Proceedings of the 16th Annual Conference of the European Association for Machine Translation

EAMT 2012

Trento|Italy, May 28th - 30th 2012

EUROPEAN ASSOCIATION FOR MACHINE TRANSLATION

Edited by Mauro Cettolo, Marcello Federico, Lucia Specia, Andy Way

Fbk - Editoria n. 5 / 5-2012

EAMT 2012

Proceedings of the 16th Annual Conference of the European Association for Machine Translation

Trento | Italy, May 28th - 30th 2012

Edited by
Mauro Cettolo, Marcello Federico, Lucia Specia, Andy Way

FONDAZIONE
BRUNO KESSLER

Table of Contents

Oral Session 1 – User Papers

Oral Session 2 – Research Papers

Poster Session 1 – User and Project Papers

User Papers:

Project Papers:

Poster Session 2 – Project Papers

Oral Session 3 – Research Papers

Poster Session 3 – Research and Project Papers

Research Papers:

Poster Session 4 – Research Papers

Oral Session 4 – Research Papers

Oral Session 5 – Research Papers

Foreword

The European Association for Machine Translation (EAMT) organised its first Workshop/ Conference back in 1996, in Austria. Up until 2009, when I became EAMT President, events had been held in Denmark, Switzerland, the Czech Republic, Slovenia, the UK, Hungary, Ireland, Malta, Norway and Germany.

When I took over, I was very keen as EAMT President to see our conferences take place in countries that we hadn't visited before. In 2009, we went to Spain, France in 2010, and last year we went to one of the Benelux countries, namely Belgium. All three events were fantastically organised, and proved to be very successful.

This year, we are continuing this trend. I am very pleased that this year we are holding the EAMT annual conference for the first time in Italy, where MT has thrived for quite some time now. I am also pleased to say that this is the first EAMT conference held since I became President of the International Association of MT (IAMT), a role I am honoured to fulfil.

This is the 16th 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 database version of the MT Compendium, sponsoring R&D activities, an extension of our activities to the MENA region, Best Paper Award etc. Note also that since its inception in 1997, the EAMT has not raised its Membership rates, and we will continue to hold the cost of membership for 2012. Joining us really is great value, especially in a year like 2012, where more than one IAMT-affiliated event takes place (EAMT here, and AMTA later in the year in San Diego: http://amta2012.amtaweb.org/), especially now that with the help of the Presidents of the other regional associations, Alon Lavie and Hitoshi Isahara, we have arranged for conference discounts to benefit all IAMT members, no matter which regional association you have joined.

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 all very fortunate to have such a strong body of colleagues representing our Association.

In addition to all this, EAMT conferences continue 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 16th Conference is no exception, and in particular I would like to thank my Programme Co-Chair Lucia Specia, together with the overall Conference Chair, Marcello Federico, for helping me assemble a very attractive programme, comprising of Research and User tracks, poster sessions, and a terrific Invited Speaker in Don DePalma. As in the past two years, a special session has been

organised where some prominent FP7 projects are featured, so this too will be a really interesting session.

Last but not least, I would especially like to thank our local organizers, Marcello Federico and Mauro Cettolo, who very generously volunteered to hold the meeting in Trento. We are very grateful to Marcello and his team for their excellent organization of this event.

Finally, thanks to all of you for coming. I 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.

Andy Way
Director of Language Technology

Applied Language Solutions,
Delph, Saddleworth, UK

President of the EAMT

andy.way@appliedlanguage.com

Message from the Conference Chair

It is a great pleasure to welcome you at the Fondazione Bruno Kessler (FBK) for the 16th Conference of the European Association for Machine Translation. This is the first time that the EAMT annual conference has been organized in Italy and I'm very thankful to the Board of EAMT for giving me the opportunity to host the 2012 edition in Trento.

Given the increasing popularity of the EAMT conference over the last few years, the board of EAMT decided to organize the 2012 conference in two and a half days instead of two. This choice was indeed rewarded by an unexpectedly high number of paper submissions this year, 30% more than in 2011. Hence, I really hope you will enjoy the technical program and will find yourself comfortable with the conference venue.

The conference will be held at the technological and scientific hub of FBK, on the hill of Povo, a suburb of Trento. Morning sessions and coffee breaks will be hosted in the conference room of the main building. Lunches will be instead served in the large hall of the North building, where also the afternoon poster sessions will take place. In the same building, there will be extra rooms available for informal meetings as well as a cafeteria.

I hope you will enjoy the two social events that we have organized: the welcome reception in the Sass underground archaeological area in Trento, and the conference banquet along the lake in Riva del Garda. I hope these two occasions will give you a taste of the architectural, cultural and natural beauty of Trentino and Italy.

