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 Hajčová, 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 Hajicová (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 Xavier Carreras Bruno Gaume
Eneko Agirre Justine Cassell Robert Gaizauskas
Lars Ahrenberg Eric Castelli Michel Galley
Akiho Aizawa Jason Chang Aldo Gangemi
Manuel Alcantara Ming-Wei Chang Jianfeng Gao
Beatrice Alex Eugene Charniak Claire Gardent
Galen Andrew Hsin-Hsi Chen Eric Gaussier
Ash Asudeh Keh-jiann Chen Monte George
Bogdan Babych Colin Cherry Ulrich Germann
Jason Baldridge David Chiang Jonathan Ginzburg
Carmen Banea Philipp Cimiano Roxana Girju
Srinivas Bangalore Silvie Cinková
Roy Bar-Haim Stephen Clark John Goldsmith
Marco Baroni Kevin Cohen Sharon Goldwater
Regina Barzilay Trevor Cohn Julio Gonzalo
Roberto Basili Nigel Collier Ralph Grishman
Núria Bel Michael Collins Claire Grover
Linda Bell Cleo Condoravdi Joakim Gustafsson
Anja Belz Michael Connor Nizar Habash
Pushpak Bhattacharyya Robin Cooper Benjamin Hachey
Dan Bikel James Curran Udo Hahn
Alan Black Ido Dagan Jan Hajčíč
Patrick Blackburn Robert Dale Eva Hajčíková
Phil Blunsom Mary Dalrymple Keith Hall
Christian Boitet Hal Daumé Sanda Harabagiu
Kalina Bontcheva Quang Xuan Do Mary Harper
Johan Bos Bill Dolan Tingting He
Chris Brew Chris Dyer Xiaodong He
Ted Briscoe Marc Dymetman William Headden
Chris Brockett Markus Egg Mary Hearne
Paul Buitelaar Andreas Eisele Marti Hearst
Razvan Bunescu Jason Eisner James Henderson
Harry Bunt Mark Ellison Mark Hepple
Stephan Busemann Micha Elsner Andy Hickl
Miriam Butt Alex Chenyu Fang Keikichi Hirose
Bill Byrne Christiane Fellbaum Julia Hirschberg
Janet Cahn Charles Fillmore Graeme Hirst
Chris Callison-Burch George Foster Jerry Hobbs
Nicoletta Calzolari Gil Francopolou Julia Hockenmaier
Nicola Cancedda Alex Fraser Merle Horne
Sander Canisius Dayne Freitag David House
Rolf Carlson Atsushi Fujita Shu-Kai Hsieh
Marine Carpuat Adam Funk Chu-Ren Huang
<table>
<thead>
<tr>
<th>Name</th>
<th>Name</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kentaro Inui</td>
<td>Diane Litman</td>
<td>Cécile Paris</td>
</tr>
<tr>
<td>Pierre Isabelle</td>
<td>Bing Liu</td>
<td>Jeff Pelletier</td>
</tr>
<tr>
<td>David Israel</td>
<td>Yang Liu</td>
<td>Gerald Penn</td>
</tr>
<tr>
<td>Abe Itycheriah</td>
<td>Adam Lopez</td>
<td>Marco Pennachiotti</td>
</tr>
<tr>
<td>Donghong Ji</td>
<td>Qin Lu</td>
<td>Wim Peters</td>
</tr>
<tr>
<td>Su Jian</td>
<td>Kim Teng Lua</td>
<td>Slav Petrov</td>
</tr>
<tr>
<td>Jing Jiang</td>
<td>Klaus Macherey</td>
<td>Anselmo Peñas</td>
</tr>
<tr>
<td>Howard Johnson</td>
<td>Wolfgang Macherey</td>
<td>Manfred Pinkal</td>
</tr>
<tr>
<td>Mark Johnson</td>
<td>Kikuo Maekawa</td>
<td>Massimo Poesio</td>
</tr>
<tr>
