

7th Workshop on Statistical Machine Translation 2012

(WMT 2012)

**Montreal, Quebec, Canada
7 – 8 June 2012**

ISBN: 978-1-62276-592-8

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© (2012) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2012)

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

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
Web: www.proceedings.com

Table of Contents

<i>Putting Human Assessments of Machine Translation Systems in Order</i> Adam Lopez	1
<i>Findings of the 2012 Workshop on Statistical Machine Translation</i> Chris Callison-Burch, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut and Lucia Specia	10
<i>Semantic Textual Similarity for MT evaluation</i> Julio Castillo and Paula Estrella	52
<i>Improving AMBER, an MT Evaluation Metric</i> Boxing Chen, Roland Kuhn and George Foster	59
<i>TerrorCat: a Translation Error Categorization-based MT Quality Metric</i> Mark Fishel, Rico Sennrich, Maja Popović and Ondřej Bojar	64
<i>Class error rates for evaluation of machine translation output</i> Maja Popovic	71
<i>SPEDE: Probabilistic Edit Distance Metrics for MT Evaluation</i> Mengqiu Wang and Christopher Manning	76
<i>Quality estimation for Machine Translation output using linguistic analysis and decoding features</i> Eleftherios Avramidis	84
<i>Black Box Features for the WMT 2012 Quality Estimation Shared Task</i> Christian Buck	91
<i>Linguistic Features for Quality Estimation</i> Mariano Felice and Lucia Specia	96
<i>PRHLT Submission to the WMT12 Quality Estimation Task</i> Jesús González-Rubio, Alberto Sanchís and Francisco Casacuberta	104
<i>Tree Kernels for Machine Translation Quality Estimation</i> Christian Hardmeier, Joakim Nivre and Jörg Tiedemann	109
<i>LORIA System for the WMT12 Quality Estimation Shared Task</i> David Langlois, Sylvain Raybaud and Kamel Smaïli	114
<i>Quality Estimation: an experimental study using unsupervised similarity measures</i> Erwan Moreau and Carl Vogel	120
<i>The UPC Submission to the WMT 2012 Shared Task on Quality Estimation</i> Daniele Pighin, Meritxell González and Lluís Màrquez	127

<i>Morpheme- and POS-based IBM1 and language model scores for translation quality estimation</i>	
Maja Popovic	133
<i>DCU-Symantec Submission for the WMT 2012 Quality Estimation Task</i>	
Raphael Rubino, Jennifer Foster, Joachim Wagner, Johann Roturier, Rasul Samad Zadeh Kaljahi and Fred Hollowood	138
<i>The SDL Language Weaver Systems in the WMT12 Quality Estimation Shared Task</i>	
Radu Soricut, Nguyen Bach and Ziyuan Wang	145
<i>Regression with Phrase Indicators for Estimating MT Quality</i>	
Chunyang Wu and Hai Zhao	152
<i>Non-Linear Models for Confidence Estimation</i>	
Yong Zhuang, Guillaume Wisniewski and François Yvon	157
<i>Combining Quality Prediction and System Selection for Improved Automatic Translation Output</i>	
Radu Soricut and Sushant Narsale	163
<i>Match without a Referee: Evaluating MT Adequacy without Reference Translations</i>	
Yashar Mehdad, Matteo Negri and Marcello Federico	171
<i>Comparing human perceptions of post-editing effort with post-editing operations</i>	
Maarit Koponen	181
<i>Review of Hypothesis Alignment Algorithms for MT System Combination via Confusion Network Decoding</i>	
Antti-Veikko Rosti, Xiaodong He, Damianos Karakos, Gregor Leusch, Yuan Cao, Markus Freitag, Spyros Matsoukas, Hermann Ney, Jason Smith and Bing Zhang	191
<i>On Hierarchical Re-ordering and Permutation Parsing for Phrase-based Decoding</i>	
Colin Cherry, Robert C. Moore and Chris Quirk	200
<i>CCG Syntactic Reordering Models for Phrase-based Machine Translation</i>	
Dennis Nolan Mehay and Christopher Hardie Brew	210
<i>Using Categorial Grammar to Label Translation Rules</i>	
Jonathan Weese, Chris Callison-Burch and Adam Lopez	222
<i>Using Syntactic Head Information in Hierarchical Phrase-Based Translation</i>	
Junhui Li, Zhaopeng Tu, Guodong Zhou and Josef van Genabith	232
<i>Fully Automatic Semantic MT Evaluation</i>	
Chi-kiu Lo, Anand Karthik Tumuluru and Dekai Wu	243
<i>Probes in a Taxonomy of Factored Phrase-Based Models</i>	
Ondřej Bojar, Bushra Jawaid and Amir Kamran	253
<i>The CMU-Avenue French-English Translation System</i>	
Michael Denkowski, Greg Hanneman and Alon Lavie	261

