Foreword by the President of the IAMT and the EAMT

Céad mile fáilte romhaibh!

It’s a pleasure for me to warmly welcome you all to the 17th Machine Translation Summit.

Every two years, the International Association for Machine Translation (IAMT), an umbrella organization comprising the Asian Association for Machine Translation (AAMT), the Association for Machine Translation in the Americas (AMTA), and the European Association for Machine Translation (EAMT), jointly call everyone related in some way or another to machine translation and translation technologies to the most inclusive MT conference in the world, a real Summit. It brings together senior and junior researchers, developers, vendors and all kinds of users, coming from academia, industry, or even as freelancers, to share and become aware of any new developments in the field.

This is the sixth such summit held in Europe, after having visited Munich (1989), Luxembourg (1995), Santiago de Compostela (2001), Copenhagen (2007), and Nice (2013).

The organizers have assembled an excellent programme; after two days with a wide offer of tutorials and workshops, the main conference features four tracks: the research track, the users’ track, a new translators’ track, and the usual projects track, and includes three invited talks, poster sessions and oral sessions. Everyone will find something going on that interests them throughout the event.

Every six years, the EAMT organizes the MT Summit in Europe. The EAMT is a growing association, which organizes a yearly conference, sponsors research, development and community outreach initiatives, and annually grants a Best Thesis Award. Individuals, institutions and companies from Europe, Africa and the Middle East can join the EAMT for a modest fee and benefit from all these activities. In addition to that, EAMT members (as AMTA and AAMT members) enjoy attractive discounted fees when attending EAMT, AMTA and AAMT conferences. This is possible thanks to our members but especially to my colleagues in the EAMT Executive Committee —coming from both academia and industry— who work hard to make it all happen.

The XVII MT Summit would not be possible without the hard work of our local organizers, headed by my predecessor as EAMT president and current Executive Board member, Andy Way, who have, with the help of other MT actors from the Adapt Centre and the professional collaboration of Abbey Conference & Events, put together an excellent conference. I am very thankful for their hard work and for having put their local MT expertise at the disposal of the EAMT (and the IAMT).

Bainígí taitneamh as, that is, enjoy. Enjoy the programme, the company, and the city. Ten years ago, I lived here and worked here for a year and I’ll miss it every day of my life. And I’ll tell you something: it is especially the local people that makes Dublin —and all of Ireland— one of the best places in the world to hold a conference like our Summit. I’m sure you’ll bring home sweet memories of it!

Baile Átha Cliath/Dublin, Lúnasa/August 2019

Mikel L. Forcada
President of the IAMT and the EAMT
Professor of Computer Languages and Systems Universitat d’Alacant
Alacant, Valencian Country, Spain.
Email: mlf@ua.es
Foreword by MT Summit 2019 Conference Chair

Back in 2017, on behalf of the International Association for Machine Translation (IAMT), the European Association for Machine Translation (EAMT) entrusted me with hosting this conference that you are currently attending.

While I was grateful for the trust shown in me, as a previous IAMT/EAMT president, I was acutely aware of the need to deliver; compared to our annual EAMT conferences, MT Summits provide us with the opportunity to show our Asian and American friends and colleagues that we can put on an event that all three regional associations and the IAMT can be duly proud of; if you mess up, Europe has to wait 6 years to try to put it right!

After two years of hard work, I can say with some confidence that we have achieved this. One of the first things I did was put together a very strong support team. I would really like to thank our seven co-chairs of the four tracks, namely:

- Research track co-chairs: Barry Haddow & Rico Sennrich (University of Edinburgh, UK)
- User track co-chairs: John Tinsley (Iconic Translation Machines, Ireland) & Dimitar Shterionov (ADAPT Centre, Dublin City University, Ireland)
- Translator track co-chairs: Celia Rico (Universidad Europea de Madrid, Spain) & Federico Gaspari (ADAPT Centre, Dublin City University, Ireland)
- Projects chair: Mikel L. Forcada (Universitat d’Alacant, Spain)

I am also very grateful to Laura Rossi (Lexis Nexis, The Netherlands) and Antonio Toral (University of Groningen, The Netherlands) for acting as excellent Tutorials and Workshops Chairs, respectively. I hope you all benefited from attending these pre-conference events!

For the most part, it is these 9 individuals who have put together the programme assembled before you. Each of them will comment on their Tracks later in these proceedings, but they all deserve our heartfelt thanks, as do the panels of reviewers they assembled which helped improve all our papers. From a personal point of view, I am delighted that we have – for the first but surely not the last time – included a Translator track; I have advocated for some time now that it is only through dialogue that MT developers and the translator community can advance our field. I have been very keen to take up a number of recent opportunities to speak at translator conferences, so I am especially pleased to welcome translators to this event; thank you for coming!

