EACL-2006

11th Conference
of the European Chapter of the
Association for Computational Linguistics

Proceedings of the Conference

April, 3rd – 7th, 2006
Trento, Italy

The conference, the workshop and the tutorials are sponsored by:

Center for the Evaluation of Language and Communication Technologies

Celct
c/o BIC, Via dei Solteri, 38
38100 Trento, Italy
http://www.celct.it

Xerox Research Centre Europe
6 Chemin de Maupertuis
38240 Meylan, France
http://www.xrce.xerox.com

CELI s.r.l.
Corso Moncalieri, 21
10131 Torino, Italy
http://www.celi.it

Thales
45 rue de Villiers
92526 Neuilly-sur-Seine Cedex, France
http://www.thalesgroup.com

EACL-2006 is supported by

Trentino S.p.a. and Metalsistem Group

© April 2006, Association for Computational Linguistics

ISBN 1-932432-59-0

Order copies of ACL proceedings from:
Priscilla Rasmussen,
Association for Computational Linguistics (ACL),
3 Landmark Center,
East Stroudsburg, PA 18301 USA

Phone +1-570-476-8006
Fax +1-570-476-0860
E-mail: acl@aclweb.org
On-line order form: http://www.aclweb.org/

ISBN 1-932432-59-0
90000>
9 781932 432596

Preface

This volume contains the papers of the main session of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL). The conference also contains tutorials, workshops, posters, demos and a student workshop, which have companion volumes to these proceedings. We are proud to present two invited speakers at the main conference: Alfonso Caramazza, of the Cognitive Neuropsychology Laboratory, Harvard University, and Kevin Knight, of the Information Sciences Institute, University of Southern California. We are grateful to both for agreeing to deliver the keynote talks.

The main conference for EACL 2006 had a 46% increase in submitted papers compared to the 10th EACL which was held in 2003. This year we had 264 papers submitted from 35 countries in five continents (according to the affiliation of the corresponding author). The large number of papers submitted, and the broad geographical basis, demonstrate the growth in our field and considerable interest in EACL as a truly international conference. We would like to thank all authors of submitted papers for selecting EACL for presentation of their work.

Given the very substantial number of submitted papers, selection of those to appear at the conference, and therefore in these proceedings, has been a very challenging task. All papers were read by a minimum of three people, with discussions where there were differences of opinion. The final selections were made at a one and a half-day program committee meeting in Brighton. Of the 264 submissions, 52 papers were accepted giving an acceptance rate of 19.7%. We appreciate that this constitutes a considerably more competitive acceptance rate than EACL 2003, inevitable in light of the great increase in the number of submissions. The variety of work in the accepted papers highlights the great breadth of work in our field. We really hope that you both enjoy attending the conference and perusing these proceedings.

There are a great many people who have collaborated with us in the organisation of this conference. We are extremely grateful to the substantial effort put in by our 10 area chairs: Paul Buitelaar, Jason Eisner, Marcello Federico, Rob Gaizauskas, Mirella Lapata, Massimo Poesio, Ehud Reiter, Giorgio Satta, Antal van den Bosch and Andy Way. We are also indebted to the 240 reviewers for their expertise and hard work. Their careful scrutiny of papers and considered opinions are vital to the quality of the conference. We are indebted to Gertjan van Noord, the EACL President, for his advice and guidance. We benefited from great collaboration with the local organizers (Alberto Lavelli and Bernardo Magnini, co-chairs), who welcome us all to the beautiful city of Trento, and the publication chairs (Eneko Agirre, Sergio Balari and Emanuele Pianta), who have had the hard work of preparing the proceedings. We have also enjoyed collaboration with chairs for the poster & demo sessions (Frank Keller and Gabor Proszkey), workshops (Maarten de Rijke and Caroline Sporleder), student session (Sebastian Padó, Jonathon Read and Violeta Seretan) and tutorials (Alexis Nasr, Kemal Oflazer and Miles Osborne), all of whom have been essential in putting together what we hope will be a truly stimulating program together.

We hope you have a very enjoyable time at EACL 2006.