<td>Michael Johnston</td>
<td>Bernardo Magnini</td>
<td>Thierry Poibeau</td>
</tr>
<tr>
<td>Kristina Jokinen</td>
<td>Inderjeet Mani</td>
<td>Ana-Maria Popescu</td>
</tr>
<tr>
<td>Gareth Jones</td>
<td>Daniel Marcu</td>
<td>Robert Porzel</td>
</tr>
<tr>
<td>Aravind Joshi</td>
<td>Katja Markert</td>
<td>Richard Power</td>
</tr>
<tr>
<td>Asanee Kawtrakul</td>
<td>Erwin Marsi</td>
<td>John Prager</td>
</tr>
<tr>
<td>Noriko Kando</td>
<td>Andrew McCallum</td>
<td>Ian Pratt-Hartman</td>
</tr>
<tr>
<td>Kyoko Kanzaki</td>
<td>Diana McCarthy</td>
<td>Judita Preiss</td>
</tr>
<tr>
<td>Jussi Karlgren</td>
<td>David McClosky</td>
<td>Laurent Prévot</td>
</tr>
<tr>
<td>Junichi Kazama</td>
<td>Majorie McShane</td>
<td>Stephen Pulman</td>
</tr>
<tr>
<td>Jin-Dong Kim</td>
<td>Arul Menezes</td>
<td>Vasin Punyakanok</td>
</tr>
<tr>
<td>Tracy King</td>
<td>Paola Merlo</td>
<td>James Pustejovsky</td>
</tr>
<tr>
<td>Ewan Klein</td>
<td>Rada Mihalcea</td>
<td>Chris Quirk</td>
</tr>
<tr>
<td>Alexandre Klementiev</td>
<td>Yusuke Miyao</td>
<td>Owen Rambow</td>
</tr>
<tr>
<td>Kevin Knight</td>
<td>Monica Monachini</td>
<td>Alian Ramsay</td>
</tr>
<tr>
<td>Alistair Knott</td>
<td>Simonetta Montemagni</td>
<td>Ari Rappoport</td>
</tr>
<tr>
<td>Philipp Koehn</td>
<td>Raymond Mooney</td>
<td>Lev Ratinov</td>
</tr>
<tr>
<td>Hasida Koiti</td>
<td>Robert Moore</td>
<td>Roi Reichart</td>
</tr>
<tr>
<td>Alexander Koller</td>
<td>Paul Morarescu</td>
<td>Ehud Reiter</td>
</tr>
<tr>
<td>Emiel Krahmer</td>
<td>Alessandro Moschitti</td>
<td>Stefan Riezler</td>
</tr>
<tr>
<td>Ivana Kruijff-Korbayova</td>
<td>Reinhard Muskens</td>
<td>Ellen Riloff</td>
</tr>
<tr>
<td>Geert-Jan Kruijff</td>
<td>Tetsuya Nasukawa</td>
<td>Hae-Chang Rim</td>
</tr>
<tr>
<td>Roland Kuhn</td>
<td>Ani Nenkova</td>
<td>Brian Roark</td>
</tr>
<tr>
<td>Shankar Kumar</td>
<td>Günter Neumann</td>
<td>Barbara Rosario</td>
</tr>
<tr>
<td>Sadao Kurohashi</td>
<td>Hwee Tou Ng</td>
<td>Mathias Rossignol</td>
</tr>
<tr>
<td>Philippe Langlais</td>
<td>Vincent Ng</td>
<td>Dan Roth</td>
</tr>
<tr>
<td>Stefan Larsson</td>
<td>Thi Minh Huyen Nguyen</td>
<td>Rachel Roxas</td>
</tr>
<tr>
<td>Alberto Lavelli</td>
<td>Nicolas Nicolov</td>
<td>Vasile Rus</td>
</tr>
<tr>
<td>Claudia Leacock</td>
<td>Joakim Nivre</td>
<td>Kenji Sagaie</td>
</tr>
<tr>
<td>Matt Lease</td>
<td>Franz Och</td>
<td>Mark Sammons</td>
</tr>
<tr>
<td>Gary Geunbae Lee</td>
<td>Kemal Oflazer</td>
<td>Rajeev Sangal</td>
</tr>
<tr>
<td>Kiyong Lee</td>
<td>Naoaki Okazaki</td>
<td>Anoop Sarkar</td>
</tr>
<tr>
<td>Lillian Lee</td>
<td>Manabu Okumura</td>
<td>Len Schubert</td>
</tr>
<tr>
<td>Gregor Leusch</td>
<td>Alessandro Oltramari</td>
<td>Holger Schwenk</td>
</tr>
<tr>
<td>Gina Levow</td>
<td>Miles Osborne</td>