I am of course grateful to everyone who submitted a paper; whether your work was selected for presentation or not, if no-one had submitted, we wouldn’t have had a conference. For those of you whose work was selected for presentation, many thanks for coming to Dublin, and to DCU, which have been my home and workplace now for 28 years, half my life. All of you would have interacted via EasyChair, and I am grateful to Carol Scarton, EAMT secretary, for her effort in setting up the various accounts which enabled the submission and notification processes to run so smoothly.

When you act as IAMT/EAMT president, or edit the Machine Translation journal, or act as track chair at major conferences, sometimes you have to be a bit of a pain, because you are often asking busy people to do things, mostly for free! Having been around the block a few times, I have lots of contacts in the industry, so I made myself responsible for bringing in sponsorship. I know they will say I was close to pestering them on many occasions, but I am truly grateful for the hugely generous support we obtained from our sponsors from the translation and CAT industry:

- Silver sponsors: Microsoft, and STAR
- Bronze sponsors: Pangeanic, text & form, CrossLang, Flitto, VistaTec
Other sponsors: Welocalize, Iconic, XTM, Unbabel, DCU, ELRA, Tilde, Springer, Apertium

I am also extremely grateful to Fáilte Ireland for their generous support of this conference, which helped my ADAPT@DCU colleagues Joss Moorkens and Sharon O’Brien present our bid in Nagoya in 2017, as well as supporting our excellent invited speakers: Laura Casanellas, Helena Moniz, and Arianna Bisazza. With many women in our team, it’s extremely important to have strong female role models, and we could not have asked for better from Laura, Arianna and Helena; many thanks to all of you for agreeing to share your expertise with us!

We took the decision a while back to try to be as green a conference as possible. You will already have noticed that, in order to reduce waste, there is no delegate bag. To reduce paper, we are not producing paper proceedings, and the normal programme booklet has been replaced by a smaller ‘bradge’ which doubles as a name badge. We are hoping to have a tree-planting ceremony during the conference in order to reduce the carbon footprint of the Summit. To reduce transport costs, we are using onsite accommodation at DCU, and will promote the use of public transport to the off-site events. Thanks to DCU Sustainability Manager Sam Fahy for her support in these efforts.

While we decided not to produce printed proceedings, they still needed to be put together in electronic form. I am grateful to Jenny Walsh for producing such an excellent conference logo, but huge thanks are due to Alberto Poncelas for putting together the proceedings, and for helping workshop chairs to produce theirs. Alberto has also liaised with Matt Post to ensure that your papers are indexed in perpetuity on the ACL Anthology!

I have two final people to thank. Firstly, I am very grateful to Grainne McQuaid and her team in Abbey Conference and Events for their professional support of the conference. You will have met them at registration, and they are available throughout the event to ensure your needs are met. We have been engaging with them for 2 years now, and this has been a true partnership that has made this journey an enjoyable one. Secondly, I am especially grateful to my colleague Jane Dunne, for managing the planning of the conference, and for managing me too. Jane has done this over and above her work on a European project, and I could not have chosen a better, nicer person to engage with over these past two years – thank you Jane on behalf of everyone; we are all deeply grateful for your huge effort in getting us to where we all are today!

Finally, I really hope that 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. I am fortunate indeed that many of my very best friends are in the MT community, and I hope to meet as many of you as possible during the event.

Andy Way
Chair, MT Summit 2019
Deputy Director ADAPT Centre School of Computing Dublin City University
Dublin, Ireland.
Email: Andy.Way@adaptcentre.ie
Foreword by the Research Track Program Chairs

It is with great pleasure that we present the proceedings of the research track at MT Summit XVII. We solicited full papers that present novel research contributions across all areas of MT, particularly encouraging submissions that are oriented towards building robust and practical systems.

We received 57 submissions for the research track. Out of these, 1 was withdrawn before the notification of acceptance, and 31 were accepted (54%). 4 were withdrawn after the notification of acceptance, resulting in 27 papers being presented at the MT Summit.

The selected papers span a wide range of topics in machine translation. Major themes include translation from other modalities than text, especially speech, the analysis of MT models and automatic translations, and the integration of MT into translation workflows, with studies on post-editing and constrained/controllable MT.

We would like to thank the members of the Program Committee for their timely reviews.