Diana McCarthy and Shuly Wintner (Program co-chairs)

EACL 2006 Conference Organizers

Program chairs: Diana McCarthy, *University of Sussex, UK*; Shuly Wintner, *University of Haifa, Israel*.

Tutorial chairs: Alexis Nasr, *University of Paris 7, France*; Kemal Oflazer, *Sabancy University, Turkey*; Miles Osborne, *University of Edinburgh, UK*.

Workshop chairs: Maarten de Rijke, *University of Amsterdam, Netherlands*; Caroline Sporleder, *Tilburg University, The Netherlands*.

Posters and Demonstrations chairs: Frank Keller, *University of Edinburgh, UK*; Gabor Proszeky, *Morphologic, Hungary*.

Student Research Workshop chairs: Sebastian Padó, *Saarland University, Germany*; Jonathon Read, *University of Sussex, UK*; Violeta Seretan, *University of Geneva, Switzerland*.

Publication chairs: Eneko Agirre, *Basque Country University*; Sergio Balari Ravera, *Universitat Autònoma de Barcelona*; Emanuele Pianta, *ITC-irst, Italy*.

Sponsorships and Exhibitions chairs: Geert-Jan M. Kruijff, *DFKI, Germany*; Frédérique Segonde, *Xerox Research Centre Europe, France*.

Local organizers: Alberto Lavelli (chair), *ITC-irst, Italy*; Bernardo Magnini (chair), *ITC-irst, Italy*; Christian Girardi, *ITC-irst, Italy*; Barbara Mezzabotta, *ITC-irst, Italy*; Matteo Negri, *ITC-irst, Italy*; Susana Otero Díaz, *ITC-irst, Italy*; Emanuele Pianta, *ITC-irst, Italy*; Carlo Strapparava, *ITC-irst, Italy*; Alessandro Tuccio, *ITC, Italy*.

Conference Webmaster: Christian Girardi, *ITC-irst, Italy*.

EACL Executive Committee: Gertjan van Noord, *University of Groningen, The Netherlands*; Anette Frank, *DFKI, Germany*; Michael Rosner, *University of Malta, Malta*; Alex Lascarides, *University of Edinburgh, UK*.

EACL 2006 Program Committee

Program Co-chairs: Diana McCarthy, *University of Sussex, UK*; Shuly Wintner, *University of Haifa, Israel*.

Area chairs: Antal van den Bosch (Tilburg, NL)—Machine learning, algorithms for NLP; Paul Buitelaar (DFKI, DE)—Lexicon, lexical databases, ontologies, language resources; Jason Eisner (Johns Hopkins, US)—Phonology, morphology, finite-state technology, tagging, word segmentation; Marcello Federico (ITC, IT)—Speech, language modelling, spoken dialog systems; Rob Gaizauskas (Sheffield, UK)—Question answering, information extraction, information retrieval; Mirella Lapata (Edinburgh, UK)—Semantics, including lexical semantics, formal semantics and logic, Textual entailment; Massimo Poesio (Essex, UK)—Pragmatics, dialog systems, discourse; Ehud Reiter (Aberdeen, UK)—Summarization, generation; Giorgio Satta (Padua, IT)—Syntax, parsing, formalisms; Andy Way (Dublin, IE)—Machine translation.

