

# **8th Workshop on Statistical Machine Translation**

**(WMT 2013)**

**Sofia, Bulgaria  
8 - 9 August 2013**

ISBN: 978-1-62748-986-7

**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© (2013) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2013)

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>Findings of the 2013 Workshop on Statistical Machine Translation</i> Ondřej Bojar, Christian Buck, Chris Callison-Burch, Christian Federmann, Barry Haddow, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut and Lucia Specia .....	1
<i>Results of the WMT13 Metrics Shared Task</i> Matouš Macháček and Ondřej Bojar .....	45
<i>The Feasibility of HMEANT as a Human MT Evaluation Metric</i> Alexandra Birch, Barry Haddow, Ulrich Germann, Maria Nadejde, Christian Buck and Philipp Koehn .....	52
<i>LIMSI @ WMT13</i> Alexander Allauzen, Nicolas Pécheux, Quoc Khanh Do, Marco Dinarelli, Thomas Lavergne, Aurélien Max, Hai-Son Le and François Yvon .....	62
<i>The CMU Machine Translation Systems at WMT 2013: Syntax, Synthetic Translation Options, and Pseudo-References</i> Waleed Ammar, Victor Chahuneau, Michael Denkowski, Greg Hanneman, Wang Ling, Austin Matthews, Kenton Murray, Nicola Segall, Alon Lavie and Chris Dyer .....	70
<i>Feature Decay Algorithms for Fast Deployment of Accurate Statistical Machine Translation Systems</i> Ergun Biciçi .....	78
<i>CUni Multilingual Matrix in the WMT 2013 Shared Task</i> Karel Bílek and Daniel Zeman .....	85
<i>Chimera – Three Heads for English-to-Czech Translation</i> Ondřej Bojar, Rudolf Rosa and Aleš Tamchyna .....	92
<i>Yandex School of Data Analysis Machine Translation Systems for WMT13</i> Alexey Borisov, Jacob Dlougach and Irina Galinskaya .....	99
<i>The Karlsruhe Institute of Technology Translation Systems for the WMT 2013</i> Eunah Cho, Thanh-Le Ha, Mohammed Mediani, Jan Niehues, Teresa Herrmann, Isabel Slawik and Alex Waibel .....	104
<i>TÜBİTAK-BİLGEM German-English Machine Translation Systems for W13</i> Ilknur Durgar El-Kahlout and Coşkun Mermer .....	109
<i>Edinburgh’s Machine Translation Systems for European Language Pairs</i> Nadir Durrani, Barry Haddow, Kenneth Heafield and Philipp Koehn .....	114
<i>Munich-Edinburgh-Stuttgart Submissions of OSM Systems at WMT13</i> Nadir Durrani, Alexander Fraser, Helmut Schmid, Hassan Sajjad and Richárd Farkas .....	122
<i>Towards Efficient Large-Scale Feature-Rich Statistical Machine Translation</i> Vladimir Eidelman, Ke Wu, Ferhan Ture, Philip Resnik and Jimmy Lin .....	128
<i>The TALP-UPC Phrase-Based Translation Systems for WMT13: System Combination with Morphology Generation, Domain Adaptation and Corpus Filtering</i> Lluís Formiga, Marta R. Costa-jussà, José B. Mariño, José A. R. Fonollosa, Alberto Barrón-Cedeño and Lluís Marquez .....	134

