Proceedings of

## SSST-2

Second Workshop on

# Syntax and Structure in Statistical Translation

David Chiang and Dekai Wu (editors)

ACL-08: HLT Columbus, Ohio, USA 20 June 2008 Production and Manufacturing by Omnipress Inc. 2600 Anderson Street Madison, WI 53707 USA

©2008 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 978-1-932432-16-9

## Introduction

The Second Workshop on Syntax and Structure in Statistical Translation (SSST-2) was held on 20 June 2008 following the ACL-08: HLT conference hosted by Ohio State University in Columbus, Ohio. Like the first SSST workshop in 2007, it aimed to bring together researchers from different communities working in the rapidly growing field of statistical, tree-structured models of natural language translation.

We selected eleven papers for this year's workshop. There was a strong emphasis this year on the use of explicit syntactic information: constituent structures (Yamamoto, Okuma and Sumita; Zhou, Xiang, Zhu and Gao; Elming; Clark, Frederking and Levin; Lavie, Parlikar and Ambati), dependency structures (Nikoulina and Dymetman; Ma, Ozdowska, Sun and Way), part-of-speech tags (Tillmann), and combinations thereof (Ge, Ittycheriah and Papineni). These papers applied syntactic information to grammar-based models as well as to phrase-based and word-based models. The program was rounded out by papers describing new decoding techniques (Li and Khudanpur) and machine-learning techniques (Subotin) for grammar-based translation models.

We would like to thank our authors and our Program Committee for making this year's SSST workshop another success.

David Chiang and Dekai Wu<sup>1</sup>

<sup>&</sup>lt;sup>1</sup>This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under GALE Contract Nos. HR0011-06-C-0023, subcontract SRI International (Dekai Wu) and HR0011-06-C-0022, subcontract BBN Technologies 9500008412 (David Chiang). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency.

#### **Organizers:**

David CHIANG, USC Information Sciences Institute, USA Dekai WU, Hong Kong University of Science and Technology (HKUST), Hong Kong

#### **Program Committee:**

Srinivas BANGALORE, AT&T Research, USA Marine CARPUAT, Hong Kong University of Science and Technology (HKUST), Hong Kong Pascale FUNG, Hong Kong University of Science and Technology (HKUST), Hong Kong Daniel GILDEA, University of Rochester, USA Kevin KNIGHT, USC Information Sciences Institute, USA Jonas KUHN, University of Potsdam, Germany Yang LIU, Institute of Computing Technology, Chinese Academy of Sciences, China Daniel MARCU, USC Information Sciences Institute, USA Yuji MATSUMOTO, Nara Institute of Science and Technology, Japan Hermann NEY, RWTH Aachen, Germany Owen RAMBOW, Columbia University, USA Philip RESNIK, University of Maryland, USA Stefan RIEZLER, Google Inc., USA Libin SHEN, BBN Technologies, USA Christoph TILLMANN, IBM T. J. Watson Research Center, USA Stephan VOGEL, Carnegie Mellon University, USA Taro WATANABE, NTT Communication Science Laboratories, Japan Andy WAY, Dublin City University, Ireland Yuk-Wah WONG, Google Inc., USA Richard ZENS, Google Inc., USA

## **Table of Contents**

Imposing Constraints from the Source Tree on ITG Constraints for SMT   Hirofumi YAMAMOTO, Hideo OKUMA and Eiichiro SUMITA
A Scalable Decoder for Parsing-Based Machine Translation with Equivalent Language Model State Maintenance Zhifei LI and Sanjeev KHUDANPUR
Prior Derivation Models For Formally Syntax-Based Translation Using Linguistically Syntactic Parsing and Tree Kernels Bowen ZHOU, Bing XIANG, Xiaodan ZHU and Yuqing GAO
Generalizing Local Translation Models   Michael SUBOTIN 28
A Rule-Driven Dynamic Programming Decoder for Statistical MT Christoph TILLMANN
Syntactic Reordering Integrated with Phrase-Based SMT Jakob Elming
Experiments in Discriminating Phrase-Based Translations on the Basis of Syntactic Coupling Features Vassilina NIKOULINA and Marc DYMETMAN
Multiple Reorderings in Phrase-Based Machine Translation   Niyu GE, Abe ITTYCHERIAH and Kishore PAPINENI
Improving Word Alignment Using Syntactic Dependencies Yanjun MA, Sylwia OZDOWSKA, Yanli SUN and Andy WAY
Inductive Detection of Language Features via Clustering Minimal Pairs: Toward Feature-Rich Gram- mars in Machine Translation Jonathan H. CLARK, Robert FREDERKING and Lori LEVIN
Syntax-Driven Learning of Sub-Sentential Translation Equivalents and Translation Rules from Parsed Parallel Corpora Alon LAVIE, Alok PARLIKAR and Vamshi AMBATI

## Workshop Program

## Friday, June 20, 2008

9:00–9:05	Opening Remarks
9:05–9:30	<i>Imposing Constraints from the Source Tree on ITG Constraints for SMT</i> Hirofumi YAMAMOTO, Hideo OKUMA and Eiichiro SUMITA
9:30–9:55	A Scalable Decoder for Parsing-Based Machine Translation with Equivalent Lan- guage Model State Maintenance Zhifei LI and Sanjeev KHUDANPUR
9:55–10:20	Prior Derivation Models For Formally Syntax-Based Translation Using Linguisti- cally Syntactic Parsing and Tree Kernels Bowen ZHOU, Bing XIANG, Xiaodan ZHU and Yuqing GAO
10:30-11:00	Coffee Break
11:00-11:25	Generalizing Local Translation Models Michael SUBOTIN
11:25–12:15	Invited Talk
12:15–13:45	Lunch
13:45–14:10	A Rule-Driven Dynamic Programming Decoder for Statistical MT Christoph TILLMANN
14:10-14:35	Syntactic Reordering Integrated with Phrase-Based SMT Jakob Elming
14:40-15:05	Experiments in Discriminating Phrase-Based Translations on the Basis of Syntactic Coupling Features Vassilina NIKOULINA and Marc DYMETMAN
15:05–15:30	Multiple Reorderings in Phrase-Based Machine Translation Niyu GE, Abe ITTYCHERIAH and Kishore PAPINENI
15:30-16:00	Coffee Break

#### Friday, June 20, 2008 (continued)

16:00-16:25	Improving Word Alignment Using Syntactic Dependencies
	Yanjun MA, Sylwia OZDOWSKA, Yanli SUN and Andy WAY

- 16:25–16:50 Inductive Detection of Language Features via Clustering Minimal Pairs: Toward Feature-Rich Grammars in Machine Translation Jonathan H. CLARK, Robert FREDERKING and Lori LEVIN
- 16:50–17:15 Syntax-Driven Learning of Sub-Sentential Translation Equivalents and Translation Rules from Parsed Parallel Corpora Alon LAVIE, Alok PARLIKAR and Vamshi AMBATI
- 17:15–17:45 Panel Discussion