COLING 2014

**The 25th International Conference
on Computational Linguistics**

Proceedings of COLING 2014: Tutorial Abstracts

August 23-29, 2014
Dublin, Ireland

Papers marked with a Creative Commons or other specific license statement are copyright the respective authors (or their employers).

Preface

This volume contains papers from the 25th International Conference on Computational Linguistics (Coling 2014) held in Dublin, Ireland. The conference is organized by the Centre for Global Intelligent Content (CNGL) and held at the Helix Conference Centre at Dublin City University (DCU) from 25–29 August 2014, under the auspices of the International Committee on Computational Linguistics (ICCL).

COLING is almost 50 years old, its first gathering having taken place in New York in 1965. It has been organized once every two years, initially in odd years and then in even years, after COLING 1976 in Ottawa. Throughout its long history, COLING's aspiration to provide an amicable forum for participants with broad backgrounds to present and share their ideas remains the same. We believe that the inherent complexity of language is worthy of study from diverse perspectives and that COLING provides a venue for fruitful interdisciplinary interaction.

We accepted 217 papers (138 oral presentations and 79 poster presentations) from 685 effective submissions, having received 705 submissions in total. Regardless of the format of presentation, all of the accepted papers were allocated 12 pages in the proceedings.

The review process of a large conference such as COLING is always complex and occasionally encounters difficulties. The program committee has to cope with the challenges of selecting which papers to accept among a large quantity of high quality submissions. The task of choosing 217 papers from 685 strong submissions covering the ever broadening fields of computational linguistics was not an easy one.

To cope with the anticipated difficulties, we asked six senior colleagues to join the Scientific Advisory Board (SAB) and help us through all stages of reviewing papers. They are: Ralph Grishman (New York University, USA), Yuji Matsumoto (NAIST, Japan), Joakim Nivre (Uppsala Univ., Sweden), Michael Picheny (IBM TJ Watson Research Center, USA), Donia Scott (Univ. of Sussex, UK), and Chengqing Zong (CAS, China).

We had 20 thematic areas and each area was chaired by two or more area chairs. Thanks to over 800 responsive reviewers, the review process proceeded in a very smooth manner, and each paper was read at least by three reviewers. In some cases, papers and their reviews were carefully assessed by Area Co-Chairs, one of the SAB members and by us, in our roles as Program Committee Co-Chairs. We are extremely happy with the very strong set of papers that has been accepted for presentation at the conference. It is, however, with regret that we had no choice but to reject a large number of high quality papers, due to the sheer volume of submissions received.

We would like to thank the SAB members and the Program Committee Area Chairs for their dedicated and efficient review work, and our reviewers for their professionalism in delivering high quality reviews. We also thank the authors of all the papers for submitting their fruits of labour to COLING. Although we were only able to accept a small subset of the submitted papers, we do hope that all authors and reviewers have benefited from this process of indirect dialogue.

Last but not least, we would like to thank the people who made COLING 2014 and this volume possible. We thank General Chairs, Josef van Genabith (Universität des Saarlandes/DFKI) and Andy Way (CNGL, DCU), and the chairs of the Local Organizing Committee, Cara Green (CNGL, DCU) and John Judge (CNGL/NCLT, DCU), for their tireless work. We are especially grateful to the Publications Chairs, Joachim Wagner (CNGL, DCU), Liadh Kelly (CNGL, DCU) and Lorraine Goeuriot (CNGL, DCU), for their hard work in preparing the proceedings.

Prof. Jan Hajic (Charles University, Czech Republic)
Prof. Junichi Tsujii (Microsoft Research, China)
COLING 2014 Program Committee Co-Chairs
July 8, 2014

Welcome from the General Chairs

We are very pleased indeed to welcome you all to COLING 2014, the 25th International Conference on Computational Linguistics. We are particularly proud that the ICCL selected Dublin City University (DCU) as the location of COLING 2014.

DCU and its National Centre for Language Technology (NCLT) have a long track record in NLP. Unlike India, the previous COLING host country, Ireland is a very small country. A unique feature of the Irish University landscape is that universities team up with industry partners and each other to pool expertise to form large research centres. DCU is a founding member of CNGL, the Centre for Gobal Intelligent Content. COLING 2014 is organised by DCU in partnership with the CNGL, and as General Chairs we are proud to represent both DCU and CNGL.

The conference is taking place at the Helix Conference Centre, a stunning building added to the DCU campus in 2002. DCU is a young, dynamic and ambitious university; since admitting its first students in 1980, DCU has grown in both student numbers and size and now occupies a 72-acre site in Glasnevin, just to the north of Dublin city centre. To date almost 50,000 students have graduated from DCU and are now playing significant roles in enterprise and business globally. Today in 2014, DCU delivers more than 200 programmes to over 12,000 students across its four faculties — Humanities and Social Sciences, Science and Health, Engineering and Computing and DCU Business School. DCU's excellence is recognised internationally and it is ranked among the top-50 young Universities worldwide (QS 'Top 50 under 50' 2013). In the last eight years, DCU has twice been named Sunday Times 'University of the Year'.