<i>PhraseFix: Statistical Post-Editing of TectoMT</i>	
Petra Galuščáková, Martin Popel and Ondřej Bojar . . . . .	141
<i>Feature-Rich Phrase-based Translation: Stanford University's Submission to the WMT 2013 Translation Task</i>	
Spence Green, Daniel Cer, Kevin Reschke, Rob Voigt, John Bauer, Sida Wang, Natalia Silveira, Julia Neidert and Christopher D. Manning . . . . .	148
<i>Factored Machine Translation Systems for Russian-English</i>	
Stéphane Huet, Elena Manishina and Fabrice Lefèvre . . . . .	154
<i>Omnifluent English-to-French and Russian-to-English Systems for the 2013 Workshop on Statistical Machine Translation</i>	
Evgeny Matusov and Gregor Leusch . . . . .	158
<i>Pre-Reordering for Machine Translation Using Transition-Based Walks on Dependency Parse Trees</i>	
Antonio Valerio Miceli Barone and Giuseppe Attardi . . . . .	164
<i>Edinburgh's Syntax-Based Machine Translation Systems</i>	
Maria Nadejde, Philip Williams and Philipp Koehn . . . . .	170
<i>Shallow Semantically-Informed PBSMT and HPBSMT</i>	
Tsuyoshi Okita, Qun Liu and Josef van Genabith . . . . .	177
<i>Joint WMT 2013 Submission of the QUAERO Project</i>	
Stephan Peitz, Saab Mansour, Matthias Huck, Markus Freitag, Hermann Ney, Eunah Cho, Teresa Herrmann, Mohammed Mediani, Jan Niehues, Alex Waibel, Alexander Allauzen, Quoc Khanh Do, Bianka Buschbeck and Tonio Wandmacher . . . . .	185
<i>The RWTH Aachen Machine Translation System for WMT 2013</i>	
Stephan Peitz, Saab Mansour, Jan-Thorsten Peter, Christoph Schmidt, Joern Wuebker, Matthias Huck, Markus Freitag and Hermann Ney . . . . .	193
<i>The University of Cambridge Russian-English System at WMT13</i>	
Juan Pino, Aurelien Waite, Tong Xiao, Adrià de Gispert, Federico Flego and William Byrne . . . . .	200
<i>Joshua 5.0: Sparser, Better, Faster, Server</i>	
Matt Post, Juri Ganitkevitch, Luke Orland, Jonathan Weese, Yuan Cao and Chris Callison-Burch . . . . .	206
<i>The CNGL-DCU-Prompsit Translation Systems for WMT13</i>	
Raphael Rubino, Antonio Toral, Santiago Cortés Vaíllo, Jun Xie, Xiaofeng Wu, Stephen Doherty and Qun Liu . . . . .	213
<i>QCRI-MES Submission at WMT13: Using Transliteration Mining to Improve Statistical Machine Translation</i>	
Hassan Sajjad, Svetlana Smekalova, Nadir Durrani, Alexander Fraser and Helmut Schmid . . . . .	219
<i>Tunable Distortion Limits and Corpus Cleaning for SMT</i>	
Sara Stymne, Christian Hardmeier, Jörg Tiedemann and Joakim Nivre . . . . .	225
<i>Munich-Edinburgh-Stuttgart Submissions at WMT13: Morphological and Syntactic Processing for SMT</i>	
Marion Weller, Max Kisselew, Svetlana Smekalova, Alexander Fraser, Helmut Schmid, Nadir Durrani, Hassan Sajjad and Richárd Farkas . . . . .	232

<i>Coping with the Subjectivity of Human Judgements in MT Quality Estimation</i> Marco Turchi, Matteo Negri and Marcello Federico .....	240
<i>Online Polylingual Topic Models for Fast Document Translation Detection</i> Kriste Krstovski and David A. Smith .....	252
<i>Combining Bilingual and Comparable Corpora for Low Resource Machine Translation</i> Ann Irvine and Chris Callison-Burch .....	262
<i>Generating English Determiners in Phrase-Based Translation with Synthetic Translation Options</i> Yulia Tsvetkov, Chris Dyer, Lori Levin and Archana Bhatia .....	271
<i>Dramatically Reducing Training Data Size Through Vocabulary Saturation</i> William Lewis and Sauleh Eetemadi .....	281
<i>Multi-Task Learning for Improved Discriminative Training in SMT</i> Patrick Simianer and Stefan Riezler .....	292
<i>Online Learning Approaches in Computer Assisted Translation</i> Prashant Mathur, Cettolo Mauro and Marcello Federico .....	301
<i>Length-Incremental Phrase Training for SMT</i> Joern Wuebker and Hermann Ney .....	309
<i>Positive Diversity Tuning for Machine Translation System Combination</i> Daniel Cer, Christopher D. Manning and Dan Jurafsky .....	320
<i>Selecting Feature Sets for Comparative and Time-Oriented Quality Estimation of Machine Translation Output</i> Eleftherios Avramidis and Maja Popovic .....	329
<i>SHEF-Lite: When Less is More for Translation Quality Estimation</i> Daniel Beck, Kashif Shah, Trevor Cohn and Lucia Specia .....	337
<i>Referential Translation Machines for Quality Estimation</i> Ergun Bicici .....	343
<i>FBK-UEdin Participation to the WMT13 Quality Estimation Shared Task</i> José Guilherme Camargo de Souza, Christian Buck, Marco Turchi and Matteo Negri .....	352
<i>The TALP-UPC Approach to System Selection: Asiya Features and Pairwise Classification Using Random Forests</i> Lluís Formiga, Meritxell González, Alberto Barrón-Cedeño, José A. R. Fonollosa and Lluís Marquez .....	359
<i>Quality Estimation for Machine Translation Using the Joint Method of Evaluation Criteria and Statistical Modeling</i> Aaron Li-Feng Han, Yi Lu, Derek F. Wong, Lidia S. Chao, Liangye He and Junwen Xing .....	365
<i>MT Quality Estimation: The CMU System for WMT'13</i> Silja Hildebrand and Stephan Vogel .....	373
<i>LORIA System for the WMT13 Quality Estimation Shared Task</i> David Langlois and Kamel Smaili .....	380

