

COLING 2012

**24th International Conference on
Computational Linguistics**

**Proceedings of COLING 2012:
Technical Papers**

**Program chairs:
Martin Kay and Christian Boitet**

**8-15 December 2012
Mumbai, India**

Diamond sponsors

Tata Consultancy Services
Linguistic Data Consortium for Indian Languages (LDC-IL)

Gold Sponsors

Microsoft Research
Beijing Baidu Netcon Science Technology Co. Ltd.

Silver sponsors

IBM, India Private Limited
Crimson Interactive Pvt. Ltd.
Yahoo
Easy Transcription & Software Pvt. Ltd.

Proceedings of COLING 2012: Technical Papers
Martin Kay and Christian Boitet (eds.)
Revised preprint edition, 2012

Published by The COLING 2012 Organizing Committee
Indian Institute of Technology Bombay,
Powai,
Mumbai-400076
India
Phone: 91-22-25764729
Fax: 91-22-2572 0022
Email: pb@cse.iitb.ac.in

This volume © 2012 The COLING 2012 Organizing Committee.
Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike*
3.0 Nonported license.

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

Some rights reserved.

Contributed content copyright the contributing authors.
Used with permission.

Also available online in the ACL Anthology at <http://aclweb.org>

Introduction from the Program Chairs

The members of the International Committee on Computational Linguistics were immensely privileged to be able to accept the invitation to hold our 24th COLING Conference here in India, a country which can justly be thought of as the center of the linguistic world. This is where Panini wrote the first formal grammar in the 6th century BC and where a linguistic diversity flourishes today that is nothing short of astounding to the rest of us. This conference has received twice as many submissions as any of its predecessors and, in many ways, is twice as rich because of the high proportion of contributions by teachers, researchers and, above all, students. Many are from India and other countries, such as Iran, with long and diverse linguistic traditions. There are challenges here for linguists of all varieties, most especially for those who put their faith in n-grams and machine learning.

The 195 full-length technical papers in 5 parallel tracks, 138 posters, and 66 demonstrations that will be presented still constitute no more than a quarter of the total number of submissions. The chairs of some of the 26 program subcommittees were overwhelmed with both their number and their quality. The International Committee is always greatly indebted to the area chairs and reviewers for the invaluable work that they do. Never so much as on this occasion.

Our greatest debt is clearly to our colleagues here in Mumbai, as will become clear to all as the week proceeds. They were even less well equipped than we on the permanent committee to predict what they were getting into, but they have risen to the occasion in every way and you will find them to be immensely warm, helpful, and resourceful hosts.

COLING's founding fathers wanted these conferences to be more than learned presentations. They wanted them to be opportunities to meet, and talk and delight in the company of other who share our fascination with language and the processes that make it work. Some call this the COLING spirit. There is nowhere that could nurture this spirit more effectively than here in India.

*Martin Kay
Christian Boitet
(Program chairs)
December 2012, Mumbai*

Introduction from the Organizing Chairs

It is a matter of great pride that the 24th International Conference on Computational Linguistics (COLING 2012) is taking place in India, the land of multilinguality and multicultural. The organization of an event of COLING's scale takes enormous energy, planning and time. Two years back, in Beijing, when COLING was awarded to India, we knew that the task will be demanding, and happily for us, the NLP team at IIT Bombay, the organizing institute, has risen to the occasion.

At the time of going to press, the total number of registrants in COLING has exceeded 700. With delegates coming from 60 countries, COLING 2012 will witness a colourful diversity of language and culture, and fittingly so. Conforming to current practices of international conferences, there are two days of workshops and tutorials before the main conference and one day of workshop immediately after. 15 focussed and topical workshops will be attended by about 300 delegates, as will be 6 high quality tutorials of contemporary interest.

Social events include a reception in the evening of 10th Dec, a banquet in a nearby 7 star hotel on 11th Dec, an excursion to the famed Bhaja caves on 12th Dec and a cultural evening of Indian classical music on 13th Dec. Bhaja Caves, built in the period 3rd century BC to 2nd century CE, is a set of Buddhist monastery-caves near the hill station of Lonavala, nestled in the Sahayadri mountain ranges, about 90 km to the south-east of Mumbai. There will be cultural evening on the fourth day of the conference, featuring a solo performance on "tabla", the representative of Indian percussion instruments, and another solo on Sitar, an instrument that drew world's attention Indian classical music tradition.

Indian Institute of Technology Bombay is fittingly the host of COLING 2012. IITs have, over the years, emerged as the premier institutes of technology in India. The Computer Science and Engineering Department at IIT Bombay is one of the largest and oldest Departments of CSE in the country. Each and every member of the 40 strong NLP group at IIT Bombay is toiling hard to make COLING 2012 a resounding success.

The Government and industries have been our generous sponsors. All their names and logos are to be found in printed and USB proceedings. We thank them wholeheartedly.

Technology Development in Indian Languages (TDIL) project of Department of IT, Ministry of Communication and Information Technology, has been the harbinger of growth of NLP in India. COLING happening in India is a result of this long history of active patronage.

Logistics wise, the “large events” – inauguration, invited speeches, reception and the cultural program – are in the convocation hall of IIT Bombay. Oral presentations are all in the newly constructed Victor Menezes Convention Center (VMCC) about 200 mtrs from the convocation hall. Poster presentations are in the convocation hall, except on the first day, when it is VMCC.

A very competent team of volunteers will be available for any assistance. We hope COLING participants will have a memorable time in India.

Pushpak Bhattacharyya
Rajeev Sangal
(Organizing chairs)
December 2012, Mumbai

Program Committee:

Program Chair: Martin Kay (Stanford University)
Program Co-chair: Christian Boitet (University of Grenoble)
Workshop Chair: Prof. Laurent Besacier
Tutorial Chair: Prof. Sadao Kuroshashi

Area Chairs:

Indian language technology: Dipti Sharma (IIIT Hyderabad)
Underresourced languages: Vincent Berment (C&S, Paris & LIG, GETALP, Grenoble)
Morphology & POS Tagging: Gábor Prószték (MorphoLogic & Pázmány Péter Catholic University, Budapest)
Grammar and formalisms: Hans Uszkoreit (DFKI, Saarbrücken)
Parsing: Mark Johnson (Macquarie Univ., Sydney)
Semantics: Igor Boguslavsky (RAS, IPPI/IITP, Moscow & UPM, Madrid)
Discourse and pragmatics: Eva Hajičová (Charles Univ., Prague)
Coreference analysis resolution: Alexander Gelbukh (Instituto Politécnico Nacional, Mexico-city)
Ontologies and terminology: Christophe Roche (Univ. de Savoie, Chambéry)
Textual Entailment: Lauri Karttunen (Stanford University)
Resources and annotation: Nicoletta Calzolari (CNR – ILC, Pisa)
Psychological and neurological modelling: Véronique Aubergé (CNRS – LIG lab Grenoble) & Rohit Manchanda (IITB, Mumbai)
Empirical Machine Translation: Philipp Koehn (University of Edinburgh)
Expert or Hybrid Machine Translation: Mandel Shi (Xiamen University)
Hybrid man+machine architectures & human factors: Hervé Blanchon (Université de Grenoble (LIG, GETALP))
Information Retrieval: Jian-Yun Nie (Univ. de Montréal (RALI))
Summarization: Horacio Saggion (Universitat Pompeu Fabra) & Sivaji Bandyopadhyaya (Jadavpur University)
Named Entity recognition: Sergei Nirenburg (Univ. of Maryland, Baltimore)
Word Sense Disambiguation: Mathieu Lafourcade (Univ. de Montpellier II (LIRMM))
Sentiment and text classification: Yorick Wilks (Univ. of Sheffield)
Information & content extraction, text mining: Junichi Tsujii (MSRA, Beijing)
Question Answering: Constantin Orasan (University of Wolverhampton)
Speech recognition and synthesis: Roland Kuhn (National Research Council of Canada, Ottawa)
Software internationalization & localization: Andy Way (Univ. of Dublin and Applied Language)
Deployment of NLP-based applications, software integration & quality: Rajeev Sangal (IIIT Hyderabad) & Christian Boitet (Université de Grenoble (UJF, LIG, GETALP))
Natural Language Generation: Donia Scott (Univ. of Sussex)

Invited Speakers:

Prof. Paul Kiparsky (Stanford University)
Prof. Makoto Nagao (Kyoto University)
Prof. Dipti Misra (Sharma IIIT-Hyderabad)
Prof. Barbara Moser-Mercer (University of Geneva)

Organizing Committee:

Organizing Chair: Pushpak Bhattacharria

Organizing Co-Chair: Rajeev Sangal

Vasant Zende

Rupash Modak

Ganesh Ramakrishnan

Balamurali A R

N. Vasudevan

Rahul Sharnagat

Raj dabre

Jaya Sarswati

Rajita Shukla

Gajanan Rane

Samir Soni

Ritesh Shah (CDAC Mumbai)

Deepak Jagtap

Anup Kunchukuttan

Abhijeet Mishra

Brijesh Bhatt

Swapnil Choudhary

Laxmi Kashyap

Kahsyap Popat

Manish Shrivastava

Publication committee:

Publication chair: Roger Evans (University of Brighton)

Ms. Sudha (IIT Bombay)

Mrs. Rajita Shukla (IIT Bombay)

Mr. Rahul Sharnagat (IIT Bombay)

Mr. Kashyap Popat (IIT Bombay)

List of the 723 reviewers for COLING-2012, Mumbai

Eneko Agirre, University of the Basque Country, Spain.
Lupe Aguado, Universidad Politécnica de Madrid, Spain.
Khurshid Ahmad, Trinity College, University of Dublin, Ireland.
Rania Al-Sabbagh, University of Illinois at Urbana-Champaign, United States.
Vicente Alabau, Universitat Politècnica de València, Spain.
Inaki Alegria, University of the Basque Country, Spain.
Laura Alonso Alemany, Universidad Nacional de Córdoba, Argentina.
Le An Ha, Univ. Wolverhampton, United Kingdom.
Sophia Ananiadou, University of Manchester, United Kingdom.
R Ananthkrishnan, IBM Research - India, India.
Ion Androutsopoulos, Athens University of Economics and Business, Greece.
Gabor Angeli, Stanford, United States.
Marianna Apidianaki, LIMSI-CNRS, France.
Eiji Aramaki, The university of Tokyo, Japan.
Karunesh Arora, CDAC, India.
Nicholas Asher, CNRS, IRIT, Université Paul Sabatier, Toulouse, France.
Corinne Astesano, Univeristé de Toulouse, UTM, Laboratoire Octogone-Lordat, France.
Véronique Aubergé, CNRS, Grenoble (LIG, GETALP), France.
Tania Avgustinova, Saarland University, Germany.
Julia Aymerich, Pan American Health Organization, United States.
Wilker Aziz, University of Wolverhampton, United Kingdom.
Bruno Bachimont, Université de Technologie de Compiègne, France.
B Lakshmi Bai, IIIT Hyderabad, India.
Collin Baker, International Computer Science Institute, United States.
Timothy Baldwin, Affiliation unknown, United States.
Kalika Bali, Microsoft Research Labs India, India.
Rafael E Banchs, Institute for Infocomm Research, Singapore.
Sivaji Bandyopadhyay, Jadavpur University, India.
Sivaji Bandyopadhyay, Jadavpur University, India.
Carmen Banea, University of North Texas, United States.
Srinivas Bangalore, ATT, India.
Eva Banik, Computational Linguistics Ltd, United Kingdom.
Regina Barzilay, MIT, United States.
Núria Bel, Universitat Pompeu Fabra, Spain.
Valérie Bellyneck, Université de Grenoble (UJF, LIG, GETALP), France.
Emily M. Bender, University of Washington, United States.
Shane Bergsma, Johns Hopkins University, United States.
Vincent Berment, CS, Paris & Inalco, Paris & GETALP (LIG, UJF), Grenoble, France.
Nicola Bertoldi, FBK, Italy.
Laurent Besacier, Université de Grenoble (UJF, LIG, GETALP), France.
Steven Bethard, University of Colorado Boulder, United States.
Rajesh Bhatt, UMass Amherst, United States.
Pushpak Bhattacharyya, CFIIT, IIT Bombay, India.
Kristín Bjarnadóttir, The Arni Magnússon Institute for Icelandic Studies, Iceland.
Patrick Blackburn, Roskilde University, Denmark.
Herve Blanchon, Univ. of Grenoble (UPMF, LIG, GETALP), France.
Victor Bocharov, Saint-Petersburg State University, Russian Federation.
Igor Boguslavsky, RAS, IPPI/IITP, Moscow & UPM, Madrid, Russian Federation.
Christian Boitet, Université de Grenoble (LIG, GETALP), France.

