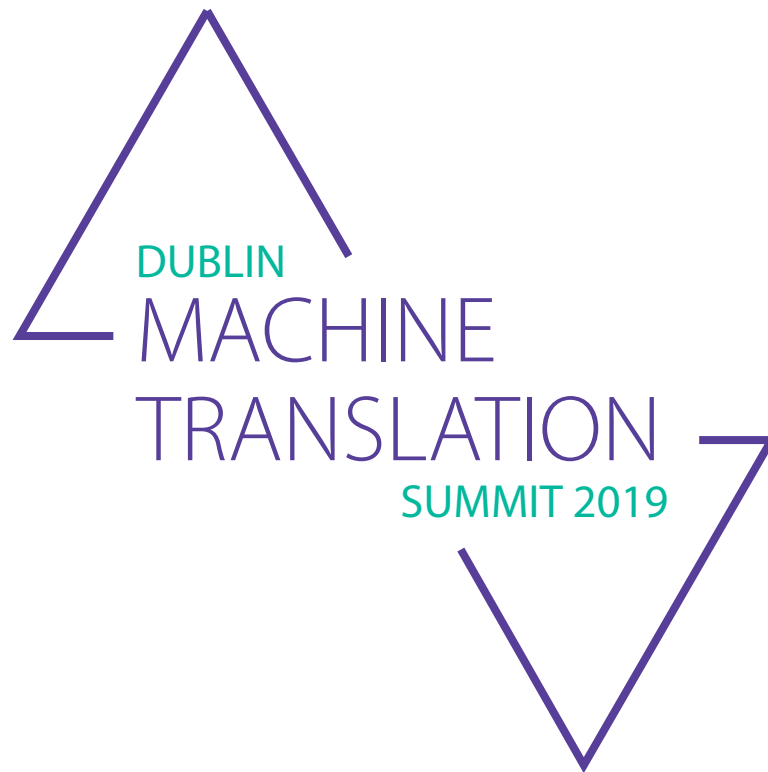


# Machine Translation Summit XVII



The 2nd Workshop on Technologies for MT of Low  
Resource Languages (LoResMT 2019)

<https://sites.google.com/view/loresmt/>

20 August, 2019  
Dublin, Ireland

# The 2nd Workshop on Technologies for MT of Low Resource Languages (LoResMT 2019)

<https://sites.google.com/view/loresmt/>

20 August, 2019  
Dublin, Ireland



© 2019 The authors. These articles are licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

## Preface from the co-chairs of the workshop

Machine translation (MT) technologies have been improved significantly in the last two decades, with the developments on phrased-based statistical MT (SMT) and recently the neural MT (NMT). However, most of these methods rely on the availability of large parallel data (millions to tens of millions sentence pairs) in the training, which are resources that do not exist in many language pairs. The development of monolingual MT is a recent approach that enables building MT systems without parallel data. However, a large amount of monolingual corpus is still required to train this kind of MT systems.

The workshop solicits papers on MT systems/methods for low resource languages in general. We hope to provide a forum for researchers working on MT for low resource languages and relevant NLP tools from our community. This year the LoResMT proceeding archives MT research on languages from all over the world, e.g. Crimean Tatar, Dravidian, Irish, Malayalam, Shipibo-Konibo, Torwali, Turkish. In addition, research on no-resource situation and the use of NMT for low resource languages will also be presented in LoResMT. Shared Tasks on MT for Bhojpuri, Magahi, Sindhi and Latvian (to and from English) is also organized under LoResMT, by Atul Kr. Ojha, Valentin Malykh, Pinkey Nainwani, Varvara Logacheva and Chao-Hong Liu. Two system description papers are archived in the proceeding.

In the LoResMT workshop, two corpus-building research teams will introduce their on-going work and resulting linguistic resources. They are DARPA LORELEI Program team led by Jennifer Tracey that curates corpora of low resource languages globally, and Minzu University of China team led by Xiaobing Zhao that builds corpora of minority languages in China, e.g. Mongolian, Tibetan and Uyghur. Researchers, led by Suo-nancairang, who work on the Tibetan language from Qinghai Normal University will also present their research at LoResMT.

We would like to express our sincere gratitude to the many researchers who helped as organizers, and reviewers and made the workshop successful. We are especially thankful to MT Summit organizers Andy Way, Antonio Toral, Jane Dunne for their continuous help on the workshop, and Alberto Poncelas for his various helps including the preparation of the proceeding. We are very grateful to the authors who submitted their work to the workshop and come to exchange their research at the venue. Thank you so much!