Nothing of this conference could have been organized without good teamwork. Hence, I wish to thank the Program co-Chairs, Lucia Specia and Andy Way, for managing the unexpectedly high number of submissions and for arranging the conference program. Thank you also to the staff at FBK, which worked with impressive professional dedication to set-up this event. In particular, I wish to express my gratitude to Mauro Cettolo, Local Organization Chair, Silvia Malesardi, venue and the social events, Francesca Guerzoni, website, Moira Osti, marketing, Luigi Massimiliano Cordisco, correspondence, Barbara Gazzoli and Adalberta Bragagna, editing of proceedings, and to our student volunteers, Prashant Mathur, Jose Camargo de Souza, and Nick Ruiz.

The organization of a conference is the sum of many important parts. Among them there are also the sponsors, which generously supported EAMT 2012. In particular, I would like to acknowledge support of the Superintendence for Cultural and Archaeological Heritage of the Province of Trento, for hosting the welcome reception, and of Springer, for sponsoring the best paper award. Last but not least, I wish to publicly thank our silver sponsor, Microsoft Translator, and our bronze sponsor Virtus.

I wish you a very successful conference and pleasant stay in Trento!

Marcello Federico, Fondazione Bruno Kessler, IT

EAMT-2012 Conference Chair

Message from the Programme Chairs

It is a great pleasure for us to welcome you to the 16th Conference of the European Association for Machine Translation (EAMT) in Trento. We have been happy to serve as programme co-chairs of a conference that has become the yearly reference conference for European machine translation developers, researchers and users, and keeps growing year by year. A sign of this growth is that the conference was extended from 2 to 2 ½ days in order to keep to the single track format – which makes EAMT events very homely for regulars and newcomers alike.

As in previous years, the conference has three main tracks: (i) a research track, where researchers report about significant research results in any aspect of machine translation and related areas, with a substantial evaluation component, (ii) a user track, where users report their experiences with machine translation in business, government, or NGOs, and (iii) a projects track to publicize EU and international projects and initiatives. We also introduced a technology showcase for product demonstrations in order to encourage participation from developers/industry. In order to encourage submissions for the user track, we changed the format of these submissions: short papers with 2-4 pages. For projects/product demonstrations, both submissions only required a 1-page abstract.

We received a record number of submissions - a total of 102 papers: 57 in the research track, 17 in the user track, and 28 project/product descriptions. Most of the latter were accepted, but were reformulated by the project participants to conform to the conference style-guide. As far as research and user papers are concerned, after double-blind review by at least three leading MT reviewers, 40 of them (54%) were accepted and found their way into the proceedings: 29 research papers (51%) – 12 for oral presentation and 17 for poster presentation – and 11 user papers (64%), 4 for oral presentation and 7 for poster presentation. Poster presenters will also have the opportunity to showcase their work in a one-minute poster boaster oral session. As expected, submissions come mainly from Europe, with a large number of submissions also received from the US this year. We also received papers with authors from the Japan, Canada, Brazil, Hong Kong, India, Kazakhstan, Mexico and Singapore.

We are in debt to the members of the programme committee and to the secondary reviewers they appointed for some of their papers. As the number of papers received was even higher than usual, they had an unusually large workload: we especially thank them for their invaluable help, which most of them completed on time, which made our lives easier!

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; and 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 you have in your conference pack, truly represents the best of machine translation research, development and real-world usage.

As an opener, we will enjoy an invited talk by Don DePalma, from Common Sense Advisory, which we hope will appeal to both our research and our user audience. To close the conference, we will have a presentation by the winner of the EAMT Best Thesis Award, Abby Levenberg, completed under the supervision of Dr. Miles Osborne at the University of Edinburgh, on *Stream-based Statistical Machine Translation*.

We thank you all: authors, presenters, members of the programme committee, reviewers and secondary reviewers, and attendees, for helping us to make EAMT-2012 a success: we hope you enjoy the programme that we have prepared for you.

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 Trento, headed by Marcello Federico. It has been great to work with them, and we send them a special thank you!

Lucia Specia, University of Sheffield, UK
Andy Way, Applied Language Solutions, UK

EAMT-2012 co-programme chairs

Committees

Conference Chair

Marcello Federico (FBK-irst, Italy)

Programme Chairs

Research Track

Lucia Specia (University of Sheffield, UK)

User Track

Andy Way (Applied Language Solutions, UK)

Local Organization Chair

Mauro Cettolo (FBK-irst, Italy)

Programme Committee

Research Track

Wilker Aziz (University of Wolverhampton, UK)
Loïc Barrault (Université du Maine, France)
Nicola Bertoldi (Fondazione Bruno Kessler, Italy)
Laurent Besacier (Université J. Fourier, France)
Alexandra Birch (University of Edinburgh, UK)
Hervé Blanchon (l'Université Pierre Mendès - Grenoble 2, France)
Ondrej Bojar (Charles University, Prague)
Ralf Brown (Carnegie Mellon University, USA)
Antal van den Bosch (Universiteit van Tilburg, Netherlands)
Bill Byrne (Cambridge University, UK)
Nicola Cancedda (Xerox Research Centre, France)
Michael Carl (IAI Saarbrücken, Germany)
Marine Carpuat (NRC Institute for Information Technology, Canada)
Helena Caseli (Universidade Federal de São Carlos, Brazil)
Marta Costa-jussà (Barcelona Media, Spain)
Jinhua Du (Xi'an University of Technology, China)
Marc Dymetman (Xerox Research Centre, France)
Andreas Eisele (European Commission, Luxembourg)
Mark Fishel (University of Zurich, Switzerland)