Barry Haddow and Rico Sennrich
Organizers

President of the IAMT and the EAMT
Mikel Forcada University of Alicante

MT Summit 2019
Conference Chair
Andy Way ADAPT Centre

Programme Chairs

Research track co-chairs
Barry Haddow The University of Edinburgh
Rico Sennrich The University of Edinburgh

User track co-chairs
John Tinsley Iconic Translation Machines
Dimitar Shterionov Dublin City University

Translator track co-chairs
Celia Rico Universidad Europea de Madrid
Federico Gaspari Dublin City University

Projects track co-chair
Mikel L. Forcada University of Alicante

Program Committee

Research Track
Alberto Poncelas DCU
Aizhan Imankulova Tokyo Metropolitan University
Aleš Tamchyna Memsource a. s.
Alon Lavie Carnegie Mellon University
Anabela Barreiro INESC-ID
Andreas Maletti Universität Leipzig
Andrei Popescu-Belis HEIG-VD / HES-SO
Ankur Bapna Google
Ann Clifton Simon Fraser University
Annette Rios Gonzales University of Zurich
Anoop Kunchukuttan IIT Bombay
Antonio Toral University of Groningen
Antonio Valerio Miceli Barone University of Pisa
Arturo Oncevay The University of Edinburgh
Arul Menezes Microsoft
Arya McCarthy Johns Hopkins University
Atsushi Fujita National Institute of Information and Communications Technology
Helena Caseli, Federal University of São Carlos (UFSCar)
Helmut Schmid, Ludwig Maximilian University of Munich
Hiroshi Echizenya, Hokkai-Gakuen University
Houda Bouamor, Carnegie Mellon University
Huda Khayrallah, Johns Hopkins University
Iacer Calixto, University of Amsterdam
Isao Goto, NHK
Jan Niehues, Maastricht University
Jesús González-Rubio, WebInterpret
Jiajun Zhang, Institute of Automation Chinese Academy of Sciences
Joachim Daiber, Apple Inc.
John Henderson, The MITRE Corporation
John Tinsley, Iconic Translation
Joke Daems, Ghent University
Jonathan Mallinson, The University of Edinburgh
Joost Bastings, University of Amsterdam
Jörg Tiedemann, University of Helsinki
José G. C. de Souza, ebay
Josep Crego, SYSTRAN
Joss Moorkens, ADAPT Centre
Juan Antonio Pérez-Ortiz, Universitat d’Alacant; Departament de Llenguatges i Sistemes Informàtics
Julia Ive, King’s College London
Julia Kreutzer, Heidelberg University
Julian Schamper, DeepL GmbH
Katsuhito Sudoh, Nara Institute of Science and Technology
Ke Hu, ADAPT Centre; Dublin City University
Kenji Imamura, National Institute of Information and Communications Technology
Kenton Murray, Carnegie Mellon University School of Computer Science
Laura Jehl, Institut für Computerlinguistik; Universität Heidelberg
Laurent Besacier, Laboratoire d’Informatique de Grenoble
Lemao Liu, NICT
Lexi Birch, The University of Edinburgh
Lieve Macken, Ghent University
Linfeng Song, University of Rochester
Luisa Bentivogli, FBK
Maja Popovic, ADAPT Centre; DCU
Manny Rayner, University of Geneva
Mara Chinea Rios, Universitat Politècnica de València