Ondrej Bojar, Charles University, Prague, Czech Republic.
Danushka Bollegala, University of Tokyo, Japan.
Anastasia Bonch-Osmolovskaya, Higher School of Economics, philology department Moscow, Russian Federation.
Francis Bond, Nanyang Technological University, Singapore.
Kalina Bontcheva, University of Sheffield, United Kingdom.
Lars Borin, University of Gothenburg, Sweden.
Vladimir Borshev, University of Massachusetts, United States.
Johan Bos, University of Groningen, Netherlands.
Nadjet Bouayad-Agha, University Pompeu Fabra, Spain.
Florian Boudin, Université de Nantes, France.
Mohand Boughanem, IRIT, CNRS, France.
Gosse Bouma, University of Groningen, Netherlands.
Paolo Bouquet, University of Trento, Italy.
António Branco, University of Lisbon, Portugal.
Pavel Braslavski, Kontur labs, Russian Federation.
François Brown de Colstoun, Lingua et Machina, France.
Rebecca Bruce, University of North Carolina @ Asheville, United States.
Gerhard Budin, University of Vienna, Austria.
Paul Buitelaar, DERI, National University of Ireland, galway, Ireland.
Harry Bunt, Tilburg University, Netherlands.
Miriam Butt, University of Konstanz, Germany.
Lynne Cahill, University of Brighton, United Kingdom.
Shu Cai, USC/ISI, United States.
Nicoletta Calzolari, CNR, ILC, Pisa, Italy.
Erik Cambria, National University of Singapore, Singapore.
Nick Campbell, TCD, Ireland.
Yunbo Cao, Microsoft Research Asia, China.
Jesus Cardeñosa, Universidad Politecnica de Madrid (Spain), Spain.
Claire Cardie, Cornell University, United States.
Michael Carl, Copenhagen Business School, Denmark.
Marine Carpuat, National Research Council Canada, Canada.
Francisco Casacuberta, Universitat Politècnica de València, Spain.
Eric Castelli, International Research Institute MICA - CNRS, Viet Nam.
Daniel Cer, Stanford - NLP Group, United States.
Özlem Çetinoğlu, IMS, University of Stuttgart, Germany.
Vineet Chaitanya, IIIT Hyderabad, India.
Baobao Chang, Peking University, China.
Eugene Charniak, Brown University, United States.
Jacques Chauché, IIRMM Montpellier France, France.
Wanxiang Che, Harbin Institute of Technology, China.
Yu-N Cheah, Universiti Sains Malaysia, Malaysia.
Ying Chen, China Agricultural University, China.
Wenliang Chen, I2R, Singapore, Singapore.
Jiajun Chen, Nanjing University, China.
Boxing Chen, National Research Council, Canada.
Yidong Chen, Xiamen University, China.
Colin Cherry, National Research Council Canada, Canada.
Jean-Pierre Chevallet, Université de Grenoble (UJF, LIG, MRIM), France.
Jean-Pierre Chevrot, Lidilem, Université Stendhal, Institut Universitaire de France, France.
David Chiang, USC/ISI, United States.
Jen-Tzung Chien, National Chiao Tung University, Taiwan.

Manoj Chinnakotla, Relevance and Data Sciences Team, Bing, Microsoft, India.
Raymond Chiong, Swinburne University of Technology, Australia.
Key-sun Choi, KAIST, Republic of Korea.
Jinho Choi, University of Massachusetts Amherst, United States.
Monojit Choudhury, Microsoft Research Lab India, India.
Khalid Choukri, ELRA/ELDA, France.
Ken Church, IBM, United States.
Massimiliano Ciaramita, Google, Switzerland.
Philipp Cimiano, University of Bielefeld, Germany.
Michael Collins, Columbia University, United States.
Sherri L Condon, The MITRE Corporation, United States.
José Carlos Cortizo Pérez, BrainSINS, Spain.
Rute Costa, CLUNL - Universidade Nova de Lisboa, Portugal.
Dan Cristea, Al. I. Cuza University of Iasi, Romania.
Xiaodong Cui, IBM T. J. Watson Research Center, United States.
Aron Culotta, Northeastern Illinois University, United States.
Hamish Cunningham, University of Sheffield, United Kingdom.
Iria da Cunha, Universitat Pompeu Fabra, Spain.
Walter Daelemans, CLIPS, University of Antwerp, Belgium.
Ido Dagan, Bar-Ilan University, Israel.
Beatrice Daille, Université de Nantes - LINA, France.
Om Damani, IIT Bombay, India.
Luc Damas, Université de Savoie, France.
Sandipan Dandapat, CNGL, School of Computing, Dublin City University, Ireland.
Laurence Danlos, University Paris Diderot, France.
Kareem M Darwish, QF, Qatar.
Amitava Das, NTNU, Norway.
Marie-Catherine de Marneffe, Stanford University, United States.
Maarten de Rijke, University of Amsterdam, Netherlands.
Koenraad De Smedt, ULB, Norway.
Thierry Declerck, DFKI, Language Technology Lab, Germany.
Steve DeNeefe, SDL Language Weaver, United States.
Pascal Denis, INRIA, France.
Tejaswini Deoskar, University of Edinburgh, United Kingdom.
Heidi Depraetere, CrossLang NV, Belgium.
Leon Derczynski, University of Sheffield, United Kingdom.
Alain Désilets, National Research Council of Canada, Canada.
Barbara Di Eugenio, Univ. of Illinois at Chicago, United States.
Gaël Dias, University of Caen Basse-Normandie, France.
Fernando Diaz, Microsoft, United States.
Alberto Diaz, Universidad Complutense de Madrid, Spain.
Mike Dillinger, TOPs Globalization Consulting, United States.
Bill Dolan, Microsoft Research, United States.
Zhendong Dong, Canada Keentime Inc., Canada.
Sophie Donnadieu, Université de Savoie, France.
Justin Dornescu, University of Wolverhampton, United Kingdom.
Mark Dras, Macquarie University, Australia.
Markus Dreyer, SDL Language Weaver, United States.
Jinhua Du, Faculty of Automation and Information Engineering, Xi'an University of Technology, China.
Xiangyu Duan, Institute for Infocomm Research, Singapore, Singapore.
Pablo Duboue, Les Laboratoires Foulab / Université de Montreal, Canada.

Jonathan Dunn, Purdue University, United States.
Georges Dupret, Yahoo! Labs, United States.
Chris Dyer, Carnegie Mellon University, United States.
Kurt Eberle, Lingenio GmbH, Germany.
Terumasa Ehara, Yamanashi Eiwa College, Japan.
Jacob Eisenstein, Georgia Institute of Technology, United States.
Jason Eisner, Johns Hopkins University, United States.
Asif Ekbal, IIT Patna, India.
Michael Elhadad, Ben Gurion University, Israel.
Jeremy Ellman, Northumbria University, United Kingdom.
Jakob Elming, Copenhagen Business School, Denmark.
Micha Elsner, Ohio State University, United States.
Brigitte Endres-Niggemeyer, Noapps, Germany.
Chantal Enguehard, LINA, University of Nantes, France.
Tomaz Erjavec, Jozef Stefan Institute, Slovenia.
Katrín Erk, University of Texas at Austin, United States.
Xavier Blanco Escoda, Universitat Autònoma de Barcelona, Spain.
Emmanuelle Esperança-Rodier, UJF LIG Getalp, France.
Jérôme Euzenat, INRIA & LIG, France.
Roger Evans, University of Brighton, United Kingdom.
Achille Falaise, LIG-GETALP, France.
Ji Fang, Medallia.com, United States.
Atefeh Farzindar, NLP Technologies Inc., Canada.
Christiane Fellbaum, Princeton University, United States.
Zhiwei FENG, Institute of Applied Linguistics, The Ministry of Education, China.
Oscar Ferrandez, Nuance Communications, Inc., United States.
Antonio Ferrandez, University of Alicante, Spain.
Gabriela Ferraro, Universitat Pompeu Fabra, Spain.
Dan Flickinger, CSLI, Stanford University, United States.
Radu Florian, IBM, United States.
Ray Flournoy, Adobe, United States.
Corina Forascu, Univ. A.I.I. Cuza of Iasi, Faculty of Computer Science, Romania.
George Foster, National Research Council Canada, Canada.
Gil Francopoulo, CNRS-LIMSI-IMMI + Tagmatica, France.
Anette Frank, Universität Heidelberg, Germany.
Guohong Fu, School of Computer Science and Technology, Heilongjiang University, China.
Piotr Fuglewicz, TiP Sp. z o. o., Poland.
Atsushi Fujii, Tokyo Institute of Technology, Japan.
Robert Gaizauskas, University of Sheffield, United Kingdom.
Michel Galley, Microsoft, United States.
Suryakanth Gangashetty, International Institute of Information Technology Hyderabad, India.
Jianfeng Gao, Microsoft Research, Redmond, United States.
Claire Gardent, CNRS/LORIA UMR 7503 Nancy, France.
Éric Gaussier, Univ. J. Fourier, France.
Alexander Gelbukh, Instituto Politécnico Nacional, Mexico.
Josef Van Genabith, Dublin City University, Ireland.
Kim Gerdes, Sorbonne Nouvelle & Chinese Academy of Sciences, France.
Salvatore Giammarresi, PayPal, United States.
George Giannakopoulos, NCSR Demokritos, Greece.
Dafydd Gibbon, Universität Bielefeld, Germany.
Daniel Gildea, University of Rochester, United States.
Kevin Gimpel, Carnegie Mellon University, United States.

Filip Ginter, University of Turku, Finland.
Corina R Girju, UIUC, United States.
Oren Glickman, Unaffiliated, Israel.
Asunción Gómez Pérez, Universidad Politecnica de Madrid, Spain.
Jerome Goulian, LIG - GETALP, France.
Cyril Goutte, National Research Council Canada, Canada.
Vishal Goyal, Department of Computer Science, Punjabi University Patiala, India.
Jorge Gracia, Universidad Politécnica de Madrid, Spain.
Didier Grandjean, University of Geneva, Switzerland.
Daniel Grasmick, Lucy Software and Services GmbH, Germany.
Brigitte Grau, LIMSI-CNRS, France.
Gregory Grefenstette, 3DS Exalead, France.
Nikolai Grigoriev, Yandex, Russian Federation.
Ralph Grishman, New York University, United States.
Iryna Gurevych, UKP Lab, Technische Universität Darmstadt and DIPF, Germany.
Louise Guthrie, University of Brighton, United Kingdom.
Barry Haddow, University of Edinburgh, United Kingdom.
Eva Hajicova, Charles University, Prague, Czech Republic.
Tanmay Haldankar, OC, IITB, India.
Keith Hall, Google Research, United States.
Rejwanul Haque, Applied Language Solutions, India.
Sanda Harabagiu, University of Texas at Dallas, United States.
Christian Hardmeier, Uppsala University, Sweden.
Tony Hartley, Tokyo University of Foreign Studies, Japan.
Sven Hartrumpf, SEMPRIA GmbH, 40237 Düsseldorf, Germany.
chikara hashimoto, NICT, Japan.
Jun Hatori, Apple Inc., United States.
Petter Haugereid, University of Haifa, Israel.
Katsuhiko Hayashi, NAIST, Japan.
Xiaodong He, Microsoft Research Redmond, United States.
Jing He, University of Montreal, Canada.
Kenneth Heafield, University of Edinburgh, Carnegie Mellon University, United States.
Sigrún Helgadóttir, The Árni Magnússon Institute for Icelandic Studies, Iceland.
Christian Hempelmann, Texas A&M-Commerce, United States.
James Henderson, Xerox Research Centre Europe, France.
Iris Hendrickx, Centro de Linguística da Universidade de Lisboa, Portugal.
Jesus M. Hermida, University of Alicante, Spain.
Nicolas Hernandez, University of Nantes, France.
Huang Heyan, Department of Computer Science and Technology, Beijing Institute of Technology, China.
Amanda Hicks, University at Buffalo, United States.
Djoerd Hiemstra, University of Twente, Netherlands.
Erhard Hinrichs, Tuebingen University, Germany.
Hieu Hoang, University of Edinburgh, United Kingdom.
Thomas Hoar, Precision Translation Tools Co., Ltd., Canada.
Julia Hockenmaier, University of Illinois at Urbana-Champaign, United States.
Jim Hogan, Queensland University of Technology, Australia.
Kristy Hollingshead, Department of Defense, United States.
Fred Hollowood, Symantec, CNGL, Ireland.
Matthew Honnibal, Macquarie University, Australia.
Ales Horak, Masaryk University, Czech Republic.
Véronique Hoste, University College Ghent, Belgium.

