

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Alexander Gelbukh (Ed.)

Computational Linguistics and Intelligent Text Processing

12th International Conference, CICLing 2011
Tokyo, Japan, February 20-26, 2011
Proceedings, Part II



Springer

Volume Editor

Alexander Gelbukh
Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Col. Nueva Industrial Vallejo, CP 07738, Mexico D.F., Mexico
E-mail: gelbukh@gelbukh.com

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-19436-8 e-ISBN 978-3-642-19437-5
DOI 10.1007/978-3-642-19437-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011921814

CR Subject Classification (1998): H.3, H.4, F.1, I.2, H.5, H.2.8, I.5

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

CICLing 2011 was the 12th Annual Conference on Intelligent Text Processing and Computational Linguistics. The CICLing conferences provide a wide-scope forum for the discussion of the art and craft of natural language processing research as well as the best practices in its applications.

This set of two books contains four invited papers and a selection of regular papers accepted for presentation at the conference. Since 2001, the proceedings of the CICLing conferences have been published in Springer's *Lecture Notes in Computer Science* series as volume numbers 2004, 2276, 2588, 2945, 3406, 3878, 4394, 4919, 5449, and 6008.

The set has been structured into 13 sections:

- Lexical resources
- Syntax and parsing
- Part-of-speech tagging and morphology
- Word sense disambiguation
- Semantics and discourse
- Opinion mining and sentiment analysis
- Text generation
- Machine translation and multilingualism
- Information extraction and information retrieval
- Text categorization and classification
- Summarization and recognizing textual entailment
- Authoring aid, error correction, and style analysis
- Speech recognition and generation

The 2011 event received a record high number of submissions. A total of 298 papers by 658 authors from 48 countries were submitted for evaluation by the International Program Committee, see Tables 1 and 2. This two-volume set contains revised versions of 74 papers, by 227 authors, selected for presentation; thus the acceptance rate for this set was 25%.

The books feature invited papers by

- Christopher Manning, Stanford University, USA
- Diana McCarthy, Lexical Computing Ltd., UK
- Jun'ichi Tsujii, U. of Tokyo, Japan, and U. of Manchester and NacTeM, UK
- Hans Uszkoreit, Saarland University and DFKI, Germany

who presented excellent keynote lectures at the conference. Publication of extended full-text invited papers in the proceedings is a distinctive feature of the CICLing conferences. Furthermore, in addition to the presentation of their invited papers, the keynote speakers organized separate vivid informal events; this is also a distinctive feature of this conference series.

Table 1. Statistics of submissions and accepted papers by country or region

Country or region	Authors		Papers ¹	Country or region	Authors		Papers ¹
	Subm.	Subm.	Accp.		Subm.	Subm.	Accp.
Australia	17	7	3	Korea (South)	10	4.29	1
Austria	2	1.33	0.33	Macao	4	1	–
Belgium	4	2	1	Malaysia	5	2	–
Brazil	5	2	1	Mexico	13	6.92	2
Canada	11	6.33	2	Myanmar	7	2	–
China	47	17.67	5.67	Nigeria	3	1	–
Colombia	3	2	–	Norway	7	2.17	–
Croatia	3	2	–	Pakistan	6	3.57	–
Cuba	2	0.67	–	Peru	2	0.50	–
Czech Rep.	14	8.50	3	Poland	2	2	–
Egypt	9	2.67	1.67	Portugal	25	9.67	2
Finland	3	2	–	Romania	7	3.33	–
France	36	16.68	4.83	Russia	8	2.33	–
Georgia	1	1	–	Saudi Arabia	1	1	–
Germany	29	12.58	3.50	Singapore	7	2.50	1
Greece	6	2	1	Spain	39	14.30	4.30
Hong Kong	5	2	1	Sweden	5	1.39	1
India	85	41.75	6.42	Taiwan	13	5	–
Iran	23	18	3	Thailand	6	3	1
Ireland	14	7	1	Tunisia	9	3.15	–
Israel	3	1.75	–	Turkey	8	4.17	1
Italy	17	6.25	2.25	UK	13	6.67	0.50
Japan	71	29.67	14	USA	39	17.87	4.53
Jordan	1	0.50	–	Viet Nam	8	4.33	1
				<i>Total:</i>	658	298	74

¹ By the number of authors: e.g., a paper by two authors from the USA and one from the UK is counted as 0.67 for the USA and 0.33 for the UK.

