

# **Joint 5th Workshop on Statistical Machine Translation and MetricsMATR 2010**

**(WMT10)**

**Held at ACL 2010**

**Uppsala, Sweden  
15-16 July 2010**

**ISBN: 978-1-61738-823-1**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2010) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2010)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>A Semi-Supervised Word Alignment Algorithm with Partial Manual Alignments</i> Qin Gao, Nguyen Bach and Stephan Vogel .....	1
<i>Fast Consensus Hypothesis Regeneration for Machine Translation</i> Boxing Chen, George Foster and Roland Kuhn .....	11
<i>Findings of the 2010 Joint Workshop on Statistical Machine Translation and Metrics for Machine Translation</i> Chris Callison-Burch, Philipp Koehn, Christof Monz, Kay Peterson, Mark Przybocki and Omar Zaidan .....	17
<i>LIMSI's Statistical Translation Systems for WMT'10</i> Alexandre Allauzen, Josep M. Crego, İlknur Durgar El-Kahlout and Francois Yvon .....	54
<i>2010 Failures in English-Czech Phrase-Based MT</i> Ondrej Bojar and Kamil Kos .....	60
<i>An Empirical Study on Development Set Selection Strategy for Machine Translation Learning</i> Hui Cong, Zhao Hai, Lu Bao-Liang and Song Yan .....	67
<i>The University of Maryland Statistical Machine Translation System for the Fifth Workshop on Machine Translation</i> Vladimir Eidelman, Chris Dyer and Philip Resnik .....	72
<i>Further Experiments with Shallow Hybrid MT Systems</i> Christian Federmann, Andreas Eisele, Yu Chen, Sabine Hunsicker, Jia Xu and Hans Uszkoreit ..	77
<i>Improved Features and Grammar Selection for Syntax-Based MT</i> Greg Hanneman, Jonathan Clark and Alon Lavie .....	82
<i>FBK at WMT 2010: Word Lattices for Morphological Reduction and Chunk-Based Reordering</i> Christian Hardmeier, Arianna Bisazza and Marcello Federico .....	88
<i>The RWTH Aachen Machine Translation System for WMT 2010</i> Carmen Heger, Joern Wuebker, Matthias Huck, Gregor Leusch, Saab Mansour, Daniel Stein and Hermann Ney .....	93
<i>Using Collocation Segmentation to Augment the Phrase Table</i> Carlos A. Henríquez Q., Marta Ruiz Costa-jussà, Vidas Daudaravicius, Rafael E. Banchs and José B. Mariño .....	98
<i>The RALI Machine Translation System for WMT 2010</i> Stéphane Huet, Julien Bourdaillet, Alexandre Patry and Philippe Langlais .....	103
<i>Exodus - Exploring SMT for EU Institutions</i> Michael Jellinghaus, Alexandros Poulis and David Kolovratník .....	110
<i>More Linguistic Annotation for Statistical Machine Translation</i> Philipp Koehn, Barry Haddow, Philip Williams and Hieu Hoang .....	115
<i>LIUM SMT Machine Translation System for WMT 2010</i> Patrik Lambert, Sadaf Abdul-Rauf and Holger Schwenk .....	121

<i>Lessons from NRC's Portage System at WMT 2010</i>	
Samuel Larkin, Boxing Chen, George Foster, Ulrich Germann, Eric Joanis, Howard Johnson and Roland Kuhn .....	127
<i>Joshua 2.0: A Toolkit for Parsing-Based Machine Translation with Syntax, Semirings, Discriminative Training and Other Goodies</i>	
Zhifei Li, Chris Callison-Burch, Chris Dyer, Juri Ganitkevitch, Ann Irvine, Sanjeev Khudanpur, Lane Schwartz, Wren Thornton, Ziyuan Wang, Jonathan Weese and Omar Zaidan .....	133
<i>The Karlsruhe Institute for Technology Translation System for the ACL-WMT 2010</i>	
Jan Niehues, Teresa Herrmann, Mohammed Mediani and Alex Waibel .....	138
<i>MATREX: The DCU MT System for WMT 2010</i>	
Sergio Penkale, Rejwanul Haque, Sandipan Dandapat, Pratyush Banerjee, Ankit K. Srivastava, Jinhua Du, Pavel Pecina, Sudip Kumar Naskar, Mikel L. Forcada and Andy Way .....	143
<i>The Cunei Machine Translation Platform for WMT '10</i>	
Aaron Phillips .....	149
<i>The CUED HiFST System for the WMT10 Translation Shared Task</i>	
Juan Pino, Gonzalo Iglesias, Adrià de Gispert, Graeme Blackwood, Jamie Brunning and William Byrne .....	155
<i>The LIG Machine Translation System for WMT 2010</i>	
Marion Potet, Laurent Besacier and Hervé Blanchon .....	161
<i>Linear Inversion Transduction Grammar Alignments as a Second Translation Path</i>	
Markus Saers, Joakim Nivre and Dekai Wu .....	167
<i>UPV-PRHLT English–Spanish System for WMT10</i>	
Germán Sanchis-Trilles, Jesús Andrés-Ferrer, Guillem Gascó, Jesús González-Rubio, Pascual Martínez-Gómez, Martha-Alicia Rocha, Joan-Andreu Sánchez and Francisco Casacuberta .....	172
<i>Reproducible Results in Parsing-Based Machine Translation: The JHU Shared Task Submission</i>	
Lane Schwartz .....	177
<i>Vs and OOVs: Two Problems for Translation between German and English</i>	
Sara Stymne, Maria Holmqvist and Lars Ahrenberg .....	183
<i>To Cache or Not To Cache? Experiments with Adaptive Models in Statistical Machine Translation</i>	
Jörg Tiedemann .....	189
<i>Applying Morphological Decompositions to Statistical Machine Translation</i>	
Sami Virpioja, Jaakko Väyrynen, Andre Mansikkaniemi and Mikko Kurimo .....	195
<i>Maximum Entropy Translation Model in Dependency-Based MT Framework</i>	
Zdeněk Žabokrtský, Martin Popel and David Mareček .....	201
<i>UCH-UPV English–Spanish System for WMT10</i>	
Francisco Zamora-Martinez and Germán Sanchis-Trilles .....	207
<i>Hierarchical Phrase-Based MT at the Charles University for the WMT 2010 Shared Task</i>	
Daniel Zeman .....	212
<i>Incremental Decoding for Phrase-Based Statistical Machine Translation</i>	
Baskaran Sankaran, Ajeet Grewal and Anoop Sarkar .....	216

