Sponsorship
The COLING 2010 very gratefully acknowledges the following commitments in sponsorship:

Platinum Sponsors
- National Natural Science Foundation of China
- Department of Language Information Administration, Ministry of Education, PRC

Gold Sponsor
BaiDu

Silver Sponsors
Google
Fujitsu R&D Center CO., LTD.
Microsoft Research
Beijing TRS Information Technology Co., Ltd
Shenyang Globla Envoy software Co.,Ltd.

Supporters
Asian Federation of Natural Language Processing
Institute of Automation
Chinese Academy of Sciences
Institute of Computing Technology
Chinese Academy of Sciences
Harbin Institute of Technology
Peking University
Tsinghua University
Preface

You will find in this volume papers from the 23rd International Conference on Computational Linguistics (COLING 2010) held in Beijing, China on August 23-27, 2010 under the auspices of the International Committee on Computational Linguistics (ICCL), and organized by the Chinese Information Processing Society (CIPS) of China. For this prestigious natural language processing conference to be held in China is a significant event for computational linguistics and for colleagues in China, demonstrating both the maturity of our field and the development of academic areas in China.

COLING started as a friendly gathering in New York in 1965, and has grown steadily since. Yet COLING’s aspiration to be a different conference remains the same. COLING strives to maintain its key qualities of embracing different theories and encouraging young scholars in spite of its growing size. A new component introduced at COLING 2010 underlines this quality. A RefreshINGenious (RING) session, organized by Aravind Joshi, our General Chair, allows new and un-orthodox ideas to be presented before they are fully developed in order to generate more discussion and stimulate other new ideas. We hope that this can become an important feature of COLING in the future.

The 155 oral papers included in the hardcopy proceedings published by Tsinghua University Press, as well as the 334 papers included in the electronic proceedings (the same 155 oral papers plus 179 poster papers) are selected from among 815 effective submissions among the more than 840 submissions received. The very selective acceptance rate of 19.02% for oral presentations (155/815 submissions) indicates the extremely high quality of the papers. An additional 21.96% (179/815) are selected for poster presentations to bring the overall acceptance rate to 40.98% (334/815).

We would like to thank the program committee area chairs for their dedicated and efficient review work, and our 738 reviewers for giving us very high quality reviews with a very short turnaround time, allowing us to maintain both the review quality and schedule even given the extraordinary number of submissions. Of course we thank the authors of the 840 papers for submitting their labor of love to COLING. Although we were only able to accept a minority of the submitted papers, we do hope that all authors and reviewers benefit from this process of indirect dialogue. We are especially grateful to the incredibly hard-working team of Stanford volunteers Jenny Finkel, Adam Vogel, and Mengqiu Wang, and HIT volunteers Sam Liang and Lemon Liu, who provided timely and efficient support for the two program chairs at every step of the review and publication processes.

Last but not least, we would like to thank the people who made COLING 2010 and this volume possible. We thank local arrangement committee co-chairs Professor Chengqing Zong and Professor Le Sun for their tireless work which will make COLING-2010 a sure success. Our special appreciation goes to the Chinese Information Processing Society (CIPS) and Professor Youqi Cao for their generous support as the COLING 2010 organizer. Lastly, Professor Qin Lu and Professor Tiejun Zhao should be recognized for their meticulous preparation for editing and publication, which brought this volume to reality.

Chu-Ren Huang and Dan Jurafsky,
COLING 2010 Program Committee Co-chairs

July 8, 2010
COLING 2010 is organized by the Chinese Information Processing Society of China (CIPS) and under the auspices of the International Committee on Computational Linguistics (ICCL).