Reviewers: Steve Abney, Eneko Agirre, Jan Alexandersson, Rie Kubota Ando, Ion Androutsopoulos, Paolo Baggia, TimBaldwin, Srinivas Bangalore, Roberto Basili, John Bateman, AnjaBelz, Nicola Bertoldi, DanBikel, Bran Boguraev, Johan Bos, Pierre Boullier, ThorstenBrants, Chris Brew, Ralf Brown, Paul Buitelaar, Stephan Busemann, Donna Byron, Charles Callaway, Sander Canisius, Michael Carl, John Carroll, Mauro Cettolo, Jean-Cédric Chappelier, David Chiang, Grace Chung, Ilyas Cicekli, Philipp Cimiano, Fabio Ciravegna, Alexander Clark, Stephen Clark, MichaelCollins, Ann Copestake, Koby Crammer, Silviu Cucerzan, James Curran, James Cussens, Walter Daelemans, Ido Dagan, Maarten de Rijke, Barbara Di Eugenio, Bonnie Dorr, Myroslava Dzikovska, Phil Edmonds, Jason Eisner, Michael Elhadad, Tomaz Erjavec, Roger Evans, Stefan Evert, MarcelloFederico, Radu Florian, Mikel Forcada, Bob Frank, Tamas Gaal, Rob Gaizauskas, AldoGangemi, Jianfeng Gao, Claire Gardent, Eric Gaussier, Dale Gerdemann, Daniel Gildea, Jonathan Ginzburg, Julio Gonzalo, Cyril Goutte, GregGrefenstette, Trond Grenager, Ralph Grishman, Claire Grover, Nizar Habash, Udo Hahn, Patrick Hanks, Sanda Harabagiu, Daniel Hardt, Henk Harkema, Mary Hearne, Marti Hearst, Uli Heid, James Henderson, Mark Hepple, Graeme Hirst, John Hutchins, Kristiina Jokinen, Laura Kallmeyer, Stephan Kanthak, Nikiforos Karamanis, Dimitar Kazakov, Andre Kempe, Dorothy Kenny, Adam Kilgarriff, Katrin Kirchhoff, Dan Klein, Ewan Klein, Kevin Knight, Philipp Koehn, Rob Koeling, Alexander Koller, Andras Kornai, Kimmo Koskenniemi, Emiel Krahmer, Steven Krauwer, Hans-Ulrich Krieger, Ivana Kruijff-Korbayova, Udo Kruschwitz, Taku Kudo, Shankar Kumar, Nicholas Kushmerick, Irene Langkilde-Geary, Philippe Langlais, Mirella Lapata, Eric Laporte, Alex Lascarides, Alberto Lavelli, Alon Lavie, Oliver Lemon, Alessandro Lenci, Gina-Anne Levow, Marc Light, Dekang Lin, Jimmy Lin, Ying Lin, Diane Litman, Klaus Macherey, Bernardo Magnini, Rob Malouf, Daniel Marcu, José B. Mariño, Katja Markert, Lluís Márquez, Erwin Marsi, Yuji Matsumoto, John Maxwell, Michael Maxwell, Diana McCarthy, Ryan McDonald, Dan Melamed, Chris Mellish, Paola Merlo, Rada Mihalcea, Stoyan Mihov, David Milward, Teruko Mitamura, Sien Moens, Dan Moldovan, Christof Monz, Alessandro Moschitti, Larry Moss, Stefan Mueller, Yang Muyung, Vivi Nastase, Claire Nedellec, Mark-Jan Nederhof, Gunter Neumann, Hermann Ney, Hwee Tou Ng, Joakim Nivre, Elena Not, Jon Oberlander, Franz Och, Kemal Oflazer, Miles Osborne, Sebastian Padó, Karel Pala, Martha Palmer, Patrick Pantel, Ted Pedersen, Fuchun Peng, Gerald Penn, Manfred Pinkal, Massimo Poesio, Robert Porzel, John Prager, Detlef Prescher, Matthew Purver, Owen Rambow, Ehud Reiter, Jean-Michel Renders, Germán Rigau, Graeme Ritchie, Brian Roark, Jim Rogers, Dan Roth, Stefan Rueping, Horacio Saggion, Anna Sagvall Hein, Anoop Sarkar, Giorgio Satta, David Schlangen,

Tanja Schultz, Holger Schwenk, Satoshi Sekine, Stuart Shieber, Khalil Simaan, Harold Somers, Karen Sparck-Jones, Caroline Sporleder, Richard Sproat, Peter Spyns, Patrick Saint Dizier, Ed Stabler, Manfred Stede, Mark Stevenson, Suzanne Stevenson, Oliver Streiter, Michael Strube, Tomek Strzlkowski, Jian Su, Key-Yih Su, Pasi Tapanainen, Louis ten Bosch, Simone Teufel, Stefanie Tomko, Kristina Toutanova, David Traum, Harald Trost, Peter Turney, Kees van Deemter, Antal van den Bosch, Gertjan van Noord, Menno van Zaanen, Paola Velardi, Renata Vieira, Eric Villemonte de la Clergerie, Martin Volk, Piek Vossen, Atro Voutilainen, Chao Wang, Andy Way, Bonnie Webber, David Weir, Michael White, Richard Wicentowski, Shuly Wintner, Dekai Wu, Peng Xu.

