

Message from the Programme Committee Chairs

We are delighted to welcome you to the International Workshop on Using Linguistic Information for Hybrid Machine Translation in Barcelona.

To begin with, a few statistics: we received 15 submissions, of which 10 were accepted, providing an overall acceptance rate of 67%. Submissions came from 9 different countries, including 2 beyond Europe's borders. The countries providing most papers were France, Germany and Spain with 3 each.

As programme chairs, we are of course indebted to the panel of reviewers, whose names are listed elsewhere. The 22 reviewers looked at 1 or 2 papers each, thus ensuring that each submission had three separate reviews. We asked reviewers to work to a tight schedule, and with few exceptions they got their reviews in on time, which in turn meant that we could notify authors of acceptance in a timely fashion. We hope that authors 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, the deadline was generally met, and we were able to avoid the usual panic and scramble associated with getting the document off to the printers on time.

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. At this point we hand matters 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.

Finally, thanks to all authors, presenters and attendees for making this a successful workshop.

Gorka Labaka and David Farwell
LIHMT 2011 Programme Committee co-chairs

Message from the Local Organising Committee

It gives us great pleasure to welcome you to the LIHMT 2011, the International Workshop on Using Linguistic Information for Hybrid Machine Translation, here on the Campus Nord of the Universitat Politècnica de Catalunya in Barcelona, Spain. We have tried to make all the necessary arrangements to ensure that your participation in the workshop 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, shops 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 presentations that are at the forefront of Machine Translation research and development.

In organising the workshop we have received significant financial support from the Universitat Politècnica de Catalunya, the Generalitat de Catalunya and the Spanish Ministry for Science and Innovation. We also would like to thank the Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (TALP), the European Association for Machine Translation (EAMT) and European Language Resources Association (ELRA) for their generous sponsorship. Finally, we have also had the unselfish assistance of local staff and students. In particular we wish to thank Gorika Labaka and Meritxell González for their many hours of effort, especially in maintaining the workshop web site.

So without further ado, welcome and enjoy the conference.

Local Organising Committee:

David Farwell

Lluís Màrquez

Meritxell González

Cristina España-Bonet

Daniele Piguin

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla

Universitat Politècnica de Catalunya

OpenMT-2 Workshop on Using Linguistic Information for Hybrid Machine Translation

Scientific Committee

Co-Chair: David Farwell (Technical University of Catalonia, TALP, Barcelona)

Co-Chair: Gorka Labaka (University of the Basque Country, Donostia)

Iñaki Alegria (University of the Basque Country, Donostia)

Ondřej Bojar (Charles University, Czech Republic)

Josep M. Crego (LIMSI/CNRS, France)

Arantza Díaz de Ilarraza (University of the Basque Country, Donostia)

Chris Dyer (Carnegie Mellon University, US)

Cristina España-Bonet (Technical University of Catalonia, TALP, Barcelona)

Marcello Federico (Fondazione Bruno Kessler, Italy)

Mikel Forcada (University of Alacant, Alicante)

Adrià de Gispert (University of Cambridge, UK)

Meritxell González (Technical University of Catalonia, TALP, Barcelona)

Kevin Knight (Information Sciences Institute, US)

Philipp Koehn (University of Edinburgh, UK)

Patrik Lambert (Université du Maine, France)

José Mariño (Technical University of Catalonia, TALP, Barcelona)

Lluís Màrquez (Technical University of Catalonia, TALP, Barcelona)

Daniele Pighin (Technical University of Catalonia, TALP, Barcelona)

Aarne Ranta (Chalmers University of Technology, Gothenburg, Sweden)

Marta R. Costa-jussà (Barcelona Media, Spain)

Felipe Sánchez-Martínez (University of Alacant, Alicante)

Kepa Sarasola (University of the Basque Country, Donostia)

About the OpenMT-2 Project

(<http://ixa.si.ehu.es/openmt2>)

The main goal of the OpenMT-2 project is the development of Open Source Machine Translation Architectures based on hybrid models and advanced semantic processors. These architectures will be open-source systems combining the three main Machine Translation frameworks – Rule-Based MT (RBMT), Statistical MT (SMT) and Example-Based MT (EBMT) – into hybrid systems. Implemented architectures and systems will be Open Source, so it will allow rapid system adaptation or development of new advanced Machine Translations systems for other languages. We will test system functionality for different languages: English, Spanish, Catalan and Basque; thus evaluating such architectures in different contexts. While there are many corpus resources for English and Spanish, there are not so many for Catalan and Basque. While the structure of some of those languages is very similar (Catalan and Spanish), others are very different (English and Basque). Basque is an agglutinative and highly inflecting language, unlike English, Catalan and Spanish.

In parallel there has been extensive work on developing an automatic Evaluation platform that supports the introduction of linguistically motivated morphological, syntactic and semantic metrics into the design of MT Evaluation methodologies. It also supports the development and testing of concrete, linguistically-based evaluation techniques.

The main innovative points of the OpenMT-2 project are:

- The design of hybrid systems combining traditional linguistic rules, example-based methods and statistical methods.
- The development of MT evaluation methods based on linguistically motivated metrics.
- The implementation of Open Source Systems.
- The use of advanced syntactic and semantic processing in MT.

Tentative Programme

Friday, November 18, 2011.

- 8:00: Registration
- 9:00: Opening
- 9:10: *Improved Statistical Machine Translation Using MultiWord Expressions*, Dhouha Bouamor, Nasredine Semmar and Pierre Zweigenbaum
- 9:35: *Using Apertium linguistic data for tokenization to improve Moses SMT performance*, Santiago Cortés Vaíllo and Sergio Ortiz Rojas
- 10:00: Plenary Session – Alon Lavie: *Statistical MT with Syntax and Morphology: Challenges and Some Solutions*
- 10:45: Coffee
- 11:15: *A Radically Simple, Effective Annotation and Alignment Methodology for Semantic Frame Based SMT and MT Evaluation*, Chi-Kiu Lo and Dekai Wu
- 11:40: *Reordering by Parsing*, Jakob Elming and Martin Haulrich
- 12:25: *Comparing CBMT Approaches Using Restricted Resources*, Monica Gavrila and Natalia Elita
- 12:45: Plenary Session – Ondřej Bojar: *Rich Morphology and What Can We Expect from Hybrid Approaches to MT*
- 13:30: Lunch
- 15:00: *A New Hybrid Machine Translation Approach Using Cross-Language Information Retrieval and Only Target Text Corpora*, Nasredine Semmar and Dhouha Bouamor
- 15:25: *Word Translation Disambiguation without Parallel Texts*, Erwin Marsi, André Lynum, Lars Bungum and Björn Gambäck
- 15:50: Plenary Session – Lucia Specia: *Linguistic Indicators for Quality Estimation of Machine Translation*
- 16:35: Coffee
- 17:05: *Deep evaluation of hybrid architectures: simple metrics correlated with human judgements*, Gorka Labaka, Arantza Diaz De Ilarraza, Cristina España-Bonet, Lluís Màrquez and Kepa Sarasola
- 17:30: *VERTa: Exploring a Multidimensional Linguistically-Motivated Metric*, Elisabet Comelles, Jordi Atserias, Victoria Arranz, Irene Castellon and Olivier Hamon
- 17:55: Round Table – Recapping the Central Issues
- 18:20: Closing