ACL-IJCNLP 2015

The 53rd Annual Meeting of the
Association for Computational Linguistics and the
7th International Joint Conference on Natural Language Processing

Proceedings of the Fourth Workshop on Hybrid Approaches to Translation (HyTra)

July 31, 2015
Beijing, China

©2015 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 978-1-941643-67-9

Introduction

Welcome to the Fourth Workshop on Hybrid Approaches to Translation (HyTra-4) held in conjunction with ACL-2015, Beijing !

The workshop series on Hybrid Approaches to Translation aims at providing a communication platform, building a research community and informing research agenda around theoretical and practical issues of Hybrid MT, and specifically – the problems, methodologies, resources and theoretical ideas which originate outside the mainstream MT paradigm, but have potential to enhance the quality of state-of-the-art MT systems. The workshop series fills in a gap in the current paradigm allowing researchers to explore new pathways of bringing together a diverse range of technologies, methods and tools into MT domain.

The current Fourth Workshop on Hybrid Approaches to Translation - HyTra-4 - builds on a successful series of past events held in conjunction with international conferences (up to this year all of them so far took place in Europe):

HyTra-1 was held (together with the ESIRMT workshop) as a joint 2-day workshop at EACL 2012, Avignon, France: http://www-lium.univ-lemans.fr/esirmt-hytra/

HyTra-2 was a 1-day workshop at ACL 2013, Sophia, Bulgaria: http://hytra.barcelonamedia.org/hytra 2013/

HyTra-3 was a 1-day workshop at the EACL 2014, Gothenburg, Sweden. This workshop for the first time included an Industry Session – with invited talks of representatives from several companies, such as BMMT, SDL, Systran, Tilde, Lingenio company representatives, which highlighted an emerging industrial uptake of the Hybrid MT field by major developers of industrial MT systems: http://parles.upf.edu/llocs/plambert/hytra/hytra2014/

HyTra workshops have attracted a good number of submissions and participants each time, and included invited talks, full papers, and poster sessions. The invited speakers were Philipp Koehn, Hermann Ney, Will Lewis and Chris Quirk, Hans Uszkoreit and Joakim Nivre. The range of topics covered addresses all the areas of linguistic analysis relevant to MT, such as morphology, syntax, discourse, named entity recognition, etc., and a range of underlying MT architectures – statistical and rule-based. The workshops allow sufficient time for panel discussions which take form of exploratory brainstorming sessions and address further pathways of the development and integration of the hybrid MT technologies.

For the HyTra-4 workshop we have accepted 9 papers, which appear in this volume. The workshop hosts the invited talk by Prof. Dr. Hinrich Schuetze, the Chair of Computational Linguistics and the Director of the Center for Information and Language Processing of the Ludwig Maximilian University of Munich, Germany. HyTra-4 also includes the Industry Session, which brings together academic and industrial researchers and practitioners, and is now becoming a traditional part of the workshop.

We hope HyTra-4 will become a successful continuation of the HyTra workshop series and will result in interesting discussions, ideas and collaborations.

Bogdan Babych, University of Leeds
Kurt Eberle, Lingenio GmbH, Heidelberg
Patrik Lambert, Pompeu Fabra University, Barcelona
Reinhard Rapp, University of Mainz
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Marta R. Costa-jussà, Instituto Politécnico Nacional, Mexico

Organizers:

Bogdan Babych, University of Leeds
Kurt Eberle, Lingenio GmbH, Heidelberg
Patrik Lambert, Pompeu Fabra University, Barcelona
Reinhard Rapp, University of Mainz
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Marta R. Costa-jussà, Instituto Politécnico Nacional, Mexico

Program Committee:

Ahmet Aker, University of Sheffield, UK
Bogdan Babych, University of Leeds, UK
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Alexey Baytin, Yandex, Moscow, Russia
Núria Bel, Universitat Pompeu Fabra, Barcelona, Spain
Anja Belz, University of Brighton, UK
Pierrette Bouillon, ISSCO/TIM/ETI, University of Geneva, Switzerland
Michael Carl, Copenhagen Business School, Denmark
Marta R. Costa-jussà, Instituto Politécnico Nacional, Mexico
Oliver Culo, University of Mainz, Germany
Kurt Eberle, Lingenio GmbH, Heidelberg, Germany
Andreas Eisele, DGT (European Commission), Luxembourg
Marcello Federico, Fondazione Bruno Kessler, Trento, Italy
Christian Federmann, Microsoft Research, Seattle, USA
Maxim Khalilov, BMMT, Germany
Udo Kruschwitz, University of Essex, UK
Patrik Lambert, Pompeu Fabra University, Barcelona, Spain
Yannick Parmentier, Orleans University, France
Reinhard Rapp, University of Mainz, Germany
Serge Sharoff, University of Leeds, UK
George Tambouratzis, Institute for Language and Speech Processing, Athens, Greece
Jörg Tiedemann, University of Uppsala, Sweden

