Coling 2008

22nd International Conference on Computational Linguistics

Proceedings of the Conference Volume 1

Programme chairs:
Donia Scott and Hans Uszkoreit

18-22 August 2008
Manchester, UK

©2008 The Coling 2008 Organizing Committee

Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Nonported* license
http://creativecommons.org/licenses/by-nc-sa/3.0/
Some rights reserved

Order copies of this and other Coling 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 978-1-905593-44-6

Design by Chimney Design, Brighton, UK
Production and manufacture by One Digital, Brighton, UK

Preface

COLING 2008, the 22nd International Conference on Computational Linguistics, is the first COLING conference in the UK, a country with a rich history and lively research scene in Computational Linguistics. The great response to the call for papers may have been caused by this location or it may just have been a consequence of the rapid growth of our discipline. Anyway, the 600 submissions of high average quality we received made it relatively easy for the programme committee to put together an excellent programme.

After a thorough reviewing process including a period of interactive deliberation, the programme committee selected 145 full papers and 35 poster presentations. The central criterion for the selection was scientific quality rather than geographic balance or the desirable spread across subareas. We tried to apply a multidimensional concept of quality that does not exclusively favour technically sound engineering papers but also yields some space for challenging scientific insights and first reports on novel approaches.

Looking at the distribution of the papers among subfields of CL, we made a few observations. One concerns the central theme of machine learning. Although the term machine learning only appears in the name of one single session, machine learning actually transcends nearly all represented subfields of our discipline.

After decades of hibernation, the area of machine translation has again become a central field of research. Almost all of the MT related submissions are on statistical translation but a growing number of papers describe clever combinations of methods from different paradigms. Compared with MT, the area of natural language generation is much less represented, which may partially be due to this year's International Language Generation Conference in Ohio.

The area of information extraction still keeps growing. With subareas such as opinion mining, sentiment detection and event extraction it has become rather diversified.

A special observation concerns specialized types of phrase disambiguation or classification that cannot easily be subsumed under IR or IE since the described methods could also be utilized for summarization, paraphrasing or other application types. In general it has become harder to assign method papers to just one traditional technology area. This is nicely reflected in the authors' choice of multiple keywords from different areas.

We received only few submissions on speech technologies, in our opinion even less than in earlier COLING conferences. Although this development might simply be attributed to the inevitable and ever progressing differentiation of the human language technologies, it may also be the case that the meeting market in this area is well covered by the well known speech conferences. This year's ACL conference also had just a single speech processing session.

We hope that our colleagues will forgive us for having been rather strict with double submissions. In several cases accepted submissions were finally turned down because a paper with largely overlapping contents has appeared or is scheduled to appear elsewhere. We believe that our field has to find a proper way of dealing with the increasing number of professional conferences without sacrificing the basic principles of scientific publishing.

After deciding to enrich COLING 2008 with a Best Paper Award, we received an offer to support the award from the renowned scientific publishing house Springer. This prize will first be conferred at this Conference. We are grateful to Springer for this generous donation and thank especially Olga Chiarcos for her efforts in this case.

Together with Olga Chiarcos we also thought about other ways to make COLING even more attractive and visible. Olga proposed a special book publication of extended versions of selected ground-breaking COLING papers. This is an excellent idea which we are going to implement already for this COLING conference.

Finally, we want to thank the people who were essential to this academic programme. There are the area chairs who have with great commitment and dedication steered the reviewing process to a successful end: Paul Buitelaar, Robert Dale, Mary Dalrymple, Bill Dolan, Robert Gaizauskas, Eva Hajičová, Julia Hirschberg, Chu-Ren Huang, Pierre Isabelle, Mark Johnson, Miles Osborne, Stephen Pulman, Dan Roth, Jun'ichi Tsujii. We also wish to gratefully acknowledge the successful work of our numerous reviewers who are listed on pages v-viii. Our special gratitude goes to Roger Evans and Christian Spurk. Roger has worked hard and uncompromisingly on these proceedings; he has been a very thoughtful and creative publication chair. Christian has played a central role in organizing the technical basis for the online reviewing and in the communication with authors, area chairs, reviewers and organizers.

We would also like to thank the local organizer Harold Somers for his valuable collaboration.

