne Levow

9:50–10:15   *Learning Pronunciation Dictionaries: Language Complexity and Word Selection Strategies*
John Kominek and Alan W Black

10:15–10:45   Break

Machine Translation II

10:45–11:10   *Relabeling Syntax Trees to Improve Syntax-Based Machine Translation Quality*
Bryant Huang and Kevin Knight

11:10–11:35   *Grammatical Machine Translation*
Stefan Riezler and John T. Maxwell III

11:35–12:00   *Synchronous Binarization for Machine Translation*
Hao Zhang, Liang Huang, Daniel Gildea and Kevin Knight

Dialogue

10:45–11:10   *Modelling User Satisfaction and Student Learning in a Spoken Dialogue Tutoring System with Generic, Tutoring, and User Affect Parameters*
Kate Forbes-Riley and Diane Litman

11:10–11:35   *Comparing the Utility of State Features in Spoken Dialogue Using Reinforcement Learning*
Joel Tetreault and Diane Litman

11:35–12:00   *Backoff Model Training using Partially Observed Data: Application to Dialog Act Tagging*
Gang Ji and Jeff Bilmes

Relation Extraction

10:45–11:10   *Exploring Syntactic Features for Relation Extraction using a Convolution Tree Kernel*
Min Zhang, Jie Zhang and Jian Su

11:10–11:35   *Integrating Probabilistic Extraction Models and Data Mining to Discover Relations and Patterns in Text*
Aron Culotta, Andrew McCallum and Jonathan Betz

11:35–12:00   *Preemptive Information Extraction using Unrestricted Relation Discovery*
Yusuke Shinyama and Satoshi Sekine

12:00–1:30   Lunch

Tuesday, June 6 (continued)

Best Paper And Plenary Demo Presentations

1:30–2:00   *Probabilistic Context-Free Grammar Induction Based on Structural Zeros*
Mehryar Mohri and Brian Roark

2:00–2:30   *Prototype-Driven Learning for Sequence Models*
Aria Haghighi and Dan Klein

2:30–3:00   Plenary demos

InfoMagnets: Making Sense of Corpus Data
Jamie Arguello and Carolyn Rose

Question Answering with Web, Mobile and Speech Interfaces
Edward Whittaker, Joanna Mrozinski, and Sadaoki Furui

From Pipedreams to Products and Promise!
Janet Baker and Patri Pugliese

3:00–3:15   Break

Posters and Demos

3:15–5:15   Posters

A Comparison of Tagging Strategies for Statistical Information Extraction
Christian Siefkes

A Finite-State Model of Georgian Verbal Morphology
Olga Gurevich

A Maximum Entropy Framework that Integrates Word Dependencies and Grammatical Relations for Reading Comprehension
Kui Xu, Helen Meng and Fuliang Weng

Answering the question you wish they had asked: The impact of paraphrasing for Question Answering
Pablo Duboue and Jennifer Chu-Carroll

Comparing the roles of textual, acoustic and spoken-language features on spontaneous-conversation summarization
Xiaodan Zhu and Gerald Penn

Evolving optimal inspectable strategies for spoken dialogue systems
Dave Toney, Johanna Moore and Oliver Lemon

Exploiting Variant Corpora for Machine Translation
Michael Paul and Eiichiro Sumita

Gesture Improves Coreference Resolution
Jacob Eisenstein and Randall Davis

Illuminating Trouble Tickets with Sublanguage Theory
Svetlana Symonenko, Steven Rowe and Elizabeth D. Liddy

Improved Affinity Graph Based Multi-Document Summarization
Xiaojun Wan and Jianwu Yang

Investigating Cross-Language Speech Retrieval for a Spontaneous Conversational Speech Collection
Diana Inkpen, Muath Alzghool, Gareth Jones and Douglas Oard

Measuring Semantic Relatedness Using People and WordNet
Beata Beigman Klebanov

MMR-based Active Machine Learning for Bio Named Entity Recognition
Seokhwan Kim, Yu Song, Kyungduk Kim, Jeong-Won Cha and Gary Geunbae Lee