Marc DyNETman, Xerox Research Centre Europe
Marcello Federico, Amazon AI
Marcin Junczys-Dowmunt, Microsoft
Marco Turchi, Fondazione Bruno Kessler
Marianna Apidianaki, CNRS
Marija Brkic, Department of Informatics; University of Rijeka
Marion Weller-Di Marco, University of Amsterdam
Mark Fishel, University of Tartu
<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rico Sennrich</td>
<td>The University of Edinburgh</td>
</tr>
<tr>
<td>Roland Kuhn</td>
<td>National Research Council of Canada</td>
</tr>
<tr>
<td>Roman Grundkiewicz</td>
<td>The University of Edinburgh; School of Informatics</td>
</tr>
<tr>
<td>Rudolf Rosa</td>
<td>Charles University</td>
</tr>
<tr>
<td>Saab Mansour</td>
<td>RWTH Aachen University</td>
</tr>
<tr>
<td>Sadao Kurohashi</td>
<td>Kyoto University</td>
</tr>
<tr>
<td>Sameen Maruf</td>
<td>Monash University</td>
</tr>
<tr>
<td>Sameer Bansal</td>
<td>The University of Edinburgh</td>
</tr>
<tr>
<td>Samuel Läubli</td>
<td>University of Zurich</td>
</tr>
<tr>
<td>Sarah Ebling</td>
<td>University of Zurich</td>
</tr>
<tr>
<td>Sergio Penkale</td>
<td>Lingo24</td>
</tr>
<tr>
<td>Shahram Khadivi</td>
<td>eBay</td>
</tr>
<tr>
<td>Shankar Kumar</td>
<td>Google</td>
</tr>
<tr>
<td>Sharon O’Brien</td>
<td>Dublin City University</td>
</tr>
<tr>
<td>Sheila Castilho</td>
<td>Dublin City University/ADAPT Centre</td>
</tr>
<tr>
<td>Stephan Peitz</td>
<td>Apple</td>
</tr>
<tr>
<td>Surafe Melaku Lakew</td>
<td>University of Trento</td>
</tr>
<tr>
<td>Tamer Alkhouli</td>
<td>RWTH Aachen University</td>
</tr>
<tr>
<td>Taro Watanabe</td>
<td>NICT</td>
</tr>
<tr>
<td>Teresa Herrmann</td>
<td>Fujitsu</td>
</tr>
<tr>
<td>Tim Anderson</td>
<td>Wright-Patterson Air Force Research Laboratory</td>
</tr>
<tr>
<td>Tomáš Musil</td>
<td>Charles University in Prague</td>
</tr>
<tr>
<td>Toshiaki Nakazawa</td>
<td>The University of Tokyo</td>
</tr>
<tr>
<td>Tsuyoshi Okita</td>
<td>Kyushu Institute of Technology</td>
</tr>
<tr>
<td>Ulrich Germann</td>
<td>The University of Edinburgh</td>
</tr>
<tr>
<td>Víctor M. Sánchez-Cartagena</td>
<td>Transducens Research Group; Departament de Llenguatges i Sistemes Informàtics; Universitat d’Alacant</td>
</tr>
<tr>
<td>Viktor Hangya</td>
<td>Ludwig Maximilian University of Munich</td>
</tr>
<tr>
<td>Vincent Vandeghinste</td>
<td>Instituut voor de Nederlandse Taal, Centre for Computational Linguistics, KU Leuven</td>
</tr>
<tr>
<td>Vishal Chowdhary</td>
<td>Microsoft</td>
</tr>
<tr>
<td>Vu Hoang</td>
<td>The University of Melbourne</td>
</tr>
<tr>
<td>Wei Wang</td>
<td>Google</td>
</tr>
<tr>
<td>Xing Niu</td>
<td>University of Maryland</td>
</tr>
<tr>
<td>Yinfei Yang</td>
<td>Redfin Inc.</td>
</tr>
<tr>
<td>Yuki Arase</td>
<td>Osaka University</td>
</tr>
<tr>
<td>Yunsu Kim</td>
<td>RWTH Aachen University</td>
</tr>
<tr>
<td>Yvette Graham</td>
<td>Dublin City University</td>
</tr>
</tbody>
</table>