Less connected with this volume but essential for the overall success of the conference programme were the tutorial chair, Philipp Koehn; the workshops chair, Mark Stevenson, and all the workshop organisers; the demo chairs, Allan Ramsay and Kalina Bontcheva; the people who solicited the urgently needed sponsorships, John Tait and Anne de Roeck; as well as the colleague who recruited the student helpers, Paul Bennett.

But our greatest thanks, of course, go to the authors for their excellent contributions.

Donia Scott and Hans Uszkoreit
Programme Chairs of COLING 2008

Organizers:

Programme: Donia Scott (Open University) and Hans Uszkoreit (Universität des Saarlandes/DFKI)
Local organization: Harold Somers (University of Manchester)
Workshops: Mark Stevenson (University of Sheffield)
Tutorials: Philipp Koehn (University of Edinburgh)
Publications: Roger Evans (University of Brighton)
Demos: Allan Ramsay (University of Manchester) and Kalina Bontcheva (University of Sheffield)
Sponsorship: John Tait (IRF, Vienna) and Anne de Roeck (Open University)
Student helpers: Paul Bennett (University of Manchester)

Programme Chairs:

Donia Scott (Open University, UK)
Hans Uszkoreit (Universität des Saarlandes/DFKI, Germany)

Area Chairs:

Paul Buitelaar (DFKI, Germany)
Robert Dale (Macquarie University, Australia)
Mary Dalrymple (University of Oxford, UK)
Bill Dolan (Microsoft Research, USA)
Robert Gaizauskas (University of Sheffield, UK)
Eva Hajičová (Univerzita Karlova v Praze, Czech Republic)
Julia Hirschberg (Columbia University, USA)
Chu-Ren Huang (Academia Sinica, Taiwan)
Pierre Isabelle (NRC Institute for Information Technology, Canada)
Mark Johnson (Brown University, USA)
Miles Osborne (University of Edinburgh, UK)
Stephen Pulman (University of Oxford, UK)
Dan Roth (University of Illinois at Urbana-Champaign, USA)
Jun'ichi Tsujii (Tokyo Daigaku, Japan and University of Manchester, UK)

Invited speakers:

Dr Elizabeth Shriberg, Senior Research Psycholinguist, Speech Technology & Research Laboratory, SRI International, Menlo Park CA and International Computer Science Institute, Berkeley CA

Prof John Shawe-Taylor, Centre for Computational Statistics and Machine Learning, University College London

Programme Committee:

Eugene Agichtein
Eneko Agirre
Lars Ahrenberg
Akiko Aizawa
Manuel Alcantara
Beatrice Alex
Galen Andrew
Ash Asudeh
Bogdan Babych
Jason Baldridge
Carmen Banea
Srinivas Bangalore
Roy Bar-Haim
Marco Baroni
Regina Barzilay
Roberto Basili
Núria Bel
Linda Bell
Anja Belz
Pushpak Bhattacharyya
Dan Bikel
Alan Black
Patrick Blackburn
Phil Blunsom
Christian Boitet
Kalina Bontcheva
Johan Bos
Chris Brew
Ted Briscoe
Chris Brockett
Paul Buitelaar
Razvan Bunescu
Harry Bunt
Stephan Busemann
Miriam Butt
Bill Byrne
Janet Cahn
Chris Callison-Burch
Nicoletta Calzolari
Nicola Cancedda
Sander Canisius
Rolf Carlson
Marine Carpuat
Xavier Carreras
Justine Cassell
Eric Castelli
Jason Chang
Ming-Wei Chang
Eugene Charniak
Hsin-Hsi Chen
Keh-jiann Chen
Colin Cherry
David Chiang
Philipp Cimiano
Silvie Cinková
Stephen Clark
Kevin Cohen
Trevor Cohn
Nigel Collier
Michael Collins
Cleo Condoravdi
Michael Connor
Robin Cooper
James Curran
Ido Dagan
Robert Dale
Mary Dalrymple
Hal Daumé
Quang Xuan Do
Bill Dolan
Chris Dyer
Marc Dymetman
Markus Egg
Andreas Eisele
Jason Eisner
Mark Ellison
Micha Elsner
Alex Chenyu Fang
Christiane Fellbaum
Charles Fillmore
George Foster
Gil Francopoulo
Alex Fraser
Dayne Freitag
Atsushi Fujita
Adam Funk
Bruno Gaume
Robert Gaizauskas
Michel Galley
Aldo Gangemi
Jianfeng Gao
Zhao-Ming Gao
Claire Gardent
Eric Gaussier
Monte George
Ulrich Germann
Jonathan Ginzburg
Roxana Girju
John Goldsmith
Sharon Goldwater
Julio Gonzalo
Ralph Grishman
Claire Grover
Joakim Gustafson
Nizar Habash
Benjamin Hachey
Udo Hahn
Jan Hajič
Eva Hajičová
Keith Hall
Sanda Harabagiu
Mary Harper
Tingting He
Xiaodong He
William Headden
Mary Hearne
Marti Hearst
James Henderson
Mark Hepple
Andy Hickl
Keikichi Hirose
Julia Hirschberg
Graeme Hirst
Jerry Hobbs
Julia Hockenmaier
Merle Horne
David House
Shu-Kai Hsieh
Chu-Ren Huang

