

# Foreword

The European Association for Machine Translation (EAMT) organised its first Workshop/ Conference back in 1996, in Austria. Since then, these events have been held in Denmark, Switzerland, the Czech Republic, Slovenia, the UK, Hungary, Ireland, Malta, Norway, Germany, Spain, and last year in France.

I am very keen as EAMT President to see our conferences take place in new venues, so I am pleased that this year the EAMT annual conference will take place for the first time in one of the Benelux countries.

This is the 15th Annual Meeting of the EAMT, which as an organisation continues to grow and thrive. The numbers of student, individual, institutional and corporate members continue to rise, partly due to improved membership packages, but also because of the range of new initiatives that the Association has recently undertaken, including the Best PhD Thesis Award, the new database version of the MT Compendium, sponsoring research and development activities, an extension of our activities to the MENA region etc.

As last year, I would like to thank my colleagues on the EAMT Committee, who continue to provide me with invaluable support. They work tirelessly on behalf of all of us, and we are very fortunate to have such a strong body of colleagues representing our Association.

In addition to all this, the Conferences that the EAMT has organised have continued to improve in quality, with the result that ever larger audiences have been attracted to our events, to the extent that the annual EAMT Conference is now a must for many protagonists in the field, and not just from Europe. This 15th Conference is no exception, and in particular I would like to thank our co-Chairs Mikel Forcada and Heidi Depraetere for assembling a very attractive programme, comprising of Research and User tracks, poster sessions, and a great choice of Invited Speaker in Fred Hollowood. As last year, a special session has been organised where some of the prominent FP7 projects will be featured; these were only starting up at this point in 2010, but all have made significant progress in the interim, so this should be a really interesting session.

We would not have a programme if it weren't for the co-Chairs and the programme committee who worked so hard to put together the programme before you now, but we would not have a Conference if our local organizer, Vincent Vandeghinste had not volunteered to hold the meeting in the historic city of Leuven. We are very grateful to Vincent and his team for their excellent organization of this event.

Finally, thanks to all of you for coming. I wish all of you an enjoyable conference, that you benefit from the excellent programme that has been assembled, and that you go away from here having made new friends.

Andy Way  
President of the EAMT  
Centre for Next Generation Localisation  
School of Computing  
Dublin City University Ireland  
and  
Applied Language Solutions  
Delph, Saddleworth  
UK

# Message from the Programme Chairs

It is a great pleasure for us to welcome you to the 15th Conference of the European Association for Machine Translation (EAMT) in Leuven. It has been an honor for both of us to have been invited by the EAMT to serve as co-chairs of a conference that has become the yearly reference conference for European machine translation (MT) developers, researchers and users.

As in 2010, the conference has two tracks: a research track, where researchers report about significant research results in any aspect of machine translation and related areas, with a substantial evaluation component, and a user track, where users report their experiences with MT, in business, government, or NGOs.

We received 69 papers: 58 in the research track and 11 in the user track. After double-blind review, generally by three (occasionally two) anonymous reviewers among the best MT experts in the world, 40 of them, that is, 58%, have been accepted and have found their way into the proceedings: 31 research papers (53%) and 9 user papers (82%).

We are in debt with the members of the programme committee and with the secondary reviewers they appointed for some of their papers. As the number of papers received was unusually high, they had an unusually large workload: we especially thank them for their invaluable help, which most of them completed on time (yes, we had to twist some arms after the deadline). We hope that the reviewers' comments were useful and constructive and helped all authors: for those whose papers weren't accepted, by increasing their chance in a later submission somewhere else; for those whose papers got in, to improve their manuscripts. We know we didn't give them a lot of time to do so, and we thank authors for sending their camera-ready versions on time.

We hope that the resulting selection of papers, which have in your hands, truly represents the best of machine translation research, development and real-world usage. Papers come mainly from Europe, but we also have papers with authors from Japan, Canada, Hong Kong or Singapore. The countries with the largest number of authors in accepted papers are: Ireland (15), Spain (14), Germany (12), and France (11).