**Chao-Hong Liu and Alina Karakanta**

# Organizers

## Workshop Chairs

Alina Karakanta	Fondazione Bruno Kessler (FBK)
Atul Kr. Ojha	Panlingua Language Processing LLP, New Delhi
Chao-Hong Liu	ADAPT Centre, Dublin City University
Jonathan Washington	Swarthmore College
Nathaniel Oco	National University (Phillippines)
Surafel Melaku Lakew	Fondazione Bruno Kessler (FBK)
Valentin Malykh	Huawei, Moscow Institute of Physics and Technology
Xiaobing Zhao	Minzu University of China

## Program Committee

Alberto Poncelas	ADAPT Centre, Dublin City University
Alina Karakanta	Fondazione Bruno Kessler (FBK)
Atul Kr. Ojha	Panlingua Language Processing LLP, New Delhi
Chao-Hong Liu	ADAPT Centre, Dublin City University
Daria Dzendzik	ADAPT Centre, Dublin City University
Eva Vanmassenhove	ADAPT Centre, Dublin City University
Koel Dutta Chowdhury	Saarland University
Majid Latifi	Universitat Politècnica de Catalunya
Meghan Dowling	ADAPT Centre, Dublin City University
Nathaniel Oco	National University (Phillippines)
Sangjee Dondrub	Qinghai Normal University
Santanu Pal	Saarland University
Surafel Melaku Lakew	Fondazione Bruno Kessler (FBK)
Tewodros Abebe	Addis Ababa University
Thepchai Supnithi	NECTEC
Valentin Malykh	Huawei, Moscow Institute of Physics and Technology
Vinit Ravishankar	University of Oslo
Yalemisew Abgaz	ADAPT Centre, Dublin City University

# Contents

<b>Finite State Transducer based Morphology analysis for Malayalam Language</b>	<b>1</b>
<i>Santhosh Thottingal</i>	
<b>A step towards Torwali machine translation: an analysis of morphosyntactic challenges in a low-resource language</b>	<b>6</b>
<i>Naeem Uddin and Jalal Uddin</i>	
<b>Workflows for kickstarting RBMT in virtually No-Resource Situation</b>	<b>11</b>
<i>Tommi A Pirinen</i>	
<b>A Continuous Improvement Framework of Machine Translation for Shipibo-Konibo</b>	<b>17</b>
<i>Héctor Erasmo Gómez Montoya, Kervy Dante Rivas Rojas and Arturo Oncevay</i>	
<b>Machine Translation for Crimean Tatar to Turkish</b>	<b>24</b>
<i>Memduh Gökırmak, Francis Tyers and Jonathan Washington</i>	
<b>Developing a Neural Machine Translation system for Irish</b>	<b>32</b>
<i>Arne Defauw, Sara Szoc, Tom Vanallemeersch, Anna Bardadym, Joris Brabers, Frederic Everaert, Kim Scholte, Koen Van Winckel and Joachim Van den Bogaert</i>	
<b>Sentence-Level Adaptation for Low-Resource Neural Machine Translation</b>	<b>39</b>
<i>Aaron Mueller and Yash Kumar Lal</i>	
<b>Corpus Building for Low Resource Languages in the DARPA LORELEI Program</b>	<b>48</b>
<i>Jennifer Tracey, Stephanie Strassel, Ann Bies, Zhiyi Song, Michael Arrigo, Kira Grifitt, Dana Delgado, Dave Graff, Seth Kulick, Justin Mott and Neil Kuster</i>	
<b>Multilingual Multimodal Machine Translation for Dravidian Languages utilizing Phonetic Transcription</b>	<b>56</b>
<i>Bharathi Raja Chakravarthi, Ruba Priyadharshini, Bernardo Stearns, Arun Jayapal, Sridevy S, Mihael Arcan, Manel Zarrouk and John P McCrae</i>	
<b>A3-108 Machine Translation System for LoResMT 2019</b>	<b>64</b>
<i>Saumitra Yadav, Vandan Mujadia and Manish Shrivastava</i>	
<b>Factored Neural Machine Translation at LoResMT 2019</b>	<b>68</b>
<i>Saptarashmi Bandyopadhyay</i>	
<b>JHU LoResMT 2019 Shared Task System Description</b>	<b>72</b>
<i>Paul McNamee</i>	

