XIII ANNUAL CONFERENCE

EAMT

14-15 May 2009 Barcelona, Spain
Universitat Politècnica de Catalunya

Proceedings of the
13th Annual Conference of the European Association for Machine Translation

14–15 May 2009
Universitat Politècnica de Catalunya
Barcelona, Spain

European Association for Machine Translation

UPC

TALP

Published by
Universitat Politècnica de Catalunya
C. Jordi Girona, 31
08034 Barcelona
Spain

ISBN:

Printed by
SERVIS GRÀFICS COPISTERIA IMATGE, S.L.
Cinca, 8
08030 Barcelona
Spain

Conference Sponsors

EUROPEAN LANGUAGE RESOURCES ASSOCIATION
ELRA

Springer
the language of science

Order copies of this and other EAMT proceedings from:

EAMT Secretariat
c/o A. Clarke
Schützenweg 57
CH-4123 Allschwil
Switzerland
`secretariat@eamt.org`
`http://www.eamt.org/`

Foreword

The European Association for Machine Translation (EAMT) has a long tradition for organising annual workshops and conferences with the purpose of bringing together people and companies with a professional interest in machine translation and other tools for translation, be it users, researchers, developers, or providers who want to follow latest developments.

The term “machine translation”, MT, is interpreted in its widest sense at the EAMT conferences. This means that MT refers not only to fully automatic translation, but also to all other kinds of tools for translation or multilinguality, be it translation memory, parallel corpora and other resources for translations, alignment, terminology tools, etc.

This year, the focus of the EAMT conference is on how to develop translation technologies for and among languages having smaller speech communities or limited digital resources. Actually most of the world’s languages are of this category, and we are sure that researchers, users and providers have an interest in such technologies.

The fact that researchers and users with explicit needs are brought together, and that there is a fair number of research presentations as well as some more practically oriented presentations, provide excellent opportunities for mutual feedback. Through our annual conferences we have been able to create an environment for interesting discussions and maybe even for the creation of new partnerships, and we are sure that this conference will contribute to the continued success of the EAMT conferences.

EAMT is deeply dependent on colleagues willing to take upon themselves the tasks of programme committee work and local organization.

So, before closing, I first want to thank all of the programme committee for their invaluable contribution, not least the programme committee co-chairs Lluís Màrquez and Harold Somers.

Secondly, I want to thank David Farwell, Adrián R. Fonollosa, José Mariño and other colleagues at the Centre for Speech and Language Applications and Technologies (TALP) at the Universitat Politècnica de Catalunya for the local organization. The MT research group at TALP kindly invited the EAMT to have the conference in Barcelona, and it has been a great pleasure for the EAMT Executive Committee to collaborate with them.

Finally: I wish all of you an excellent and enjoyable conference!

Bente Maegaard
Center for Sprogteknologi,
University of Copenhagen, Denmark
President of the EAMT

Message from the Programme Committee Chairs

We are delighted to welcome you to the 13th Annual Conference of the European Association for Machine Translation in Barcelona. Starting out as occasional workshops, I hope you will agree that the EAMT Annual Conference has now established itself in the calendar of important events in the field of MT.

As may be appropriate these days in our field, let us start with some statistics: we received 54 submissions, of which 14 were accepted as full papers, 18 as posters, providing a healthy overall acceptance rate of 59%. Submissions came from 24 different countries, including 9 beyond Europe's borders. The two countries providing most papers were Spain (10) and Ireland (8).

The decision to designate submissions as papers or posters was taken purely as a reflection of the most suitable form of presentation of the work, as determined by an explicit request to reviewers. While there are still some who think of posters as "second-class" papers, we would like to assure all presenters that this is far from the case. For a start, all papers are afforded equal space and prominence in these Proceedings. Furthermore, as more and more people are discovering, presenting one's work as a poster can have serious benefits and advantages over an oral presentation, not least of which are the possibility of engaging on a personal level with a select audience which has explicitly chosen to seek out your presentation and to participate in it. More than a few presenters are now coming to the idea that a poster can actually be more effective than the traditional but impersonal lecture style of presentation.