Eduard Hovy, Carnegie Mellon University, United States.
Chang Hu, University of Maryland, United States.
Yunhua Hu, Affiliation unknown, China.
Liang Huang, CUNY, United States.
Shujian Huang, Nanjing University, China.
Jimmy Huang, York Univ., Canada.
Chu-Ren Huang, Affiliation unknown, China.
Matthias Huck, RWTH Aachen University, Germany.
Samar Husain, University of Potsdam, Germany.
Sarmad Hussain, KICS, Pakistan.
Nancy Ide, Vassar College, United States.
Adrian Ifte, Al. I. Cuza University of Iasi, Faculty of Computer Science, Romania.
Ryu Iida, Tokyo Institute of Technology, Japan.
Kentaro Inui, Tohoku University, Japan.
Leonid Iomdin, Institute for Information Transmission Problems, RAS, Russian Federation.
Boris Iomdin, Russian Language Institute, RAS, Russian Federation.
Radu Ion, Research Institute for Artificial Intelligence, Romanian Academy, Romania.
Pierre Isabelle, National Research Council, Canada.
Hitoshi Isahara, NiCT, Japan.
Mustafa Jarrar, Mustafa Jarrar, Sina Institute, Birzeit University, Palestine, Palestine.
Girish Jha, Jawaharlal Nehru University, New Delhi, India.
Heng Ji, City University of New York, United States.
Jie Jiang, Applied Language Solutions, United Kingdom.
Wenbin Jiang, Institute of Computing Technology, CAS, China.
Jing Jiang, Singapore Management University, Singapore.
Narsu JIN, Hangzhou Normal University, China.
Mark Johnson, Macquarie univ., Sydney, Australia.
Howard Johnson, National Research Council, Canada.
Kristiina Jokinen, University of Helsinki and University of Tartu, Finland.
Gareth Jones, Dublin City University, Ireland.
Pamela Jordan, University of Pittsburgh, United States.
Aravind K Joshi, University of Pennsylvania, United States.
Alain Joubert, LIRMM, University of Montpellier, France.
Heiki-Jaan Kaalep, University of Tartu, Estonia.
Sylvain Kahane, Modyco, Université Paris Ouest Nanterre, France.
Nobuhiro KAJI, Tokyo University, Japan.
Hiroshi Kanayama, IBM Research - Tokyo, Japan.
Lauri Karttunen, Stanford University, United States.
Graham Katz, CACI, Inc., United States.
Daisuke Kawahara, Kyoto University, Japan.
Martin Kay, Stanford University, United States.
Junichi Kazama, NiCT, Japan.
Maxim Khalilov, TAUS, Netherlands.
Mitesh Khapra, IBM Research India, India.
Genichiro KIKUI, Okayama Prefectural University, Japan.
Adam Kilgarriff, Lexical Computing Ltd, United Kingdom.
Tracy Holloway King, eBay, United States.
Ewan Klein, University of Edinburgh, United Kingdom.
Kevin Knight, USC/ISI, United States.
Philipp Koehn, University of Edinburgh, United Kingdom.
Natalia Konstantinova, University of Wolverhampton, RIILP, United Kingdom.
Ruud Koolen, Tilburg University, Netherlands.

Valia Kordoni, DFKI GmbH and Saarland University, Germany.
Ioannis Korkontzelos, National Centre for Text Mining, School of Computer Science, The University of Manchester, United Kingdom.
Kimmo Koskenniemi, University of Helsinki, Finland.
Leila Kosseim, Concordia University, Canada.
Milen Kouylekov, Celi S.R.L., Italy.
Zornitsa Kozareva, Information Sciences Institute/University of Southern California, United States.
Elena Kozerenko, Institute of Informatics Problems of the RAS, Russian Federation.
Emiel Kraahmer, Tilburg University, Netherlands.
Roland Kuhn, National Research Council, Canada.
Amba Kulkarni, University of Hyderabad, India.
Shankar Kumar, Google, United States.
Oren Kurland, Technion, Israel.
Mathieu Lafourcade, Univ. de Montpellier II (LIRMM), France.
Sobha Lalitha Devi, AU-KBC Research Centre, India.
Mounia Lalmas, Yahoo! Labs, United States.
Lori Lamel, LIMSI-CNRS, France.
Terry Langendoen, National Science Foundation, United States.
Philippe Langlais, Université de Montréal (RALI), Canada.
Guy Lapalme, Université de Montréal (RALI), Canada.
Lynda Tamine Lechani, Institut de Recherche en Informatique (IRIT), France, France.
Alain Lecomte, Université Paris 8, France.
Gary Geunbae Lee, POSTECH, Republic of Korea.
G S Lehal, Punjabi University, Patiala, India.
Alessandro Lenci, University of Pisa, Italy.
Yves Lepage, IPS, Waseda university, Japan.
Lori Levin, Carnegie Mellon University, United States.
Roger Levy, UCSD, United States.
Will Lewis, Microsoft Research, United States.
Haizhou Li, Institute for Infocomm Research, Singapore.
Juanzi Li, Tsinghua University, China.
Fennie Liang, Guy's and St Thomas' Hospital, United Kingdom.
Donghui Lin, Department of Social Informatics, Kyoto University, Japan.
Krister Linden, University of Helsinki, Finland.
Kenneth C Litkowski, CL Research, United States.
Qun Liu, Dublin City University, Ireland & Chinese Academy of Sciences, China, China.
Kang Liu, Institute of Automation, Chinese Academy of Sciences, China.
Xiaohua Liu, Microsoft Research Asia, China.
Yang Liu, University of Texas at Dallas, United States.
Elena Lloret, University of Alicante, Spain.
N. V. Loukachevitch, Lomonosov Moscow State University, Russian Federation.
António Lucas Soares, INESC Porto and University of Porto - Faculty of Engineering, Portugal.
Robert Luk, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China.
YanJun Ma, Baidu, China.
Bill MacCartney, Google, United States.
Bente Maegaard, University of Copenhagen, Denmark.
Mathew Magimai Doss, Idiap Research Institute, Martigny, Switzerland.
Bernardo Magnini, FBK, Italy.
Gudrun Magnusdottir, Chairperson EStEAM AB, Germany.
Prasenjit Majumdar, IIIT Hyderabad, India.
M W Mak, The Hong Kong Polytechnic University, China.

Brian Mak, The Hong Kong University of Science and Technology, Hong Kong Special Administrative Region of China.
Abbas Malik, Faculty of Computing and IT, King Abdulaziz University, Jeddah, Saudi Arabia.
Radhika Mamidi, IIIT Hyderabad, India.
Suresh Manandhar, University of York, UK, United States.
Rohit Manchanda, IITB, Mumbai, India.
Mathieu -nagata Mangeot, GETALP-LIG Laboratory, France.
Prashanth Mannem, LTRC, IIIT-Hyderabad, India.
Christopher Manning, Stanford University, United States.
Stephane Marchand-Maillet, Dept of Computer Science - University of Geneva, Switzerland.
Daniel Marcu, ISI/USC, United States.
Joseph Mariani, LIMSI-CNRS & IMMI, France.
Jose Mariño, Universitat Politècnica de Catalunya, Spain.
Ronaldo Martins, UNDL Foundation, Switzerland.
Roche Mathieu, LIRMM, CNRS, Univ. Montpellier 2, France.
Spyros Matsoukas, Raytheon BBN Technologies, United States.
Yuji Matsumoto, Nara Institute of Science and Technology, Japan.
Kazuyuki Matsumoto, The University of Tokushima, Japan.
Takuya Matsuzaki, National Institute of Informatics, Japan.
Arne Mauser, Google, United States.
Aurélien Max, LIMSI-CNRS, France.
Mark Maybury, Department of Defense, United States.
James Mayfield, Johns Hopkins University, United States.
Diana Maynard, University of Sheffield, United Kingdom.
Joseph F McCarthy, University of Washington Bothell, United States.
David McClosky, Stanford University / IBM Research, United States.
John McCrae, University of Bielefeld, Germany.
Kathleen McKeown, Columbia University, United States.
Paul McNamee, Johns Hopkins University, United States.
Marjorie McShane, University of Maryland Baltimore County, United States.
Karine Megerdoomian, MITRE, United States.
Beáta Megyesi, Uppsala University, Sweden.
Arul Menezes, Microsoft, United States.
Elisabeth Metais, CNAM, France.
Detmar Meurers, Universität Tübingen, Germany.
Haitao Mi, IBM Watson Research Center, United States.
Rada Mihalcea, University of North Texas, United States.
Márton Miháلتz, Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary.
Eleni Miltsakaki, University of Pennsylvania, United States.
Jean-luc Minel, Modyco, Université Paris Ouest Nanterre La Défense, France.
Jiri Mirovsky, Charles University, Prague, Czech Republic.
Hemant Misra, Philips Electronics India Limited, India.
Ruslan Mitkov, University of Wolverhampton, United Kingdom.
Mandar Mitra, Indian Statistical Institute, India.
Makoto Miwa, National Centre for Text Mining, The University of Manchester, Japan.
Yusuke Miyao, NII, Japan.
Riichiro Mizoguchi, Osaka University, Japan.
Marie-francine Moens, Hogent, Belgium.
jacques Moeschler, University of Geneva, Switzerland.
Abdel-rahman Mohamed, University of Toronto, Canada.
Karo Moilanen, Department of Computer Science, University of Oxford, United Kingdom.
Dan Moldovan, Univ. Dallas, United States.