**General Chair:**
Aravind K. Joshi (University of Pennsylvania)

**Program Chairs:**
Chu-Ren Huang (The Hong Kong Polytechnic University)
Dan Jurafsky (Stanford University)

**Advisors to Organizing Committee:**
Youqi Cao (The Chinese Information Processing Society of China)
Zhendong Dong (The Chinese Information Processing Society of China)
Changning Huang (Microsoft Research Asia)
Sheng Li (Harbin Institute of Technology)
Tianshun Yao (Northeastern University)
Shiwen Yu (Peking University)
Zhiwei Feng (Institute of Applied Linguistics, Ministry of Education)
Kaiying Liu (Shanxi University)

**Organization Chairs:**
Chengqing Zong (Institute of Automation, Chinese Academy of Sciences)
Le Sun (Institute of Software, Chinese Academy of Sciences)

**Publication Chairs:**
Qin Lu (The Hong Kong Polytechnic University)
Tiejun Zhao (Harbin Institute of Technology)

**Tutorial Chairs:**
Dan Gildea (University of Rochester)
Xuanjing Huang (Fudan University)

**Workshop Chairs:**
Noah Smith (Carnegie Mellon University)
Takenobu Tokunaga (Tokyo Institute of Technology)
Haifeng Wang (BaiDu)

**Publicity Chairs:**
Hal Daumé III (University of Utah)
Bin Wang (Institute of Computing Technology, Chinese Academy of Sciences)
Minghui Dong (Institute for Infocomm Research)
Monica Monachini (Institute for Computational Linguistics)
Aline Villavicencio (Federal University of Rio Grande do Sul)

Sponsorship Chairs:
Tingting He (Huazhong Normal University)
Hinrich Schütze (University of Stuttgart)
Key-Sun Choi (Korea Advanced Institute of Science and Technology)
Rion Snow (Stanford University)
Takehito Utsuro (University of Tsukuba)
Renata Vieira (Pontifical Catholic University of Rio Grande do Sul)
Hao Yu (Fujitsu(China) R&D Center)

Demo Chairs:
Ting Liu (Harbin Institute of Technology)
Yang Liu (The University of Texas at Dallas)

Program Committee Members/AreaChairs:
Nianwen Xue (Brandeis University)
Rajeev Sangal (India Institute of Information Technology)
Roger Levy (University of California, San Diego)
Justine Cassell (Northwestern University)
Caroline Sporleder (Saarland University)
Gary Geunbae Lee (Pohang University of Science and Technology)
Rosie Jones (Yahoo! Research)
Nicoletta Calzolari (Istituto di Linguistica Computazionale)
Chris Callison-Burch (Johns Hopkins University)
Qun Liu (Institute of Computing Technology)
Pierre Isabelle (National Research Council of Canada)
Sadao Kurohashi (Kyoto University)
Sebastian Padó (Universität Stuttgart)
Takenobu Tokunaga (Tokyo Institute of Technology)
Bo Pang (Yahoo! Research)
Haizhou Li (Institute for Infocomm Research)
David A Smith (University of Massachusetts, Amherst)
Donia Scott (University of Sussex)
Emily Bender (University of Washington)
Ming Zhou (Microsoft Research Asia)

Organization Committee Members:
Juanzi Li (Tsinghua University)
KWONG Olivia (Hong Kong City University)
Bin Sun (Peking University)
Houfeng Wang (Peking University)
Xiaojie Wang (Beijing University of Posts and Telecommunications)
Endong Xun (Beijing Language and Culture University)
Erhong Yang (Beijing Language and Culture University)
Jun Zhao (Institute of Automation, Chinese Academy of Sciences)
Jingbo Zhu (Northeastern University)
Program Committee Members/Reviewers:

Rob Abbott  Jordan Boyd-Graber  Yejin Choi
Takeshi Abekawa  S. R. K. Branavan  Kenneth Church
Omri Abend  Antonio Branco  Massimiliano Ciaramita
Steven Abney  Eric Breck  Philipp Cimiano
Eyten Adar  Chris Brew  Shay Cohen
Guadalupe Aguado  Sabine Buchholz  Kevyn Collins-Thompson
Takeshi Abekawa  Stefan Buettcher  John Conroy
Cecilia Ovesdotter Alm  Hung Bui  Bonaventura Coppola
Omar Alonso  Razvan Bunescu  Ed Cormany
Bharat Ram Ambati  Aljoscha Burchardt  Dan Cristea
Massih-Reza Amini  Stephan Busemann  Bruce Croft
Xiandong An  Miriam Butt  Dick Crouch
Sophia Ananiadou  William Byrne  Montse Cuadros
Jaime Arguello  Lynne Cahill  Silviu-Petru Cucerzan
Masayuki Asahara  Jamie Callan  Oliver Čulo
Nicholas Asher  Bin Cao  Aron Culotta
Tania Avgustinova  Yunbo Cao  Beatrice Daille
Necip Fazil Ayan  Giuseppe Carenini  Hercules Dalianis
Lakshmi Bai  Jean Carletta  Sandipan Dandapat
Jason Baldridge  Marine Carpuat  Hoa Dang
Timothy Baldwin  Xavier Carreras  Dana Dannells
Carmen Banea  John Carroll  Dipanjan Das
Srinivas Bangalore  Ben Carterette  Sajib Dasgupta
Colin Bannard  Francisco Casacuberta  Hal Daume
Ken Barker  Steve Cassidy  Dmitry Davidov
Marco Baroni  Eric Castelli  Adria de Gispert
Regina Barzilay  Alexandru Ceausu  Eric De La Clergerie
John Bateman  Nick Cercone  Valeria de Paiva
Beata Beigman Klebanov  Jeong-Won Cha  Maarten de Rijke
Daisuke Bekki  Vineet Chaitanya  Thierry Declerck
Nria Bel  Yllias Chali  Vera Demberg
Yinon Bentor  Nate Chambers  Steve DeNeefe
Sabine Bergler  Baobao Chang  John DeNero
Aditya Bhargava  Pi-Chuan Chang  Pascal Denis
Pushpak Bhattacharyya  Wanxiang Che  Mona Diab
Timothy Bickmore  Ciprian Chelba  Georgiana Dinu
Klinton Bicknell  Aitao Chen  Pinar Donmez
Alexandra Birch  Bin Chen  Justin Dornescu
Philippe Blache  Boxing Chen  Ascander Dost
Alan Black  Chien Chin Chen  Antoine Doucet
John Blitzer  Harr Chen  Doug Downey
Michael Bloodgood  Hsin-Hsi Chen  Markus Dreyer
Phil Blunsom  Jinying Chen  Gregory Druck
Ondrej Bojar  Keh-Jiann Chen  Xiangyu Duan
Gemma Boleda  Wenliang Chen  Amit Dubey
Francis Bond  Xiao Chen  Kevin Duh
Johan Bos  Charibeth K. Cheng  Michael Dukes
Marisa Boston  Colin Cherry  Chris Dyer
Pierrette Bouillon  Tee Kiah Chia  Marc Dymetman
Julien Bourdailliet  David Chiang  Markus Egg
# Table of Contents

**Testing SDRT’s Right Frontier**  
Stergos Afantenos and Nicholas Asher .................................................. 1

**Identifying Multi-word Expressions by Leveraging Morphological and Syntactic Idiosyncrasy**  
Hassan Al-Haj and Shuly Wintner ...................................................... 10

**Robust Measurement and Comparison of Context Similarity for Finding Translation Pairs**  
Daniel Andrade, Tetsuya Nasukawa and Junichi Tsujii .......................... 19

**Multilingual Subjectivity: Are More Languages Better?**  
Carmen Banea, Rada Mihalcea and Janyce Wiebe ................................. 28

**Plagiarism Detection across Distant Language Pairs**  
Alberto Barrón-Cedeño, Paolo Rosso, Eneko Agirre and Gorka Labaka ........... 37

**Automatic Detection of Non-deverbal Event Nouns for Quick Lexicon Production**  
Nuria Bel, Maria Coll and Gabriela Resnik ........................................ 46

**Fast and Accurate Arc Filtering for Dependency Parsing**  
Shane Bergsma and Colin Cherry ..................................................... 53

**A Hierarchical Classifier Applied to Multi-way Sentiment Detection**  
Adrian Bickerstaffe and Ingrid Zukerman ......................................... 62

**Fluency Constraints for Minimum Bayes-Risk Decoding of Statistical Machine Translation Lattices**  
Graeme Blackwood, Adrià de Gispert and William Byrne ........................ 71

**Self-Annotation for fine-grained geospatial relation extraction**  
Andre Blessing and Hinrich Schütze ............................................... 80