With this event we introduced a new policy of giving preference to papers with verifiable and reproducible results: we encouraged the authors to provide, in electronic form, a proof of their claims or a working description of the suggested algorithm, in addition to the verbal description given in the paper. If the paper claimed experimental results, we encouraged the authors to make available to the community all the input data necessary to verify and reproduce these results; if it claimed to advance human knowledge by introducing an algorithm, we encouraged the authors to make the algorithm itself, in some programming language, available to the public. This additional electronic material will be permanently stored on CICLing’s server, www.CICLing.org, and will be available to the readers of the corresponding paper for download under a license that permits its free use for research purposes.

In the long run we expect that computational linguistics will have verifiability and clarity standards similar to those of mathematics: in mathematics, each claim is accompanied by a complete and verifiable proof (usually much

Table 2. Statistics of submissions and accepted papers by topic²

Accepted	Submitted	% accepted	Topic
13	40	33	Lexical resources
13	47	28	Practical applications
11	39	28	Clustering and categorization
11	44	25	Other
10	28	36	Acquisition of lexical resources
10	29	34	Statistical methods (mathematics)
10	52	19	Machine translation & multilingualism
9	25	36	Syntax and chunking (linguistics)
9	31	29	Semantics and discourse
9	58	16	Information extraction
7	46	15	Text mining
6	12	50	Symbolic and linguistic methods
6	50	12	Information retrieval
5	13	38	Parsing algorithms (mathematics)
5	16	31	Noisy text processing and cleaning
5	18	28	Summarization
4	11	36	Text generation
4	16	25	Opinion mining
4	17	24	POS tagging
3	7	43	Speech processing
3	8	38	Cross-language information retrieval
3	15	20	Word sense disambiguation
3	20	15	Formalisms and knowledge representation
2	6	33	Emotions and humor
2	13	15	Named entity recognition
1	5	20	Spelling and grammar checking
1	7	14	Anaphora resolution
1	7	14	Textual entailment
1	8	12	Question answering
1	11	9	Natural language interfaces
1	12	8	Morphology
–	6	0	Computational terminology

² As indicated by the authors. A paper may belong to several topics.

greater in size than the claim itself); each theorem—and not just its description or general idea—is completely and precisely presented to the reader. Electronic media allow computational linguists to provide material analogous to the proofs and formulas in mathematics in full length—which can amount to megabytes or gigabytes of data—separately from a 12-page description published in a book. A more detailed argumentation for this new policy can be found on www.CICLing.org/why_verify.htm.

To encourage the authors to provide algorithms and data along with the published papers, we introduced a new Verifiability, Reproducibility, and Working Description Award. The main factors in choosing the awarded submission were

technical correctness and completeness, readability of the code and documentation, simplicity of installation and use, and exact correspondence to the claims of the paper. Unnecessary sophistication of the user interface was discouraged; novelty and usefulness of the results were not evaluated—those parameters were evaluated for the paper itself and not for the data.

The following papers received the Best Paper Awards, the Best Student Paper Award, as well as the Verifiability, Reproducibility, and Working Description Award, correspondingly (the best student paper was selected from the papers of which the first author was a full-time student, excluding the papers that received a Best Paper Award):

- 1st Place: *Co-related Verb Argument Selectional Preferences*, by Hiram Calvo, Kentaro Inui, and Yuji Matsumoto;
- 2nd Place: *Self-Adjusting Bootstrapping*, by Shoji Fujiwara and Satoshi Sekine;
- 3rd Place: *Effective Use of Dependency Structure for Bilingual Lexicon Creation*, by Daniel Andrade, Takuya Matsuzaki, and Jun'ichi Tsujii;
- Student: *Incorporating Coreference Resolution into Word Sense Disambiguation*, by Shangfeng Hu and Chengfei Liu;
- Verifiability: *Improving Text Segmentation with Non-systematic Semantic Relation*, by Viet Cuong Nguyen, Le Minh Nguyen, and Akira Shimazu.

