

Proceedings of

SSST-6

Sixth Workshop on

Syntax, Semantics and Structure in Statistical Translation

Marine Carpuat, Lucia Specia and Dekai Wu (editors)

ACL 2012 / SIGMT / SIGLEX Workshop
50th Annual Meeting of the Association for Computational

©2012 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 978-1-937284-38-1

Introduction

The Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-6) was held on 12 July 2012 following the ACL 2012 conference in Jeju, Korea. Like the first five SSST workshops in 2007, 2008, 2009, 2010, and 2011, 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 13 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. Semantic SMT research includes context-dependent WSD (word sense disambiguation) for SMT (Carpuat and Wu 2007, 2008; Chan, Ng and Chiang 2007; Giménez and Màrquez 2007); SRL (semantic role labeling) for SMT (Wu and Fung 2009); and SRL for MT evaluation (Lo and Wu 2010, 2011). In the second year since "Semantics" was explicitly added to the workshop name, the work exploring SMT's connections to semantics, predicate-argument structure, and deep syntax has continued to grow.

There is increased interest in modeling semantic and deep syntactic structure in translation. Quernheim and Knight lay a foundation for modeling semantics in SMT by developing weighted acceptors and transducers for feature structures. Haugereid and Bond show how semantic transfer rules for rule-based MT can be extracted from corpora using SMT phrase aligners. Han, Sudoh, Wu, Duh, Tsukada and Nagata introduce source reordering rules for Chinese-Japanese translation, based HPSG deep parses of Chinese sentences.

Semantic and syntactic models continue to provide rich models of source context. Apidianaki, Wisniewski, Sokolov, Max and Yvon use word sense disambiguation models as n-best reranking features and local language models to improve translation quality. Wang, Osenova and Simov integrate rich morphological and grammar-based information in a factored SMT framework. Source preprocessing is also used to model verbal constructs in English-Hindi translation (Arora and Sinha), zero pronoun resolution in Japanese-English translation (Taira, Sudoh and Nagata), and clause structure (Koeva, Leseva, Stoyanova, Dekova, Genov, Rizov, Dimitrova, Tarpomanova and Kukova). Wetzel and Bond build training examples designed to improve the translation of negated sentences.

Conversely, existing SMT systems and resources can be used to enrich existing semantic resources. Arcan, Federmann and Buitelaar show how SMT can be used in combination with other techniques to translate the vocabulary of a domain-specific ontology.

The challenges of correctly evaluating the semantics of MT output are also explored. Lo and Wu show how to automate the tuning of semantic role based MT evaluation metrics for English. Bojar and Wu investigate the porting of SRL based MT metrics to a very different language, Czech. Rosa, Dušek, Mareček and Popel show how to improve rule-based correction of SMT output by designing a parser specifically for that task.

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

Marine Carpuat, Lucia Specia, and Dekai Wu

Acknowledgements

This work was supported in part by the National Research Council Institute for Information Technology (Marine Carpuat); by the Department of Computer Science at the University of Sheffield (Lucia Specia); and the Defense Advanced Research Projects Agency (DARPA) under BOLT contract no. HR0011-12-C-0016 and GALE contract nos. HR0011-06-C-0022 and HR0011-06-C-0023, the European Union under the FP7 grant agreement no. 287658, and the Hong Kong Research Grants Council (RGC) research grant GRF621008 (Dekai Wu). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the RGC, EU, or DARPA.

Organizers:

Marine CARPUAT, National Research Council (NRC), Canada
Lucia SPECIA, University of Wolverhampton, UK
Dekai WU, Hong Kong University of Science and Technology (HKUST), Hong Kong

Program Committee:

Marianna APIDIANAKI, Alpage, INRIA and University Paris 7, France
Wilker AZIZ, University of Wolverhampton, UK
Srinivas BANGALORE, AT&T Research, USA
David CHIANG, USC ISI, USA
Colin CHERRY, National Research Council (NRC), Canada
Mona DIAB, Columbia University, USA
Alexander FRASER, University of Stuttgart, Germany
Daniel GILDEA, University of Rochester, USA
Nizar HABASH, Columbia University, USA
Yifan HE, Dublin City University, Ireland
Kevin KNIGHT, USC ISI, USA
Philipp KOEHN, University of Edinburgh, UK
Alon LAVIE, Carnegie Mellon University, USA
Yanjun MA, Baidu, China
Daniel MARCU, USC ISI and Language Weaver, USA
Lluís MÀRQUEZ, Universitat Politècnica de Catalunya, Spain
Sudip Kumar NASKAR, Dublin City University, Ireland
Hwee Tou NG, National University of Singapore, Singapore
Daniel PIGHIN, Universitat Politècnica de Catalunya, Spain
Markus SAERS, Hong Kong University of Science and Technology (HKUST), Hong Kong
Libin SHEN, IBM, USA
Matthew SNOVER, City University of New York, USA
John TINSLEY, Dublin City University, Ireland
Stephan VOGEL, Qatar Computing Research Institute, Qatar
Taro WATANABE, NICT, Japan
Deyi XIONG, National University of Singapore, Singapore
François YVON, Université Paris Sud 11, France