Conference Program

First Day: Wedneseday, April 5, 2006

9:00-9:30 Opening Session
9:30-10:30 **Invited Talk 1:** *Isn't Linguistic Structure Important, Asked the Engineer* by Kevin Knight

10:30-11:00 Coffee Break

Session 1: Lexicon

11:00-11:30 *Inheritance and the CCG Lexicon*
Mark McConville

11:30-12:00 *Using Encyclopedic Knowledge for Named entity Disambiguation*
Razvan Bunescu and Marius Paşca

12:00-12:30 *Weakly Supervised Approaches for Ontology Population*
Hristo Tanev and Bernardo Magnini

12:30-14:30 Lunch

Session 2: Machine Translation

14:30-15:00 *Computational Complexity of Statistical Machine Translation*
Raghavendra Udupa U. and Hemanta K. Maji

15:00-15:30 *Computing Consensus Translation for Multiple Machine Translation Systems Using Enhanced Hypothesis Alignment*
Evgeny Matusov, Nicola Ueffing and Hermann Ney

15:30-16:00 *Phrase-Based Backoff Models for Machine Translation of Highly Inflected Languages*
Mei Yang and Katrin Kirchhoff

16:00-16:30 Coffee Break

Session 3: Dialog (spoken)

16:30-17:00 *Automatic Detection of Nonreferential It in Spoken Multi-Party Dialog*
Christoph Müller

17:00-17:30 *Generating Statistical Language Models from Interpretation Grammars in Dialogue Systems*
Rebecca Jonson

17:30-18:00 *Information Presentation in Spoken Dialogue Systems*
Vera Demberg and Johanna D. Moore

First Day: Wedneseday, April 5, 2006 (*cont.*)

Session 4: Parsing

11:00-11:30 *Constraints on Non-Projective Dependency Parsing*
Joakim Nivre

11:30-12:00 *Online Learning of Approximate Dependency Parsing Algorithms*
Ryan McDonald and Fernando Pereira

12:00-12:30 *Statistical Dependency Parsing for Turkish*
Gülşen Eryiğit and Kemal Oflazer

Session 5: Machine Learning

14:30-15:00 *Generalized Hebbian Algorithm for Incremental Singular Value Decomposition in Natural Language Processing*
Genevieve Gorrell

15:00-15:30 *Improving Probabilistic Latent Semantic Analysis with Principal Component Analysis*
Ayman Farahat and Francine Chen

15:30-16:00 *Making Tree Kernels Practical for Natural Language Learning*
Alessandro Moschitti

Session 6: Word Sense

16:30-17:00 *Determining Word Sense Dominance Using a Thesaurus*
Saif Mohammad and Graeme Hirst

17:00-17:30 *Experiments on the Validation of Sense Annotations Assisted by Lexical Chains*
Roberto Navigli

17:30-18:00 *Word Sense Induction: Triplet-Based Clustering and Automatic Evaluation*
Stefan Bordag

Session 7: Demonstrations 1

11:00-12:30

Session 8: Demonstrations 2

14:30-16:00

Session 9: Demonstrations 3

16:30-18:00

Second Day: Thursday, April 6, 2006

9:30-10:30 **Invited Talk 2:** *Objects, Actions, Nouns and Verbs: The View from Cognitve Neuroscience*
by Alfonso Caramazza

10:30-11:00 Coffee Break

Session 10: Alignment

11:00-11:30 *A Comparison of Syntactically Motivated Word Alignment Spaces*
Colin Cherry and Dekang Lin

11:30-12:00 *Improved Lexical Alignment by Combining Multiple Reified Alignments*
Dan Tufiş, Radu Ion, Alexandru Ceauşu and Dan Ştefănescu