Invited Speakers:

Hinrich Schütze, Ludwig Maximilian University of Munich, Germany
Gerard de Melo, Tsinghua University Beijing, China

Invited Companies of the Industrial Session:

Baidu
CCID TransTech
Lingenio GmbH

Table of Contents

Bootstrapping a hybrid deep MT system
João Silva, João Rodrigues, Luís Gomes and António Branco 1

Multi-system machine translation using online APIs for English-Latvian
Matīss Rikters 6

What a Transfer-Based System Brings to the Combination with PBMT
Aleš Tamchyna and Ondrej Bojar 11

Establishing sentential structure via realignments from small parallel corpora
George Tambouratzis and Vassiliki Pouli 21

Passive and Pervasive Use of Bilingual Dictionary in Statistical Machine Translation
Liling Tan 30

Automated Simultaneous Interpretation: Hints of a Cognitive Framework for Machine Translation
Rafael E. Banchs 35

A fuzzier approach to machine translation evaluation: A pilot study on post-editing productivity and automated metrics in commercial settings
Carla Parra Escartín and Manuel Arcedillo 40

A Methodology for Bilingual Lexicon Extraction from Comparable Corpora
Reinhard Rapp 46

Ongoing Study for Enhancing Chinese-Spanish Translation with Morphology Strategies
Marta R. Costa-jussà 56

Baidu Translate: Research and Products
Zhongjun He 61

On Improving the Human Translation Process by Using MT Technologies under a Cognitive Framework
Geng Xinhui 63

Towards a shared task for shallow semantics-based translation (in an industrial setting)
Kurt Eberle 64

Workshop Program

Friday, July 31, 2015

9:15–10:30 Introduction and Keynote Speech I

9:15–9:30 *Welcome and introduction*

9:30–10:30 *Invited talk*
Hinrich Schütze

10:30–11:00 Coffee Break

11:00–12:30 Hybrid MT system design

11:00–11:15 *Bootstrapping a hybrid deep MT system*
João Silva, João Rodrigues, Luís Gomes and António Branco

11:15–11:30 *Multi-system machine translation using online APIs for English-Latvian*
Matīss Rikters

11:30–11:55 *What a Transfer-Based System Brings to the Combination with PBMT*
Aleš Tamchyna and Ondrej Bojar

11:55–12:20 *Establishing sentential structure via realignments from small parallel corpora*
George Tambouratzis and Vassiliki Pouli

12:20–14:00 Lunch Break

Friday, July 31, 2015 (continued)

14:00–15:30 Resources and evaluation of hybrid MT systems

14:00–14:15 *Passive and Pervasive Use of Bilingual Dictionary in Statistical Machine Translation*
Liling Tan

14:15–14:30 *Automated Simultaneous Interpretation: Hints of a Cognitive Framework for Machine Translation*
Rafael E. Banchs

14:30–14:45 *A fuzzier approach to machine translation evaluation: A pilot study on post-editing productivity and automated metrics in commercial settings*
Carla Parra Escartín and Manuel Arcedillo

14:45–15:10 *A Methodology for Bilingual Lexicon Extraction from Comparable Corpora*
Reinhard Rapp

15:10–15:25 *Ongoing Study for Enhancing Chinese-Spanish Translation with Morphology Strategies*
Marta R. Costa-jussà

15:30–16:00 Coffee Break

16:00–17:00 Keynote Speech II

16:00–17:00 *Invited talk*
Gerard de Melo

17:00–18:15 Industry applications and Hybrid MT

17:00–17:30 *Baidu Translate: Research and Products*
Zhongjun He

17:30–18:00 *On Improving the Human Translation Process by Using MT Technologies under a Cognitive Framework*
CCID Trans Tech

18:00–18:15 *Towards a shared task for shallow semantics-based translation (in an industrial setting)*
Kurt Eberle

18:15–18:20 Conclusions