

Proceedings of SSST-8

**Eighth Workshop on**

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

Dekai Wu, Marine Carpuat,  
Xavier Carreras and Eva Maria Vecchi (editors)

EMNLP 2014 / SIGMT / SIGLEX Workshop  
25 October 2014  
Doha, Qatar

Production and Manufacturing by  
*Taberg Media Group AB*  
*Box 94, 562 02 Taberg*  
*Sweden*

©2014 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](mailto:acl@aclweb.org)

ISBN 978-1-937284-96-1

## Introduction

The Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8) was held on 25 October 2014 preceding the EMNLP 2014 conference in Doha, Qatar. Like the first seven SSST workshops in 2007, 2008, 2009, 2010, 2011, 2012 and 2013, 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 special theme focused on compositional distributional semantics, distributed representations, and continuous vector space models in MT.

We selected 19 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, and vector space representations to support these approaches.

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

Dekai Wu, Marine Carpuat, Xavier Carreras and Eva Maria Vecchi

**Organizers:**

Dekai Wu, Hong Kong University of Science and Technology (HKUST)  
Marine Carpuat, National Research Council (NRC) Canada  
Xavier Carreras, Xerox Research Centre Europe  
Eva Maria Vecchi, Cambridge University

**Program Committee:**

Timothy Baldwin, University of Melbourne  
Srinivas Bangalore, AT&T Labs Research  
Phil Blunsom, Oxford University  
Colin Cherry, National Research Council Canada  
David Chiang, USC/ISI  
Shay B. Cohen, University of Edinburgh  
Georgiana Dinu, University of Trento  
Chris Dyer, Carnegie Mellon University  
Marc Dymetman, Xerox Research Centre Europe  
Philipp Koehn, University of Edinburgh  
Alon Lavie, Carnegie Mellon University  
Chi-kiu Lo, HKUST  
Markus Saers, HKUST  
Khalil Sima'an, University of Amsterdam  
Ivan Vulić, University of Leuven  
Taro Watanabe, NICT  
François Yvon, LIMSI  
Ming Zhou, Microsoft Research Asia

**Invited Speaker:**

Timothy Baldwin, University of Melbourne

## Table of Contents

|                                                                                                                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Vector Space Models for Phrase-based Machine Translation</i><br>Tamer Alkhouli, Andreas Guta and Hermann Ney . . . . .                                                                                           | 1   |
| <i>Bilingual Markov Reordering Labels for Hierarchical SMT</i><br>Gideon Maillette de Buy Wenniger and Khalil Sima'an . . . . .                                                                                     | 11  |
| <i>Better Semantic Frame Based MT Evaluation via Inversion Transduction Grammars</i><br>Dekai Wu, Chi-kiu Lo, Meriem Beloucif and Markus Saers . . . . .                                                            | 22  |
| <i>Rule-based Syntactic Preprocessing for Syntax-based Machine Translation</i><br>Yuto Hatakoshi, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura . . . . .                                         | 34  |
| <i>Applying HMEANT to English-Russian Translations</i><br>Alexander Chuchunkov, Alexander Tarelkin and Irina Galinskaya . . . . .                                                                                   | 43  |
| <i>Reducing the Impact of Data Sparsity in Statistical Machine Translation</i><br>Karan Singla, Kunal Sachdeva, Srinivas Bangalore, Dipti Misra Sharma and Diksha Yadav . . . . .                                   | 51  |
| <i>Expanding the Language model in a low-resource hybrid MT system</i><br>George Tambouratzis, Sokratis Sofianopoulos and Marina Vassiliou . . . . .                                                                | 57  |
| <i>Syntax and Semantics in Quality Estimation of Machine Translation</i><br>Rasoul Kaljahi, Jennifer Foster and Johann Roturier . . . . .                                                                           | 67  |
| <i>Overcoming the Curse of Sentence Length for Neural Machine Translation using Automatic Segmentation</i><br>Jean Pouget-Abadie, Dzmitry Bahdanau, Bart van Merriënboer, Kyunghyun Cho and Yoshua Bengio . . . . . | 78  |
| <i>Ternary Segmentation for Improving Search in Top-down Induction of Segmental ITGs</i><br>Markus Saers and Dekai Wu . . . . .                                                                                     | 86  |
| <i>A CYK+ Variant for SCFG Decoding Without a Dot Chart</i><br>Rico Sennrich . . . . .                                                                                                                              | 94  |
| <i>On the Properties of Neural Machine Translation: Encoder–Decoder Approaches</i><br>Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau and Yoshua Bengio . . . . .                                             | 103 |
| <i>Transduction Recursive Auto-Associative Memory: Learning Bilingual Compositional Distributed Vector Representations of Inversion Transduction Grammars</i><br>Kartteek Addanki and Dekai Wu . . . . .            | 112 |
| <i>Transformation and Decomposition for Efficiently Implementing and Improving Dependency-to-String Model In Moses</i><br>Liangyou Li, Jun Xie, Andy Way and Qun Liu . . . . .                                      | 122 |
| <i>Word’s Vector Representations meet Machine Translation</i><br>Eva Martínez García, Jörg Tiedemann, Cristina España-Bonet and Lluís Màrquez . . . . .                                                             | 132 |
| <i>Context Sense Clustering for Translation</i><br>João Casteleiro, Gabriel Lopes and Joaquim Silva . . . . .                                                                                                       | 135 |