The authors of the awarded papers (except for the Verifiability Award) were given extra time for their presentations. In addition, the Best Presentation Award and the Best Poster Award winners were selected by a ballot among the attendees of the conference.

Besides its high scientific level, one of the success factors of CICLing conferences is their excellent cultural program. The attendees of the conference had a chance to visit Kamakura—known for the Kamakura period of ancient history of Japan—where they experienced historical Japanese cultural heritage explained by highly-educated local volunteers and saw Shinto (traditional religion of Japan) shrines and old Buddhist temples characteristic of Japan. They recalled recent history at the Daigo Fukuryu Maru Exhibition Hall, which tells the story of a Japanese boat exposed to and contaminated by nuclear fallout from a thermonuclear device test in 1954. Finally, the participants familiarized themselves with modern Japanese technology during guided tours to Toshiba Science Museum and Sony Square; the latter can only be accessed by special invitation from Sony. And of course they enjoyed Tokyo, the largest city in the world, futuristic and traditional at the same time, during an excursion to the Japanese-style East Gardens of the Imperial Palace and a guided tour of the city, by bus and boat (see photos on www.CICLing.org).

I would like to thank all those involved in the organization of this conference. In the first place these are the authors of the papers that constitute this book: it is the excellence of their research work that gives value to the book and sense to the work of all other people. I thank all those who served on the Program Committee, Software Reviewing Committee, Award Selection Committee, as

well as the additional reviewers, for their hard and very professional work. Special thanks go to Ted Pedersen, Grigori Sidorov, Yasunari Harada, Manuel Vilares Ferro, and Adam Kilgariff, for their invaluable support in the reviewing process.

I thank the School of Law and Media Network Center of Waseda University, Japan, for hosting the conference; the Institute for Digital Enhancement of Cognitive Development (DECODE) of Waseda University for valuable collaboration in its organization; and Waseda University for providing us with the best conference facilities. With deep gratitude I acknowledge the support of Professor Waichiro Iwashi, the dean of the School of Law of Waseda University, and Professor Toshiyasu Matsushima, Dean and Director of Media Network Center of Waseda University. I express my most cordial thanks to the members of the local Organizing Committee for their enthusiastic and hard work. The conference would not have been a success without the kind help of Professor Mieko Ebara, Ms. Mayumi Kawamura, Dr. Kyoko Kanzaki, and all the other people involved in the organization of the conference and cultural program activities.

My most special thanks go to Professor Yasunari Harada, Director of DECODE, for his great enthusiasm and infinite kindness and patience; countless nights without sleep, after a whole day of teaching and meetings, spent on the organization of the conference, from the strategic planning to the finest details. I feel very lucky to have had the opportunity to collaborate with this prominent scientist, talented organizer, and caring friend. From him I learnt a lot about human relationships as well as about planning and organization.

With great gratitude I acknowledge the financial support of the Kayamori Foundation of Information Science Advancement, which greatly helped us to keep the fees low. I would like to express my gratitude to the Kamakura Welcome Guide Association for making our visit to this historical city of Japan a memorable and enjoyable one. Thanks are also due to Sony and Totsusangyo Corporation, Toshiba Science Museum, and Daigo Fukuryu Maru Exhibition Hall, for arranging special visits and guided tours for CICLing 2011 participants. I would like to specifically recognize the help of Mr. Masahiko Fukakushi, Executive Officer and Corporate Senior Vice President of Toshiba Corporation, in arranging our visit to Toshiba Science Museum and the help of Dr. Atsushi Ito, Distinguished Research Engineer at KDDI R&D Laboratories, in providing wireless Internet access to the attendees of the conference.