**Top Accuracy and Fast Dependency Parsing is not a Contradiction**  
Bernd Bohnet .................................................................................. 89

**Broad Coverage Multilingual Deep Sentence Generation with a Stochastic Multi-Level Realizer**  
Bernd Bohnet, Leo Wanner, Simon Mill and Alicia Burga .......................... 98

**Towards an optimal weighting of context words based on distance**  
Bernard Brousseau-Villeneuve, Jian-Yun Nie and Noriko Kando .................. 107

**Measuring the Non-compositionality of Multiword Expressions**  
Fan Bu, Xiaoyan Zhu and Ming Li .................................................. 116

**A Utility-Driven Approach to Question Ranking in Social QA**  
Razvan Bunescu and Yunfeng Huang ............................................. 125

**Simultaneous Ranking and Clustering of Sentences: A Reinforcement Approach to Multi-Document Summarization**  
Xiaoyan Cai, Wenjie Li, You Ouyang and Hong Yan .............................. 134
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-to-End Coreference Resolution via Hypergraph Partitioning</td>
<td>143</td>
</tr>
<tr>
<td>Exploiting Background Knowledge for Relation Extraction</td>
<td>152</td>
</tr>
<tr>
<td>Jointly Modeling WSD and SRL with Markov Logic</td>
<td>161</td>
</tr>
<tr>
<td>Bipolar Person Name Identification of Topic Documents Using Principal Component Analysis</td>
<td>170</td>
</tr>
<tr>
<td>Emotion Cause Detection with Linguistic Constructions</td>
<td>179</td>
</tr>
<tr>
<td>A Twin-Candidate Based Approach for Event Pronoun Resolution using Composite Kernel</td>
<td>188</td>
</tr>
<tr>
<td>Unsupervised Synthesis of Multilingual Wikipedia Articles</td>
<td>197</td>
</tr>
<tr>
<td>Simplicity is Better: Revisiting Single Kernel PPI Extraction</td>
<td>206</td>
</tr>
<tr>
<td>An ontology-driven system for detecting global health events</td>
<td>215</td>
</tr>
<tr>
<td>Constraining robust constructions for broad-coverage parsing with precision grammars</td>
<td>223</td>
</tr>
<tr>
<td>Local lexical adaptation in Machine Translation through triangulation: SMT helping SMT</td>
<td>232</td>
</tr>
<tr>
<td>Automated Translation of Semantic Relationships</td>
<td>241</td>
</tr>
<tr>
<td>Comparison of different algebras for inducing the temporal structure of texts</td>
<td>250</td>
</tr>
<tr>
<td>Generating Learner-Like Morphological Errors in Russian</td>
<td>259</td>
</tr>
<tr>
<td>Resolving Object and Attribute Coreference in Opinion Mining</td>
<td>268</td>
</tr>
<tr>
<td>Entity Disambiguation for Knowledge Base Population</td>
<td>277</td>
</tr>
<tr>
<td>A Discriminative Latent Variable-Based &quot;DE&quot; Classifier for Chinese-English SMT</td>
<td>286</td>
</tr>
</tbody>
</table>
A Novel Reordering Model Based on Multi-layer Phrase for Statistical Machine Translation
Yanqing He, Yu Zhou, Chengqing Zong and Huilin Wang .................................................... 447

Standardizing Wordnets in the ISO Standard LMF: Wordnet-LMF for GermaNet
Verena Henrich and Erhard Hinrichs ............................................................ 456

Normal-form parsing for Combinatory Categorial Grammars with generalized composition and type-raising
Julia Hockenmaier and Yonatan Bisk ............................................................ 465

An Empirical Study on Web Mining of Parallel Data
Gumwon Hong, Chi-Ho Li, Ming Zhou and Hae-Chang Rim ............................... 474

Enhancing Cross Document Coreference of Web Documents with Context Similarity and Very Large Scale Text Categorization
Jian Huang, Pucktada Treeratpituk, Sarah Taylor and C. Lee Giles .................... 483

Feature-Rich Discriminative Phrase Rescoring for SMT
Fei Huang and Bing Xiang ........................................................................... 492