12:00-12:30 *Towards Robust Context-Sensitive Sentence Alignment for Monolingual Corpora*
Rani Nelken and Stuart M. Shieber

12:30-14:30 Lunch

Session 11: Student Research Workshop 1

14:30-16:00

16:00-16:30 Coffee Break

Session 12: Dialog I

16:30-17:00 *Addressee Identification in Face-to-Face Meetings*
Natasa Jovanovic, Rieks op den Akker and Anton Nijholt

17:00-17:30 *Identifying Repair Targets in Action Control Dialogue*
Kotaro Funakoshi and Takenobu Tokunaga

17:30-18:00 *Keeping the Initiative: An Empirically-Motivated Approach to Predicting User-Initiated Dialogue Contribution in HCI*
Kerstin Fischer and John A. Bateman

Session 13: Sentiment & Opinion

11:00-11:30 *Determining Term Subjectivity and Term Orientation for Opinion Mining*
Andrea Esuli and Fabrizio Sebastiani

11:30-12:00 *Latent Variable Models for Semantic Orientations of Phrases*
Hiroya Takamura, Takashi Inui and Manabu Okumura

12:00-12:30 *Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses*
Alina Adreevskaia and Sabine Bergler

Second Day: Thursday, April 6, 2006 (*cont.*)

Session 14: Student Research Workshop 2

14:30-16:00

Session 15: Web Corpora

16:30-17:00 *A Figure of Merit for the Evaluation of Web-Corpus Randomness*
Massimiliano Ciaramita and Marco Baroni

17:00-17:30 *Compiling French-Japanese Terminologies from the Web*
Xavier Robitaille, Yasuhiro Sasaki, Masatsugu Tonoike, Satoshi Sato and Takehito Utsuro

17:30-18:00 *Web Text Corpus for Natural Language Processing*
Vinci Liu and James R. Curran

Session 16: Posters 1

11:00-12:30

Session 17: Student Research Workshop 3

14:30-16:00

Session 18: Posters 2

16:30-18:10

Third Day: Friday, April 7, 2006

Session 19: MT Evaluation

9:30-10:00 *CDER: Efficient MT Evaluation Using Block Movements*
Gregor Leusch, Nicola Ueffing and Hermann Ney

10:00-10:30 *Re-evaluation the Role of Bleu in Machine Translation Research*
Chris Callison-Burch, Miles Osborne and Philipp Koehn

10:30-11:00 Coffee Break

Session 20: Lexicon & Tagging

11:00-11:30 *Adaptive Transformation-Based Learning for Improving Dictionary Tagging*
Burcu Karagol-Ayan, David Doermann and Amy Weinberg

11:30-12:00 *From Detecting Errors to Automatically Correcting Them*
Markus Dickinson

12:00-12:30 EACL Business Meeting
12:30-14:30 Lunch

Session 21: Dialog II

15:00-15:30 *Automatic Segmentation of Multiparty Dialogue*
Pei-yun Hsueh, Johanna D. Moore and Steve Renals

15:30-16:00 *Recognizing Textual Parallelisms with Edit Distance and Similarity Degree*
Marie Guégan and Nicolas Hernandez

16:00-16:30 *Using Reinforcement Learning to Build a Better Model of Dialogue State*
Joel R. Tetreault and Diane J. Litman

16:30-17:00 Plenary Closing Session

Session 22: Summarization

9:30-10:00 *Discriminative Sentence Compression with Soft Syntactic Evidence*
Ryan McDonald

10:00-10:30 *Multi-Document Summarization of Evaluative Text*
Giuseppe Carenini, Raymond Ng and Adam Pauls

Third Day: Friday, April 7, 2006 (*cont.*)

Session 23: Generation

11:00-11:30 *Comparing Automatic and Human Evaluation of NLG Systems*
Anja Belz and Ehud Reiter

11:30-12:00 *Structuring Knowledge for Reference Generation: A Clustering Algorithm*
Albert Gatt