Saeedeh Momtazi, Ubiquitous Knowledge Processing Lab (UKP-DIPF), German Institute for Educational Research and Educational Information, Germany.
 Monica Monachini, CNR, ILC, Pisa, Italy.
 Elena Montiel-Ponsoda, Universidad Politécnica de Madrid, Spain.
 Christof Monz, University of Amsterdam, Netherlands.
 Robert C Moore, Google, United States.
 Lidia Moreno, Universidad Politécnica de Valencia, Spain.
 Asuncion Moreno, Universitat Politècnica de Catalunya, Spain.
 Alessandro Moschitti, University of Trento, Italy.
 Isabelle Moulinier, Thomson Reuters, United States.
 Stefan Müller, Freie Universität Berlin, Germany.
 Rafael Muñoz, DSI, Univ. Alicante, Spain.
 Masaki Murata, Tottori University, Japan.
 Yugo Murawaki, Kyoto University, Japan.
 Didier Nakache, Affiliation unknown, France.
 Preslav Nakov, QCRI, Qatar Foundation, Qatar.
 Tahira Naseem, MIT, United States.
 Costanza Navarretta, University of Copenhagen, Denmark.
 Roberto Navigli, Sapienza University of Rome, Italy.
 Anja Nedoluzko, Charles University, Prague, Czech Republic.
 Matteo Negri, Fondazione Bruno Kessler - irst, Italy.
 Ani Nenkova, University of Pennsylvania, United States.
 John Nerbonne, University of Groningen and Freiburg Institute of Advanced Studies, Netherlands.
 Guenter Neumann, DFKI, Germany.
 Nelson Ng, eBay, United States.
 Hwee Tou Ng, National University of Singapore, Singapore.
 Vincent Ng, University of Texas at Dallas, United States.
 Hong-Thai Nguyen, PolySpot, France.
 Tu Anh T. Nguyen, The Open University, UK, United Kingdom.
 Jian-Yun Nie, Université de Montréal (RALI), Canada.
 Takashi Ninomiya, Ehime University, Japan.
 Sergei Nirenburg, Univ. of Maryland, Baltimore, United States.
 Malvina Nissim, University of Bologna, Italy.
 Zhengyu Niu, Baidu, China.
 Joakim Nivre, Uppsala University, Sweden.
 Attila Novák, MTA-PPKE Language Technology Research Group – Pázmány Péter Catholic University, Faculty of Information Technology, Budapest, Hungary.
 Morgan O'Brien, McAfee, Ireland.
 Douglas Oard, University of Maryland, United States.
 J E J M Odiijk, Utrecht University, Netherlands.
 Stephan Oepen, Universitetet i Oslo, Norway.
 Kemal Oflazer, Carnegie Mellon University, Qatar.
 Jong-hoon OH, NICT, Japan.
 Naoaki Okazaki, Tohoku University, Japan.
 Manabu Okumura, Tokyo Institute of Technology, Japan.
 Constantin Orasan, University of Wolverhampton, United Kingdom.
 Csaba Oravecz, Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary.
 Daniel Ortiz, Polytechnic University of Valencia, Spain.
 Petya Osenova, Sofia University "St. Kl. Ohridski", Bulgaria.
 Ekaterina Ovchinnikova, USC ISI, United States.
 Sebastian Pado, Heidelberg University, Germany.
 Karel Pala, NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic.

Matteo Palmonari, University of Milan-Bicocca, Italy.
 Manuel Palomar, University of Alicante, Spain.
 Thiago Pardo, University of São Paulo, Brazil.
 Jong C Park, KAIST, Republic of Korea.
 Barbara Partee, University of Massachusetts Amherst, United States.
 Ranjani Parthasarathi, Anna University, Chennai, India, India.
 Marius Pasca, Google Inc., United States.
 Siddharth Patwardhan, IBM T. J. Watson Research Center, United States.
 Soma Paul, International Institution of Information Technology (IIIT), India.
 Bolette Sandford Pedersen, University of Copenhagen, Denmark.
 Anselmo Peñas, NLP & IR Group, UNED, Spain.
 Sergio Penkale, Applied Language Solutions, United Kingdom.
 Gerald Penn, University of Toronto, Canada.
 Bhaskararao Peri, International Institute of Information Technology, India.
 Wim Peters, University of Sheffield, United Kingdom.
 Maxim Petrenko, The International Monetary Fund, Machine Translation Section, United States.
 Slav Petrov, Google, Germany.
 Christian Pietsch, Bielefeld University, Germany.
 Maria Mercedes Piñango, Yale University, United States.
 Prasad Pingali, Guest faculty, IIIT Hyderabad and CEO, SETU Software Systems, India.
 Mārcis Pinnis, Tilde, Latvia.
 Stelios Piperidis, ILSP/Athena RC, Greece.
 Tommi Pirinen, University of Helsinki, Finland.
 Vito Pirrelli, Institute for Computational Linguistics CNR Pisa, Italy.
 Paul Piwek, The Open University, United Kingdom.
 Laura Plaza, Universidad Complutense de Madrid, Spain.
 Mirko Plitt, Autodesk, Switzerland.
 Massimo Poesio, University of Essex, United Kingdom.
 Thierry Poibeau, LATTICE-CNRS, France.
 Simone Paolo Ponzetto, Università di Roma "La Sapienza", Italy.
 Hoifung Poon, Microsoft Research, China.
 Kashyap Popat, OC, IITB, India.
 Octavian Popescu, FBK-HLT, Italy.
 Andrei Popescu-Belis, Idiap Research Institute, Switzerland.
 Maja Popovic, DFKI, Germany.
 Fred Popowich, Simon Fraser University, Canada.
 Matt Post, Johns Hopkins University, United States.
 Christopher Potts, Stanford, United States.
 R Power, The Open University, United Kingdom.
 Sameer Pradhan, Raytheon BBN Technologies, United States.
 John Prager, IBM T.J. Watson Research Ctr., United States.
 Kishore Prahallad, IIIT Hyderabad, India.
 Rashmi Prasad, University of Wisconsin-Milwaukee, United States.
 Violaine Prince, LIRMM-CNRS/Univ. Montpellier 2, France.
 Jean-Philippe Prost, Affiliation unknown, France.
 Gábor Proszéky, MorphoLogic & Pázmány Péter Catholic University, Hungary.
 Adam Przepiórkowski, Institute of Computer Science, Polish Academy of Sciences, Poland.
 James Pustejovsky, Brandeis University, United States.
 Sampo Pyysalo, NaCTeM and University of Manchester, United Kingdom.
 Vahed Qazvinian, University of Michigan, Ann Arbor, United States.
 Chris Quirk, Microsoft Research, United States.
 Valeria Quochi, Istituto di Linguistica Computazionale "A. Zampolli", CNR, Italy.

Marta R. Costa-jussà, Barcelona Media Innovation Center, Spain.
Achla Raina, Indian Institute of Technology Kanpur, India.
Owen Rambow, Columbia University, United States.
Balisoamanandray Ranaivo-Malançon, Universiti Malaysia Sarawak, Malaysia.
Aarne Ranta, University of Gothenburg, Sweden.
Umamaheshwara G Rao, IIIT Hyderabad, India.
Victor Raskin, Purdue University, United States.
Paul Rayson, Lancaster University, United Kingdom.
Marta Recasens, Stanford University, United States.
Steve Richardson, The Church of Jesus Christ of Latter-day Saints, United States.
Stefan Riezler, Heidelberg University, Germany.
German Rigau, University of the Basque Country, Spain.
Luca Rigazio, Panasonic, United States.
Brian Roark, OHSU, United States.
Christophe Roche, Université de Savoie, France.
Horacio Rodriguez, UPC, Spain.
Eiríkur Rögnvaldsson, University of Iceland, Iceland.
Hannah Rohde, University of Edinburgh, Linguistics & English Language, United Kingdom.
Laurent Romary, Inria & HUB, France.
Paolo Rosso, Universitat Politècnica de València, Spain.
Dan Roth, University of Illinois at Urbana-Champaign, United States.
Johann Roturier, Symantec, France.
David Rouquet, LIG-GETALP, France.
Bogdan Sacaleanu, IBM Research Ireland, Ireland.
Kenji Sagae, University of Southern California, United States.
Horacio Saggion, Universitat Pompeu Fabra, Spain.
Yoshinori Sagisaka, Waseda University & NICT, Japan.
Tetsuya Sakai, Microsoft Research Asia, China.
Mark Sammons, University of Illinois at Urbana-Champaign, United States.
Felipe Sánchez-Martínez, Universitat d'Alacant, Spain.
Rajeev Sangal, IIIT Hyderabad, India.
Diana Santos, University of Oslo, Norway.
Sudeshna Sarkar, IIT Kharagpur, India.
Anoop Sarkar, SFU, Canada.
Felix Sasaki, DFKI / W3C Fellow, Germany.
Roser Sauri, Barcelona Media Innovation Center, Spain.
Jacques Savoy, University of Neuchatel, Switzerland.
Kevin Scannell, Saint Louis University, United States.
Dag Schmidtke, Microsoft, Germany.
Klaus-Dirk Schmitz, Cologne University of Applied Sciences, Germany.
Didier Schwab, Université de Grenoble, France.
Lane Schwartz, Air Force Research Lab, United States.
Holger Schwenk, LIUM, Université du Mans, France.
Donia Scott, Univ. of Sussex, United Kingdom.
Satoshi Sekine, New York University, United States.
Vladimir Selegey, Russian State University for the Humanities, Russian Federation.
Mark Seligman, Spoken Translation, Inc., United States.
Jean Senellart, Systran, France.
Gilles Sérasset, GETALP-LIG, Université Joseph Fourier, Grenoble 1, France.
Sam Sethserey, Institute of Technology of Cambodia, Cambodia.
Ritesh Shah, OC, IITB, India.
Dipti Sharma, IIIT Hyderabad, India.

Serge Sharoff, University of Leeds, United Kingdom.
Bernadette Sharp, Staffordshire University, United Kingdom.
Wade Shen, MIT, United States.
Svetlana Sheremetyeva, LanA Consulting ApS, Denmark.
Shuming SHI, Microsoft, China.
Xiaodong SHI, Xiamen University, China.
Tomohide Shibata, Kyoto University, Japan.
Hiroyuki Shindo, NTT Communication Science Laboratories, Japan.
Eyal Shnarch, Bar Ilan University, Israel.
Luo Si, Purdue University, United States.
Advait Siddharthan, University of Aberdeen, United Kingdom.
Grigori Sidorov, Instituto Politecnico Nacional (IPN), Mexico, Mexico.
Khe Chai SIM, National University of Singapore, Singapore.
Khalil Simaan, Institute for Logic, Language and Computation, University of Amsterdam, Netherlands.
Kiril Simov, IICT, Bulgarian Academy of Sciences, Bulgaria.
Navjyoti Singh, IIIT Hyderabad, India.
Smriti Singh, Insideview Technologies (India) Pvt. Ltd., India.
Nandini Chatterjee Singh, National Brain Research Centre, India.
Mukul Sinha, Expert Software Consultants Ltd., India.
Ravi Sinha, University of North Texas, United States.
Milena Slavcheva, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Bulgaria.
Raymond Slyh, AFRL, United States.
David Smith, Northeastern University, United States.
Alexei Sokirko, Yandex, Russian Federation.
Ruihua Song, Microsoft Research Asia, China.
Fei Song, University of Guelph, Canada.
Claudia Soria, CNR, ILC, Pisa, Italy.
Virach Sornlertlamvanich, NECTEC, Thailand.
Lucia Specia, University of Sheffield, United Kingdom.
Marcus Spies, Professor, Chair of Knowledge Management, LMU University of Munich, Germany.
Valentin I Spitkovsky, Stanford University and Google Inc., United States.
Bangalore Srinivas, ATT, India.
Narayanan Srinivasan, CBCS, India.
Sanja Stajner, University of Wolverhampton, United Kingdom.
Efsthathios Stamatatos, University of the Aegean, Greece.
Manfred Stede, University of Potsdam, Germany.
Mark Steedman, University of Edinburgh, United Kingdom.
Ralf Steinberger, European Commission - Joint Research Centre (JRC), Italy.
Josef Steinberger, Joint Research Centre, Italy.
Mark Stevenson, University of Sheffield, United Kingdom.
Matthew Stone, Rutgers, United States.
Veselin Stoyanov, Johns Hopkins University, United States.
Carlo Strapparava, FBK-irst, Italy.
Michael Strube, HITS gGmbH, Germany.
Tomek Strzalkowski, SUNY Albany, United States.
Keh-Yih Su, Behavior Design Corp., Taiwan.
Jian Su, Institute for Infocomm Research, Singapore, Singapore.
Mari Carmen Suárez-Figueroa, Ontology Engineering Group (OEG), Universidad Politécnica de Madrid (UPM), Spain.
Le Sun, ISCAS, China.
Bapi Raju Surampudi, University of Hyderabad, India, India.

