

RANLP 2017

**The First Workshop on Human-Informed Translation and
Interpreting Technology (HiT-IT)**

Proceedings of the Workshop

September 7th, 2017

Varna, Bulgaria

Proceedings of the Workshop
Human-Informed Translation and Interpreting Technology
associated with RANLP 2017

Varna, Bulgaria
7 September 2017

ISBN 978-954-452-042-7

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Introduction

Human Translation and Machine Translation (MT) aim to solve the same problem (i.e. translating from one language into another) in two seemingly different ways.

There are many Natural Language Processing (NLP)/Computational linguistics efforts towards improving the work of translators and interpreters (for example Computer-Assisted Translation (CAT) tools, electronic dictionaries, concordancers, spell-checkers, terminological databases and terminology extraction tools, translation memories, partial machine translation of template documents, speech recognition systems for automatic subtitling, etc.). In turn, the NLP field makes use of the work and the knowledge of professional translators and interpreters to build models for automatic translation - e.g. by using parallel aligned text and speech corpora for text and speech machine translation learning, human evaluators of machine translation output, human annotations for automatic MT post-editing or using eye-tracking for learning editing patterns of professional translators, etc.

While there have been many workshops and conferences representing both sides: 1) Machine Translation in NLP (e.g. WMT, EAMT conferences), and 2) Automatic tools for translators and interpreters in Translation/Interpreting studies (e.g. Translating and The Computer, and the MT Summit conferences), there has not been a common publication & discussion venue for both sides. What makes our workshop unique is that it is a unified workshop which welcomes the contributions of both fields towards each other.

This workshop addresses BOTH the most recent developments in contributions of NLP to translation/interpreting and the contributions of translation/interpreting to NLP/MT. In this way it addresses the interests of researchers & specialists in both areas and their joint collaborations, aiming for example to improve their own tasks with the techniques & knowledge of the other field or to help the development of the other field with their own techniques & knowledge.

This year, we have received 11 high quality submissions, among which 8 have been accepted. Two articles have been accepted as full papers, and the rest six as short papers. Each submission has been reviewed by at least 2 reviewers, who were highly specialized experts either in human contributions to Machine Translation, or in Interpreting and Translation technologies, both from the computer science and Translation/Interpreting studies fields. The papers covered a broad spectrum of topics, such as:

- Automatic text simplification for translators
- Arabic dialects corpora creation for translators, interpreters and machine translation
- Extracting interpreting strategies, in order to improve speech-to-text machine translation
- Comparing machine translation and human translation
- NLP approaches & systems for building educational tools & resources for interpreters
- NLP approaches & systems for building educational tools & resources for translators
- Computer-assisted translation tools, such as translation memories, machine translation, etc.
- Translation resources, such as corpora, terminological databases, dictionaries
- Computer-assisted interpreting software, such as interpreters workbench, etc.
- Interpreting resources, such as corpora, terminological databases, dictionaries

- User requirements for interpreting and translation tools
- Methodologies for collecting user requirements
- Human accuracy metrics and human evaluation of machine translation
- Theoretical papers with translators/interpreters views on how machine translation should work/what output should produce
- Pre-editing and post-editing of machine translation
- Theoretical papers and practical applications on applying translation techniques & knowledge to NLP and machine translation
- Theoretical papers and practical applications on applying interpreting techniques & knowledge to NLP and machine translation

Our authors' nationalities & affiliations covered a wide range of countries, including: United Kingdom, Austria, Spain, Bulgaria, Jordan, Egypt, Qatar, Italy, Germany, Morocco, and United States.

The workshop also featured two invited talks on the topics of “The translation world and beyond... What's next? Don't get me wrong...”, by Prof. Dr. Ruslan Mitkov, Director of the Research Institute in Information and Language Processing at the University of Wolverhampton, UK, and “Why are Computers (Still) Bad Translators?”, provided by Dr. Preslav Nakov, Senior Scientist at the Qatar Computing Research Institute, HBKU. The workshop concluded with a vibrant round-table discussion.

We would like to thank the authors for submitting their articles to the HiT-IT Workshop, the members of the Programme Committee for their efforts to provide exhaustive reviews, even if the workshop was highly interdisciplinary, and the conference RANLP 2017, which has accepted us as a workshop and hosted us. We hope that all the participants received valuable feedback about their research as well as found their place in the fuzzy interdisciplinary field between human and machine translation.

Irina Temnikova, Constantin Orăsan, Gloria Corpas and Stephan Vogel,

Organisers of the First Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT) organised on 7 September 2017 Varna, Bulgaria.