Kentaro Inui
Pierre Isabelle
David Israel
Abe Ittycheriah
Donghong Ji
Su Jian
Jing Jiang
Howard Johnson
Mark Johnson
Michael Johnston
Kristina Jokinen
Gareth Jones
Aravind Joshi
Asanee Kawtrakul
Noriko Kando
Kyoko Kanzaki
Jussi Karlgren
Junichi Kazama
Jin-Dong Kim
Tracy King
Ewan Klein
Alexandre Klementiev
Kevin Knight
Alistair Knott
Philipp Koehn
Hasida Koiti
Alexander Koller
Emiel Krahmer
Ivana Kruijff-Korbayova
Geert-Jan Kruijff
Roland Kuhn
Shankar Kumar
Sadao Kurohashi
Philippe Langlais
Stafan Larsson
Alberto Lavelli
Claudia Leacock
Matt Lease
Gary Geunbae Lee
Kiyong Lee
Lillian Lee
Gregor Leusch
Gina Levow
Hang Li
Sujian Li
Chin-Yew Lin
Ken Litkowski
Diane Litman
Bing Liu
Yang Liu
Adam Lopez
Qin Lu
Kim Teng Lua
Klaus Macherey
Wolfgang Macherey
Kikuo Maekawa
Bernardo Magnini
Inderjeet Mani
Daniel Marcu
Katja Markert
Erwin Marsi
Andrew McCallum
Diana McCarthy
David McClosky
Majorie McShane
Arul Menezes
Paola Merlo
Rada Mihalcea
Yusuke Miyao
Monica Monachini
Simonetta Montemagni
Raymond Mooney
Robert Moore
Paul Morarescu
Alessandro Moschitti
Reinhard Muskens
Tetsuya Nasukawa
Ani Nenkova
Günter Neumann
Hwee Tou Ng
Vincent Ng
Thi Minh Huyen Nguyen
Nicolas Nicolov
Joakim Nivre
Franz Och
Kemal Oflazer
Naoaki Okazaki
Manabu Okumura
Alessandro Oltramari
Miles Osborne
Martha Palmer
Shimei Pan
Bo Pang
Soo-Min Pantel
Cécile Paris
Jeff Pelletier
Gerald Penn
Marco Pennachiotti
Wim Peters
Slav Petrov
Anselmo Peñas
Manfred Pinkal
Massimo Poesio
Thierry Poibeau
Ana-Maria Popescu
Robert Porzel
Richard Power
John Prager
Ian Pratt-Hartman
Judita Preiss
Laurent Prévot
Stephen Pulman
Vasin Punyakanok
James Pustejovsky
Chris Quirk
Owen Rambow
Allan Ramsay
Ari Rappoport
Lev Ratinov
Roi Reichart
Ehud Reiter
Stefan Riezler
Ellen Riloff
Hae-Chang Rim
Brian Roark
Barbara Rosario
Mathias Rossignol
Dan Roth
Rachel Roxas
Vasile Rus
Kenji Sagae
Mark Sammons
Rajeev Sangal
Anoop Sarkar
Len Schubert
Holger Schwenk
Donia Scott
Satoshi Sekine
Stephanie Seneff
Petr Sgall
Libin Shen