Session 24: Semantics

15:00-15:30 *A Clustering Approach for Nearly Unsupervised Recognition of Nonliteral Language*
Julia Birke and Anoop Sarkar

15:30-16:00 *Automatically Constructing a Lexicon of Verb Phrase Idiomatic Combinations*
Afsaneh Fazly and Suzanne Stevenson

16:00-16:30 *Modelling Semantic Role Pausibility in Human Sentence Processing*
Ulrike Padó, Matthew Crocker and Frank Keller

Session 25: Multimodality

9:30-10:00 *Data-Driven Generation of Emphatic Facial Displays*
Mary Ellen Foster and Jon Oberlander

10:00-10:30 *Edit Machines for Robust Multimodal Language Processing*
Srinivas Bangalore and Michael Johnston

Session 26: Parsing & Formalisms

11:00-11:30 *Parsing Arabic Dialects*
David Chiang, Mona Diab, Nizar Habash, Owen Rambow and Safiullah Shareef

11:30-12:00 *Unifying Synchronous Tree Adjoining Grammars and Tree Transducers via Bimorphisms*
Stuart M. Shieber

Session 27: IE & QA

14:30-15:00 *A Machine Learning Approach to Extract Temporal Information from Texts in Swedish and Generate Animated 3D Scenes*
Anders Berglund, Richard Johansson and Pierre Nugues

15:00-15:30 *A Probabilistic Answer Type Model*
Christopher Pinchak and Dekan Lin

15:30-16:00 *Exploiting Shallow Linguistic Information for Relation Extraction from Biomedical Literature*
Claudio Giuliano, Alberto Lavelli and Lorenza Romano

16:00-16:30 *Investigating a Generic Paraphrase-Based Approach for Relation Extraction*
Lorenza Romano, Milen Kouylekov, Idan Szpektor, Ido Dagan and Alberto Lavelli

Table of Contents

Inheritance and the CCG Lexicon
Mark McConville 1

Using Encyclopedic Knowledge for Named Entity Disambiguation
Razvan Bunescu and Marius Pasça 9

Weakly Supervised Approaches for Ontology Population
Hristo Tanev and Bernardo Magnini 17

Computational Complexity of Statistical Machine Translation
Raghavendra Udupa U. and Hemanta K. Maji 25

Computing Consensus Translation from Multiple Machine Translation Systems Using Enhanced Hypotheses Alignment
Evgeny Matusov, Nicola Ueffing and Hermann Ney 33

Phrase-Based Backoff Models for Machine Translation of Highly Inflected Languages
Mei Yang and Katrin Kirchhoff 41

Automatic Detection of Nonreferential It in Spoken Multi-Party Dialog
Christoph Müller 49

Generating Statistical Language Models from Interpretation Grammars in Dialogue Systems
Rebecca Jonson 57

Information Presentation in Spoken Dialogue Systems
Vera Demberg and Johanna D. Moore 65

Constraints on Non-Projective Dependency Parsing
Joakim Nivre 73

Online Learning of Approximate Dependency Parsing Algorithms
Ryan McDonald and Fernando Pereira 81

Statistical Dependency Parsing of Turkish
Gülşen Eryiğit and Kemal Oflazer 89

Generalized Hebbian Algorithm for Incremental Singular Value Decomposition in Natural Language Processing
Genevieve Gorrell 97

Improving Probabilistic Latent Semantic Analysis with Principal Component Analysis
Ayman Farahat and Francine Chen 105

Making Tree Kernels practical for Natural Language Learning
Alessandro Moschitti 113

Determining Word Sense Dominance Using a Thesaurus
Saif Mohammad and Graeme Hirst 121

Experiments on the Validation of Sense Annotations Assisted by Lexical Chains
Roberto Navigli 129

Word Sense Induction: Triplet-Based Clustering and Automatic Evaluation
Stefan Bordag 137

A Comparison of Syntactically Motivated Word Alignment Spaces
Colin Cherry and Dekang Lin 145

Improved Lexical Alignment by Combining Multiple Reified Alignments
Dan Tufiş, Radu Ion, Alexandru Ceauşu and Dan Ştefănescu 153