The entire submission and reviewing process was supported for free by the EasyChair system (www.EasyChair.org). Last but not least, I deeply appreciate the Springer staff's patience and help in editing this volume and getting it printed in record short time—it is always a great pleasure to work with Springer.

February 2011

Alexander Gelbukh
General Chair

Organization

CICLing 2011 was co-hosted by the School of Law and Media Network Center of Waseda University, and was organized by the CICLing 2011 Organizing Committee, in conjunction with the Natural Language and Text Processing Laboratory of the CIC (Center for Computing Research) of the IPN (National Polytechnic Institute), Mexico, and the Institute for Digital Enhancement of Cognitive Development of Waseda University, with partial financial support from the Kayamori Foundation of Information Science Advancement.

Organizing Chair

Yasunari Harada

Organizing Committee

Glen Stockwell
Ryo Otoguro
Joji Maeno
Noriaki Kusumoto
Akira Morita

Kyoko Kanzaki
Kanako Maebo
Mieko Ebara
Mayumi Kawamura
Alexander Gelbukh

Program Chair

Alexander Gelbukh

Program Committee

Sophia Ananiadou
Bogdan Babych
Sivaji Bandyopadhyay
Roberto Basili
Pushpak Bhattacharyya
Nicoletta Calzolari
Sandra Carberry
Kenneth Church
Dan Cristea
Silviu Cucerzan
Mona T. Diab
Alex Chengyu Fang

Anna Feldman
Daniel Flickinger
Alexander Gelbukh
Roxana Girju
Gregory Grefenstette
Iryna Gurevych
Nizar Habash
Sanda Harabagiu
Yasunari Harada
Ales Horak
Eduard Hovy
Nancy Ide

XII Organization

Diana Inkpen
Sylvain Kahane
Alma Kharrat
Adam Kilgarriff
Richard Kittredge
Sandra Kübler
Krister Lindén
Aurelio Lopez
Bernardo Magnini
Cerstin Mahlow
Christopher Manning
Yuji Matsumoto
Diana McCarthy
Rada Mihalcea
Ruslan Mitkov
Dunja Mladenic
Marie-Francine Moens
Dan Moldovan
Masaki Murata
Smaranda Muresan
Roberto Navigli
Kjetil Nørvåg
Kemal Oflazer
Constantin Orasan
Maria Teresa Pazienza
Ted Pedersen

Viktor Pekar
Anselmo Peñas
Irina Prodanof
James Pustejovsky
Fuji Ren
German Rigau
Fabio Rinaldi
Vasile Rus
Horacio Saggion
Franco Salvetti
Hassan Sawaf
Satoshi Sekine
Serge Sharoff
Thamar Solorio
Benno Stein
Vera Lúcia Strube De Lima
Jun Suzuki
Christoph Tillmann
George Tsatsaronis
Jun'ichi Tsujii
Karin Verspoor
Manuel Vilares Ferro
Hai Feng Wang
Leo Wanner
Annie Zaenen