NER Systems that Suit User's Preferences: Adjusting the Recall-Precision Trade-off for Entity Extraction
Einat Minkov, Richard Wang, Anthony Tomasic and William Cohen

OntoNotes: The 90% Solution
Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw and Ralph Weischedel

Sentence Planning for Realtime Navigational Instruction
Laura Stoia, Donna Byron, Darla Shockley and Eric Fosler-Lussier

Syntactic Kernels for Natural Language Learning: the Semantic Role Labeling Case
Alessandro Moschitti

Temporal Classification of Text and Automatic Document Dating
Angelo Dalli

The MILE Corpus for Less Commonly Taught Languages
Alison Alvarez, Lori Levin, Robert Frederking, Simon Fung, Donna Gates and Jeff Good

Using Semantic Authoring for Blissymbols Communication Boards
Yael Netzer and Michael Elhadad

Weblog Classification for Fast Splog Filtering: A URL Language Model Segmentation Approach
Franco Salvetti and Nicolas Nicolov

Tuesday, June 6 (continued)

Posters and Demos

3:15–5:15   Demonstrations

InfoMagnets: Making sense of Corpus Data
Jaime Arguello and Carolyn Rose

From Pipedreams to Products and Promise
Janet Baker and Patri Pugliese

SmartNotes: Implicit Labeling of Meeting Data through User Note-Taking and Browsing
Satanjeev Banerjee and Alexander Rudnicky

MTTK: An Alignment Toolkit for Statistical Machine Translation
Yonggang Deng and William Byrne

AquaLog: An Ontology-driven Question Answering System to interface the Semantic Web
Vanessa Lopez Garcia, Enrico Motta, and Victoria Uren

Knowtator: A Protege plug-in for Annotated Corpus Construction
Philip V. Ogren

Automating the Creation of Interactive Glyph-supplemented Scatterplots for Visualizing Algorithm Results
Dinoj Surendran

Automatic Cluster Stopping with Criterion Functions and the Gap Statistic
Ted Pedersen and Anagha Kulkarni

SconeEdit: a Text-guided Domain Knowledge Editor
Alicia Tribble, Benjamin Lambert, and Scott E. Fahlman

Factoid Question Answering with Web, Mobile, and Speech Interfaces
E.W.D. Whittaker, J. Mrozinski, and S. Furui

Automated Quality Monitoring for Call Centers using Speech and NLP Technologies
G. Zweig, O. Siohan, G. Saon, B. Ramabhadran, D. Povey, L. Mangu, and B. Kingsbury

7:00   Banquet

Wednesday, June 7

9:00–10:00 **Keynote Speaker II: Diane Litman**

10:00–10:30 Break

Morphology/Grammar Induction

10:30–10:55 *Learning Morphological Disambiguation Rules for Turkish*
Deniz Yuret and Ferhan Ture

10:55–11:20 *Cross-Entropy and Estimation of Probabilistic Context-Free Grammars*
Anna Corazza and Giorgio Satta

11:20–11:45 *Estimation of Consistent Probabilistic Context-free Grammars*
Mark-Jan Nederhof and Giorgio Satta

11:45–12:10 *A Better N-Best List: Practical Determinization of Weighted Finite Tree Automata*
Jonathan May and Kevin Knight

Generation/Summarization/Question Answering

10:30–10:55 *Aggregation via Set Partitioning for Natural Language Generation*
Regina Barzilay and Mirella Lapata

10:55–11:20 *Incorporating Speaker and Discourse Features into Speech Summarization*
Gabriel Murray, Steve Renals, Jean Carletta and Johanna Moore

11:20–11:45 *Nuggeteer: Automatic Nugget-Based Evaluation using Descriptions and Judgements*
Gregory Marton and Alexey Radul

11:45–12:10 *Will Pyramids Built of Nuggets Topple Over?*
Jimmy Lin and Dina Demner-Fushman

Information Retrieval

10:30–10:55 *Creating a Test Collection for Citation-based IR Experiments*
Anna Ritchie, Simone Teufel and Stephen Robertson