At the time of writing, the total number of people registered to attend COLING has exceeded 675. With delegates from 58 countries, COLING 2014 will witness a colourful diversity of language and culture, which is appropriate given that Dublin is known as the localisation capital of the world. Some evidence for this comes from our sponsors, to whom we are extremely grateful: Baidu, eBay, Microsoft, Symantec and Google.

We are very pleased with the programme that has been assembled for you, comprising of four days for the main conference with a total of 138 oral presentations, 79 posters and a special track with 28 demo presentations, two days of workshops and tutorials before the main conference, and other satellite workshops immediately after. 18 topical workshops with a sharp focus on issues of key interest today will be attended by about 191 delegates, and the 6 high-quality tutorials are sure to attract large crowds. Social events include a welcome reception on the evening of 24th August, the conference banquet in the Guinness Storehouse on 26th, and excursions to some beautiful places of interest on 27th.

When DCU was awarded COLING two years ago, our own personal situations were quite different. One of us was away working in the translation industry in the UK, while the other was leading the Science Foundation Ireland and Industry-funded CNGL research center. Over the past few months, we have changed countries, and jobs: Andy is back as Deputy Director of the CNGL's Centre for Intelligent Content, while Josef has moved to Saarbrücken to take up a Chair and a Scientific Directorship at DFKI.

While these changes were taking place, we both had the backing of a remarkable team. The organization of an event on the scale of COLING takes enormous energy, planning and commitment from a large number of individuals. We have assembled a large, competent team of volunteers who are available to assist you while you are here in Dublin. We are sure that all of you participating at COLING — at tutorials, workshops, or the main conference — will enjoy the time you spend here in Ireland, and will look back on the event as one of the most memorable that you attend. Finally, thanks to all of you for coming. We hope you all enjoy the conference, that you benefit from the excellent programme that has been assembled, and that you go away from here having made new friends.

Prof. Josef Van Genabith (Universität des Saarlandes/DFKI, Germany)
Prof. Andy Way (CNGL, DCU, Ireland)

Organisers

General Chairs

Prof. Andy Way, CNGL, Dublin City University, Ireland

Prof. Josef van Genabith, Universität des Saarlandes/DFKI, Germany

Programme Chairs

Prof. Junichi Tsujii, Microsoft Research, China

Prof. Jan Hajic, Charles University, Czech Republic

Workshops Chairs

Dr. Jennifer Foster, CNGL, Dublin City University, Ireland

Prof. Dan Gildea, University of Rochester, USA

Prof. Tim Baldwin, University of Melbourne, Australia

Publication Chairs

Dr. Joachim Wagner, CNGL, Dublin City University, Ireland

Dr. Liadh Kelly, CNGL, Dublin City University, Ireland

Dr. Lorraine Goeuriot, CNGL, Dublin City University, Ireland

Publicity Chairs

Dr. Dorothy Kenny, SALIS, Dublin City University, Ireland

Prof. Seong-Bae Park, Kyungpook National University, Korea

Local Chairs

Dr. Cara Greene, CNGL, Dublin City University, Ireland

Dr. John Judge, CNGL / NCLT, Dublin City University, Ireland

Scientific Advisory Board

Prof. Joakim Nivre, Uppsala Univ., Sweden

Prof. Yuji Matsumoto, NAIST, Japan

Dr. Michael Picheny, IBM TJ Watson Research Center, USA

Prof. Donia Scott, Univ. of Sussex, UK

Prof. Chengqing Zong, CAS, China

Prof. Ralph Grishman, New York University, USA

Tutorial Chairs

Prof. Qun Liu, CNGL, Dublin City University, Ireland

Prof. Fei Xia, University of Washington, USA

Demo Chairs

Dr. Lamia Tounsi, CNGL, Dublin City University, Ireland

Dr. Rafal Rak, NaCTeM, University of Manchester, UK

Sponsorship Chairs

Dr. Páraic Sheridan, CNGL, Dublin City University, Ireland

Prof. Hans Uszkoreit, DFKI, Germany

Dr. Huaping Zhang, Beijing Institute of Technology, China

Dr. Hiromi Nakaiwa, NTT, Japan

Table of Contents

Conference Program

Saturday, 23rd of August 2014

Biomedical/Clinical NLP
Ozlem Uzuner, Meliha Yetisgen and Amber Stubbs

Using Neural Networks for Modeling and Representing Natural Languages
Tomas Mikolov

Sunday, 24th of August 2014

Multilingual Word Sense Disambiguation and Entity Linking
Roberto Navigli and Andrea Moro

Automated Grammatical Error Correction for Language Learners
Joel Tetreault and Claudia Leacock

Selection Bias, Label Bias, and Bias in Ground Truth
Anders Søgaard, Barbara Plank and Dirk Hovy

Dependency Parsing: Past, Present, and Future
Wenliang Chen, Zhenghua Li and Min Zhang