

DiscoMT 2019

**The Fourth Workshop on Discourse
in Machine Translation**

Proceedings of the Workshop

November 3, 2019
Hong Kong, China

©2019 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-950737-74-1

Preface

The DiscoMT series of workshops explores the challenges and opportunities that appear when translating entire texts, as opposed to sentences in isolation. Started in 2013, the workshops are a forum for discussing novel and prospective strategies to take advantage of inter-sentential context when performing machine translation. Already a stimulating research question in the days of phrase-based statistical MT systems, the use of text-level context is both a necessity, as it enables systems to make correct translation choices, and an opportunity, as it may provide crucial information that is not available locally.

For these reasons, when translating entire texts, one cannot ignore text-level properties. This is becoming increasingly clear in neural machine translation (NMT), where text-level aspects of translation may be one of the obstacles to high-quality automatic translation for high-resource languages, after the advances in translation quality observed in recent years. Indeed, while MT of sentences removed from their contexts may seem to have reached quality levels comparable to human translations, experts still clearly prefer entire texts from human translators, as several recent evaluation studies have shown.

The first three editions of DiscoMT – held every two years – have helped to consolidate a small but thriving community of researchers. Considerable effort has been expended recently on document-level MT, such that now, several individuals and/or groups are working on similar or overlapping problems. Notable efforts include work on document-level influences on lexical choice in SMT and NMT, methods and annotated resources for discourse-level MT, discourse-sensitive assessment metrics, and specific discourse phenomena in SMT and NMT.

As exemplified by the papers presented at previous editions of DiscoMT and this year’s main NLP conferences, specific research topics in document-level MT are: NMT extensions taking into consideration context from multiple sentences or entire documents; pronoun translation between languages which differ in pronoun usage; explicitation/implicitation in translating discourse connectives; context-aware translation of ambiguous terms; assessing document-level properties of MT output, including coherence; and preserving document-level properties characteristic of register, genre, and other types of text variation.

In addition to the invited talks, DiscoMT 2019 will feature oral and poster presentations of studies at the intersection of machine translation (under any of its paradigms) and discourse, from a variety of perspectives. Along with the peer-reviewed articles submitted and accepted to DiscoMT, the workshop has also invited a number of posters from EMNLP-IJCNLP, to diversify and enrich the poster session. The program thus includes a variety of MT models, especially neural ones, that consider larger contexts than state-of-the-art ones do, along with assessments of their capabilities to correctly translate discourse-level dependencies.

We hope that workshops such as this one will continue to stimulate work on Discourse and Machine Translation, in a wide range of discourse phenomena and MT architectures.

We would like to thank all the authors who submitted papers to the workshop, as well as all the members of the Program Committee who reviewed the submissions and delivered thoughtful, informative reviews.

The Chairs
October 5, 2019

Chairs

Christian Hardmeier, Uppsala University, Sweden and University of Edinburgh, UK
Sharid Loáiciga, University of Gothenburg, Sweden
Andrei Popescu-Belis, HEIG-VD/HES-SO, Switzerland
Deyi Xiong, Tianjin University, China

Program Committee:

Marine Carpuat, University of Maryland, USA
Marta Costa-jussà, Universitat Politècnica de Catalunya, Spain
Zhengxian Gong, Soochow University, Suzhou, China
Francisco Guzmán, Facebook, USA
Yulia Grishina, University of Postdam and Amazon, Berlin, Germany
Gholamreza Haffari, Monash University, Clayton, Australia
Shafiq Joty, Nanyang Technological University, Singapore
Ekaterina Lapshinova-Koltunski, Saarland University, Saarbrücken, Germany
Lesly Miculicich Werlen, Idiap Research Institute, Martigny, Switzerland
Preslav Nakov, Qatar Computing Research Institute, Doha, Qatar
Michal Novak, Charles University, Prague, Czech Republic
Nikolaos Pappas, Idiap Research Institute, Martigny, Switzerland
Lucie Poláková, Charles University, Prague, Czech Republic
Maja Popovic, Adapt Centre, Dublin City University, Ireland
Annette Rios, University of Zurich, Switzerland
Carol Scarton, University of Sheffield, UK
Rico Sennrich, University of Edinburgh, UK and University of Zurich, Switzerland
Jörg Tiedemann, University of Helsinki, Finland
Yannick Versley, Independent consultant, IBM Services, Germany
Martin Volk, University of Zurich, Switzerland
Bonnie Webber, University of Edinburgh, UK
Kellie Webster, Google, New York, USA
Min Zhang, Soochow University, Suzhou, China
Sandrine Zufferey, University of Bern, Switzerland

Table of Contents

<i>Analysing Coreference in Transformer Outputs</i>	
Ekaterina Lapshinova-Koltunski, Cristina España-Bonet and Josef van Genabith	1
<i>Context-Aware Neural Machine Translation Decoding</i>	
Eva Martínez Garcia, Carles Creus and Cristina España-Bonet	13
<i>When and Why is Document-level Context Useful in Neural Machine Translation?</i>	
Yunsu Kim, Duc Thanh Tran and Hermann Ney	24
<i>Data augmentation using back-translation for context-aware neural machine translation</i>	
Amane Sugiyama and Naoki Yoshinaga	35
<i>Context-aware Neural Machine Translation with Coreference Information</i>	
Takumi Ohtani, Hidetaka Kamigaito, Masaaki Nagata and Manabu Okumura	45
<i>Analysing concatenation approaches to document-level NMT in two different domains</i>	
Yves Scherrer, Jörg Tiedemann and Sharid Loáiciga	51

Conference Program

Sunday, November 3, 2019

9:00–9:05 *Welcome*

9:05–9:50 **Session 1: Keynote Talk**

Document-level Machine Translation: the Current State and the Challenges
Professor Qun Liu, Noah’s Ark Lab, Huawei

9:50–10:30 **Session 2: Poster Spotlights**

10:30–11:00 *Coffee Break*

11:00–12:30 **Session 3: Posters**

Analysing Coreference in Transformer Outputs
Ekaterina Lapshinova-Koltunski, Cristina España-Bonet and Josef van Genabith

Context-Aware Neural Machine Translation Decoding
Eva Martínez Garcia, Carles Creus and Cristina España-Bonet

When and Why is Document-level Context Useful in Neural Machine Translation?
Yunsu Kim, Duc Thanh Tran and Hermann Ney

Data Augmentation using Back-translation for Context-aware Neural Machine Translation
Amane Sugiyama and Naoki Yoshinaga

Context-aware Neural Machine Translation with Coreference Information
Takumi Ohtani, Hidetaka Kamigaito, Masaaki Nagata and Manabu Okumura

Analysing Concatenation Approaches to Document-level NMT in two Different Domains
Yves Scherrer, Jörg Tiedemann and Sharid Loáiciga

Additional invited posters from the main conference