|                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------|-----|
| <i>Evaluating Word Order Recursively over Permutation-Forests</i>                                               |     |
| Miloš Stanojević and Khalil Sima'an .....                                                                       | 138 |
| <i>Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based Statistical Machine Translation</i> |     |
| Matthias Huck, Hieu Hoang and Philipp Koehn .....                                                               | 148 |
| <i>How Synchronous are Adjuncts in Translation Data?</i>                                                        |     |
| Sophie Arnoult and Khalil Sima'an .....                                                                         | 157 |

# Conference Program

**Saturday, October 25, 2014**

**9:00–10:30**    **Session 1: Morning Orals**

9:00–9:10    *Opening Remarks*

Dekai Wu, Marine Carpuat, Xavier Carreras, Eva Maria Vecchi

9:10–9:30    *Vector Space Models for Phrase-based Machine Translation*

Tamer Alkhouli, Andreas Guta and Hermann Ney

9:30–9:50    *Bilingual Markov Reordering Labels for Hierarchical SMT*

Gideon Maillette de Buy Wenniger and Khalil Sima'an

9:50–10:10    *Better Semantic Frame Based MT Evaluation via Inversion Transduction Grammars*

Dekai Wu, Chi-kiu Lo, Meriem Beloucif and Markus Saers

10:10–10:30    *Rule-based Syntactic Preprocessing for Syntax-based Machine Translation*

Yuto Hatakoshi, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura

**10:30–11:00**    *Coffee break*

**11:00–12:00**    **Invited talk by Timothy Baldwin**

11:00–12:00    *Composed, Distributed Reflections on Semantics and Statistical Machine Translation*

Timothy Baldwin

**Saturday, October 25, 2014 (continued)**

**12:00–12:30 Session 2: Morning Spotlights**

- 12:00–12:05 *Applying HMEANT to English-Russian Translations*  
Alexander Chuchunkov, Alexander Tarelkin and Irina Galinskaya
- 12:05–12:10 *Reducing the Impact of Data Sparsity in Statistical Machine Translation*  
Karan Singla, Kunal Sachdeva, Srinivas Bangalore, Dipti Misra Sharma and Diksha Yadav
- 12:10–12:15 *Expanding the Language model in a low-resource hybrid MT system*  
George Tambouratzis, Sokratis Sofianopoulos and Marina Vassiliou
- 12:15–12:20 *Syntax and Semantics in Quality Estimation of Machine Translation*  
Rasoul Kaljahi, Jennifer Foster and Johann Roturier
- 12:20–12:25 *Overcoming the Curse of Sentence Length for Neural Machine Translation using Automatic Segmentation*  
Jean Pouget-Abadie, Dzmitry Bahdanau, Bart van Merriënboer, Kyunghyun Cho and Yoshua Bengio
- 12:25–12:30 *Ternary Segmentation for Improving Search in Top-down Induction of Segmental ITGs*  
Markus Saers and Dekai Wu

**12:30–14:00 Lunch break**

**14:00–15:30 Session 3: Afternoon Orals and Spotlights**

- 14:00–14:20 *A CYK+ Variant for SCFG Decoding Without a Dot Chart*  
Rico Sennrich
- 14:20–14:40 *On the Properties of Neural Machine Translation: Encoder–Decoder Approaches*  
Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau and Yoshua Bengio
- 14:40–15:00 *Transduction Recursive Auto-Associative Memory: Learning Bilingual Compositional Distributed Vector Representations of Inversion Transduction Grammars*  
Karttek Addanki and Dekai Wu
- 15:00–15:20 *Transformation and Decomposition for Efficiently Implementing and Improving Dependency-to-String Model In Moses*  
Liangyou Li, Jun Xie, Andy Way and Qun Liu



**Saturday, October 25, 2014 (continued)**

15:20–15:25 *Word's Vector Representations meet Machine Translation*  
Eva Martinez Garcia, Jörg Tiedemann, Cristina España-Bonet and Lluís Màrquez

15:25–15:30 *Context Sense Clustering for Translation*  
João Casteleiro, Gabriel Lopes and Joaquim Silva

**15:30–16:00** *Coffee break*

**16:00–16:15** **Session 4: Afternoon Spotlights**

16:00–16:05 *Evaluating Word Order Recursively over Permutation-Forests*  
Miloš Stanojević and Khalil Sima'an

16:05–16:10 *Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based Statistical Machine Translation*  
Matthias Huck, Hieu Hoang and Philipp Koehn

16:10–16:15 *How Synchronous are Adjuncts in Translation Data?*  
Sophie Arnoult and Khalil Sima'an

**16:15–17:30** **Poster session**

16:15–17:30 *Poster session of all workshop papers*  
All workshop presenters