<i>LIG System for WMT13 QE Task: Investigating the Usefulness of Features in Word Confidence Estimation for MT</i>	
Ngoc Quang Luong, Benjamin Lecouteux and Laurent Besacier . . . . .	386
<i>DCU-Symantec at the WMT 2013 Quality Estimation Shared Task</i>	
Raphael Rubino, Joachim Wagner, Jennifer Foster, Johann Roturier, Rasoul Samad Zadeh Kaljahi and Fred Hollowood . . . . .	392
<i>LIMSI Submission for the WMT'13 Quality Estimation Task: an Experiment with N-Gram Posteriors</i>	
Anil Kumar Singh, Guillaume Wisniewski and François Yvon . . . . .	398
<i>Ranking Translations using Error Analysis and Quality Estimation</i>	
Mark Fishel . . . . .	405
<i>Are ACT's Scores Increasing with Better Translation Quality?</i>	
Najeh Hajlaoui . . . . .	408
<i>A Description of Tunable Machine Translation Evaluation Systems in WMT13 Metrics Task</i>	
Aaron Li-Feng Han, Derek F. Wong, Lidia S. Chao, Yi Lu, Liangye He, Yiming Wang and Jiayi Zhou . . . . .	414
<i>MEANT at WMT 2013: A Tunable, Accurate yet Inexpensive Semantic Frame Based MT Evaluation Metric</i>	
Chi-kiu Lo and Dekai Wu . . . . .	422
<i>An Approach Using Style Classification Features for Quality Estimation</i>	
Erwan Moreau and Raphael Rubino . . . . .	429
<i>DCU Participation in WMT2013 Metrics Task</i>	
Xiaofeng Wu, Hui Yu and Qun Liu . . . . .	435
<i>Efficient Solutions for Word Reordering in German-English Phrase-Based Statistical Machine Translation</i>	
Arianna Bisazza and Marcello Federico . . . . .	440
<i>A Phrase Orientation Model for Hierarchical Machine Translation</i>	
Matthias Huck, Joern Wuebker, Felix Rietig and Hermann Ney . . . . .	452
<i>A Dependency-Constrained Hierarchical Model with Moses</i>	
Yvette Graham . . . . .	464
<i>Investigations in Exact Inference for Hierarchical Translation</i>	
Wilker Aziz, Marc Dymetman and Sriram Venkatapathy . . . . .	472
<i>Evaluating (and Improving) Sentence Alignment under Noisy Conditions</i>	
Omar Zaidan and Vishal Chowdhary . . . . .	484
<i>Multi-Rate HMMs for Word Alignment</i>	
Elif Eyigöz, Daniel Gildea and Kemal Oflazer . . . . .	494
<i>Hidden Markov Tree Model for Word Alignment</i>	
Shuhei Kondo, Kevin Duh and Yuji Matsumoto . . . . .	503
<i>An MT Error-Driven Discriminative Word Lexicon using Sentence Structure Features</i>	
Jan Niehues and Alex Waibel . . . . .	512