

Proceedings of SSST-9

Ninth Workshop on

**Syntax, Semantics and Structure
in Statistical Translation**

Dekai Wu, Marine Carpuat,
Eneko Agirre and Nora Aranberri (editors)

NAACL HLT 2015 / SIGMT / SIGLEX Workshop
4 June 2015
Denver, Colorado, USA

This workshop is partially funded by the European Union QTLep project (FP7-ICT-2013-10-610516).



©2015 The Association for Computational Linguistics

Order print-on-demand copies from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571
USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

ISBN 978-1-941643-41-9

Introduction

The Ninth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-9) was held on 4 June 2015 following the NAACL HLT 2015 conference in Denver, Colorado. Like the first eight SSST workshops in 2007, 2008, 2009, 2010, 2011, 2012, 2013 and 2014, it aimed to bring together researchers from different communities working in the rapidly growing field of structured statistical models of natural language translation.

This year's SSST featured an award for best paper to advance statistical MT using lexical semantics and deep language processing. The €500 prize was sponsored by the European Union QTLeap project (<http://qt leap.eu>, FP7-ICT-2013.4.1-610516), which aims to research and deliver an articulated methodology for machine translation that explores deep language engineering approaches in view of breaking the way to translations of higher quality.

We selected 13 papers and extended abstracts for this year's workshop, many of which reflect statistical machine translation's movement toward not only tree-structured and syntactic models incorporating stochastic synchronous/transduction grammars, but also increasingly semantic models and the closely linked issues of deep syntax and shallow semantics, vector space representations to support these approaches, and semantic evaluation methodologies.

Thanks are due once again to our authors and our Program Committee for making the ninth SSST workshop another success.

Dekai Wu, Marine Carpuat, Eneko Agirre and Nora Aranberri

Organizers:

Dekai Wu, Hong Kong University of Science and Technology (HKUST)
Marine Carpuat, University of Maryland
Eneko Agirre, University of the Basque Country (UPV/EHU)
Nora Aranberri, University of the Basque Country (UPV/EHU)

Program Committee:

Timothy Baldwin, University of Melbourne
Ondřej Bojar, Charles University in Prague
Aljoscha Burchardt, German Research Centre for Artificial Intelligence (DFKI)
Francisco Casacuberta, Universitat Politècnica de València
Colin Cherry, National Research Council (NRC) Canada
David Chiang, USC/ISI
Stephen Clark, University of Cambridge
Kevin Duh, Nara Institute of Science and Technology (NAIST)
Marc Dymetman, Xerox Research Centre Europe
Daniel Gildea, University of Rochester
Nal Kalchbrenner, University of Oxford
Philipp Koehn, University of Edinburgh
Gorka Labaka, University of the Basque Country (UPV/EHU)
Alon Lavie, Carnegie Mellon University
Chi-kiu Lo, Hong Kong University of Science and Technology (HKUST)
Markus Saers, Hong Kong University of Science and Technology (HKUST)
Khalil Sima'an, University of Amsterdam
Ivan Vulić, University of Leuven
Taro Watanabe, Google
Andy Way, Dublin City University
Deyi Xiong, Soochow University
François Yvon, LIMSI

Table of Contents

<i>Harmonizing word alignments and syntactic structures for extracting phrasal translation equivalents</i> Dun Deng, Nianwen Xue and Shiman Guo	1
<i>Non-projective Dependency-based Pre-Reordering with Recurrent Neural Network for Machine Translation</i> Antonio Valerio Miceli Barone and Giuseppe Attardi	10
<i>Translating Negation: Induction, Search And Model Errors</i> Federico Fancellu and Bonnie Webber	21
<i>SMT error analysis and mapping to syntactic, semantic and structural fixes</i> Nora Aranberri	30
<i>Unsupervised False Friend Disambiguation Using Contextual Word Clusters and Parallel Word Alignments</i> Maryam Aminian, Mahmoud Ghoneim and Mona Diab	39
<i>METEOR-WSD: Improved Sense Matching in MT Evaluation</i> Marianna Apidianaki and Benjamin Marie	49
<i>Analyzing English-Spanish Named-Entity enhanced Machine Translation</i> Mikel Artetxe, Eneko Agirre, Iñaki Alegria and Gorka Labaka	52
<i>Predicting Prepositions for SMT</i> Marion Weller, Alexander Fraser and Sabine Schulte im Walde	55
<i>Translation reranking using source phrase dependency features</i> Antonio Valerio Miceli Barone	57
<i>Semantics-based pretranslation for SMT using fuzzy matches</i> Tom Vanallemeersch and Vincent Vandeghinste	61
<i>What Matters Most in Morphologically Segmented SMT Models?</i> Mohammad Salameh, Colin Cherry and Grzegorz Kondrak	65
<i>Improving Chinese-English PropBank Alignment</i> Shumin Wu and Martha Palmer	74

Conference Program

2015/06/04

08:55–09:00 *Opening Remarks*

09:00–10:30 **Session 1**

09:00–10:00 *Invited Talk*
Philipp Koehn

10:00–10:30 *Harmonizing word alignments and syntactic structures for extracting phrasal translation equivalents*
Dun Deng, Nianwen Xue and Shiman Guo

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2**

11:00–11:30 *Non-projective Dependency-based Pre-Reordering with Recurrent Neural Network for Machine Translation*
Antonio Valerio Miceli Barone and Giuseppe Attardi

11:30–12:00 *Translating Negation: Induction, Search And Model Errors*
Federico Fancellu and Bonnie Webber

12:00–12:30 *SMT error analysis and mapping to syntactic, semantic and structural fixes*
Nora Aranberri

2015/06/04 (continued)

12:30–13:55 Lunch Break

13:55–14:30 Session 3

13:55–14:00 *QTLep Best Paper Award*

14:00–14:30 *Unsupervised False Friend Disambiguation Using Contextual Word Clusters and Parallel Word Alignments*

Maryam Aminian, Mahmoud Ghoneim and Mona Diab

14:30–15:30 Session 4: Posters

14:30–14:35 *METEOR-WSD: Improved Sense Matching in MT Evaluation*

Marianna Apidianaki and Benjamin Marie

14:35–14:40 *Analyzing English-Spanish Named-Entity enhanced Machine Translation*

Mikel Artetxe, Eneko Agirre, Iñaki Alegria and Gorka Labaka

14:40–14:45 *Predicting Prepositions for SMT*

Marion Weller, Alexander Fraser and Sabine Schulte im Walde

14:45–14:50 *Translation reranking using source phrase dependency features*

Antonio Valerio Miceli Barone

14:50–14:55 *Semantics-based pretranslation for SMT using fuzzy matches*

Tom Vanallemeersch and Vincent Vandeghinste

2015/06/04 (continued)

15:30–16:00 Coffee Break

16:00–17:00 Session 5

16:00–16:30 *What Matters Most in Morphologically Segmented SMT Models?*
Mohammad Salameh, Colin Cherry and Grzegorz Kondrak

16:30–17:00 *Improving Chinese-English PropBank Alignment*
Shumin Wu and Martha Palmer