10:55–11:20 *A Machine Learning based Approach to Evaluating Retrieval Systems*
Huyen-Trang Vu and Patrick Gallinari

11:20–11:45 *Language Model Information Retrieval with Document Expansion*
Tao Tao, Xuanhui Wang, Qiaozhu Mei and ChengXiang Zhai

11:45–12:10 *Towards Spoken-Document Retrieval for the Internet: Lattice Indexing For Large-Scale Web-Search Architectures*
Zheng-Yu Zhou, Peng Yu, Ciprian Chelba and Frank Seide

12:10–1:40   Lunch

1:40–2:30   NAACL Business Meeting

Short Papers: Morphology/Syntax

2:30–2:45   *Subword-based Tagging by Conditional Random Fields for Chinese Word Segmentation*
               Ruiqiang Zhang, Kikui Genichiro and sumita eiichiro

2:45–3:00   *Accurate Parsing of the Proposition Bank*
               Gabriele Musillo and Paola Merlo

3:00–3:15   *Early Deletion of Fillers In Processing Conversational Speech*
               Matthew Lease and Mark Johnson

3:15–3:30   *Parser Combination by Reparsing*
               Kenji Sagae and Alon Lavie

Short Papers: Semantics

2:30–2:45   *Unsupervised Induction of Modern Standard Arabic Verb Classes*
               Neal Snider and Mona Diab

2:45–3:00   *Word Domain Disambiguation via Word Sense Disambiguation*
               Antonio Sanfilippo, Stephen Tratz and Michelle Gregory

3:00–3:15   *Evaluation of Utility of LSA for Word Sense Discrimination*
               Esther Levin, Mehrbod Sharifi and Jerry Ball

3:15–3:30   *Semi-supervised Relation Extraction with Label Propagation*
               Jinxiu Chen, Donghong Ji, Chew Lim Tan and Zhengyu Niu

Short Papers: Speech and Video Processing

2:30–2:45   *Initial Study on Automatic Identification of Speaker Role in Broadcast News Speech*
               Yang Liu

2:45–3:00   *Extracting Salient Keywords from Instructional Videos Using Joint Text, Audio and Visual
               Cues*
               Youngja Park and Ying Li

3:00–3:15   *Class Model Adaptation for Speech Summarisation*
               Pierre Chatain, Edward Whittaker, Joanna Mrozinski and Sadaoki Furui

3:15–3:30   *Summarizing Speech Without Text Using Hidden Markov Models*
               Sameer Maskey and Julia Hirschberg

3:30–4:00   Break

Semantics II

4:00–4:25   *A fast finite-state relaxation method for enforcing global constraints on sequence decoding*
Roy Tromble and Jason Eisner

4:25–4:50   *Semantic role labeling of nominalized predicates in Chinese*
Nianwen Xue

4:50–5:15   *Learning for Semantic Parsing with Statistical Machine Translation*
Yuk Wah Wong and Raymond Mooney

Evaluation

4:00–4:25   *ParaEval: Using Paraphrases to Evaluate Summaries Automatically*
Liang Zhou, Chin-Yew Lin, Dragos Stefan Munteanu and Eduard Hovy

4:25–4:50   *Paraphrasing for Automatic Evaluation*
David Kauchak and Regina Barzilay

4:50–5:15   *An Information-Theoretic Approach to Automatic Evaluation of Summaries*
Chin-Yew Lin, Guihong Cao, Jianfeng Gao and Jian-Yun Nie

Processing in/for Language Models

4:00–4:25   *Cross Linguistic Name Matching in English and Arabic*
Andrew Freeman, Sherri Condon and Christopher Ackerman

4:25–4:50   *Language Model-Based Document Clustering Using Random Walks*
Gunes Erkan

4:50–5:15   *Unlimited vocabulary speech recognition for agglutinative languages*
Mikko Kurimo, Antti Puurula, Ebru Arisoy, Vesa Siivola, Teemu Hirsimki, Janne Pylkknen, Tanel Alume and Murat Saraclar