Software Reviewing Committee

Ted Pedersen
Florian Holz
Miloš Jakubíček

Sergio Jiménez Vargas
Ronald Winnemöller

Award Committee

Alexander Gelbukh
Eduard Hovy
Rada Mihalcea

Ted Pedersen
Yorick Wiks

Additional Referees

Naveed Afzal
 Rodrigo Agerri
 Alexandre Agustini
 Laura Alonso Alemany
 Rania Al-Sabbagh
 Maik Anderka
 Paolo Annesi
 Eiji Aramaki
 Jordi Atserias
 Wilker Aziz
 João B. Rocha-Junior
 Nguyen Bach
 Vít Baisa
 Jared Bernstein
 Pinaki Bhaskar
 Arianna Bisazza
 Eduardo Blanco
 Bernd Bohnet
 Nadjet Bouayad-Agha
 Elena Cabrio
 Xavier Carreras
 Miranda Chong
 Danilo Croce
 Amitava Das
 Dipankar Das
 Jan De Belder
 Diego Decao
 Iustin Dornescu
 Kevin Duh
 Oana Frunza
 Caroline Gasperin
 Jesús Giménez
 Marco Gonzalez
 Amir H. Razavi
 Masato Hagiwara
 Laritza Hernández
 Ryuichiro Higashinaka
 Dennis Hoppe
 Adrian Iftene
 Iustina Ilisei
 Chris Irwin Davis
 Yasutomo Kimura
 Levi King
 Olga Kolesnikova

Natalia Konstantinova
 Sudip Kumar Naskar
 Alberto Lavelli
 Yulia Ledeneva
 Chen Li
 Nedim Lipka
 Lucelene Lopes
 Christian M. Meyer
 Kanako Maebo
 Sameer Maskey
 Yashar Mehdad
 Simon Mille
 Michael Mohler
 Manuel Montes y Gómez
 Rutu Mulkar-Mehta
 Koji Murakami
 Vasek Nemcik
 Robert Neumayer
 Zuzana Neverilova
 Ryo Otoguro
 Partha Pakray
 Santanu Pal
 Michael Piotrowski
 Natalia Ponomareva
 Martin Potthast
 Heri Ramampiaro
 Luz Rello
 Stefan Rigo
 Alvaro Rodrigo
 Alex Rudnick
 Ruhi Sarikaya
 Gerold Schneider
 Leanne Seaward
 Ravi Sinha
 Amber Smith
 Katsuhito Sudoh
 Xiao Sun
 Irina Temnikova
 Wren Thornton
 Masato Tokuhisa
 Sara Tonelli
 Diana Trandabat
 Stephen Tratz
 Yasushi Tsubota

Jordi Turmo
Kateryna Tymoshenko
Sriram Venkatapathy
Renata Vieira
Nina Wacholder
Dina Wonsever

Clarissa Xavier
Jian-ming Xu
Caixia Yuan
Rong Zhang
Bing Zhao

Website and Contact

The website of the CICLE conference series is www.CICLE.org. It contains information about the past CICLE conferences and their satellite events, including published papers or their abstracts, photos, video recordings of keynote talks, as well as the information about the forthcoming CICLE conferences and the contact options.

Table of Contents – Part II

Machine Translation and Multilingualism

Ontology Based Interlingua Translation	1
<i>Leonardo Lesmo, Alessandro Mazzei, and Daniele P. Radicioni</i>	
Phrasal Equivalence Classes for Generalized Corpus-Based Machine Translation	13
<i>Rashmi Gangadharaiyah, Ralf D. Brown, and Jaime Carbonell</i>	
A Multi-view Approach for Term Translation Spotting	29
<i>Raphaël Rubino and Georges Linarès</i>	
ICE-TEA: In-Context Expansion and Translation of English Abbreviations	41
<i>Waleed Ammar, Kareem Darwish, Ali El Kahki, and Khaled Hafez</i>	
Word Segmentation for Dialect Translation	55
<i>Michael Paul, Andrew Finch, and Eiichiro Sumita</i>	
TEP: Tehran English-Persian Parallel Corpus	68
<i>Mohammad Taher Pilevar, Heshaam Faili, and Abdol Hamid Pilevar</i>	
Effective Use of Dependency Structure for Bilingual Lexicon Creation (Best Paper Award, Third Place)	80
<i>Daniel Andrade, Takuya Matsuzaki, and Jun'ichi Tsujii</i>	
Online Learning via Dynamic Reranking for Computer Assisted Translation	93
<i>Pascual Martínez-Gómez, Germán Sanchis-Trilles, and Francisco Casacuberta</i>	