Of those 40 papers, 25 will be presented as regular 25-minute oral presentations, and 15 as posters. Poster presenters will, for the first time in EAMT-2011, have the opportunity to get in the spotlight and use two minutes to try and convince attendees to visit their poster during a "poster booster" oral session.

Both oral and poster papers have the same format and extension in the proceedings. To facilitate the dissemination of these results, we have chosen an open license (Creative Commons - Attribution - NonDerivatives) for papers, which can now be freely distributed in their Proceedings form.

As in last year's edition, international projects and initiatives with European funding will also have the chance to present themselves in a poster session, preceded by a "poster booster" session too.

As an opener, we will enjoy an invited talk by Fred Hollowood, of Symantec, "A Tale of Two Technologies"; we hope it will appeal to both our research and our user audience. To close the conference, we will have a presentation by the winner of the 2010 EAMT Best Thesis Award,

Gonzalo Iglesias from the Universidade de Vigo, on “Hierarchical Phrase-based Translation with Weighted Finite-State Transducers”.

We thank you all: authors, presenters, members of the programme committee, reviewers and secondary reviewers, and attendees, for helping us to make EAMT-2011 a success: we hope you enjoy the programme that we have prepared for you. We know it’s dense, but we hope it’s worth it!

As these proceedings are being finalized, our job is almost finished, and the conference is now in good hands: those of the local organizers in Leuven, headed by Vincent Vandeghinste. It has been great to work with them, and we send them a special thank-you!

Heidi Depraetere, CrossLanguage n.v., Belgium

Mikel L. Forcada, Universitat d’Alacant, Spain  
EAMT-2011 co-chairs

# Message from the Local Organizing Committee

We welcome you to the EAMT 2011, the 15th Annual Conference of the European Association for Machine Translation. This year's conference is being held at the Faculty of Arts of the Katholieke Universiteit Leuven in Belgium, the oldest university in the Low Countries, founded in 1425.

We hope you will enjoy the arrangements we have made for you and have an interesting and productive conference, benefiting from the opportunity to meet with MT researchers and users from all over Europe and the rest of the world.

While in Leuven, we recommend that you see our historical center; enjoy the magnificent Gothic architecture of the town hall, have a beer on the Oude Markt (Old Market), and stroll through the Groot Begijnhof (Grand Béguinage), a Unesco World Heritage site near the conference dinner.

We would like to thank the Dutch, German, and Computational Linguistics research unit, the Computational Linguistics in Flanders (CLIF) research community, and the Dutch-Flemish STEVIN research programme for providing financial support for this conference.

We would also like to thank our sponsors for their generous support: the European Language Resources Association (ELRA), PROMT, Crosslanguage, and Onliner Translations.

Welcome to Leuven and enjoy the conference.

Local Organizing Committee:

Vincent Vandeghinste  
Joachim Van den Bogaert  
Frank Van Eynde  
Ineke Schuurman  
Scott Martens  
Tom Vanallemeersch  
Liesbeth Augustinus

Centre for Computational Linguistics, KULeuven



# Programme Committee

## **Editors and Programme Committee Chairs:**

Research Track: Mikel L. Forcada, Universitat d'Alacant

User Track: Heidi Depraetere, Crosslanguage n.v.