Jun Suzuki, NTT CS Lab., Japan.
 Stan Szpakowicz, School of Electrical Engineering & Computer Science, University of Ottawa, Canada.
 Marko Tadić, University of Zagreb, Faculty of Humanities and Social Sciences, Croatia.
 Hirotoishi Taira, NTT Communication Science Laboratories, Japan.
 John Tait, Johntait.net Ltd., United Kingdom.
 Koichi Takeda, IBM Research, Japan.
 Songbo Tan, Institute of Computing Technology, China.
 Tien Ping Tan, Universiti Sains Malaysia, Malaysia.
 Hristo Tanev, Joint Research Centre, United Kingdom.
 Enya Kong Tang, Linton University College, Malaysia.
 Kristian Tangsgaard Hvelplund, University of Copenhagen, Denmark.
 Julia Taylor, Purdue University, United States.
 Joel Tetreault, Educational Testing Service, France.
 Michael Thelwall, University of Wolverhampton, United Kingdom.
 Mariët Theune, University of Twente, Netherlands.
 Joerg Tiedemann, Uppsala University, Sweden.
 Christoph Tillmann, IBM Research, Germany.
 Lana Timoshenko, Institute for Information Transmission Problems, Russian Academy of Sciences, Russian Federation.
 Tomoki Toda, Nara Institute of Science and Technology, Japan.
 Takenobu Tokunaga, TITECH, Japan.
 Sara Tonelli, Fondazione Bruno Kessler, Italy.
 kentaro Torisawa, National Institute of Information and Communications Technology (NICT), Japan.
 Juan-manuel Torres, LIA Université d'Avignon, France.
 Kristina Toutanova, Microsoft, United States.
 Dolf Trieschnigg, University of Twente, Netherlands.
 Harald Trost, Medical University of Vienna, Austria.
 Andrew Trotman, University of Otago, New Zealand.
 José A Troyano, University of Seville, Spain.
 Richard Tzong-Han Tsai, Yuan Ze University, Taiwan.
 Reut Tsarfaty, Uppsala University, Sweden.
 Daniel Tse, University of Sydney, Australia.
 Jun-ichi Tsujii, MSRA, Beijing, China.
 Yoshimasa Tsuruoka, The University of Tokyo, Japan.
 Zhaopeng Tu, Institute of Computing Technology, Chinese Academy of Sciences, China.
 Francis Tyers, Universitat d'Alacant, Spain.
 Hiroshi Uchida, UNDL, UNU, Japan.
 Raghavendra Udupa, Microsoft Research, India.
 Olga Uryupina, University of Trento, Italy.
 Hans Uszkoreit, DFKI, Saarbrücken, Germany.
 Masao Utiyama, NICT, Japan.
 Takehito Utsuro, University of Tsukuba, Japan.
 Ashwini Vaidya, University of Colorado, Boulder, United States.
 Kees van Deemter, University of Aberdeen, United Kingdom.
 Antal van den Bosch, LET, Netherlands.
 Josef van Genabith, CNGL, NCLT, Dublin City University, Ireland.
 Dániel Varga, Budapest University of Technology and Economics, Hungary.
 Vasudeva Varma, IIIT Hyderabad, India.
 Sriram Venkatapathy, Xerox Research Centre Europe, France.
 Suzan Verberne, Radboud University Nijmegen, Netherlands.
 Marc Verhagen, Brandeis University, United States.

Jose Luis Vicedo, University of Alicante (Spain), Spain.
Renata Vieira, PUCRS, Brazil.
David Vilar, DFKI, Germany.
Aline Villavicencio, Federal University of Rio Grande do Sul, Brazil.
Anne Vilnat, LIMSI-CNRS et Université Paris-Sud, France.
Jorge Vivaldi, Institut Universitari de Lingüística Aplicada (UPF), Spain.
Klaus von Heusinger, University of Cologne, Germany.
Piek Vossen, VU University Amsterdam, Netherlands.
Atro Voutilainen, University of Helsinki, Finland.
Xiaojun Wan, Peking University, China.
Haifeng Wang, Baidu, China.
Haifeng Wang, Baidu, China.
Rui Wang, DFKI GmbH, Germany.
Kewen Wang, Griffith University, Australia.
Bin Wang, Institute of Computing Technology, Chinese Academy of Sciences, China.
Leo Wanner, ICREA and Pompeu Fabra University, Spain.
Taro Watanabe, NICT, Japan.
Andy Way, Univ. of Dublin and Applied Language, Ireland.
Bonnie Webber, University of Edinburgh, United Kingdom.
Éric Wehrli, University of Geneva, Switzerland.
Furu Wei, Microsoft Research Asia, China.
Ben Wellner, Brandeis University, United States.
Michael White, The Ohio State University, United States.
Jan Wiebe, University of Pittsburgh, United States.
Yorick Wilks, University of Sheffield, United Kingdom.
Sandra Williams, The Open University, United Kingdom.
Philip James Williams, University of Edinburgh, United Kingdom.
Shuly Wintner, University of Haifa, Israel.
Andreas Witt, Institut für Deutsche Sprache, Germany.
Sue Ellen Wright, Affiliation unknown, United States.
Dekai Wu, HKUST, Hong Kong Special Administrative Region of China.
Fei Xia, University of Washington, United States.
Deyi Xiong, Institute for Infocomm Research, Singapore, Singapore.
Feiyu Xu, DFKI LT-Lab, Germany.
Gu Xu, Microsoft Bing, United States.
Nianwen Xue, Brandeis University, United States.
Serge Yablonsky, St. Petersburg University, Russia, Russian Federation.
Hongfei Yan, Peking University, China.
Muyun YANG, Harbin Institute of Technology, China.
Heng Yu, Institute of Computing Technology, Chinese Academy of Sciences, China.
Francois Yvon, LIMSI/CNRS & Univ Paris Sud, France.
Zdenek Zabokrtsky, Charles University, Prague, Czech Republic.
Annie Zaenen, Affiliation unknown, United States.
Fabio Massimo Zanzotto, University of Rome "Tor Vergata", Italy.
B Y Zarin, Universiti Malaysia Sarawak, Malaysia.
Torsten Zesch, Technische Universität Darmstadt, Germany.
Luke Zettlemoyer, University of Washington, United States.
Yi Zhang, German Research Center for Artificial Intelligence, Germany.
Min Zhang, Institute for Infocomm Research, Singapore.
Yue Zhang, Singapore University of Technology and Design, Singapore.
Hui Zhang, USC/ISI, United States.
Tiejun Zhao, Harbin Institute of Technology, China.

Yinggong Zhao, State Key Laboratory for Novel Software Technology at Nanjing University, China.
Ventsislav Zhechev, Autodesk, Switzerland.
Jing Zheng, SRI International, United States.
Ming Zhou, Microsoft Research Asia, China.
Guodong Zhou, Soochow University, China.
Xiaodan Zhu, National Research Council Canada, Canada.
ZhuJB Zhu, Northeastern University, China.
Sarka Zikanova, Charles University, Prague, Czech Republic.
Chengqing Zong, Institute of Automation, CAS, China.
Pierre Zweigenbaum, LIMSI-CNRS, France.

Table of Contents

<i>Multi-Dimensional Feature Merger for Question Answering</i> Apoorv Agarwal, J William Murdock, Jennifer Chu-Carroll, Adam Lally and Aditya Kalyanpur	1
<i>Unsupervised Discovery of Relations and Discriminative Extraction Patterns</i> Alan Akbik, Larysa Visengeriyeva, Priska Herger, Holmer Hemsén and Alexander Löser	17
<i>Automatic Detection of Point of View Differences in Wikipedia</i> Khalid Al Khatib, Hinrich Schütze and Cathleen Kantner	33
<i>SpeedRead: A Fast Named Entity Recognition Pipeline</i> Rami Al-Rfou' and Steven Skiena	51
<i>Experiments with Term Translation</i> Mihael Arcan, Christian Federmann and Paul Buitelaar	67
<i>The Floating Arabic Dictionary: An Automatic Method for Updating a Lexical Database through the Detection and Lemmatization of Unknown Words</i> Mohammed Attia, Younes Samih, Khaled Shaalan and Josef van Genabith	83
<i>Contribution of Complex Lexical Information to Solve Syntactic Ambiguity in Basque</i> Aitziber Atutxa, Eneko Agirre and Kepa Sarasola	97
<i>Comparative Quality Estimation: Automatic Sentence-Level Ranking of Multiple Machine Translation Outputs</i> Eleftherios Avramidis	115
<i>Constructing Reference Semantic Predictions from Biomedical Knowledge Sources</i> Demeke Ayele, Jean-Pierre Chevallet, Million Meshesha and Getnet Kassie	133
<i>Translation Quality-Based Supplementary Data Selection by Incremental Update of Translation Models</i> Pratyush Banerjee, Sudip Kumar Naskar, Johann Roturier, Andy Way and Josef van Genabith	149
<i>Text Reuse Detection using a Composition of Text Similarity Measures</i> Daniel Bär, Torsten Zesch and Iryna Gurevych	167
<i>Deriving Paraphrases for Highly Inflected Languages from Comparable Documents</i> Kfir Bar and Nachum Dershowitz	185
<i>Harvesting Parallel Text in Multiple Languages with Limited Supervision</i> Luciano Barbosa, Vivek Kumar Rangarajan Sridhar, Mahsa Yarmohammadi and Srinivas Bangalore	201
<i>An Evaluation of Statistical Post-Editing Systems Applied to RBMT and SMT Systems</i> Hanna Béchara, Raphaël Rubino, Yifan He, Yanjun Ma and Josef van Genabith	215
<i>Prague Dependency Treebank 2.5 – a Revisited Version of PDT 2.0</i> Eduard Bejček, Jarmila Panevová, Jan Popelka, Pavel Straňák, Magda Ševčíková, Jan Štěpánek and Zdeněk Žabokrtský	231

<i>Deriving a Lexicon for a Precision Grammar from Language Documentation Resources: A Case Study of Chintang</i>	
Emily M. Bender, Robert Schikowski and Balthasar Bickel	247
<i>Quantifying Semantics using Complex Network Analysis</i>	
Chris Biemann, Stefanie Roos and Karsten Weihe	263
<i>Improvements to Training an RNN parser</i>	
Richard Billingsley and James Curran	279
<i>Thread Specific Features are Helpful for Identifying Subjectivity Orientation of Online Forum Threads</i>	
Prakhar Biyani, Sumit Bhatia, Cornelia Caragea and Prasenjit Mitra	295
<i>Natural Language Generation for Nature Conservation: Automating Feedback to Help Volunteers Identify Bumblebee Species</i>	
Steven Blake, Advait Siddharthan, Hien Nguyen, Nirwan Sharma, Anne-Marie Robinson, Elaine O'Mahony, Ben Darvill, Chris Mellish and Rene van der Wal	311
<i>Studying the Effect of Input Size for Bayesian Word Segmentation on the Providence Corpus</i>	
Benjamin Börschinger, Katherine Demuth and Mark Johnson	325
<i>Bayesian Language Modelling of German Compounds</i>	
Jan A. Botha, Chris Dyer and Phil Blunsom	341
<i>Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish</i>	
Stefan Bott, Luz Rello, Biljana Drndarevic and Horacio Saggon	357
<i>Identification of Social Acts in Dialogue</i>	
David Bracewell, Marc Tomlinson and Hui Wang	375
<i>Robust, Lexicalized Native Language Identification</i>	
Julian Brooke and Graeme Hirst	391
<i>Identifying Urdu Complex Predication via Bigram Extraction</i>	
Miriam Butt, Tina Bögel, Annette Hautli, Sebastian Sulger and Tafseer Ahmed	409
<i>Native Language Identification using Recurring n-grams – Investigating Abstraction and Domain Dependence</i>	
Serhiy Bykh and Detmar Meurers	425
<i>Analysis and Enhancement of Wikification for Microblogs with Context Expansion</i>	
Taylor Cassidy, Heng Ji, Lev-Arie Ratinov, Arkaitz Zubiaga and Hongzhao Huang ...	441
<i>On the Effectiveness of using Sentence Compression Models for Query-Focused Multi-Document Summarization</i>	
Yllias Chali and Sadid A. Hasan	457
<i>Towards Automatic Topical Question Generation</i>	
Yllias Chali and Sadid A. Hasan	475
<i>Adjective Deletion for Linguistic Steganography and Secret Sharing</i>	
Ching-Yun Chang and Stephen Clark	493