Information Extraction and Information Retrieval

Learning Relation Extraction Grammars with Minimal Human Intervention: Strategy, Results, Insights and Plans (Invited Paper)	106
<i>Hans Uszkoreit</i>	
Using Graph Based Method to Improve Bootstrapping Relation Extraction	127
<i>Haibo Li, Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka</i>	
A Hybrid Approach for the Extraction of Semantic Relations from MEDLINE Abstracts	139
<i>Asma Ben Abacha and Pierre Zweigenbaum</i>	

An Active Learning Process for Extraction and Standardisation of Medical Measurements by a Trainable FSA	151
<i>Jon Patrick and Mojtaba Sabbagh</i>	
Topic Chains for Understanding a News Corpus	163
<i>Dongwoo Kim and Alice Oh</i>	
From Italian Text to TimeML Document via Dependency Parsing	177
<i>Livio Robaldo, Tommaso Caselli, Irene Russo, and Matteo Grella</i>	
Self-adjusting Bootstrapping (Best Paper Award, Second Place)	188
<i>Shoji Fujiwara and Satoshi Sekine</i>	
Story Link Detection Based on Event Words	202
<i>Letian Wang and Fang Li</i>	
Ranking Multilingual Documents Using Minimal Language Dependent Resources	212
<i>G.S.K. Santosh, N. Kiran Kumar, and Vasudeva Varma</i>	
Measuring Chinese-English Cross-Lingual Word Similarity with HowNet and Parallel Corpus	221
<i>Yunqing Xia, Taotao Zhao, Jianmin Yao, and Peng Jin</i>	

Text Categorization and Classification

Comparing Manual Text Patterns and Machine Learning for Classification of E-Mails for Automatic Answering by a Government Agency	234
<i>Hercules Dalianis, Jonas Sjöbergh, and Eriks Sneiders</i>	
Using Thesaurus to Improve Multiclass Text Classification	244
<i>Nooshin Maghsoodi and Mohammad Mehdi Homayounpour</i>	
Adaptable Term Weighting Framework for Text Classification	254
<i>Dat Huynh, Dat Tran, Wanli Ma, and Dharmendra Sharma</i>	
Automatic Specialized vs. Non-specialized Sentence Differentiation	266
<i>Iria da Cunha, Maria Teresa Cabré, Eric SanJuan, Gerardo Sierra, Juan Manuel Torres-Moreno, and Jorge Vivaldi</i>	
Wikipedia Vandalism Detection: Combining Natural Language, Metadata, and Reputation Features	277
<i>B. Thomas Adler, Luca de Alfaro, Santiago M. Mola-Velasco, Paolo Rosso, and Andrew G. West</i>	
Costco: Robust Content and Structure Constrained Clustering of Networked Documents	289
<i>Su Yan, Dongwon Lee, and Alex Hai Wang</i>	

Summarization and Recognizing Textual Entailment

Learning Predicate Insertion Rules for Document Abstracting	301
<i>Horacio Saggion</i>	
Multi-topical Discussion Summarization Using Structured Lexical Chains and Cue Words	313
<i>Jun Hatori, Akiko Murakami, and Jun'ichi Tsujii</i>	
Multi-document Summarization Using Link Analysis Based on Rhetorical Relations between Sentences	328
<i>Nik Adilah Hanin Binti Zahri and Fumiyo Fukumoto</i>	
Co-clustering Sentences and Terms for Multi-document Summarization	339
<i>Yunqing Xia, Yonggang Zhang, and Jianmin Yao</i>	
Answer Validation Using Textual Entailment	353
<i>Partha Pakray, Alexander Gelbukh, and Sivaji Bandyopadhyay</i>	