Towards Robust Context-Sensitive Sentence Alignment for Monolingual Corpora
Rani Nelken and Stuart M. Shieber 161

Addressee Identification in Face-to-Face Meetings
Natasa Jovanovic, Rieks op den Akker and Anton Nijholt 169

Identifying Repair Targets in Action Control Dialogue
Kotaro Funakoshi and Takenobu Tokunaga 177

Keeping the Initiative: An Empirically-Motivated Approach to Predicting User-Initiated Dialogue Contributions in HCI
Kerstin Fischer and John A. Bateman 185

Determining Term Subjectivity and Term Orientation for Opinion Mining
Andrea Esuli and Fabrizio Sebastiani 193

Latent Variable Models for Semantic Orientations of Phrases
Hiroya Takamura, Takashi Inui and Manabu Okumura 201

Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses
Alina Andreevskaia and Sabine Bergler 209

A Figure of Merit for the Evaluation of Web-Corpus Randomness
Massimiliano Ciaramita and Marco Baroni 217

Compiling French-Japanese Terminologies from the Web
Xavier Robitaille, Yasuhiro Sasaki, Masatsugu Tonoike, Satoshi Sato and Takehito Utsuro 225

Web Text Corpus for Natural Language Processing
Vinci Liu and James R. Curran 233

CDER: Efficient MT Evaluation Using Block Movements
Gregor Leusch, Nicola Ueffing and Hermann Ney 241

Re-evaluating the Role of BLEU in Machine Translation Research
Chris Callison-Burch, Miles Osborne and Philipp Koehn 249

Adaptive Transformation-based Learning for Improving Dictionary Tagging
Burcu Karagol-Ayan, David Doermann and Amy Weinberg 257

From Detecting Errors to Automatically Correcting Them
Markus Dickinson 265

Automatic Segmentation of Multiparty Dialogue
Pei-yun Hsueh, Johanna D. Moore and Steve Renals 273

RecognizingTextual Parallelisms with Edit Distance and Similarity Degree
Marie Guégan and Nicolas Hernandez 281

Using Reinforcement Learning to Build a Better Model of Dialogue State
Joel R. Tetreault and Diane J. Litman 289

Discriminative Sentence Compression with Soft Syntactic Evidence
Ryan McDonald 297

Multi-Document Summarization of Evaluative Text
Giuseppe Carenini, Raymond Ng and Adam Pauls 305

Comparing Automatic and Human Evaluation of NLG Systems
Anja Belz and Ehud Reiter 313

Structuring Knowledge for Reference Generation: A Clustering Algorithm
Albert Gatt 321

A Clustering Approach for the Nearly Unsupervised Recognition of Nonliteral Language
Julia Birke and Anoop Sarkar 329

Automatically Constructing a Lexicon of Verb Phrase Idiomatic Combinations
Afsaneh Fazly and Suzanne Stevenson 337

Modelling Semantic Role Plausibility in Human Sentence Processing
Ulrike Padó, Matthew Crocker and Frank Keller 345

Data-driven Generation of Emphatic Facial Displays
Mary Ellen Foster and Jon Oberlander 353

Edit Machines for Robust Multimodal Language Processing
Srinivas Bangalore and Michael Johnston 361

Parsing Arabic Dialects
David Chiang, Mona Diab, Nizar Habash, Owen Rambow and Safiullah Shareef 369

Unifying Synchronous Tree-Adjoining Grammars and Tree Transducers via Bimorphisms
Stuart M. Shieber 377

A Machine Learning Approach to Extract Temporal Information from Texts in Swedish and Generate Animated 3D Scenes
Anders Berglund, Richard Johansson and Pierre Nugues 385

A Probabilistic Answer Type Model
Christopher Pinchak and Dekang Lin 393

Exploiting Shallow Linguistic Information for Relation Extraction from Biomedical Literature
Claudio Giuliano, Alberto Lavelli and Lorenza Romano 401

Investigating a Generic Paraphrase-based Approach for Relation Extraction
Lorenza Romano, Milen Kouylekov, Idan Szpektor, Ido Dagan and Alberto Lavelli 409