**Translator Track**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrià Martín-Mor</td>
<td>UAB</td>
</tr>
<tr>
<td>Alberto Poncelas</td>
<td>DCU</td>
</tr>
<tr>
<td>Ana Guerberof Arenas</td>
<td>DCU/ADAPT Centre</td>
</tr>
<tr>
<td>Bogdan Babych</td>
<td>University of Leeds</td>
</tr>
<tr>
<td>Carolina Scarton</td>
<td>European Association for Machine Translation</td>
</tr>
<tr>
<td>Celia Rico Perez</td>
<td>Universidad Europea de Madrid</td>
</tr>
<tr>
<td>Federico Gaspari</td>
<td>Dublin City University</td>
</tr>
<tr>
<td>Ignacio Garcia</td>
<td>University of Western Sydney</td>
</tr>
<tr>
<td>Joss Moorkens</td>
<td>Dublin City University</td>
</tr>
</tbody>
</table>
Katie Botkin  
Mary Phelan  
Minako O’Hagan  
Mirko Silvestrini  
Olga Torres-Hostench  
Oliver Czulo  
Pilar Sanchez-Gijón  
Sarah Bawa-Mason  
Sharon O’Brien  
Stephen Doherty  
Valeria Barbero  

**User Track**

Alberto Poncelas  
Aljoscha Burchardt  
Aalon Lavie  
Bram Bulté  
Bruno Pouliquen  
Carlos Collantes Fraile  
Carmen Heger  
Carolina Scarton  
Chao-Hong Liu  
Charlotte Tesselhar  
Chris Wendt  
Christian Federmann  
Christian Eisold  
Dag Schmidtke  
Daniel Stein  
David Vilar  
Declan Groves  
Dimitar Shiterionov  
Eva Martínez Garcia  
Eva Vanmassenhove  
Evgeny Matusov  
Félix Do Carmo  
Fred Blain  
Gema Ramírez-Sánchez  
Guodong Xie  
Heidi Depraetere  
Jean Senellart  
John Tinsley  
José G. C. de Souza  
Keith J. Miller  
Kim Harris  
Laurent Chevrrette  
Maxim Khalilov  
Mercedes García-Martínez  
Nathalie DeSutter  
Nicola Ueffing  

MultiLingual magazine  
Dublin City University  
The University of Auckland  
UNILINGUE  
Universitat Autònoma de Barcelona  
Universität Leipzig  
Autonomous University of Barcelona  
University of Portsmouth/Institute of Translation and Interpreting  
Dublin City University  
The University of New South Wales  
MT Summit  

DCU  
DFKI  
Carnegie Mellon University  
Katholieke Universiteit Leuven  
World Intellectual Property Organization  
TransPerfect / Universidad Complutense de Madrid  
SZ  
European Association for Machine Translation  
ADAPT Centre, Dublin City University  
LexisNexis Univentio  
Microsoft  
berns language consulting GmbH  
Microsoft  
eBay Inc.  
Amazon  
Microsoft  
Dublin City University  
Vicomtech / Universitat Politècnica de Catalunya  
DCU  
eBay  
ADAPT Centre  
The University of Sheffield  
Prompsit Language Engineering, S.L.  
ADAPT Centre, Dublin City University  
Cross Language NV  
SYSTRAN  
Iconic Translation Machines  
eBay Inc.  
The MITRE Corporation  
text&form GmbH  
Mondzo  
Unbabel  
Pangeanic SL  
Untranslate  
eBay
Olga Beregovaya  
Patrik Lambert  
Phil Ritchie  
Raj Pate  
Rohit Gupta  
Saša Hasan  
Silvio Picinini  
Steve Richardson  
Thierry Etchegoyhen  
Tony O’Dowd  
Yury Sharshov

**Project Track**
Alberto Poncelas  
Carolina Scarton  
Mikel Forcada

Welocalize  
Pompeu Fabra University  
Vistattec  
CDAC  
Iconic Translations Machines Ltd.  
Apple  
eBay  
The Church of Jesus Christ of Latter-day Saints  
Vicomtech-IK4  
Xcelerator Machine Translations Ltd.  
LexisNexis Univentio  
DCU  
The University of Sheffield  
University of Alicante
Contents

Online Sentence Segmentation for Simultaneous Interpretation using Multi-Shifted Recurrent Neural Network
   Xiaolin Wang, Masao Utiyama and Eiichiro Sumita 1

Robust Document Representations for Cross-Lingual Information Retrieval in Low-Resource Settings
   Mahsa Yarmohammadi, Xutai Ma, Sorami Hisamoto, Muhammad Rahman, Yiming Wang, Hainan Xu, Daniel Povey, Philipp Koehn and Kevin Duh 12

Enhancing Transformer for End-to-end Speech-to-Text Translation
   Mattia Antonino Di Gangi, Matteo Negri, Roldano Cattoni, Roberto Dessi and Marco Turchi 21

Debiasing Word Embeddings Improves Multimodal Machine Translation
   Tosho Hirasawa and Mamoru Komachi 32

Translator2Vec: Understanding and Representing Human Post-Editors
   António Góis and André F. T. Martins 43

Domain Adaptation for MT: A Study with Unknown and Out-of-Domain Tasks
   Hoang Cuong 55

What is the impact of raw MT on Japanese users of Word: preliminary results of a usability study using eye-tracking
   Ana Guerberof Arenas, Joss Moorkens and Sharon O’Brien 67

   Randy Scansani, Luisa Bentivogli, Silvia Bernardini and Adriano Ferraresi 78

Automatic error classification with multiple error labels
   Maja Popovic and David Vilar 87

Interactive-Predictive Neural Machine Translation through Reinforcement and Imitation
   Tsz Kin Lam, Shigehiko Schamoni and Stefan Riezler 96

An Intrinsic Nearest Neighbor Analysis of Neural Machine Translation Architectures
   Hamidreza Ghader and Christof Monz 107

Improving Neural Machine Translation Using Noisy Parallel Data through Distillation
   Praveen Dakwale and Christof Monz 118

Exploiting Out-of-Domain Parallel Data through Multilingual Transfer Learning for Low-Resource Neural Machine Translation 128