Authoring Aid, Error Correction, and Style Analysis

SPIDER: A System for Paraphrasing in Document Editing and Revision — Applicability in Machine Translation Pre-editing	365
<i>Anabela Barreiro</i>	
Providing Cross-Lingual Editing Assistance to Wikipedia Editors	377
<i>Ching-man Au Yeung, Kevin Duh, and Masaaki Nagata</i>	
Reducing Overdetections in a French Symbolic Grammar Checker by Classification	390
<i>Fabrizio Gotti, Philippe Langlais, Guy Lapalme, Simon Charest, and Éric Brunelle</i>	
Performance Evaluation of a Novel Technique for Word Order Errors Correction Applied to Non Native English Speakers' Corpus	402
<i>Theologos Athanaselis, Stelios Bakamidis, and Ioannis Dologlou</i>	
Correcting Verb Selection Errors for ESL with the Perceptron	411
<i>Xiaohua Liu, Bo Han, and Ming Zhou</i>	
A Posteriori Agreement as a Quality Measure for Readability Prediction Systems	424
<i>Philip van Oosten, Véronique Hoste, and Dries Tanghe</i>	
A Method to Measure the Reading Difficulty of Japanese Words	436
<i>Keiji Yasuda, Andrew Finch, and Eiichiro Sumita</i>	

Informality Judgment at Sentence Level and Experiments with Formality Score	446
<i>Shibamouli Lahiri, Prasenjit Mitra, and Xiaofei Lu</i>	

Speech Recognition and Generation

Combining Word and Phonetic-Code Representations for Spoken Document Retrieval	458
<i>Alejandro Reyes-Barragán, Manuel Montes-y-Gómez, and Luis Villaseñor-Pineda</i>	
Automatic Rule Extraction for Modeling Pronunciation Variation	467
<i>Zeeshan Ahmed and Julie Carson-Berndsen</i>	
Predicting Word Pronunciation in Japanese	477
<i>Jun Hatori and Hisami Suzuki</i>	
A Minimum Cluster-Based Trigram Statistical Model for Thai Syllabification	493
<i>Chonlasith Jucksriporn and Ohm Sornil</i>	
Automatic Generation of a Pronunciation Dictionary with Rich Variation Coverage Using SMT Methods	506
<i>Panagiota Karanasou and Lori Lamel</i>	
Author Index	519

Table of Contents – Part I

Lexical Resources

Influence of Treebank Design on Representation of Multiword Expressions	1
<i>Eduard Bejček, Pavel Straňák, and Daniel Zeman</i>	
Combining Contextual and Structural Information for Supersense Tagging of Chinese Unknown Words	15
<i>Likun Qiu, Yunfang Wu, and Yanqiu Shao</i>	
Identification of Conjunct Verbs in Hindi and Its Effect on Parsing Accuracy	29
<i>Rafiya Begum, Karan Jindal, Ashish Jain, Samar Husain, and Dipti Misra Sharma</i>	
Identification of Reduplicated Multiword Expressions Using CRF	41
<i>Kishorjit Nongmeikapam, Dhiraj Laishram, Naorem Bikramjit Singh, Ngariyanbam Mayekleima Chanu, and Sivaji Bandyopadhyay</i>	

Syntax and Parsing

Computational Linguistics and Natural Language Processing (Invited Paper)	52
<i>Jun'ichi Tsujii</i>	
An Unsupervised Approach for Linking Automatically Extracted and Manually Crafted LTAGs	68
<i>Heshaam Faili and Ali Basirat</i>	
Tamil Dependency Parsing: Results Using Rule Based and Corpus Based Approaches	82
<i>Loganathan Ramasamy and Zdeněk Žabokrtský</i>	
Incremental Combinatory Categorical Grammar and Its Derivations	96
<i>Ahmed Hefny, Hany Hassan, and Mohamed Bahgat</i>	
Dependency Syntax Analysis Using Grammar Induction and a Lexical Categories Precedence System	109
<i>Hiram Calvo, Omar J. Gambino, Alexander Gelbukh, and Kentaro Inui</i>	
Labelwise Margin Maximization for Sequence Labeling	121
<i>Wenjun Gao, Xipeng Qiu, and Xuanjing Huang</i>	

Co-related Verb Argument Selectional Preferences (Best Paper Award, First Place)	133
<i>Hiram Calvo, Kentaro Inui, and Yuji Matsumoto</i>	