<i>The Secret's in the Word Order: Text-to-Text Generation for Linguistic Steganography</i> Ching-Yun Chang and Stephen Clark.....	511
<i>Joint Modeling for Chinese Event Extraction with Rich Linguistic Features</i> Chen Chen and Vincent Ng.....	529
<i>A Simplification-Translation-Restoration Framework for Cross-Domain SMT Applications</i> Han-Bin Chen, Hen-Hsen Huang, Hsin-Hsi Chen and Ching-Ting Tan.....	545
<i>A Semi-Supervised Bayesian Network Model for Microblog Topic Classification</i> Yan Chen, Zhoujun Li, Liqiang Nie, Xia Hu, Xiangyu Wang, Tat-Seng Chua and Xiaoming Zhang.....	561
<i>A System for Multilingual Sentiment Learning On Large Data Sets</i> Alex Cheng and Oles Zhulyn.....	577
<i>Extraction of Russian Sentiment Lexicon for Product Meta-Domain</i> Iliia Chetviorkin and Natalia Loukachevitch.....	593
<i>Problems in Evaluating Grammatical Error Detection Systems</i> Martin Chodorow, Markus Dickinson, Ross Israel and Joel Tetreault.....	611
<i>Unsupervised and Semi-Supervised Morphological Analysis for Information Retrieval in the Biomedical Domain</i> Vincent Claveau.....	629
<i>A Hybrid Approach to Finding Phenotype Candidates in Genetic Texts</i> Nigel Collier, Mai-Vu Tran, Hoang-Quynh Le, Anika Oellrich, Ai Kawazoe, Martin Hall-May and Dietrich Rebholz-Schuhmann.....	647
<i>Using Argumentative Zones for Extractive Summarization of Scientific Articles</i> Danish Contractor, Yufan Guo and Anna Korhonen.....	663
<i>Annotation Tools and Knowledge Representation for a Text-To-Scene System</i> Bob Coyne, Alex Klapheke, Masoud Rouhizadeh, Richard Sproat and Daniel Bauer.....	679
<i>Towards Efficient HPSG Generation for German, a Non-Configurational Language</i> Berthold Crysmann and Woodley Packard.....	695
<i>A Corpus-Based Study of Edit Categories in Featured and Non-Featured Wikipedia Articles</i> Johannes Daxenberger and Iryna Gurevych.....	711
<i>A Computational Cognitive Model for Semantic Sub-Network Extraction from Natural Language Queries</i> Suman Deb Roy and Wenjun Zeng.....	727
<i>Extraction of Domain-Specific Bilingual Lexicon from Comparable Corpora: Compositional Translation and Ranking</i> Estelle Delpesch, Béatrice Daille, Emmanuel Morin and Claire Lemaire.....	745
<i>Twitter Topic Summarization by Ranking Tweets using Social Influence and Content Quality</i> Yajuan Duan, Zhumin Chen, Furu Wei, Ming Zhou and Heung-Yeung Shum.....	763
<i>S-Restricted Monotone Alignments: Algorithm, Search Space, and Applications</i> Steffen Eger.....	781

<i>Mining Words in the Minds of Second Language Learners: Learner-Specific Word Difficulty</i> Yo Ehara, Issei Sato, Hidekazu Oiwa and Hiroshi Nakagawa	799
<i>Jointly Disambiguating and Clustering Concepts and Entities with Markov Logic</i> Angela Fahrni and Michael Strube	815
<i>Flexible Structural Analysis of Near-Meet-Semilattices for Typed Unification-Based Grammar Design</i> Rouzbeh Farahmand and Gerald Penn	833
<i>Stacking of Dependency and Phrase Structure Parsers</i> Richárd Farkas and Bernd Bohnet	849
<i>Semantic Cohesion Model for Phrase-Based SMT</i> Minwei Feng, Weiwei Sun and Hermann Ney	867
<i>Comparing Taxonomies for Organising Collections of Documents</i> Samuel Fernando, Mark Hall, Eneko Agirre, Aitor Soroa, Paul Clough and Mark Stevenson	879
<i>Modeling the Complexity of Manual Annotation Tasks: a Grid of Analysis</i> Karèn Fort, Adeline Nazarenko and Sophie Rosset	895
<i>Extractive Multi-Document Summarization with Integer Linear Programming and Support Vector Regression</i> Dimitrios Galanis, Gerasimos Lampouras and Ion Androutsopoulos	911
<i>Cross-Lingual Topical Relevance Models</i> Debasis Ganguly, Johannes Leveling and Gareth Jones	927
<i>Structured Term Recognition in Medical Text</i> Michael Glass and Alfio Gliozzo	943
<i>A Dynamic Oracle for Arc-Eager Dependency Parsing</i> Yoav Goldberg and Joakim Nivre	959
<i>Statistical Mechanical Analysis of Semantic Orientations on Lexical Network</i> Takuma Goto, Yoshiyuki Kabashima and Hiroya Takamura	977
<i>Finding Thoughtful Comments from Social Media</i> Swapna Gottipati and Jing Jiang	995
<i>A Distributed Platform for Sanskrit Processing</i> Pawan Goyal, Gérard Huet, Amba Kulkarni, Peter Scharf and Ralph Bunker	1011
<i>Understanding the Performance of Statistical MT Systems: A Linear Regression Framework</i> Francisco Guzman and Stephan Vogel	1029
<i>Geolocation Prediction in Social Media Data by Finding Location Indicative Words</i> Bo Han, Paul Cook and Timothy Baldwin	1045
<i>Readability Classification for German using Lexical, Syntactic, and Morphological Features</i> Julia Hancke, Sowmya Vajjala and Detmar Meurers	1063

<i>Walk-based Computation of Contextual Word Similarity</i>	
Kazuo Hara, Ikumi Suzuki, Masashi Shimbo and Yuji Matsumoto	1081
<i>Flexible Japanese Sentence Compression by Relaxing Unit Constraints</i>	
Jun Harashima and Sadao Kurohashi	1097
<i>Approximating Theoretical Linguistics Classification in Real Data: the Case of German “nach” Particle Verbs</i>	
Boris Haselbach, Kerstin Eckart, Wolfgang Seeker, Kurt Eberle and Ulrich Heid	1113
<i>Bridging the Gap between Intrinsic and Perceived Relevance in Snippet Generation</i>	
Jing He, Pablo Duboue and Jian-Yun Nie	1129
<i>A Comparison and Improvement of Online Learning Algorithms for Sequence Labeling</i>	
Zhengyan He and Houfeng Wang	1147
<i>Creating an Extended Named Entity Dictionary from Wikipedia</i>	
Ryuichiro Higashinaka, Kugatsu Sadamitsu, Kuniko Saito, Toshiro Makino and Yoshihiro Matsuo	1163
<i>Statistical Method of Building Dialect Language Models for ASR Systems</i>	
Naoki Hirayama, Shinsuke Mori and Hiroshi G. Okuno	1179
<i>Tailored Feature Extraction for Lexical Disambiguation of English Verbs Based on Corpus Pattern Analysis</i>	
Martin Holub, Vincent Kríž, Silvie Cinková and Eckhard Bick	1195
<i>Method Mention Extraction from Scientific Research Papers</i>	
Hospice Houngho and Robert E. Mercer	1211
<i>Context-Enhanced Personalized Social Summarization</i>	
Po Hu, Donghong Ji, Chong Teng and Yujing Guo	1223
<i>Tweet Ranking Based on Heterogeneous Networks</i>	
Hongzhao Huang, Arkaitz Zubiaga, Heng Ji, Hongbo Deng, Dong Wang, Hieu Le, Tarek Abdelzaher, Jiawei Han, Alice Leung, John Hancock and Clare Voss	1239
<i>Improved Combinatory Categorical Grammar Induction with Boundary Words and Bayesian Inference</i>	
Yun Huang, Min Zhang and Chew-Lim Tan	1257
<i>Mining Rules for Rewriting States in a Transition-based Dependency Parser for English</i>	
Akihiro Inokuchi and Ayumu Yamaoka	1275
<i>Coreference Resolution with ILP-based Weighted Abduction</i>	
Naoya Inoue, Ekaterina Ovchinnikova, Kentaro Inui and Jerry Hobbs	1291
<i>N-gram Fragment Sequence Based Unsupervised Domain-Specific Document Readability</i>	
Shoaib Jameel, Xiaojun Qian and Wai Lam	1309
<i>Using Knowledge and Constraints To Find the Best Antecedent</i>	
Prateek Jindal and Dan Roth	1327
<i>Towards a Generic and Flexible Citation Classifier Based on a Faceted Classification Scheme</i>	
Charles Jochim and Hinrich Schütze	1343

<i>Semantics-Based Machine Translation with Hyperedge Replacement Grammars</i>	
Bevan Jones, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann and Kevin Knight	1359
<i>Answering Yes/No Questions via Question Inversion</i>	
Hiroshi Kanayama, Yusuke Miyao and John Prager	1377
<i>Improving Topic Classification for Highly Inflective Languages</i>	
Jurgita Kapociute-Dzikiene, Frederik Vaassen, Walter Daelemans and Algis Krupavičius	1393
<i>Generating “A for Alpha” When There Are Thousands of Characters</i>	
Hiroaki Kawasaki, Ryohei Sasano, Hiroya Takamura and Manabu Okumura	1411
<i>A Machine Learning Approach for Phenotype Name Recognition</i>	
Maryam Khordad, Robert E Mercer and Peter Rogan	1425
<i>Improving Combinatory Categorical Grammar Parse Reranking with Dependency Grammar Features</i>	
Sunghwan Mac Kim, Dominick Ng, Mark Johnson and James Curran	1441
<i>Inducing Crosslingual Distributed Representations of Words</i>	
Alexandre Klementiev, Ivan Titov and Binod Bhattacharai	1459
<i>Exploring Local and Global Semantic Information for Event Pronoun Resolution</i>	
Fang Kong and Guodong Zhou	1475
<i>Semantic Processing of Compounds in Indian Languages</i>	
Amba Kulkarni, Soma Paul, Malhar Kulkarni, Anil Kumar and Nitesh Surtani	1489
<i>Unsupervised Japanese-Chinese Opinion Word Translation using Dependency Distance and Feature-Opinion Association Weight</i>	
Guo-Hau Lai, Ying-Mei Guo and Richard Tzong-Han Tsai	1503
<i>On-line Trend Analysis with Topic Models: #twitter Trends Detection Topic Model Online</i>	
Jey Han Lau, Nigel Collier and Timothy Baldwin	1519
<i>Learning Compositional Semantics for Open Domain Semantic Parsing</i>	
Phong Le and Willem Zuidema	1535
<i>Evaluating Different Methods for Automatically Collecting Large General Corpora for Basque from the Web</i>	
Igor Leturia	1553
<i>Approximate Sentence Retrieval for Scalable and Efficient Example-Based Machine Translation</i>	
Johannes Leveling, Debasis Ganguly, Sandipan Dandapat and Gareth Jones	1571
<i>Improving Text Normalization using Character-Blocks Based Models and System Combination</i>	
Chen Li and Yang Liu	1587
<i>Update Summarization using a Multi-level Hierarchical Dirichlet Process Model</i>	
Jiwei Li, Sujian Li, Xun Wang, Ye Tian and Baobao Chang	1603
<i>Employing Morphological Structures and Sememes for Chinese Event Extraction</i>	
Peifeng Li and Guodong Zhou	1619