FactRank: Random Walks on a Web of Facts
Alpa Jain and Patrick Pantel ........................................................................ 501

Open Entity Extraction from Web Search Query Logs
Alpa Jain and Marco Pennacchiotti .............................................................. 510

Reranking Models in Fine-grained Opinion Analysis
Richard Johansson and Alessandro Moschitti ............................................... 519

Unsupervised phonemic Chinese word segmentation using Adaptor Grammars
Mark Johnson and Katherine Demuth .......................................................... 528

Data-Driven Parsing with Probabilistic Linear Context-Free Rewriting Systems
Laura Kallmeyer and Wolfgang Maier .......................................................... 537

Learning to Predict Readability using Diverse Linguistic Features
Rohit Kate, Xiaoqiang Luo, Siddharth Patwardhan, Martin Franz, Radu Florian, Raymond Mooney, Salim Roukos and Chris Welty ....................................................... 546

Value for Money: Balancing Annotation Effort, Lexicon Building and Accuracy for Multilingual WSD
Mitesh Khapra, Saurabh Sohoney, Anup Kulkarni and Pushpak Bhattacharyya .... 555

A Cross-lingual Annotation Projection Approach for Relation Detection
Seokhwan Kim, Minwoo Jeong, Jonghoon Lee and Gary Geunbae Lee ............ 564

Evaluating N-gram based Evaluation Metrics for Automatic Keyphrase Extraction
Su Nam Kim, Timothy Baldwin and Min-Yen Kan ........................................ 572

Improving the Quality of Text Understanding by Delaying Ambiguity Resolution
Doo Soon Kim, Ken Barker and Bruce Porter ............................................... 581
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic generation of inter-passage links based on semantic similarity</td>
<td>590</td>
</tr>
<tr>
<td>Petr Knoth, Jakub Novotny and Zdenek Zdrahal</td>
<td></td>
</tr>
<tr>
<td>Dependency-driven Anaphoricity Determination for Coreference Resolution</td>
<td>599</td>
</tr>
<tr>
<td>Fang Kong, Guodong Zhou, Longhua Qian and Qiaoming Zhu</td>
<td></td>
</tr>
<tr>
<td>Phrase Clustering for Smoothing TM Probabilities - or, How to Extract Paraphrases from Phrase Tables</td>
<td>608</td>
</tr>
<tr>
<td>Roland Kuhn, Boxing Chen, George Foster and Evan Stratford</td>
<td></td>
</tr>
<tr>
<td>Revisiting Context-based Projection Methods for Term-Translation Spotting in Comparable Corpora</td>
<td>617</td>
</tr>
<tr>
<td>Audrey Laroche and Philippe Langlais</td>
<td></td>
</tr>
<tr>
<td>Constituent Reordering and Syntax Models for English-to-Japanese Statistical Machine Translation</td>
<td>626</td>
</tr>
<tr>
<td>Young-Suk Lee, Bing Zhao and Xiaqian Luo</td>
<td></td>
</tr>
<tr>
<td>Sentiment Classification and Polarity Shifting</td>
<td>635</td>
</tr>
<tr>
<td>Shoushan Li, Sophia Y. M. Lee, Ying Chen, Chu-Ren Huang and Guodong Zhou</td>
<td></td>
</tr>
<tr>
<td>Improving Corpus Comparability for Bilingual Lexicon Extraction from Comparable Corpora</td>
<td>644</td>
</tr>
<tr>
<td>Bo Li and Eric Gaussier</td>
<td></td>
</tr>
<tr>
<td>Structure-Aware Review Mining and Summarization</td>
<td>653</td>
</tr>
<tr>
<td>Fangtao Li, Chao Han, Minlie Huang, Xiaoyan Zhu, Ying-Ju Xia, Shu Zhang and Hao Yu</td>
<td></td>
</tr>
<tr>
<td>Adaptive Development Data Selection for Log-linear Model in Statistical Machine Translation</td>
<td>662</td>
</tr>
<tr>
<td>Mu Li, Yinggong Zhao, Dongdong Zhang and Ming Zhou</td>
<td></td>
</tr>
<tr>
<td>Learning the Scope of Negation via Shallow Semantic Parsing</td>
<td>671</td>
</tr>
<tr>
<td>Junhui Li, Guodong Zhou, Hongling Wang and Qiaoming Zhu</td>
<td></td>
</tr>
<tr>
<td>Filtered Ranking for Bootstrapping in Event Extraction</td>
<td>680</td>
</tr>
<tr>
<td>Shasha Liao and Ralph Grishman</td>
<td></td>
</tr>
<tr>
<td>Exploring variation across biomedical subdomains</td>
<td>689</td>
</tr>
<tr>
<td>Tom Lippincott, Diarmuid O’Séaghdha, Lin Sun and Anna Korhonen</td>
<td></td>
</tr>
<tr>
<td>Semantic Role Labeling for News Tweets</td>
<td>698</td>
</tr>
<tr>
<td>Xiaohua Liu, Kuan Li, Bo Han, Ming Zhou, Long Jiang, Zhongyang Xiong and Changning Huang</td>
<td></td>
</tr>
<tr>
<td>Joint Parsing and Translation</td>
<td>707</td>
</tr>
<tr>
<td>Yang Liu and Qun Liu</td>
<td></td>
</tr>
<tr>
<td>Semantic Role Features for Machine Translation</td>
<td>716</td>
</tr>
<tr>
<td>Ding Liu and Daniel Gildea</td>
<td></td>
</tr>
<tr>
<td>TimeML Events Recognition and Classification: Learning CRF Models with Semantic Roles</td>
<td>725</td>
</tr>
<tr>
<td>Hector Llorens, Estela Saquete and Borja Navarro-Colorado</td>
<td></td>
</tr>
<tr>
<td>Exploiting Structured Ontology to Organize Scattered Online Opinions</td>
<td>734</td>
</tr>
<tr>
<td>Yue Lu, Huizhong Duan, Hongning Wang and ChengXiang Zhai</td>
<td></td>
</tr>
</tbody>
</table>
Enhancing Morphological Alignment for Translating Highly Inflected Languages
Minh-Thang Luong and Min-Yen Kan ................................................................. 743