Declan Groves (Dublin City University, Ireland)
Barry Haddow (University of Edinburgh, UK)
Yifan He (Dublin City University, Ireland)
Jie Jiang (Applied Language Solutions, UK)
Patrik Lambert (Universiteé du Maine, France)
David Langlois (Nancy University, France)
Alon Lavie (Carnegie Mellon University, USA)
Yanjun Ma (Dublin City University, Ireland)
Pavel Pecina (Dublin City University, Ireland)
Sergio Penkale (Applied Language Solutions, UK)
Juan Antonio Pérez-Ortiz (Universitat d'Alacant, Spain)
Daniele Pighin (Universitat Politècnica de Catalunya, Spain)
Maja Popović (DFKI, Bermany)
Carlos Ramish (University of Grenoble, France)
Felipe Sánchez-Martínez (Universitat d'Alacant, Spain)
Kepa Sarasola (Euskal Herriko Unibertsitatea, Spain)
Christophe Servan (Université du Maine, France)
Khalil Sima'an (Universiteit van Amsterdam, Netherlands)
Michel Simard (National Research Council, Canada)
Sara Stymne (Linköping University, Sweden)
Jörg Tiedemann (Uppsala University, Sweden)
John Tinsley (Dublin City University, Ireland)
Jun-Ichi Tsujii (Microsoft Research Asia, China)
Marco Turchi (JRC, Italy)
Vincent Vandeghinste (Katholieke Universiteit Leuven, Belgium)
François Yvon (LIMSI/CNRS, France)

User Track

Juan Alberto Alonso (Lucy Software Ibérica, Spain)
Diego Bartolome (Tau You, Spain)
Anthony Clarke (CLS Communications AG., Switzerland)
David Clarke (WeLocalize, Ireland)
Heidi Depraetere (Cross Language, Belgium)
Mike Dillinger (Translation Optimization Partners, US)
Ray Flournoy (Adobe Systems, US)
Viggo Hansen (EAMT Executive Committee)
Manuel Herranz (PangeaMT, Spain)
Fred Hollowood (Symantec, Ireland)
Daniel Grasmick (Lucy Software, Germany)
Dorothy Kenny (Dublin City University, Ireland)
Bente Maegaard, CST (University of Copenhagen, Denmark)
Enda Mcdonnell (Alchemy Software, Ireland)
Nelson Ng (Ebay, US)
Sharon O'Brien (Dublin City University, Ireland)
Sergio Ortiz-Rojas (Prompsit Language Engineering, Spain)
Mirko Plitt (Autodesk, Switzerland)
Gema Ramirez Sanchez (Prompsit Language Engineering, Spain)

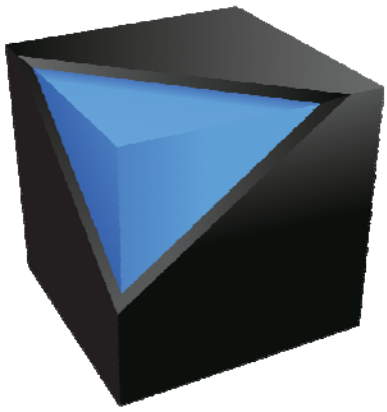
Adriane Rinsche (Language Technology Centre, UK)
Phil Ritchie (VistaTEC, Ireland)
Johann Roturier (Symantec, Ireland)
Reinhard Schaeler (University of Limerick, Ireland)
Dag Schmidtke (Microsoft Ireland, Dublin, Ireland)
Jörg Schütz (Bioloom, Germany)
Svetlana Sheremetyeva (LanA Consulting ApS, Denmark)
Svetlana Sokolova (PROMT, Russia)
Gregor Thurmair (Linguatec, Germany)
Feiyu Xu (DFKI, Germany)
Elia Yuste (PangeaMT, Spain)
Ventsislav Zhechev (Autodesk, Switzerland)

Sponsors

SILVER SPONSOR

Microsoft®
Translator

BRONZE SPONSOR



VIRTUS™

Invited Talk

The Unavoidable Adoption of Machine Translation

Donald A. DePalma

Ph.D., Chief Strategy Officer & Founder of Common Sense Advisory, Inc.

There is an inevitability to machine translation that no business, government agency, or even language service provider can avoid. It's simply a matter of the huge volume of content that organizations large and small must translate to be relevant to their global constituencies. In this presentation, DePalma will review the current state of machine translation and related technologies from a business perspective, reviewing its evolution and increasing adoption among translation buyers and suppliers. He will discuss the drivers for, obstacles to, and major trends affecting the segment. He will also look at the future of machine translation and what that means for buyers and suppliers.