As programme chairs, we are of course indebted to the panel of reviewers, whose names are listed elsewhere. 45 reviewers looked at 3 or 4 papers each, thus ensuring that each submission got three separate reviews. We asked reviewers to work to a tight schedule, and almost without exception they got their reviews in on time, which in turn meant that we could notify authors of acceptance or, regrettably, rejection, even a few days before our stated deadline. We hope that authors – whether chosen or not – have appreciated and benefited from the reviewers' comments, which were often quite extensive. Equally we are grateful to authors who were asked to prepare their final copy for these Proceedings within a fairly short deadline. Again, with only one or two exceptions, the deadline was met, and we were able to avoid the usual panic and scramble associated with this task.

As Programme chairs, our job sort of ends once we have chosen which papers to accept, and arranged them into the programme you will experience and, we hope, enjoy over the next two days. At this point we hand over to the Local Organisation Committee, but of course we have been working closely together with them since day one, a task obviously much facilitated by the fact that one of us is a "local". Nevertheless, the local organisers have been with us every step of the way, and we would like here to thank them for their support, advice and, when necessary, gentle prodding. Of great help too have been the EAMT Executive Committee, with advice on precedent, form and character, so that this conference should at the same time fit comfortably in with the EAMT conference series, meeting the needs and expectations of members of the EAMT, while, we hope, standing out as a memorable and enjoyably different conference.

Finally, thanks to all authors, presenters and attendees for making this a successful 13th Annual EAMT Conference.

Lluís Màrquez and **Harold Somers**
EAMT–2009 Programme Committee co-chairs

Message from the Local Organising Committee

It gives us great pleasure to welcome you to the EAMT 2009, the 13th Annual Conference of the European Association for Machine Translation. This year the conference is being held on the Campus Nord of the Universitat Politècnica de Catalunya in Bacrelona, Spain. We have tried to make all the necessary arrangements to ensure that your participation in the conference events is as productive and enjoyable as possible. While in Barcelona be sure to experience the special atmosphere the city has to offer: the Roman, medieval, and modernist architecture of the old city, Passeig de Gràcia and the Eixample and the wide array of excellent restaurants, theatre, music, galleries and museums. It would be unfortunate not to take in all you can while here.

Of course, we also hope that you will benefit from a strong programme of conference presentations and workshops which are at the forefront of MT and multilingual language processing research and development.

In organising the conference we have received significant financial support from the Universitat Politècnica de Catalunya and the Spanish Ministry for Science and Innovation. We also would like to thank European Languages Resources Association (ELRA) and Springer for their generous sponsorship. Finally, we have also had the unselfish assistance of local staff and students. In particular we wish to thank Coralí Planellas for her many hours of effort, especially in maintaining the conference web site.

So without further ado, welcome and enjoy the conference.

Local Organising Committee:
David Farwell
José A. R. Fonollosa
José Mariño
Lluís Màrquez

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
Universitat Politècnica de Catalunya

EAMT-2009 Organizers

Programme Committee Chairs:

Lluís Màrquez, Universitat Politècnica de Catalunya, Spain
Harold Somers, Dublin City University, Ireland

Program Committee:

Joseba Abaitua, Universidad de Deusto, Spain
Iñaki Alegria, Euskal Herriko Unibertsitatea, Spain
Juan Alonso, Translendium SL, Spain
Toni Badia, Universitat Pompeu Fabra, Spain
Rafael Banchs, Barcelona Media Innovation Centre, Spain
Pierette Bouillon, Université de Genève, Switzerland
Chris Callison-Burch, Johns Hopkins University, USA
Nicola Cancedda, Xerox Research Centre Europe, France
Michael Carl, Copenhagen Business School Handelshøjskolen, Denmark
Francisco Casacuberta, Universitat Politècnica de València, Spain
Irene Castellón, Universitat de Barcelona, Spain
Adrià de Gispert, Cambridge University, England
Arantza Díaz de Ilarraza, Euskal Herriko Unibertsitatea, Spain
Bonnie Dorr, University of Maryland, USA
Andreas Eisele, DFKI GmbH, Germany
David Farwell, Universitat Politècnica de Catalunya, Spain
Marcello Federico, Fondazione Bruno Kessler, Italy
José A. R. Fonollosa, Universitat Politècnica de Catalunya, Spain
Mikel Forcada, Universitat d'Alacant, Spain
Jesús Giménez, Universitat Politècnica de Catalunya, Spain
John Hutchins, United Kingdom
Kevin Knight, University of Southern California, USA
Philipp Koehn, University of Edinburgh, Scotland
Alon Lavie, Carnegie Mellon University, USA
Lori Levin, Carnegie Mellon University, USA
Bente Maegaard, Københavns Universitet, Denmark
Daniel Marcu, University of Southern California, USA
José B. Mariño, Universitat Politècnica de Catalunya, Spain
M. Antònia Martí, Universitat de Barcelona, Spain
Hermann Ney, Aachen University, Germany
Sharon O'Brien, Dublin City University, Ireland
Stephan Oepen, University of Oslo, Norway
Mike Rosner, University of Malta, Malta
Marta R. Costa-jussà, Universitat Politècnica de Catalunya, Spain
Anna Sågvall Hein, Uppsala universitet, Sweden
Kepa Sarasola, Euskal Herriko Unibertsitatea, Spain
Holger Schwenk, Université du Maine, France
Hisami Suzuki, Microsoft Research, USA
Cristina Vertan, Universität Hamburg, Germany
Walther von Hahn, Universität Hamburg, Germany
Andy Way, Dublin City University, Ireland
Dekai Wu, The Hong Kong University of Science and Technology, Hong Kong

Additional Reviewers:

Mauro Cettolo, Marine Carpuat, Maja Popović, Nicola Bertoldi, Felipe Sánchez-Martínez, Gonzalo Iglesias, Saab Mansour, Elisabeth Comelles

Invited Speakers:

Lori Levin, Carnegie Mellon University, USA
Nicholas Ostler, Foundation for Endangered Languages

Local Organising Committee:

David Farwell
José A. R. Fonollosa
José Mariño
Lluís Màrquez

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
Universitat Politècnica de Catalunya

Table of Contents

Invited Papers

Regular Papers

Conference Program

Thursday, May 14, 2009

9:00–9:15 Welcome

9:15–10:15 Invited Talk
“The Jungle is Neutral” – Newcomer Languages Face New Media
Nicholas Ostler

10:15–10:45 *Character-Based PSMT for Closely Related Languages*
Jörg Tiedemann

10:45–11:15 *TS3: an Improved Version of the Bilingual Concordancer TransSearch*
Stéphane Huet, Julien Bourdaillet and Philippe Langlais

11:15–11:45 Coffee break

11:45–12:15 *Estimating the Sentence-Level Quality of Machine Translation Systems*
Lucia Specia, Marco Turchi, Nicola Cancedda, Marc Dymetman and Nello Cristianini

12:15–12:45 *Evaluation-Guided Pre-Editing of Source Text: Improving MT-Tractability of Light Verb Constructions*
Bogdan Babych, Anthony Hartley and Serge Sharoff

12:45–13:15 *Learning Labelled Dependencies in Machine Translation Evaluation*
Yifan He and Andy Way

13:15–14:30 Lunch

14:30–15:00 *Improving a Catalan-Spanish Statistical Translation System using Morphosyntactic Knowledge*
Mireia Farrús, Marta R. Costa-jussà, Marc Poch, Adolfo Hernández and José B. Mariño