Automatic analysis of semantic similarity in comparable text through syntactic tree matching
Erwin Marsi and Emiel Krahmer ................................................................. 752

Controlling Listening-oriented Dialogue using Partially Observable Markov Decision Processes
Toyomi Meguro, Ryuichiro Higashinaka, Yasuhiro Minami and Kohji Dohsaka .......... 761

Recognising Entailment within Discourse
Shachar Mirkin, Jonathan Berant, Ido Dagan and Eyal Shnarch .......................... 770

Evaluating Dependency Representations for Event Extraction
Makoto Miwa, Sampo Pyysalo, Tadayoshi Hara and Jun’ichi Tsujii ......................... 779

Entity-Focused Sentence Simplification for Relation Extraction
Makoto Miwa, Rune Sætre, Yusuke Miyao and Jun’ichi Tsujii ............................ 788

Using Cross-Lingual Projections to Generate Semantic Role Labeled Annotated Corpus for Urdu - A Resource Poor Language
Smruthi Mukund, Debanjan Ghosh and Rohini Srihari ................................. 797

Recognition of Affect, Judgment, and Appreciation in Text
Alena Neviarouskaya, Helmut Prendinger and Mitsuuru Ishizuka ....................... 806

Nonparametric Word Segmentation for Machine Translation
ThuyLinh Nguyen, Stephan Vogel and Noah A. Smith .............................. 815

Automatic Discovery of Feature Sets for Dependency Parsing
Peter Nilsson and Pierre Nugues ............................................................. 824

Evaluation of Dependency Parsers on Unbounded Dependencies
Joakim Nivre, Laura Rimell, Ryan McDonald and Carlos Gómez Rodríguez .......... 833

Co-STAR: A Co-training Style Algorithm for Hyponymy Relation Acquisition from Structured and Unstructured Text
Jong-Hoon Oh, Ichiro Yamada, Kentaro Torisawa and Stijn De Saeger .............. 842