## **Programme Committee:**

### **Research Track**

Juan Alberto Alonso, Lucy Software Ibérica, Barcelona, Spain  
Hervé Blanchon, Laboratoire d'Informatique de Grenoble, France  
Ralf Brown, Carnegie Mellon University, US  
Michael Carl, Copenhagen Business School, Denmark  
Helena Caseli, Universidade Federal de São Carlos, Brasil  
Mauro Cettolo, Fondazione Bruno Kessler, Italy  
Khalid Choukri, ELDA, France  
Marta R. Costa-jussà, Barcelona Media, Spain  
Josep-Maria Crego, LIMSI/CNRS, France  
Jinhua Du, Xi'an University of Technology, China  
Andreas Eisele, European Commission, Luxembourg  
Marcello Federico, Fondazione Bruno Kessler, Italy  
Jesús Giménez, Universitat Politècnica de Catalunya, Spain  
Declan Groves, Dublin City University, Ireland  
John Hutchins, Norwich, UK  
Pierre Isabelle, National Research Council, Canada  
Jie Jiang, Dublin City University, Ireland  
Patrik Lambert, Université du Maine, France  
Alon Lavie, Carnegie-Mellon University, USA  
Yanjun Ma, Dublin City University, Ireland  
Hermann Ney, RWTH Aachen, Germany  
Stephan Oepen, Universitetet i Oslo, Norway  
Pavel Pecina, Dublin City University, Ireland  
Juan Antonio Pérez-Ortiz, Universitat d'Alacant, Spain  
Felipe Sánchez-Martínez, Universitat d'Alacant, Spain  
Kepa Sarasola, Euskal Herriko Unibertsitatea, Spain  
Holger Schwenk, Université de Le Mans, France  
Jean Senellart, Systran, France  
Svetlana Sheremetyeva, LanA Consulting ApS, Denmark  
Harold Somers, University of Manchester, UK  
Lucia Specia, University of Wolverhampton, UK  
Jörg Tiedemann, Uppsala Universitet, Sweden  
John Tinsley, Dublin City University, Ireland

Jun-Ichi Tsujii, Microsoft Research Asia, China  
Antal van den Bosch, Universiteit van Tilburg, Netherlands  
Vincent Vandeghinste, Katholieke Universiteit Leuven, Belgium  
Andy Way, Dublin City University, Ireland  
Dekai Wu, Hong Kong University of Science and Technology  
François Yvon, LIMSI/CNRS, France

### **User Track**

Andrew Bredenkamp, acrolinx, Berlin, Germany  
Anthony Clarke, Retired Partner, CLS Communications AG., Switzerland  
Mike Dillinger, Translation Optimization Partners  
Daniel Grasmick, Lucy Software, Waibstadt, Germany  
Viggo Hansen, EAMT Executive Committee  
Fred Hollowood, Symantec, Dublin, Ireland  
Bente Maegaard, CST, University of Copenhagen  
Dag Schmidtke, Microsoft Ireland, Dublin, Ireland  
Svetlana Sokolova, PROMT, Russia  
Pablo Vasquez, Cisco Systems, San Jose, USA  
Feiyu Xu, Yocoy Technologies, Berlin, Germany  
Jost Zetzsche, TM Marketplace, USA

### **Additional Reviewers**

Iñaki Alegria, Euskal Herriko Unibertsitatea, Spain  
Jonathan Clark, Carnegie-Mellon University, U.S.A.  
Michael Denkowski, Carnegie-Mellon University, U.S.A.  
Arantza Diaz de Ilarraza, Euskal Herriko Unibertsitatea, Spain  
Greg Hanneman, Carnegie-Mellon University, U.S.A.  
Yifan He, Dublin City University, Ireland  
Kenneth Heafield, Carnegie-Mellon University, U.S.A.  
Gorka Labaka, Euskal Herriko Unibertsitatea, Spain  
Aurélien Max, LIMSI, France

**Invited Speaker:**

Fred Hollowood, Symantec, Dublin, Ireland.

**Local Organising Committee:**

**Centre for Computational Linguistics, KULeuven**

Vincent Vandeghinste

Joachim Van den Bogaert

Frank Van Eynde

Liesbeth Augustinus

Ineke Schuurman

Scott Martens

Tom Van Allemeersch

**Sponsors:**



Onderzoekseenheid Nederlandse, Duitse en Computatieve Taalkunde