15:00–15:30 *Use of Rich Linguistic Information to Translate Prepositions and Grammar Cases to Basque*
Eneko Agirre, Aitziber Atutxa, Gorka Labaka, Mikel Lersundi, Aingeru Mayor and Kepa Sarasola

Thursday, May 14, 2009 (continued)

15:30–17:00 **Poster Session I**

Gappy Translation Units under Left-to-Right SMT Decoding
Josep M. Crego and François Yvon

Relevance of Different Segmentation Options on Spanish-Basque SMT
Arantza Díaz de Ilarraza, Gorka Labaka and Kepa Sarasola

English-Latvian Toponym Processing: Translation Strategies and Linguistic Patterns
Tatiana Gornostay and Inguna Skadina

An Environment for Named Entity Recognition and Translation
Filip Graliński, Krzysztof Jassem and MichałMarcińczuk

Optimal Bilingual Data for French–English PB-SMT
Sylwia Ozdowska and Andy Way

Word- and Sentence-Level Confidence Measures for Machine Translation
Sylvain Raybaud, Caroline Lavecchia, David Langlois and Kamel Smaïli

Translating Questions for Cross-Lingual QA
Jörg Tiedemann

Developing Prototypes for Machine Translation between Two Sámi Languages
Francis M. Tyers, Linda Wiechetek and Trond Trosterud

Collocations in a Rule-Based MT System: A Case Study Evaluation of their Translation Adequacy
Eric Wehrli, Violeta Seretan, Luka Nerima and Lorenza Russo

16:30–17:00 Coffee break

17:00–17:30 *Automatic Translation of Norwegian Noun Compounds*
Lars Bungum and Stephan Oepen

17:30–18:00 *Marker-Based Filtering of Bilingual Phrase Pairs for SMT*
Felipe Sánchez-Martínez and Andy Way

Friday, May 15, 2009

9:00–9:30 *Tree-Based Target Language Modeling*
Vincent Vandeghinste

9:30–10:00 *Language Model Adaptation for Difficult to Translate Phrases*
Behrang Mohit, Frank Liberato and Rebecca Hwa

10:00–11:00 Invited Talk

Adaptable, Community-Controlled, Language Technologies for Language Maintenance
Lori Levin

11:00–11:30 Coffee break

11:00–12:30 **Poster Session II**

A Phrase-Based Hidden Semi-Markov Approach to Machine Translation
Jesús Andrés-Ferrer and Alfons Juan

Building Strong Multilingual Aligned Corpora
Reza Bosagh Zadeh

A Constraint Satisfaction Approach to Machine Translation
Sander Canisius and Antal van den Bosch

Introducing the Autshumato Integrated Translation Environment
Hendrik J. Groenewald and Wildrich Fourie

A New Subtree-Transfer Approach to Syntax-Based Reordering for Statistical Machine Translation
Maxim Khalilov, José A. R. Fonollosa and Mark Dras

On Extracting Multiword NP Terminology for MT
Svetlana Sheremetyeva

Rule-Based Augmentation of Training Data in Breton-French Statistical Machine Translation
Francis M. Tyers

Can Semantic Role Labeling Improve SMT?
Dekai Wu and Pascale Fung

Are Unaligned Words Important for Machine Translation?
Yuqi Zhang, Evgeny Matusov and Hermann Ney

Friday, May 15, 2009 (continued)

12:30–13:30	EAMT Business Meeting
13:30–14:45	Lunch
14:45–15:15	*Using Supertags as Source Language Context in SMT* Rejwanul Haque, Sudip Kumar Naskar, Yanjun Ma and Andy Way
15:15–15:45	*On LM Heuristics for the Cube Growing Algorithm* David Vilar and Hermann Ney
15:45–16:15	Coffee break
16:15–16:45	*Tuning Syntactically Enhanced Word Alignment for Statistical Machine Translation* Yanjun Ma, Patrik Lambert and Andy Way
16:45–17:00	Closing session – presentation of the Springer Award for best paper