Simple and Efficient Algorithm for Approximate Dictionary Matching
Naoaki Okazaki and Jun’ichi Tsujii ......................................................... 851

Latent Mixture of Discriminative Experts for Multimodal Prediction Modeling
Derya Ozkan, Kenji Sagae and Louis-Philippe Morency ............................... 860

Text Summarization of Turkish Texts using Latent Semantic Analysis
Makbule Ozsoy, Ilyas Cicekli and Ferda Alpaslan ....................................... 869

Knowing What to Believe (when you already know something)
Jeff Pasternack and Dan Roth ................................................................. 877

Using Web-scale N-grams to Improve Base NP Parsing Performance
Emily Pitler, Shane Bergsma, Dekang Lin and Kenneth Church ....................... 886
Discriminative Induction of Sub-Tree Alignment using Limited Labeled Data
Jun Sun, Min Zhang and Chew Lim Tan ......................................................... 1047

Investigating the cross-linguistic potential of VerbNet-style classification
Lin Sun, Thierry Poibeau, Anna Korhonen and Cedric Messiant ......................... 1056

Semi-supervised dependency parsing using generalized tri-training
Anders Søgaard and Christian Rishøj .............................................................. 1065

SemanticRank: Ranking Keywords and Sentences Using Semantic Graphs
George Tsatsaronis, Iraklis Varlamis and Kjetil Nørvåg .................................... 1074

Chinese CCGbank: extracting CCG derivations from the Penn Chinese Treebank
Daniel Tse and James R. Curran ..................................................................... 1083

Dependency Forest for Statistical Machine Translation
Zhaopeng Tu, Yang Liu, Young-Sook Hwang, Qun Liu and Shouxun Lin ............. 1092

Large Scale Parallel Document Mining for Machine Translation
Jakob Uszkoreit, Jay Ponte, Ashok Popat and Moshe Dubiner ......................... 1101

Hungarian Corpus of Light Verb Constructions
Veronika Vincze and János Csirik ................................................................. 1110

Syntax Based Reordering with Automatically Derived Rules for Improved Statistical Machine Translation
Karthik Visweswariah, Jiri Navratil, Jeffrey Sorensen, Vijil Chenthamarakshan and Nandakishore Kambhatla ................................................................. 1119

Efficient Statement Identification for Automatic Market Forecasting
Henning Wachsmuth, Peter Prettenhofer and Benno Stein ........................... 1128

Towards a Unified Approach to Simultaneous Single-Document and Multi-Document Summarizations
Xiaojun Wan ................................................................................................. 1137

"Got You!": Automatic Vandalism Detection in Wikipedia with Web-based Shallow Syntactic-Semantic Modeling
William Yang Wang and Kathleen McKeown ............................................. 1146

Exploiting Salient Patterns for Question Detection and Question Retrieval in Community-based Question Answering
Kai Wang and Tat-Seng Chua ................................................................. 1155

Probabilistic Tree-Edit Models with Structured Latent Variables for Textual Entailment and Question Answering
Mengqiu Wang and Christopher Manning .................................................. 1164

A Character-Based Joint Model for Chinese Word Segmentation
Kun Wang, Chengqing Zong and Keh-Yih Su .............................................. 1173

Near-synonym Lexical Choice in Latent Semantic Space
Tong Wang and Graeme Hirst ................................................................. 1182

xix
Heterogeneous Parsing via Collaborative Decoding
Muhua Zhu, Jingbo Zhu and Tong Xiao ........................................... 1344

A Monolingual Tree-based Translation Model for Sentence Simplification
Zhemin Zhu, Delphine Bernhard and Iryna Gurevych ................................ 1353

A Minimum Error Weighting Combination Strategy for Chinese Semantic Role Labeling
Tao Zhuang and Chengqing Zong ..................................................... 1362

Detecting Speech Repairs Incrementally Using a Noisy Channel Approach
Simon Zwarts, Mark Johnson and Robert Dale .................................... 1371

Syntactic Scope Resolution in Uncertainty Analysis
Lilja Øvrelid, Erik Velldal and Stephan Oepen ...................................... 1379