<td>Donia Scott</td>
</tr>
<tr>
<td>Hang Li</td>
<td>Martha Palmer</td>
<td>Satoshi Sekine</td>
</tr>
<tr>
<td>Sujian Li</td>
<td>Shimei Pan</td>
<td>Stephanie Seneff</td>
</tr>
<tr>
<td>Chin-Yew Lin</td>
<td>Bo Pang</td>
<td>Petr Sgall</td>
</tr>
<tr>
<td>Ken Litkowski</td>
<td>Soo-Min Pantel</td>
<td>Libin Shen</td>
</tr>
<tr>
<td>Nobuyuki Shimizu</td>
<td>Simone Teufel</td>
<td>Piek Vossen</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Elizabeth Shriberg</td>
<td>Jörg Tiedemann</td>
<td>Stephen Wan</td>
</tr>
<tr>
<td>Khalil Sima’an</td>
<td>Christoph Tillmann</td>
<td>Xinglong Wang</td>
</tr>
<tr>
<td>Michel Simard</td>
<td>Takenobu Tokunaga</td>
<td>Taro Watanabe</td>
</tr>
<tr>
<td>Kevin Small</td>
<td>Kentaro Torisawa</td>
<td>Andy Way</td>
</tr>
<tr>
<td>Noah Smith</td>
<td>Kristina Toutanova</td>
<td>Bonnie Webber</td>
</tr>
<tr>
<td>Pavel Smrz</td>
<td>Isabel Trancoso</td>
<td>Davy Weissenbacher</td>
</tr>
<tr>
<td>Stephen Soderland</td>
<td>Shu-Chuan Tseng</td>
<td>Janyce Wiebe</td>
</tr>
<tr>
<td>Claudia Soria</td>
<td>Jun’ichi Tsujii</td>
<td>Yorick Wilks</td>
</tr>
<tr>
<td>Virach Sornlertlamvanich</td>
<td>Dan Tuříš</td>
<td>Kam-Fai Wong</td>
</tr>
<tr>
<td>Sofia Stamou</td>
<td>Nicola Ueffing</td>
<td>Chung-Hsien Wu</td>
</tr>
<tr>
<td>Manfred Stede</td>
<td>Hans Uszkoreit</td>
<td>Nianwen Xue</td>
</tr>
<tr>
<td>Mark Steedman</td>
<td>Takehito Utsuro</td>
<td>Roman Yangarber</td>
</tr>
<tr>
<td>Armando Stellato</td>
<td>Kees van Deemter</td>
<td>Naoki Yoshinaga</td>
</tr>
<tr>
<td>Amanda Stent</td>
<td>Josef van Genabith</td>
<td>Kun Yu</td>
</tr>
<tr>
<td>Mark Stevenson</td>
<td>Antal van den Bosch</td>
<td>Annie Zaenen</td>
</tr>
<tr>
<td>Pavel Stranak</td>
<td>Walther von Hahn</td>
<td>Fabio Massimo Zanzotto</td>
</tr>
<tr>
<td>Tomek Strzalkowski</td>
<td>Lucy Vanderwende</td>
<td>Jun Zhao</td>
</tr>
<tr>
<td>Le Sun</td>
<td>Sebastian Vargas</td>
<td>Jing Zheng</td>
</tr>
<tr>
<td>Hisami Suzuki</td>
<td>Tony Veale</td>
<td>Ming Zhou</td>
</tr>
<tr>
<td>Jun Suzuki</td>
<td>Paola Velardi</td>
<td>Michael Zock</td>
</tr>
<tr>
<td>Marc Swerts</td>
<td>Ashish Venugopal</td>
<td>Chengqing Zong</td>
</tr>
<tr>
<td>Idan Szpektor</td>
<td>Barbora Vidová Hladká</td>
<td>Richard Zens</td>
</tr>
</tbody>
</table>
# 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 Crystmann, 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
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measuring Topic Homogeneity and its Application to Dictionary-Based Word Sense Disambiguation</td>
<td>Ann Gledson and John Keane</td>
<td>273</td>
</tr>
<tr>
<td>Using Web-Search Results to Measure Word-Group Similarity</td>
<td>Ann Gledson and John Keane</td>
<td>281</td>
</tr>
<tr>
