Proceedings of SSST-7

**Seventh Workshop on**

# Syntax, Semantics and Structure in Statistical Translation

Marine Carpuat, Lucia Specia and Dekai Wu (editors)

SIGMT / SIGLEX Workshop
The 2013 Conference of the North American Chapter
of the Association for Computational Linguistics:
Human Language Technologies

©2013 The Association for Computational Linguistics

209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-47-3

# Introduction

The Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-7) was held on 13 June 2013 following the NAACL 2013 conference in Atlanta, GA, USA. Like the first six SSST workshops in 2007, 2008, 2009, 2010, 2011 and 2012, it aimed to bring together researchers from different communities working in the rapidly growing field of structured statistical models of natural language translation.

We selected 8 papers 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.

In the third year since "Semantics" was explicitly added to the workshop name, the work exploring SMT's connections to semantics has continued to grow. Carpuat shows that word sense disambiguation tasks can be viewed as a method for semantic evaluation of machine translation lexical choice. Singh studies the impact of the orthographic representation of Manipuri, a Sino-Tibetan language on the task of SMT to and from English, and explores its impact on lexical ambiguity.

Several papers deepen our understanding of theoretical and practical issues associated with structured statistical translation models.

Maillette de Buy Wenniger and Sima'an show how to extend rules in a hierarchical phrase-based system with reordering information, by defining more specific nonterminals and augmenting rules with features. Huck, Vilar, Freitag and Ney present a detailed empirical study of cube pruning for hierarchical phrase-based systems. Herrmann, Niehues and Waibel incorporate a syntactic tree-based reordering method to model long-range reorderings in a phrase-based machine translation system, and combine reordering models at different levels of linguistic representation.

Saers, Addanki and Wu present an unsupervised method for inducing an Inversion Transduction Grammar based on the Minimum Description Length principle. Maillette de Buy Wenniger and Sima'an propose a precise definition of what it means for an Inversion Transduction Grammar to cover the word alignment of a sentence, and experiment with human and machine-made alignments. Kaeshammer explores the expressiveness of synchronous linear context-free rewriting systems for machine translation by computing derivation coverage on manually word aligned parallel text.

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

Marine Carpuat, Lucia Specia, and Dekai Wu

**Organizers:**

Marine Carpuat, National Research Council Canada
Lucia Specia, University of Sheffield
Dekai Wu, Hong Kong University of Science and Technology

**Program Committee:**

Marianna Apidianak, LIMSI-CNRS
Wilker Aziz, University of Wolverhampton
Srinivas Bangalore, AT&T Labs Research
Yee Seng Chan, Raytheon BBN Technologies
Colin Cherry, National Research Council Canada
David Chiang, USC/ISI
John DeNero, Google
Marc Dymetman, Xerox Research Centre Europe
Alexander Fraser, Universität Stuttgart
Daniel Gildea, University of Rochester
Greg Hanneman, Carnegie Mellon University
Yifan He, New York University
Hieu Hoang, University of Edinburgh
Philipp Koehn, University of Edinburgh
Els Lefever, Hogeschool Gent
Chi-kiu Lo, HKUST
Daniel Marcu, SDL
Aurélien Max, LIMSI-CNRS & Univ. Paris Sud
Daniele Pighin, Google
Markus Saers, HKUST
Taro Watanabe, NICT
Deyi Xiong, Institute for Infocomm Research
Bowen Zhou, IBM Research

# Table of Contents

*A Semantic Evaluation of Machine Translation Lexical Choice*
Marine Carpuat . . . . . 1

*Taste of Two Different Flavours: Which Manipuri Script works better for English-Manipuri Language pair SMT Systems?*
Thoudam Doren Singh . . . . . 11

*Hierarchical Alignment Decomposition Labels for Hiero Grammar Rules*
Gideon Maillette de Buy Wenniger and Khalil Sima'an . . . . . 19

*A Performance Study of Cube Pruning for Large-Scale Hierarchical Machine Translation*
Matthias Huck, David Vilar, Markus Freitag and Hermann Ney . . . . . 29

*Combining Word Reordering Methods on different Linguistic Abstraction Levels for Statistical Machine Translation*
Teresa Herrmann, Jan Niehues and Alex Waibel . . . . . 39

*Combining Top-down and Bottom-up Search for Unsupervised Induction of Transduction Grammars*
Markus Saers, Karteek Addanki and Dekai Wu . . . . . 48

*A Formal Characterization of Parsing Word Alignments by Synchronous Grammars with Empirical Evidence to the ITG Hypothesis.*
Gideon Maillette de Buy Wenniger and Khalil Sima'an . . . . . 58

*Synchronous Linear Context-Free Rewriting Systems for Machine Translation*
Miriam Kaeshammer . . . . . 68

# Conference Program

**Thursday, June 13, 2013**

9:15–9:30 Opening Remarks

**Session 1**

9:30–10:00 *A Semantic Evaluation of Machine Translation Lexical Choice*
Marine Carpuat

10:00–10:30 *Taste of Two Different Flavours: Which Manipuri Script works better for English-Manipuri Language pair SMT Systems?*
Thoudam Doren Singh

10:30–11:00 Break

**Session 2**

11:00–11:30 *Hierarchical Alignment Decomposition Labels for Hiero Grammar Rules*
Gideon Maillette de Buy Wenniger and Khalil Sima'an

11:30–12:00 *A Performance Study of Cube Pruning for Large-Scale Hierarchical Machine Translation*
Matthias Huck, David Vilar, Markus Freitag and Hermann Ney

12:00–12:30 *Combining Word Reordering Methods on different Linguistic Abstraction Levels for Statistical Machine Translation*
Teresa Herrmann, Jan Niehues and Alex Waibel

12:30–2:00 Lunch

**Thursday, June 13, 2013 (continued)**

**Session 3**

2:00–3:00 Panel discussion: Meaning Representations for Machine Translation, with Jan Hajic, Kevin Knight, Martha Palmer and Dekai Wu

3:30–4:00 *Combining Top-down and Bottom-up Search for Unsupervised Induction of Transduction Grammars*
Markus Saers, Karteek Addanki and Dekai Wu

3:30–4:00 Break

**Session 4**

4:00–4:30 *A Formal Characterization of Parsing Word Alignments by Synchronous Grammars with Empirical Evidence to the ITG Hypothesis.*
Gideon Maillette de Buy Wenniger and Khalil Sima'an

4:30–5:00 *Synchronous Linear Context-Free Rewriting Systems for Machine Translation*
Miriam Kaeshammer