Nobuyuki Shimizu
Elizabeth Shriberg
Khalil Sima'an
Michel Simard
Kevin Small
Noah Smith
Pavel Smrz
Stephen Soderland
Claudia Soria
Virach Sornlertlamvanich
Sofia Stamou
Manfred Stede
Mark Steedman
Armando Stellato
Amanda Stent
Mark Stevenson
Pavel Stranak
Tomek Strzalkowski
Le Sun
Hisami Suzuki
Jun Suzuki
Marc Swerts
Idan Szpektor
Simone Teufel
Jörg Tiedemann
Christoph Tillmann
Takenobu Tokunaga
Kentaro Torisawa
Kristina Toutanova
Isabel Trancoso
Shu-Chuan Tseng
Jun'ichi Tsujii
Dan Tufiş
Nicola Ueffing
Hans Uszkoreit
Takehito Utsuro
Kees van Deemter
Josef van Genabith
Antal van den Bosch
Walther von Hahn
Lucy Vanderwende
Sebastian Varges
Tony Veale
Paola Velardi
Ashish Venugopal
Barbora Vidová Hladká
Piek Vossen
Stephen Wan
Xinglong Wang
Taro Watanabe
Andy Way
Bonnie Webber
Davy Weissenbacher
Janyce Wiebe
Yorick Wilks
Kam-Fai Wong
Chung-Hsien Wu
Nianwen Xue
Roman Yangarber
Naoki Yoshinaga
Kun Yu
Annie Zaenen
Fabio Massimo Zanzotto
Jun Zhao
Jing Zheng
Ming Zhou
Michael Zock
Chengqing Zong
Richard Zens

Table of Contents

Two-Phased Event Relation Acquisition: Coupling the Relation-Oriented and Argument-Oriented Approaches
Shuya Abe, Kentaro Inui and Yuji Matsumoto 1

A Supervised Algorithm for Verb Disambiguation into VerbNet Classes
Omri Abend, Roi Reichart and Ari Rappoport 9

On Robustness and Domain Adaptation using SVD for Word Sense Disambiguation
Eneko Agirre and Oier Lopez de Lacalle 17

An Improved Hierarchical Bayesian Model of Language for Document Classification
Ben Allison 25

Improving Alignments for Better Confusion Networks for Combining Machine Translation Systems
Necip Fazil Ayan, Jing Zheng and Wen Wang 33

Verification and Implementation of Language-Based Deception Indicators in Civil and Criminal Narratives
Joan Bachenko, Eileen Fitzpatrick and Michael Schonwetter 41

Enhancing Multilingual Latent Semantic Analysis with Term Alignment Information
Brett W. Bader and Peter A. Chew 49

Weakly Supervised Supertagging with Grammar-Informed Initialization
Jason Baldridge 57

Good Neighbors Make Good Senses: Exploiting Distributional Similarity for Unsupervised WSD
Samuel Brody and Mirella Lapata 65

A Classification of Dialogue Actions in Tutorial Dialogue
Mark Buckley and Magdalena Wolska 73

A Concept-Centered Approach to Noun-Compound Interpretation
Cristina Butnariu and Tony Veale 81

Are Morpho-Syntactic Features More Predictive for the Resolution of Noun Phrase Coordination Ambiguity than Lexico-Semantic Similarity Scores?
Ekaterina Buyko and Udo Hahn 89

ParaMetric: An Automatic Evaluation Metric for Paraphrasing
Chris Callison-Burch, Trevor Cohn and Mirella Lapata 97

Regenerating Hypotheses for Statistical Machine Translation
Boxing Chen, Min Zhang, Aiti Aw and Haizhou Li 105

Learning Reliable Information for Dependency Parsing Adaptation
Wenliang Chen, Youzheng Wu and Hitoshi Isahara 113

Other-Anaphora Resolution in Biomedical Texts with Automatically Mined Patterns
Bin Chen, Xiaofeng Yang, Jian Su and Chew Lim Tan 121

Latent Morpho-Semantic Analysis: Multilingual Information Retrieval with Character N-Grams and Mutual Information
Peter A. Chew, Brett W. Bader and Ahmed Abdelali 129

Sentence Compression Beyond Word Deletion
Trevor Cohn and Mirella Lapata 137

Mind the Gap: Dangers of Divorcing Evaluations of Summary Content from Linguistic Quality
John M. Conroy and Hoa Trang Dang 145

Hybrid Processing for Grammar and Style Checking
Berthold Crysmann, Nuria Bertomeu, Peter Adolphs, Daniel Flickinger and Tina Klüwer 153

KnowNet: Building a Large Net of Knowledge from the Web
Montse Cuadros and German Rigau 161