<i>Joint Modeling of Trigger Identification and Event Type Determination in Chinese Event Extraction</i> Peifeng Li, Qiaoming Zhu, Hongjun Diao and Guodong Zhou	1635
<i>Integrating Surface and Abstract Features for Robust Cross-Domain Chinese Word Segmentation</i> Xiaoqing Li, Kun Wang, Chengqing Zong and Keh-Yih Su	1653
<i>Code-Switch Language Model with Inversion Constraints for Mixed Language Speech Recognition</i> Ying Li and Pascale Fung	1671
<i>A Separately Passive-Aggressive Training Algorithm for Joint POS Tagging and Dependency Parsing</i> Zhenghua Li, Min Zhang, Wanxiang Che and Ting Liu	1681
<i>Graph-Based Multi-Tweet Summarization using Social Signals</i> Xiaohua Liu, Yitong Li, Furu Wei and Ming Zhou	1699
<i>Topical Word Trigger Model for Keyphrase Extraction</i> Zhiyuan Liu, Chen Liang and Maosong Sun	1715
<i>Easy-First Chinese POS Tagging and Dependency Parsing</i> Ji Ma, Tong Xiao, Jingbo Zhu and Feiliang Ren	1731
<i>Recognizing Personal Characteristics of Readers using Eye-Movements and Text Features</i> Pascual Martínez-Gómez, Tadayoshi Hara and Akiko Aizawa	1747
<i>To Exhibit is not to Loiter: A Multilingual, Sense-Disambiguated Wiktionary for Measuring Verb Similarity</i> Christian M. Meyer and Iryna Gurevych	1763
<i>Using Distributional Similarity for Lexical Expansion in Knowledge-based Word Sense Disambiguation</i> Tristan Miller, Chris Biemann, Torsten Zesch and Iryna Gurevych	1781
<i>Revising the Compositional Method for Terminology Acquisition from Comparable Corpora</i> Emmanuel Morin and Béatrice Daille	1797
<i>Is Bad Structure Better Than No Structure?: Unsupervised Parsing for Realisation Ranking</i> Yasaman Motazed, Mark Dras and François Lareau	1811
<i>Analysis of Linguistic Style Accommodation in Online Debates</i> Arjun Mukherjee and Bing Liu	1831
<i>Sentiment Analysis in Twitter with Lightweight Discourse Analysis</i> Subhabrata Mukherjee and Pushpak Bhattacharyya	1847
<i>YouCat: Weakly Supervised Youtube Video Categorization System from Meta Data & User Comments using WordNet & Wikipedia</i> Subhabrata Mukherjee and Pushpak Bhattacharyya	1865
<i>Constrained Decoding for Text-Level Discourse Parsing</i> Philippe Muller, Stergos Afantenos, Pascal Denis and Nicholas Asher	1883
<i>Incremental Learning of Affix Segmentation</i> Wondwossen Mulugeta, Michael Gasser and Baye Yimam	1901

<i>Semi-Supervised Noun Compound Analysis with Edge and Span Features</i> Yugo Murawaki and Sadao Kurohashi	1915
<i>Learning Effective and Interpretable Semantic Models using Non-Negative Sparse Embedding</i> Brian Murphy, Partha Talukdar and Tom Mitchell	1933
<i>Combining Wordnet and Morphosyntactic Information in Terminology Clustering</i> Agnieszka Mykowiecka and Malgorzata Marciniak	1951
<i>Alignment by Bilingual Generation and Monolingual Derivation</i> Toshiaki Nakazawa and Sadao Kurohashi	1963
<i>Optimizing for Sentence-Level BLEU+1 Yields Short Translations</i> Preslav Nakov, Francisco Guzman and Stephan Vogel	1979
<i>Grammarless Parsing for Joint Inference</i> Jason Naradowsky, Tim Vieira and David Smith	1995
<i>Error Mining with Suspicion Trees: Seeing the Forest for the Trees</i> Shashi Narayan and Claire Gardent	2011
<i>Structure-Driven Lexicalist Generation</i> Shashi Narayan and Claire Gardent	2027
<i>A Comparison of Syntactic Reordering Methods for English-German Machine Translation</i> Jiri Navratil, Karthik Visweswariah and Ananthkrishnan Ramanathan	2043
<i>Grounded Language Acquisition: A Minimal Commitment Approach</i> Sushobhan Nayak and Amitabha Mukerjee	2059
<i>Bayesian Text Segmentation for Index Term Identification and Keyphrase Extraction</i> David Newman, Nagendra Koilada, Jey Han Lau and Timothy Baldwin	2077
<i>Exploiting Category-Specific Information for Multi-Document Summarization</i> Jun-Ping Ng, Praveen Bysani, Ziheng Lin, Min-Yen Kan and Chew-Lim Tan	2093
<i>Improved Temporal Relation Classification using Dependency Parses and Selective Crowdsourced Annotations</i> Jun-Ping Ng and Min-Yen Kan	2109
<i>Accurate Unbounded Dependency Recovery using Generalized Categorical Grammars</i> Luan Nguyen, Marten Van Schijndel and William Schuler	2125
<i>Tibetan Base Noun Phrase Identification Framework Based on Chinese-Tibetan Sentence Aligned Corpus</i> Ming Hua Nuo, Hui Dan Liu, Wei Na Zhao, Long Long Ma, Jian Wu and Zhi Ming Ding ..	2141
<i>A Pipeline Arabic Named Entity Recognition using a Hybrid Approach</i> Mai Oudah and Khaled Shaalan	2159
<i>Attribute Extraction from Conjectural Queries</i> Marius Pasca	2177

<i>A Comprehensive Analysis of Constituent Coordination for Grammar Engineering</i> Agnieszka Patejuk and Adam Przepiórkowski	2191
<i>Simple and Effective Parameter Tuning for Domain Adaptation of Statistical Machine Translation</i> Pavel Pecina, Antonio Toral and Josef van Genabith	2209
<i>A Supervised Aggregation Framework for Multi-Document Summarization</i> Yulong Pei, Wenpeng Yin, Qifeng Fan and Lian'en Huang	2225
<i>Collective Search for Concept Disambiguation</i> Anja Pilz and Gerhard Paaß	2243
<i>Who's (Really) the Boss? Perception of Situational Power in Written Interactions</i> Vinodkumar Prabhakaran, Owen Rambow and Mona Diab	2259
<i>Bilingual Lexicon Construction from Comparable Corpora via Dependency Mapping</i> Longhua Qian, Hongling Wang, Guodong Zhou and Qiaoming Zhu	2275
<i>A MWE Acquisition and Lexicon Builder Web Service</i> Valeria Quochi, Francesca Frontini and Francesco Rubino	2291
<i>A Diverse Dirichlet Process Ensemble for Unsupervised Induction of Syntactic Categories</i> Roi Reichart, Gal Elidan and Ari Rappoport	2307
<i>From Finite-State to Inversion Transductions: Toward Unsupervised Bilingual Grammar Induction</i> Markus Saers, Karteek Addanki and Dekai Wu	2325
<i>Underspecified Query Refinement via Natural Language Question Generation</i> Hassan Sajjad, Patrick Pantel and Michael Gamon	2341
<i>Joint English Spelling Error Correction and POS Tagging for Language Learners Writing</i> Keisuke Sakaguchi, Tomoya Mizumoto, Mamoru Komachi and Yuji Matsumoto	2357
<i>Automatic Detection of Psychological Distress Indicators and Severity Assessment from Online Forum Posts</i> Shirin Saleem, Rohit Prasad, Shiv Vitaladevuni, Maciej Pacula, Michael Crystal, Brian Marx, Denise Sloan, Jennifer Vasterling and Theodore Speroff	2375
<i>Ant Colony Algorithm for the Unsupervised Word Sense Disambiguation of Texts: Comparison and Evaluation</i> Didier Schwab, Jérôme Goulian, Andon Tchechmedjiev and Hervé Blanchon	2389
<i>Learnability-Based Syntactic Annotation Design</i> Roy Schwartz, Omri Abend and Ari Rappoport	2405
<i>Improving Supervised Sense Disambiguation with Web-Scale Selectors</i> H. Andrew Schwartz, Fernando Gomez and Lyle Ungar	2423
<i>The French Social Media Bank: a Treebank of Noisy User Generated Content</i> Djamé Seddah, Benoit Sagot, Marie Candito, Virginie Mouilleron and Vanessa Combet ...	2441
<i>Initial Explorations on using CRFs for Turkish Named Entity Recognition</i> Gökhan Akin Şeker and Gülşen Eryiğit	2459

<i>Differential Evolution Based Feature Selection and Classifier Ensemble for Named Entity Recognition</i>	
Utpal Kumar Sikdar, Asif Ekbal and Sriparna Saha	2475
<i>Noun Group and Verb Group Identification for Hindi</i>	
Smriti Singh, Om P. Damani and Vaijayanthi M. Sarma	2491
<i>Named Entity Recognition System for Urdu</i>	
UmriinderPal Singh, Vishal Goyal and Gurpreet Singh Lehal	2507
<i>Easy-first Coreference Resolution</i>	
Veselin Stoyanov and Jason Eisner	2519
<i>Modeling Leadership and Influence in Multi-party Online Discourse</i>	
Tomek Strzalkowski, Samira Shaikh, Ting Liu, George Aaron Broadwell, Jenny Stromer-Galley, Sarah Taylor, Umit Boz, Veena Ravishankar and Xiaoi Ren	2535
<i>NEER: An Unsupervised Method for Named Entity Evolution Recognition</i>	
Nina Tahmasebi, Gerhard Gossen, Nattiya Kanhabua, Helge Holzmann and Thomas Risse	2553
<i>Evaluating the Translation Accuracy of a Novel Language-Independent MT Methodology</i>	
George Tambouratzis, Sokratis Sofianopoulos and Marina Vassiliou	2569
<i>Native Tongues, Lost and Found: Resources and Empirical Evaluations in Native Language Identification</i>	
Joel Tetreault, Daniel Blanchard, Aoife Cahill and Martin Chodorow	2585
<i>Inverse Document Density: A Smooth Measure for Location-Dependent Term Irregularities</i>	
Dennis Thom, Harald Bosch and Thomas Ertl	2603
<i>Efficient Discrimination Between Closely Related Languages</i>	
Jörg Tiedemann and Nikola Ljubešić	2619
<i>Semi-Supervised Semantic Role Labeling: Approaching from an Unsupervised Perspective</i>	
Ivan Titov and Alexandre Klementiev	2635
<i>Hunting for Entailing Pairs in the Penn Discourse Treebank</i>	
Sara Tonelli and Elena Cabrio	2653
<i>Implicitness of Discourse Relations</i>	
Fatemeh Torabi Asr and Vera Demberg	2669
<i>Combining Statistical Translation Techniques for Cross-Language Information Retrieval</i>	
Ferhan Ture, Jimmy Lin and Douglas Oard	2685
<i>Multi-way Tensor Factorization for Unsupervised Lexical Acquisition</i>	
Tim van de Cruys, Laura Rimell, Thierry Poibeau and Anna Korhonen	2703
<i>Sub-corpora Sampling with an Application to Bilingual Lexicon Extraction</i>	
Ivan Vulić and Marie-Francine Moens	2721
<i>The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums</i>	
Li Wang, Su Nam Kim and Timothy Baldwin	2739