Table of Contents

<i>WSD for n-best reranking and local language modeling in SMT</i> Marianna Apidianaki, Guillaume Wisniewski, Artem Sokolov, Aurélien Max and François Yvon	1
<i>Linguistically-Enriched Models for Bulgarian-to-English Machine Translation</i> Rui Wang, Petya Osenova and Kiril Simov	10
<i>Enriching Parallel Corpora for Statistical Machine Translation with Semantic Negation Rephrasing</i> Dominikus Wetzels and Francis Bond	20
<i>Towards a Predicate-Argument Evaluation for MT</i> Ondrej Bojar and Dekai Wu	30
<i>Using Parallel Features in Parsing of Machine-Translated Sentences for Correction of Grammatical Errors</i> Rudolf Rosa, Ondřej Dušek, David Mareček and Martin Popel	39
<i>Unsupervised vs. supervised weight estimation for semantic MT evaluation metrics</i> Chi-kiu Lo and Dekai Wu	49
<i>Head Finalization Reordering for Chinese-to-Japanese Machine Translation</i> Dan Han, Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada and Masaaki Nagata	57
<i>Extracting Semantic Transfer Rules from Parallel Corpora with SMT Phrase Aligners</i> Petter Haugereid and Francis Bond	67
<i>Towards Probabilistic Acceptors and Transducers for Feature Structures</i> Daniel Quernheim and Kevin Knight	76
<i>Using Domain-specific and Collaborative Resources for Term Translation</i> Mihael Arcan, Christian Federmann and Paul Buitelaar	86
<i>Improving Statistical Machine Translation through co-joining parts of verbal constructs in English-Hindi translation</i> Karunesh Kumar Arora and R. Mahesh K. Sinha	95
<i>Application of Clause Alignment for Statistical Machine Translation</i> Svetla Koeva, Svetlozara Leseva, Ivelina Stoyanova, Rositsa Dekova, Angel Genov, Borislav Rizov, Tsvetana Dimitrova, Ekaterina Tarpomanova and Hristina Kukova	102
<i>Zero Pronoun Resolution can Improve the Quality of J-E Translation</i> Hirotoshi Taira, Katsuhito Sudoh and Masaaki Nagata	111

Conference Program

Wednesday, June 29, 2005

8:45–9:00 Opening remarks

Session 1: Source language modeling

9:00–9:30 *WSD for n-best reranking and local language modeling in SMT*
Marianna Apidianaki, Guillaume Wisniewski, Artem Sokolov, Aurélien Max and François Yvon

9:30–10:00 *Linguistically-Enriched Models for Bulgarian-to-English Machine Translation*
Rui Wang, Petya Osenova and Kiril Simov

10:00–10:30 *Enriching Parallel Corpora for Statistical Machine Translation with Semantic Negation Rephrasing*
Dominikus Wetzels and Francis Bond

10:30–11:00 Coffee break

Session 2: MT output evaluation and processing

11:00–11:30 *Towards a Predicate-Argument Evaluation for MT*
Ondrej Bojar and Dekai Wu

11:30–12:00 *Using Parallel Features in Parsing of Machine-Translated Sentences for Correction of Grammatical Errors*
Rudolf Rosa, Ondřej Dušek, David Mareček and Martin Popel

12:00–12:30 *Unsupervised vs. supervised weight estimation for semantic MT evaluation metrics*
Chi-kiu Lo and Dekai Wu

12:30–2:00 Lunch

Wednesday, June 29, 2005 (continued)

Session 3: Semantic dependencies

- 2:00–2:30 *Head Finalization Reordering for Chinese-to-Japanese Machine Translation*
Dan Han, Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada and Masaaki Nagata
- 2:30–3:00 *Extracting Semantic Transfer Rules from Parallel Corpora with SMT Phrase Aligners*
Petter Haugereid and Francis Bond
- 3:00–3:30 *Towards Probabilistic Acceptors and Transducers for Feature Structures*
Daniel Quernheim and Kevin Knight

Session 4: Poster session at Coffee break

- 3:30–4:00 *Using Domain-specific and Collaborative Resources for Term Translation*
Mihael Arcan, Christian Federmann and Paul Buitelaar
- 3:30–4:00 *Improving Statistical Machine Translation through co-joining parts of verbal constructs in English-Hindi translation*
Karunesh Kumar Arora and R. Mahesh K. Sinha
- 3:30–4:00 *Application of Clause Alignment for Statistical Machine Translation*
Svetla Koeva, Svetlozara Leseva, Ivelina Stoyanova, Rositsa Dekova, Angel Genov, Borislav Rizov, Tsvetana Dimitrova, Ekaterina Tarpomanova and Hristina Kukova
- 3:30–4:00 *Zero Pronoun Resolution can Improve the Quality of J-E Translation*
Hirotoshi Taira, Katsuhito Sudoh and Masaaki Nagata
- 4:00–5:00 Panel