A Classifier-Based Approach to Preposition and Determiner Error Correction in L2 English
Rachele De Felice and Stephen G. Pulman 169

Pedagogically Useful Extractive Summaries for Science Education
Sebastian de la Chica, Faisal Ahmad, James H. Martin and Tamara Sumner 177

Looking for Trouble
Stijn De Saeger, Kentaro Torisawa and Jun'ichi Kazama 185

Re-estimation of Lexical Parameters for Treebank PCFGs
Tejaswini Deoskar 193

Representations for category disambiguation
Markus Dickinson 201

Syntactic Reordering Integrated with Phrase-Based SMT
Jakob Elming 209

Efficiently Parsing with the Product-Free Lambek Calculus
Timothy A. D. Fowler 217

A Probabilistic Model for Measuring Grammaticality and Similarity of Automatically Generated Paraphrases of Predicate Phrases
Atsushi Fujita and Satoshi Sato 225

Retrieving Bilingual Verb-Noun Collocations by Integrating Cross-Language Category Hierarchies
Fumiyo Fukumoto, Yoshimi Suzuki and Kazuyuki Yamashita 233

Mining Opinions in Comparative Sentences
Murthy Ganapathibhotla and Bing Liu 241

Integrating a Unification-Based Semantics in a Large Scale Lexicalised Tree Adjoining Grammar for French
Claire Gardent 249

Statistical Anaphora Resolution in Biomedical Texts
Caroline Gasperin and Ted Briscoe 257

Instance-Based Ontology Population Exploiting Named-Entity Substitution
Claudio Giuliano and Alfio Gliozzo 265

Measuring Topic Homogeneity and its Application to Dictionary-Based Word Sense Disambiguation
Ann Gledson and John Keane 273

Using Web-Search Results to Measure Word-Group Similarity
Ann Gledson and John Keane 281

An Algorithm for Adverbial Aspect Shift
Sabine Gründer 289

Dependency-Based N-Gram Models for General Purpose Sentence Realisation
Yuqing Guo, Josef van Genabith and Haifeng Wang 297

Homotopy-Based Semi-Supervised Hidden Markov Models for Sequence Labeling
Gholamreza Haffari and Anoop Sarkar 305

Tracking the Dynamic Evolution of Participants Salience in a Discussion
Ahmed Hassan, Anthony Fader, Michael H. Crespin, Kevin M. Quinn, Burt L. Monroe, Michael Colaresi and Dragomir R. Radev 313

Improving Statistical Machine Translation using Lexicalized Rule Selection
Zhongjun He, Qun Liu and Shouxun Lin 321

Evaluating Unsupervised Part-of-Speech Tagging for Grammar Induction
William P. Headden III, David McClosky and Eugene Charniak 329

Using Discourse Commitments to Recognize Textual Entailment
Andrew Hickl 337

Modeling Chinese Documents with Topical Word-Character Models
Wei Hu, Nobuyuki Shimizu, Hiroshi Nakagawa and Huanye Sheng 345

Non-Compositional Language Model and Pattern Dictionary Development for Japanese Compound and Complex Sentences
Satoru Ikehara, Masato Tokuhisa and Jin'ichi Murakami 353

Japanese Dependency Parsing Using a Tournament Model
Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto 361

Contents Modelling of Neo-Sumerian Ur III Economic Text Corpus
Wojciech Jaworski 369

Generating Chinese Couplets using a Statistical MT Approach
Long Jiang and Ming Zhou 377

Word Lattice Reranking for Chinese Word Segmentation and Part-of-Speech Tagging
Wenbin Jiang, Haitao Mi and Qun Liu 385

The Effect of Syntactic Representation on Semantic Role Labeling
Richard Johansson and Pierre Nugues 393

Using Hidden Markov Random Fields to Combine Distributional and Pattern-Based Word Clustering
Nobuhiro Kaji and Masaru Kitsuregawa 401

Textual Demand Analysis: Detection of Users' Wants and Needs from Opinions
Hiroshi Kanayama and Tetsuya Nasukawa 409

A Local Alignment Kernel in the Context of NLP
Sophia Katrenko and Pieter Adriaans 417

Coordination Disambiguation without Any Similarities
Daisuke Kawahara and Sadao Kurohashi 425

Generation of Referring Expressions: Managing Structural Ambiguities
Imtiaz Hussain Khan, Kees van Deemter and Graeme Ritchie 433