Organizers:

Irina Temnikova, Qatar Computing Research Institute, HBKU, Qatar
Constantin Orăsan, University of Wolverhampton, UK
Gloria Corpas Pastor, University of Malaga, Spain
Stephan Vogel, Qatar Computing Research Institute, HBKU, Qatar

Program Committee:

Ahmed Abdelali, Qatar Computing Research Institute, HBKU, Qatar
Claudio Bendazzoli, Università degli studi di Torino, Italy
Pierrette Bouillon, University of Geneva, Switzerland
Michael Carl, Copenhagen Business School, Denmark
Stephen Doherty, The University of New South Wales, Australia
Claudio Fantinuolli, University of Mainz, Germany
Marcello Federico, Fondazione Bruno Kessler, Trento
Veronique Hoste, University of Ghent, Belgium
Maarit Koponen, University of Turku, Finland
Lieve Macken, University of Ghent, Belgium
Ruslan Mitkov, University of Wolverhampton, UK
Johanna Monti, University of Naples, Italy
Hamdy Mubarak, Qatar Computing Research Institute, HBKU, Qatar
Preslav Nakov, Qatar Computing Research Institute, HBKU, Qatar
Sharon O'Brien, Dublin City University, Ireland
Santanu Pal, Saarland University, Germany
Carla Parra, ADAPT Centre, SALIS, Dublin City University, Ireland
Maja Popovic, Humboldt University of Berlin, Germany
Pablo Romero-Fresco, University of Roehampton, UK
Violeta Seretan, University of Geneva, Switzerland
Cristina Toledo, University of Córdoba, Spain
Victoria Yaneva, University of Wolverhampton, UK
Anna Zaretskaya, University of Malaga, Spain

Invited Speakers:

Ruslan Mitkov, University of Wolverhampton, UK
Preslav Nakov, Qatar Computing Research Institute, HBKU, Qatar

Table of Contents

<i>Enhancing Machine Translation of Academic Course Catalogues with Terminological Resources</i> Randy Scansani, Silvia Bernardini, Adriano Ferraresi, Federico Gaspari and Marcello Soffritti . . .	1
<i>Experiments in Non-Coherent Post-editing</i> Cristina Toledo Báez, Moritz Schaeffer and Michael Carl	11
<i>Comparing Machine Translation and Human Translation: A Case Study</i> Lars Ahrenberg	21
<i>TransBank: Metadata as the Missing Link between NLP and Traditional Translation Studies</i> Michael Ustaszewski and Andy Stauder	29
<i>Interpreting Strategies Annotation in the WAW Corpus</i> Irina Temnikova, Ahmed Abdelali, Samy Hedaya, Stephan Vogel and Aishah Al Daher	36
<i>Translation Memory Systems Have a Long Way to Go</i> Andrea Silvestre Baquero and Ruslan Mitkov	44
<i>Building Dialectal Arabic Corpora</i> Hani Elgabou and Dimitar Kazakov	52
<i>Towards Producing Human-Validated Translation Resources for the Fula language through WordNet Linking</i> Khalil Mrini and Martin Benjamin	58

Conference Program

- 9:00–9:10** *Workshop opening*
- 9:10–10:00 *Why are Computers (Still) Bad Translators?*
Preslav Nakov (invited speaker)
- 10:00–10:30 *Enhancing Machine Translation of Academic Course Catalogues with Terminological Resources*
Randy Scansani, Silvia Bernardini, Adriano Ferraresi, Federico Gaspari and Marcello Soffritti
- 10:30–11:00** *Coffee break*
- 11:00–11:30 *Experiments in Non-Coherent Post-editing*
Cristina Toledo Báez, Moritz Schaeffer and Michael Carl
- 11:30–11:50 *Comparing Machine Translation and Human Translation: A Case Study*
Lars Ahrenberg
- 11:50–12:10 *TransBank: Metadata as the Missing Link between NLP and Traditional Translation Studies*
Michael Ustaszewski and Andy Stauder
- 12:10–12:30 *Interpreting Strategies Annotation in the WAW Corpus*
Irina Temnikova, Ahmed Abdelali, Samy Hedaya, Stephan Vogel and Aishah Al Daher
- 12:30–14:00** *Lunch break*
- 14:00–14:50 *The Translation World and Beyond... What's Next? Don't Get me Wrong...*
Ruslan Mitkov (invited speaker)
- 14:50–15:10 *Translation Memory Systems Have a Long Way to Go*
Andrea Silvestre Baquero and Ruslan Mitkov
- 15:10–15:30 *Building Dialectal Arabic Corpora*
Hani Elgabou and Dimitar Kazakov
- 15:30–16:00** *Coffee break*

16:00–16:20 *Towards Producing Human-Validated Translation Resources for the Fula language through WordNet Linking*
Khalil Mrini and Martin Benjamin

16:30–17:00 *Discussion and conclusions*