<td>An Algorithm for Adverbial Aspect Shift</td>
<td>Sabine Gründers</td>
<td>289</td>
</tr>
<tr>
<td>Dependency-Based N-Gram Models for General Purpose Sentence Realisation</td>
<td>Yuqing Guo, Josef van Genabith and Haifeng Wang</td>
<td>297</td>
</tr>
<tr>
<td>Homotopy-Based Semi-Supervised Hidden Markov Models for Sequence Labeling</td>
<td>Gholamreza Haffari and Anoop Sarkar</td>
<td>305</td>
</tr>
<tr>
<td>Tracking the Dynamic Evolution of Participants Salience in a Discussion</td>
<td>Ahmed Hassan, Anthony Fader, Michael H. Crespin, Kevin M. Quinn, Burt L. Monroe, Michael Colaresi and Dragomir R. Radev</td>
<td>313</td>
</tr>
<tr>
<td>Improving Statistical Machine Translation using Lexicalized Rule Selection</td>
<td>Zhongjun He, Qun Liu and Shouxun Lin</td>
<td>321</td>
</tr>
<tr>
<td>Evaluating Unsupervised Part-of-Speech Tagging for Grammar Induction</td>
<td>William P. Headden III, David McClosky and Eugene Charniak</td>
<td>329</td>
</tr>
<tr>
<td>Using Discourse Commitments to Recognize Textual Entailment</td>
<td>Andrew Hickl</td>
<td>337</td>
</tr>
<tr>
<td>Modeling Chinese Documents with Topical Word-Character Models</td>
<td>Wei Hu, Nobuyuki Shimizu, Hiroshi Nakagawa and Huanye Sheng</td>
<td>345</td>
</tr>
<tr>
<td>Japanese Dependency Parsing Using a Tournament Model</td>
<td>Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto</td>
<td>361</td>
</tr>
<tr>
<td>Contents Modelling of Neo-Sumerian Ur III Economic Text Corpus</td>
<td>Wojciech Jaworski</td>
<td>369</td>
</tr>
<tr>
<td>Generating Chinese Couplets using a Statistical MT Approach</td>
<td>Long Jiang and Ming Zhou</td>
<td>377</td>
</tr>
<tr>
<td>Word Lattice Reranking for Chinese Word Segmentation and Part-of-Speech Tagging</td>
<td>Wenbin Jiang, Haitao Mi and Qun Liu</td>
<td>385</td>
</tr>
<tr>
<td>The Effect of Syntactic Representation on Semantic Role Labeling</td>
<td>Richard Johansson and Pierre Nugues</td>
<td>393</td>
</tr>
<tr>
<td>Using Hidden Markov Random Fields to Combine Distributional and Pattern-Based Word Clustering</td>
<td>Nobuhiro Kaji and Masaru Kitsuregawa</td>
<td>401</td>
</tr>
<tr>
<td>Textual Demand Analysis: Detection of Users’ Wants and Needs from Opinions</td>
<td>Hiroshi Kanayama and Tetsuya Nasukawa</td>
<td>409</td>
</tr>
</tbody>
</table>
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
Intiaz 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
Paweł 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, Janye 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 Improved Inference
   Xu Sun, Louis-Philippe Morency, Daisuke Okanohara, Yoshimasa Tsuruoka and Jun’ichi Tsujii 841

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-Chi̇h 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