Normalizing SMS: are Two Metaphors Better than One ?
Catherine Kobus, François Yvon and Géraldine Damnati 441

The Choice of Features for Classification of Verbs in Biomedical Texts
Anna Korhonen, Yuval Krymolowski and Nigel Collier 449

Extending a Thesaurus with Words from Pan-Chinese Sources
Oi Yee Kwong and Benjamin K. Tsou 457

Stopping Criteria for Active Learning of Named Entity Recognition
Florian Laws and Hinrich Schütze 465

Reading the Markets: Forecasting Public Opinion of Political Candidates by News Analysis
Kevin Lerman, Ari Gilder, Mark Dredze and Fernando Pereira 473

Classifying What-Type Questions by Head Noun Tagging
Fangtao Li, Xian Zhang, Jinhui Yuan and Xiaoyan Zhu 481

PNR2: Ranking Sentences with Positive and Negative Reinforcement for Query-Oriented Update Summarization
Wenjie Li, Furu Wei, Qin Lu and Yanxiang He 489

Understanding and Summarizing Answers in Community-Based Question Answering Services
Yuanjie Liu, Shasha Li, Yunbo Cao, Chin-Yew Lin, Dingyi Han and Yong Yu 497

Tera-Scale Translation Models via Pattern Matching
Adam Lopez 505

Authorship Attribution and Verification with Many Authors and Limited Data
Kim Luyckx and Walter Daelemans 513

Modeling Semantic Containment and Exclusion in Natural Language Inference
Bill MacCartney and Christopher D. Manning 521

Linguistically-Based Sub-Sentential Alignment for Terminology Extraction from a Bilingual Automotive Corpus
Lieve Macken, Els Lefever and Veronique Hoste 529

Hindi Urdu Machine Transliteration using Finite-State Transducers
M. G. Abbas Malik, Christian Boitet and Pushpak Bhattacharyya 537

Comparative Parser Performance Analysis across Grammar Frameworks through Automatic Tree Conversion using Synchronous Grammars
Takuya Matsuzaki and Jun'ichi Tsujii 545

What's the Date? High Accuracy Interpretation of Weekday Names
Pawel Mazur and Robert Dale 553

When is Self-Training Effective for Parsing?
David McClosky, Eugene Charniak and Mark Johnson 561

A Syntactic Time-Series Model for Parsing Fluent and Disfluent Speech
Tim Miller and William Schuler 569

Applying Discourse Analysis and Data Mining Methods to Spoken OSCE Assessments
Meladel Mistica, Timothy Baldwin, Marisa Cordella and Simon Musgrave 577

Random Restarts in Minimum Error Rate Training for Statistical Machine Translation
Robert C. Moore and Chris Quirk 585

Robust Similarity Measures for Named Entities Matching
Erwan Moreau, François Yvon and Olivier Cappé 593

Modeling the Structure and Dynamics of the Consonant Inventories: A Complex Network Approach
Animesh Mukherjee, Monojit Choudhury, Anupam Basu and Niloy Ganguly 601

Detecting Multiple Facets of an Event using Graph-Based Unsupervised Methods
Pradeep Muthukrishnan, Joshua Gerrish and Dragomir R. Radev 609

Investigating Statistical Techniques for Sentence-Level Event Classification
Martina Naughton, Nicola Stokes and Joe Carthy 617

Exploring Domain Differences for the Design of a Pronoun Resolution System for Biomedical Text
Ngan L.T. Nguyen and Jin-Dong Kim 625

Computer Aided Correction and Extension of a Syntactic Wide-Coverage Lexicon
Lionel Nicolas, Benoît Sagot, Miguel A. Molinero, Jacques Farré and Eric de la Clergerie 633

Parsing the SynTagRus Treebank of Russian
Joakim Nivre, Igor M. Boguslavsky and Leonid L. Iomdin 641

Semantic Classification with Distributional Kernels
Diarmuid Ó Séaghdha and Ann Copestake 649

A Discriminative Alignment Model for Abbreviation Recognition
Naoaki Okazaki, Sophia Ananiadou and Jun'ichi Tsujii 657

Semantic Role Assignment for Event Nominalisations by Leveraging Verbal Data
Sebastian Padó, Marco Pennacchiotti and Caroline Sporleder 665

Recent Advances in a Feature-Rich Framework for Treebank Annotation
Petr Pajas and Jan Štěpánek 673