<i>Implicit Discourse Relation Recognition by Selecting Typical Training Examples</i> Xun Wang, Sujian Li, Jiwei Li and Wenjie Li	2757
<i>Chinese Evaluative Information Analysis</i> You Wang, Jun'ichi Kazama, Takuya Kawada and Kentaro Torisawa	2773
<i>Harnessing the CRF Complexity with Domain-Specific Constraints. The Case of Morphosyntactic Tagging of a Highly Inflected Language</i> Jakub Waszczuk	2789
<i>A Latent Discriminative Model for Compositional Entailment Relation Recognition using Natural Logic</i> Yotaro Watanabe, Junta Mizuno, Eric Nichols, Naoaki Okazaki and Kentaro Inui... ..	2805
<i>Strategies for Mixed-Initiative Conversation Management using Question-Answer Pairs</i> Wilson Wong, Lawrence Cavedon, John Thangarajah and Lin Padgham	2821
<i>Factored Language Model based on Recurrent Neural Network</i> Youzheng Wu, Xugang Lu, Hitoshi Yamamoto, Shigeki Matsuda, Chiori Hori and Hideki Kashioka	2835
<i>Multi-View AdaBoost for Multilingual Subjectivity Analysis</i> Min Xiao and Yuhong Guo	2851
<i>Semi-supervised Representation Learning for Domain Adaptation using Dynamic Dependency Networks</i> Min Xiao, Yuhong Guo and Alexander Yates	2867
<i>Unsupervised Discriminative Induction of Synchronous Grammar for Machine Translation</i> Xinyan Xiao, Deyi Xiong, Yang Liu, Qun Liu and Shouxun Lin	2883
<i>Paraphrasing for Style</i> Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman and Colin Cherry	2899
<i>Modeling ESL Word Choice Similarities By Representing Word Intensions and Extensions</i> Huichao Xue and Rebecca Hwa	2915
<i>ISO-TimeML Event Extraction in Persian Text</i> Yadollah Yaghoobzadeh, Gholamreza Ghassem-Sani, Seyed Abolghassem Mirroshandel and Mahbaneh Eshaghzadeh	2931
<i>Measuring the Similarity between TV Programs using Semantic Relations</i> Ichiro Yamada, Masaru Miyazaki, Hideki Sumiyoshi, Atsushi Matsui, Hironori Furumiya and Hideki Tanaka	2945
<i>RelationListwise for Query-Focused Multi-Document Summarization</i> Wenpeng Yin, Lifu Huang, Yulong Pei and Lian'en Huang	2961
<i>SentTopic-MultiRank: a Novel Ranking Model for Multi-Document Summarization</i> Wenpeng Yin, Yulong Pei, Fan Zhang and Lian'en Huang	2977
<i>Language Modeling for Spoken Dialogue System based on Filtering using Predicate-Argument Structures</i> Koichiro Yoshino, Shinsuke Mori and Tatsuya Kawahara	2993

<i>Detecting Word Ordering Errors in Chinese Sentences for Learning Chinese as a Foreign Language</i> Chi-Hsin Yu and Hsin-Hsi Chen	3003
<i>Machine Translation by Modeling Predicate-Argument Structure Transformation</i> Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong	3019
<i>Tree-based Translation without using Parse Trees</i> Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong	3037
<i>Constructing Chinese Abbreviation Dictionary: A Stacked Approach</i> Longkai Zhang, Sujian Li, Houfeng Wang, Ni Sun and Xinfan Meng	3055
<i>Stacking Heterogeneous Joint Models of Chinese POS Tagging and Dependency Parsing</i> Meishan Zhang, Wanxiang Che, Ting Liu and Zhenghua Li	3071
<i>A Lazy Learning Model for Entity Linking using Query-Specific Information</i> Wei Zhang, Jian Su, Chew-Lim Tan, Yunbo Cao and Chin-Yew Lin	3089
<i>The Use of Dependency Relation Graph to Enhance the Term Weighting in Question Retrieval</i> Weinan Zhang, Zhaoyan Ming, Yu Zhang, Liqiang Nie, Ting Liu and Tat-Seng Chua	3105
<i>Long-Tail Distributions and Unsupervised Learning of Morphology</i> Qiuye Zhao and Mitch Marcus	3121
<i>User Behaviors Lend a Helping Hand: Learning Paraphrase Query Patterns from Search Log Sessions</i> Shiqi Zhao, Haifeng Wang and Ting Liu	3137
<i>Exploiting Bilingual Translation for Question Retrieval in Community-Based Question Answering</i> Guangyou Zhou, Kang Liu and Jun Zhao	3153
<i>Exploiting Lexical Dependencies from Large-Scale Data for Better Shift-Reduce Constituency Parsing</i> Muhua Zhu, Jingbo Zhu and Huizhen Wang	3171

COLING 2012

**24th International Conference on
Computational Linguistics**

Invited Speaker Abstracts

Minimum Description Length as the basis of Panini's grammar Prof. Paul Kiparsky

*Robert M. and Anne T. Bass Professor in the School of Humanities and Sciences
Department of Linguistics, Stanford University*

Abstract

Panini attempted, and to a considerable extent succeeded, in constructing the shortest possible complete grammar of Sanskrit that contains a description of its own metalanguage. Minimizing the total length of the grammar required introducing a rule or convention just in case it achieves overall economies in the grammar which outweigh the cost of stating it.

The grammar presupposes nothing beyond certain elementary relations (such as “before” and “after”) and operations (such as “replace”). Based on them, it defines a rich descriptive formalism. Simplicity dictates the inclusion of rules of grammar that encode all generalizations about Sanskrit phonology, morphology, and syntax, as well as of rules that define its grammatical categories, and of metarules that stipulate how rules of grammar apply and interact with each other. The grammar uses a fixed rule format, phonological and morphosyntactic features, rule ordering, cyclicity, blocking, the equivalent of Theta roles, inheritance hierarchies, and several hundred technical terms denoting classes of lexemes and morphemes. Four levels of representation (approximately corresponding to semantics, syntax, morphology, and phonology) emerge from the analysis.

Completeness of empirical coverage requires, among other things, the exhaustive treatment of derivational morphology, clausal syntax, variation (three degrees of optionality), and even of certain dialectal and sociolinguistic facts. Some of the abovementioned devices could be dispensed with in a less exhaustive description, as I will illustrate with the *karaka* system.

The grammar appears to be very nearly optimal. Although this has not been proved (except for certain subsystems), it appears likely, for no-one has been able to shorten the grammar in non-trivial ways (without losing content), either by modifying the rules without changing the metalanguage, or by modifying the metalanguage with additional devices and conventions, or removing some of the existing ones.

It would be anachronistic to construe the formal apparatus used by Panini as embodying a “theory”: from his perspective it merely serves to compress the grammar. But the fact that many of the same conventions and principles that modern generative grammar posits as universals of language emerge just from the attempt to construct the maximally compact description of a single language is quite remarkable. It could

be taken as a challenge to the widespread assumption that learners are innately equipped with a format for grammatical description. For if a generative grammar can be arrived at purely by minimizing description length, without relying on any further prior assumptions, might not language acquisition by humans proceed in the same way? I will argue that this is not the case, because there is no effective procedure for constructing the maximally compact representation without prior analytic bias. In fact, the near-perfection of Panini's grammar and its metalanguage required hand-crafting by many generations of grammarians. In contrast, the rapidity of normal language acquisition, and the existence of robust cross-linguistic generalizations, remains a persuasive argument for UG.

**The adaptive brain: acquiring a complex cognitive skill
in complex contexts**

Prof. Barbara Moser-Mercer

Director, Department of Interpretation

Faculty of Translation and Interpretation, University of Geneva

Abstract

Real-time human communication across language barriers relies on consecutive and simultaneous interpretation, a complex cognitive skill that can be acquired only over a certain period of time. Interpreting novices differ from interpreting experts in terms of their knowledge and knowledge organization, their analytical strategies, their use of memory processes, and the smoothness and speed with which they execute the interpreting task. In order to be able to move from comprehending a speech in one language and simultaneously interpreting that speech into another, the learner needs to make considerable adaptations to component processes of tasks already mastered, for the most part, before even being admitted to an interpreter training program. These adaptations concern mostly language comprehension and knowledge organization, component skills non-interpreters need in order to communicate. One must thus assume that significant changes occur in brain activity (functional changes or plasticity) and brain structure (structural plasticity) during the acquisition of interpreting skills that are the result of learning, knowledge re-organization, strategy acquisition, and task monitoring.

In investigating this hypothesis we have recently found evidence for brain structural plasticity in individuals training to become simultaneous interpreters as they develop expertise in this skill. We found that in interpreting students, but not in matched multilingual controls, there is an increase in gray matter volume over the course of a 15-month training program in brain regions known to be involved not only in semantic processing but also in aspects of executive function and error monitoring. Tasks involving the conversion of content from one language to another (i.e., translation and interpretation) mainly engage a left-lateralized cortico-subcortical circuit, including the basal ganglia, inferior frontal gyrus, and DLPFC. There is strong anatomical support for functional links between these regions. We propose that the evidence suggests the presence of two distinct networks contributing to the executive control of language. Although perturbing either may have superficially similar behavioral consequences, they are likely to have differing roles.

The plasticity of the brain allows for reshaping and reorganization, acquiring expertise in a task involves the generation of new neuronal connections whose survival is dependent on stimulation through extended electrical pulses that reverberate in the neural net to establish associations and connections between areas of specialized

information. The interpreter is constantly establishing logical connections on-line between what has already been comprehended and what is being heard, relying on multiple associations that have been formed while preparing for a specific assignment. The interpreter's ability to link new to existing information is thus one of the prime skills to be developed during skill acquisition: being able to associate multiple facets of data in neural networks with only a single sensory trigger firing up the entire network of associated facts emerges as one of the most important factors for successful skill acquisition. Fluency and speeded expert performance develop in a learning environment that is highly contextualized and provides multiple exposures to information so that the task can be executed efficiently as the multilingual brain adapts to complex demands.

Our understanding of skill acquisition in interpreting then informs our pedagogical approach and allows us to design learning environments for even the most extreme contexts, with socio-cultural environments characterized by political instability and conflict, where skills need to be acquired swiftly and reliably.

Digital Book, Digital Library, and Natural Language Processing

Prof. Makoto Nagao

Professor Emeritus, Kyoto University

Former President, National Diet (Congress) Library, Japan

Abstract

The following topics will be discussed with the author's experience with natural language processing and its applications to digital library.

1. Features of forth-coming digital books compared to the present-day paper books
2. Features of digital library which organizes digital books and offers highly sophisticated utilization of knowledge accumulated in digital library. This includes problems in digitization, structuring of a book according to the table of contents, varieties of retrieval methods which extract sections of a book, linking related parts of books as a hypertext structure, etc.
3. Construction of an ideal digital library based on these features
4. Natural language processing technologies which are required for the construction of future digital libraries.

NLP from Paninian Perspective
Prof. Dipti Misra Sharma
Prof. Dipti Misra Sharma

Abstract

Akshar Bharati, for several years now, has been pursuing NLP basing its linguistic models on Paninian grammatical framework. The talk will re-look at how the concepts from Panini's Grammar help in selecting and modelling linguistically informed NLP (both building resources and systems). (Akshar Bharati et al, 1995) proposed Computational Paninian Grammar models for various levels of linguistic analysis. It is noticed that it works well for relatively free word order languages. Not only, the insights from Panini's grammar help in exploiting morphological properties in computationally efficient parsing but also help in the selection of appropriate features for better machine learning.

Panini's grammar focusses on how language is used for communication. Thus, language is viewed as a system which encodes information. There are three major schools of thought in the Indian grammatical tradition, the grammarians, the logicians and the text analysts. All of these schools lay emphasis on interpretation of meaning (Shaabdabodha) from what is given in a sentence. The grammar provides ways of identification of various linguistic units, their generation processes, relations across units and the syntactic realization of these relations. The talk will look at how a principled application of the concepts and the methods given in this tradition help in developing efficient computational models.

Most parsing approaches in NLP adopt either a constituency based grammar model or a dependency based one. Conversion from one to the other, combining constituency and dependency representation and producing a hybrid tree are some of the areas that the scholars in NLP have been looking at in the direction of bringing the two approaches together. However, Paninian approach suggests that languages encode information both ways. Thus, the talk will also explore whether both constituency and dependency can be incorporated in a single model and whether this would lead to better parsing.

COLING 2012

**24th International Conference on
Computational Linguistics**

Technical Papers

