

The 26th International Conference on Computational Linguistics  
(COLING 2016)

**Sixth Workshop on Hybrid Approaches to Translation  
(HyTra 6)**

**Proceedings of the Workshop**

December 11, 2016  
Osaka, Japan

Copyright of each paper stays with the respective authors (or their employers).

ISBN 978-4-87974-713-6

## Preface

Welcome to the Sixth Workshop on Hybrid Approaches to Translation (HyTra-6) held in conjunction with COLING-2016 in Osaka!

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 a gap in the current paradigm allowing researchers to explore new pathways of bringing together a diverse range of technologies, methods and tools into the MT domain.

The current Sixth Workshop on Hybrid Approaches to Translation builds on a successful series of past events held in conjunction with international conferences:

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 took place as a full-day workshop and was co-located with ACL 2013 in Sophia, Bulgaria: <http://hytra.barcelonamedia.org/hytra2013/>

HyTra-3 was a 1-day workshop at EACL 2014 in 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, who 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/>

With HyTra-4 being held at ACL 2016 in Beijing, the HyTra workshop series for the first time moved to Asia. This edition again hosted an industrial session with representatives from Baidu, CCID TransTech and Lingenio: <http://glicom.upf.edu/hytra2015/>

The last edition of HyTra, namely HyTra-5, took place in Riga, Latvia, and was co-located with the 2016 edition of the Annual Meeting of the European Association for Machine Translation: <http://glicom.upf.edu/hytra2016/>

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 (HyTra-1), Hermann Ney, Will Lewis and Chris Quirk (HyTra-2), Hans Uszkoreit and Joakim Nivre (HyTra-3), Hinrich Schütze and Gerard de Melo (HyTra-4) and Andy Way (HyTra-5). 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-6 workshop we have accepted eight papers which appear in this volume. The workshop hosts an invited talk by Mark Seligman, CEO of Spoken Translation, Inc.

We hope HyTra-6 will become a successful continuation of the HyTra workshop series and will result in interesting discussions, ideas and collaborations.

Patrik Lambert, WebInterpret, Barcelona

Bogdan Babych, University of Leeds

Kurt Eberle, Lingenio GmbH, Heidelberg

Rafael E. Banchs, Institute for Infocomm Research, Singapore

Reinhard Rapp, University of Mainz and Magdeburg-Stendal University of Applied Sciences

Marta R. Costa-jussà, Universitat Politècnica de Catalunya, Barcelona



## **Invited Speaker**

Mark Seligman, Spoken Translation, Inc.

## **Organisers**

Patrik Lambert, WebInterpret, Barcelona  
Bogdan Babych, University of Leeds  
Kurt Eberle, Lingenio GmbH, Heidelberg  
Rafael E. Banchs, Institute for Infocomm Research, Singapore  
Reinhard Rapp, University of Mainz and Magdeburg-Stendal University of Applied Sciences  
Marta R. Costa-jussà, Universitat Politècnica de Catalunya, Barcelona

## **Programme Committee**

Bogdan Babych, University of Leeds, UK  
Rafael E. Banchs, Institute for Infocomm Research, Singapore  
Alexey Baytin, Yandex, Moscow, Russia  
Pierrette Bouillon, ISSCO/TIM/ETI, University of Geneva, Switzerland  
Marta R. Costa-jussà, Universitat Politècnica de Catalunya, Barcelona, Spain  
Josep Maria Crego, Systran, Paris, France  
Kurt Eberle, Lingenio GmbH, Heidelberg, Germany  
Cristina España, Universitat Politècnica de Catalunya, Barcelona, Spain  
Christian Federmann, Microsoft Research, Seattle, USA  
José A. R. Fonollosa, Universitat Politècnica de Catalunya, Barcelona, Spain  
Udo Kruschwitz, University of Essex, UK  
Patrik Lambert, WebInterpret, Barcelona, Spain  
Maite Melero, Pompeu Fabra University, Barcelona, Spain  
Reinhard Rapp, University of Mainz and Magdeburg-Stendal University of Applied Sciences, Germany  
Serge Sharoff, University of Leeds, UK  
Grigori Sidorov, Instituto Politécnico Nacional, Mexico  
George Tambouratzis, Institute for Language and Speech Processing, Athens, Greece  
Jörg Tiedemann, University of Uppsala, Sweden



## Table of Contents

<i>Combining fast_align with Hierarchical Sub-sentential Alignment for Better Word Alignments</i> Hao Wang and Yves Lepage .....	1
<i>Neural Network Language Models for Candidate Scoring in Hybrid Multi-System Machine Translation</i> Matīss Rikters .....	8
<i>Image-Image Search for Comparable Corpora Construction</i> Yu Hong, Liang Yao, Mengyi Liu, Tongtao Zhang, Wenxuan Zhou, Jianmin Yao and Heng Ji ...	16
<i>Predicting Translation Equivalentents in Linked WordNets</i> Krasimir Angelov and Gleb Lobanov .....	26
<i>Modifications of Machine Translation Evaluation Metrics by Using Word Embeddings</i> Haozhou Wang and Paola Merlo .....	33
<i>Verb sense disambiguation in Machine Translation</i> Roman Sudarikov, Ondřej Dušek, Martin Holub, Ondřej Bojar and Vincent Kríž .....	42
<i>Improving word alignment for low resource languages using English monolingual SRL</i> Meriem Beloucif, Markus Saers and Dekai Wu .....	51
<i>Using Bilingual Segments in Generating Word-to-word Translations</i> kavitha mahesh, Gabriel Pereira Lopes and Luís Gomes .....	61





# Workshop Program

## Sunday, December 11, 2016 – Morning Session

### 9:00–10:00 Workshop Introduction and Keynote Speech

9:00–9:10 *Welcome and Introduction*

9:10–10:00 *Keynote Speech*  
Mark Seligman

### 10:00–10:30 Coffee Break

### 10:30–12:00 Research Papers: Session 1

10:30–11:00 *Neural Network Language Models for Candidate Scoring in Hybrid Multi-System Machine Translation*  
Matīss Rikters

11:00–11:30 *Modifications of Machine Translation Evaluation Metrics by Using Word Embeddings*  
Haozhou Wang and Paola Merlo

11:30–12:00 *Verb sense disambiguation in Machine Translation*  
Roman Sudarikov, Ondřej Dušek, Martin Holub, Ondřej Bojar and Vincent Kríž

### 12:00–14:00 Lunch Break

**Sunday, December 11, 2016 – Afternoon Session**

**14:00–15:30 Research Papers: Session 2**

14:00–14:30 *Predicting Translation Equivalents in Linked WordNets*

Krasimir Angelov and Gleb Lobanov

14:30–15:00 *Using Bilingual Segments in Generating Word-to-word Translations*

Kavitha Mahesh, Gabriel Pereira Lopes and Luís Gomes

15:00–15:30 *Combining fast\_align with Hierarchical Sub-sentential Alignment for Better Word Alignments*

Hao Wang and Yves Lepage

**15:30–16:00 Coffee Break**

**16:00–17:00 Research Papers: Session 3**

16:00–16:30 *Improving word alignment for low resource languages using English monolingual SRL*

Meriem Beloucif, Markus Saers and Dekai Wu

16:30–17:00 *Image-Image Search for Comparable Corpora Construction*

Yu Hong, Liang Yao, Mengyi Liu, Tongtao Zhang, Wenxuan Zhou, Jianmin Yao and Heng Ji

**17:00–17:10 Concluding Remarks and Closing**