A Joint Information Model for N-Best Ranking
Patrick Pantel and Vishnu Vyas 681

Scientific Paper Summarization Using Citation Summary Networks
Vahed Qazvinian and Dragomir R. Radev 689

Exploiting Constituent Dependencies for Tree Kernel-Based Semantic Relation Extraction
Longhua Qian, Guodong Zhou, Fang Kong, Qiaoming Zhu and Peide Qian 697

A Method for Automatic POS Guessing of Chinese Unknown Words
Likun Qiu, Changjian Hu and Kai Zhao 705

Almost Flat Functional Semantics for Speech Translation
Manny Rayner, Pierrette Bouillon, Beth Ann Hockey and Yukie Nakao 713

Unsupervised Induction of Labeled Parse Trees by Clustering with Syntactic Features
Roi Reichart and Ari Rappoport 721

Anomalies in the WordNet Verb Hierarchy
Tom Richens 729

Translating Queries into Snippets for Improved Query Expansion
Stefan Riezler, Yi Liu and Alexander Vasserman 737

Classifying Chart Cells for Quadratic Complexity Context-Free Inference
Brian Roark and Kristy Hollingshead 745

Shift-Reduce Dependency DAG Parsing
Kenji Sagae and Jun'ichi Tsujii 753

Event Frame Extraction Based on a Gene Regulation Corpus
Yutaka Sasaki, Paul Thompson, Philip Cotter, John McNaught and Sophia Ananiadou 761

A Fully-Lexicalized Probabilistic Model for Japanese Zero Anaphora Resolution
Ryohei Sasano, Daisuke Kawahara and Sadao Kurohashi 769

Estimation of Conditional Probabilities With Decision Trees and an Application to Fine-Grained POS Tagging
Helmut Schmid and Florian Laws 777

Toward a Psycholinguistically-Motivated Model of Language Processing
William Schuler, Samir AbdelRahman, Tim Miller and Lane Schwartz 785

Metric Learning for Synonym Acquisition
Nobuyuki Shimizu, Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama and Hiroshi Nakagawa 793

Discourse Level Opinion Interpretation
Swapna Somasundaran, Janyce Wiebe and Josef Ruppenhofer 801

Acquiring Sense Tagged Examples using Relevance Feedback
Mark Stevenson, Yikun Guo and Robert Gaizauskas 809

Topic Identification for Fine-Grained Opinion Analysis
Veselin Stoyanov and Claire Cardie 817

From Words to Senses: A Case Study of Subjectivity Recognition
Fangzhong Su and Katja Markert 825

Prediction of Maximal Projection for Semantic Role Labeling
Weiwei Sun, Zhifang Sui and Haifeng Wang 833

Modeling Latent-Dynamic in Shallow Parsing: A Latent Conditional Model with Imrpoved Inference
Xu Sun, Louis-Philippe Morency, Daisuke Okanohara, Yoshimasa Tsuruoka and Jun'ichi Tsujii841

Learning Entailment Rules for Unary Templates
Idan Szpektor and Ido Dagan 849

Experiments with Reasoning for Temporal Relations between Events
Marta Tatu and Munirathnam Srikanth . . . 857

The Ups and Downs of Preposition Error Detection in ESL Writing
Joel R. Tetreault and Martin Chodorow . . . 865

A Framework for Identifying Textual Redundancy
Kapil Thadani and Kathleen McKeown . . . 873

Emotion Classification Using Massive Examples Extracted from the Web
Ryoko Tokhisa, Kentaro Inui and Yuji Matsumoto . . . 881

Relational-Realizational Parsing
Reut Tsarfaty and Khalil Sima'an . . . 889

Training Conditional Random Fields Using Incomplete Annotations
Yuta Tsuboi, Hisashi Kashima, Shinsuke Mori, Hiroki Oda and Yuji Matsumoto . . . 897

A Uniform Approach to Analogies, Synonyms, Antonyms, and Associations
Peter Turney . . . 905

Tighter Integration of Rule-Based and Statistical MT in Serial System Combination
Nicola Ueffing, Jens Stephan, Evgeny Matusov, Loïc Dugast, George Foster, Roland Kuhn, Jean Senellart and Jin Yang . . . 913

Class-Driven Attribute Extraction
Benjamin Van Durme, Ting Qian and Lenhart Schubert . . . 921