<i>How to Avoid Burning Ducks: Combining Linguistic Analysis and Corpus Statistics for German Compound Processing</i>	
Fabienne Fritzingler and Alexander Fraser	224
<i>Chunk-Based Verb Reordering in VSO Sentences for Arabic-English Statistical Machine Translation</i>	
Arianna Bisazza and Marcello Federico	235
<i>Head Finalization: A Simple Reordering Rule for SOV Languages</i>	
Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada and Kevin Duh	244
<i>Aiding Pronoun Translation with Co-Reference Resolution</i>	
Ronan Le Nagard and Philipp Koehn	252
<i>Jane: Open Source Hierarchical Translation, Extended with Reordering and Lexicon Models</i>	
David Vilar, Daniel Stein, Matthias Huck and Hermann Ney	262
<i>MANY: Open Source MT System Combination at WMT'10</i>	
Loïc Barrault	271
<i>Adaptive Model Weighting and Transductive Regression for Predicting Best System Combinations</i>	
Ergun Biciçi and S. Serdar Kozat	276
<i>L1 Regularized Regression for Reranking and System Combination in Machine Translation</i>	
Ergun Biciçi and Deniz Yuret	282
<i>An Augmented Three-Pass System Combination Framework: DCU Combination System for WMT 2010</i>	
Jinhua Du, Pavel Pecina and Andy Way	290
<i>The UPV-PRHLT Combination System for WMT 2010</i>	
Jesús González-Rubio, Germán Sanchis-Trilles, Joan-Andreu Sánchez, Jesús Andrés-Ferrer, Guillem Gascó, Pascual Martínez-Gómez, Martha-Alicia Rocha and Francisco Casacuberta	296
<i>CMU Multi-Engine Machine Translation for WMT 2010</i>	
Kenneth Heafield and Alon Lavie	301
<i>CMU System Combination via Hypothesis Selection for WMT'10</i>	
Almut Silja Hildebrand and Stephan Vogel	307
<i>JHU System Combination Scheme for WMT 2010</i>	
Sushant Narsale	311
<i>The RWTH System Combination System for WMT 2010</i>	
Gregor Leusch and Hermann Ney	315
<i>BBN System Description for WMT10 System Combination Task</i>	
Antti-Veikko Rosti, Bing Zhang, Spyros Matsoukas and Richard Schwartz	321
<i>LRscore for Evaluating Lexical and Reordering Quality in MT</i>	
Alexandra Birch and Miles Osborne	327
<i>Document-Level Automatic MT Evaluation based on Discourse Representations</i>	
Elisabet Comelles, Jesus Gimenez, Lluís Marquez, Irene Castellon and Victoria Arranz	333
<i>METEOR-NEXT and the METEOR Paraphrase Tables: Improved Evaluation Support for Five Target Languages</i>	
Michael Denkowski and Alon Lavie	339

<i>Normalized Compression Distance Based Measures for MetricsMATR 2010</i>	
Marcus Dobrinkat, Tero Tapiovaara, Jaakko Väyrynen and Kimmo Kettunen .....	343
<i>The DCU Dependency-Based Metric in WMT-MetricsMATR 2010</i>	
Yifan He, Jinhua Du, Andy Way and Josef van Genabith .....	349
<i>TESLA: Translation Evaluation of Sentences with Linear-Programming-Based Analysis</i>	
Chang Liu, Daniel Dahlmeier and Hwee Tou Ng .....	354
<i>The Parameter-Optimized ATEC Metric for MT Evaluation</i>	
Billy Wong and Chunyu Kit .....	360
<i>A Unified Approach to Minimum Risk Training and Decoding</i>	
Abhishek Arun, Barry Haddow and Philipp Koehn .....	365
<i>N-Best Reranking by Multitask Learning</i>	
Kevin Duh, Katsuhito Sudoh, Hajime Tsukada, Hideki Isozaki and Masaaki Nagata .....	375
<i>Taming Structured Perceptrons on Wild Feature Vectors</i>	
Ralf Brown .....	384
<i>Translation Model Adaptation by Resampling</i>	
Kashif Shah, Loïc Barrault and Holger Schwenk .....	392
<i>Integration of Multiple Bilingually-Learned Segmentation Schemes into Statistical Machine Translation</i>	
Michael Paul, Andrew Finch and Eiichiro Sumita .....	400
<i>Improved Translation with Source Syntax Labels</i>	
Hieu Hoang and Philipp Koehn .....	409
<i>Divide and Translate: Improving Long Distance Reordering in Statistical Machine Translation</i>	
Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, Tsutomu Hirao and Masaaki Nagata .....	418
<i>Decision Trees for Lexical Smoothing in Statistical Machine Translation</i>	
Rabih Zbib, Spyros Matsoukas, Richard Schwartz and John Makhoul .....	428