Combining Diverse Word-Alignment Symmetrizations Improves Dependency Tree Projection	144
<i>David Mareček</i>	

An Analysis of Tree Topological Features in Classifier-Based Unlexicalized Parsing	155
<i>Samuel W.K. Chan, Mickey W.C. Chong, and Lawrence Y.L. Cheung</i>	

Part of Speech Tagging and Morphology

Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics? (Invited Paper)	171
<i>Christopher D. Manning</i>	

Ripple Down Rules for Part-of-Speech Tagging	190
<i>Dat Quoc Nguyen, Dai Quoc Nguyen, Son Bao Pham, and Dang Duc Pham</i>	

An Efficient Part-of-Speech Tagger for Arabic	202
<i>Selçuk Köprü</i>	

An Evaluation of Part of Speech Tagging on Written Second Language Spanish	214
<i>M. Pilar Valverde Ibañez</i>	

Onoma: A Linguistically Motivated Conjugation System for Spanish Verbs	227
<i>Luz Rello and Eduardo Basterrechea</i>	

Word Sense Disambiguation

Measuring Similarity of Word Meaning in Context with Lexical Substitutes and Translations (Invited Paper)	238
<i>Diana McCarthy</i>	

A Quantitative Evaluation of Global Word Sense Induction	253
<i>Marianna Apidianaki and Tim Van de Cruys</i>	

Incorporating Coreference Resolution into Word Sense Disambiguation (Best Student Paper Award)	265
<i>Shangfeng Hu and Chengfei Liu</i>	

Semantics and Discourse

Deep Semantics for Dependency Structures	277
<i>Paul Bédaride and Claire Gardent</i>	
Combining Heterogeneous Knowledge Resources for Improved Distributional Semantic Models	289
<i>György Szarvas, Torsten Zesch, and Iryna Gurevych</i>	
Improving Text Segmentation with Non-systematic Semantic Relation (Verifiability Award)	304
<i>Viet Cuong Nguyen, Le Minh Nguyen, and Akira Shimazu</i>	
Automatic Identification of Cause-Effect Relations in Tamil Using CRFs	316
<i>S. Menaka, Pattabhi R.K. Rao, and Sobha Lalitha Devi</i>	
Comparing Approaches to Tag Discourse Relations	328
<i>Shamima Mithun and Leila Kosseim</i>	
Semi-supervised Discourse Relation Classification with Structural Learning	340
<i>Hugo Hernault, Danushka Bollegala, and Mitsuru Ishizuka</i>	
Integrating Japanese Particles Function and Information Structure	353
<i>Akira Ohtani</i>	
Assessing Lexical Alignment in Spontaneous Direction Dialogue Data by Means of a Lexicon Network Model	368
<i>Alexander Mehler, Andy Lücking, and Peter Menke</i>	

Opinion Mining and Sentiment Detection

Towards Well-Gounded Phrase-Level Polarity Analysis	380
<i>Robert Remus and Christian Hähnig</i>	
Implicit Feature Identification via Co-occurrence Association Rule Mining	393
<i>Zhen Hai, Kuiyu Chang, and Jung-jae Kim</i>	
Construction of Wakamono Kotoba Emotion Dictionary and Its Application	405
<i>Kazuyuki Matsumoto and Fuji Ren</i>	
Temporal Analysis of Sentiment Events – A Visual Realization and Tracking	417
<i>Dipankar Das, Anup Kumar Kolya, Asif Ekbal, and Sivaji Bandyopadhyay</i>	

Text Generation

Highly-Inflected Language Generation Using Factored Language Models	429
<i>Eder Miranda de Novais, Ivandré Paraboni, and Diogo Takaki Ferreira</i>	
Prenominal Modifier Ordering in Bengali Text Generation	439
<i>Sumit Das, Anupam Basu, and Sudeshna Sarkar</i>	
Bootstrapping Multiple-Choice Tests with THE-MENTOR	451
<i>Ana Cristina Mendes, Sérgio Curto, and Luísa Coheur</i>	
Author Index	463