Using Three Way Data for Word Sense Discrimination
Tim Van de Cruys . . . 929

Source Language Markers in EUROPARL Translations
Hans van Halteren . . . 937

A Fluid Knowledge Representation for Understanding and Generating Creative Metaphors
Tony Veale and Yanfen Hao . . . 945

Using Syntactic Information for Improving Why-Question Answering
Suzan Verberne, Lou Boves, Nelleke Oostdijk and Peter-Arno Coppen . . . 953

Coreference Systems Based on Kernels Methods
Yannick Versley, Alessandro Moschitti, Massimo Poesio and Xiaofeng Yang . . . 961

CollabRank: Towards a Collaborative Approach to Single-Document Keyphrase Extraction
Xiaojun Wan and Jianguo Xiao . . . 969

Investigating the Portability of Corpus-Derived Cue Phrases for Dialogue Act Classification
Nick Webb and Ting Liu . . . 977

Extractive Summarization Using Supervised and Semi-Supervised Learning
Kam-Fai Wong, Mingli Wu and Wenjie Li . . . 985

Domain Adaptation for Statistical Machine Translation with Domain Dictionary and Monolingual Corpora
Hua Wu, Haifeng Wang and Chengqing Zong . . . 993

Exploiting Graph Structure for Accelerating the Calculation of Shortest Paths in Wordnets
Holger Wunsch 1001

Linguistically Annotated BTG for Statistical Machine Translation
Deyi Xiong, Min Zhang, Aiti Aw and Haizhou Li 1009

Bayesian Semi-Supervised Chinese Word Segmentation for Statistical Machine Translation
Jia Xu, Jianfeng Gao, Kristina Toutanova and Hermann Ney 1017

Switching to Real-Time Tasks in Multi-Tasking Dialogue
Fan Yang, Peter A. Heeman and Andrew Kun 1025

Chinese Term Extraction Using Minimal Resources
Yuhang Yang, Qin Lu and Tiejun Zhao 1033

Measuring and Predicting Orthographic Associations: Modelling the Similarity of Japanese Kanji
Lars Yencken and Timothy Baldwin 1041

Chinese Dependency Parsing with Large Scale Automatically Constructed Case Structures
Kun Yu, Daisuke Kawahara and Sadao Kurohashi 1049

OntoNotes: Corpus Cleanup of Mistaken Agreement Using Word Sense Disambiguation
Liang-Chih Yu, Chung-Hsien Wu and Eduard Hovy 1057

An Integrated Probabilistic and Logic Approach to Encyclopedia Relation Extraction with Multiple Features
Xiaofeng Yu and Wai Lam 1065

Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Text
Taras Zagibalov and John Carroll 1073

Extracting Synchronous Grammar Rules From Word-Level Alignments in Linear Time
Hao Zhang, Daniel Gildea and David Chiang 1081

Sentence Type Based Reordering Model for Statistical Machine Translation
Jiajun Zhang, Chengqing Zong and Shoushan Li 1089

Grammar Comparison Study for Translational Equivalence Modeling and Statistical Machine Translation
Min Zhang, Hongfei Jiang, Haizhou Li, Aiti Aw and Sheng Li 1097

Automatic Generation of Parallel Treebanks
Ventsislav Zhechev and Andy Way 1105

A Hybrid Generative/Discriminative Framework to Train a Semantic Parser from an Un-annotated Corpus
Deyu Zhou and Yulan He 1113

Diagnostic Evaluation of Machine Translation Systems Using Automatically Constructed Linguistic Check-Points
Ming Zhou, Bo Wang, Shujie Liu, Mu Li, Dongdong Zhang and Tiejun Zhao 1121

Multi-Criteria-Based Strategy to Stop Active Learning for Data Annotation
Jingbo Zhu, Huizhen Wang and Eduard Hovy 1129

Active Learning with Sampling by Uncertainty and Density for Word Sense Disambiguation and Text Classification
Jingbo Zhu, Huizhen Wang, Tianshun Yao and Benjamin K. Tsou 1137

A Systematic Comparison of Phrase-Based, Hierarchical and Syntax-Augmented Statistical MT
Andreas Zollmann, Ashish Venugopal, Franz Och and Jay Ponte 1145

Choosing the Right Translation: A Syntactically Informed Classification Approach
Simon Zwarts and Mark Dras 1153