<i>Formemes in English-Czech Deep Syntactic MT</i>	
Ondřej Dušek, Zdeněk Žabokrtský, Martin Popel, Martin Majliš, Michal Novák and David Mareček	
	267
<i>The TALP-UPC phrase-based translation systems for WMT12: Morphology simplification and domain adaptation</i>	
Lluís Formiga, Carlos A. Henríquez Q., Adolfo Hernández, José B. Mariño, Enric Monte and José A. R. Fonollosa	
	275
<i>Joshua 4.0: Packing, PRO, and Paraphrases</i>	
Juri Ganitkevitch, Yuan Cao, Jonathan Weese, Matt Post and Chris Callison-Burch	
	283
<i>Syntax-aware Phrase-based Statistical Machine Translation: System Description</i>	
Ulrich Germann	
	292
<i>QCRI at WMT12: Experiments in Spanish-English and German-English Machine Translation of News Text</i>	
Francisco Guzman, Preslav Nakov, Ahmed Thabet and Stephan Vogel	
	298
<i>The RWTH Aachen Machine Translation System for WMT 2012</i>	
Matthias Huck, Stephan Peitz, Markus Freitag, Malte Nuhn and Hermann Ney	
	304
<i>Machine Learning for Hybrid Machine Translation</i>	
Sabine Hunsicker, Chen Yu and Christian Federmann	
	312
<i>Towards Effective Use of Training Data in Statistical Machine Translation</i>	
Philipp Koehn and Barry Haddow	
	317
<i>Joint WMT 2012 Submission of the QUAERO Project</i>	
Freitag Markus, Peitz Stephan, Huck Matthias, Ney Hermann, Niehues Jan, Herrmann Teresa, Waibel Alex, Hai-son Le, Lavergne Thomas, Allauzen Alexandre, Buschbeck Bianka, Crego Joseph Maria and Senellart Jean	
	322
<i>LIMSI @ WMT12</i>	
Hai-Son Le, Thomas Lavergne, Alexandre Allauzen, Marianna Apidianaki, Li Gong, Aurélien Max, Artem Sokolov, Guillaume Wisniewski and François Yvon	
	330
<i>UPM system for WMT 2012</i>	
Verónica López-Ludeña, Rubén San-Segundo and Juan M. Montero	
	338
<i>PROMT DeepHybrid system for WMT12 shared translation task</i>	
Alexander Molchanov	
	345
<i>The Karlsruhe Institute of Technology Translation Systems for the WMT 2012</i>	
Jan Niehues, Yuqi Zhang, Mohammed Mediani, Teresa Herrmann, Eunah Cho and Alex Waibel	
	349
<i>Kriya - The SFU System for Translation Task at WMT-12</i>	
Majid Razmara, Baskaran Sankaran, Ann Clifton and Anoop Sarkar	
	356

<i>DEPFIx: A System for Automatic Correction of Czech MT Outputs</i> Rudolf Rosa, David Mareček and Ondřej Dušek	362
<i>LIUM's SMT Machine Translation Systems for WMT 2012</i> Christophe Servan, Patrik Lambert, Anthony Rousseau, Holger Schwenk and Loïc Barrault . .	369
<i>Selecting Data for English-to-Czech Machine Translation</i> Aleš Tamchyna, Petra Galuščáková, Amir Kamran, Miloš Stanojević and Ondřej Bojar	374
<i>DFKI's SMT System for WMT 2012</i> David Vilar	382
<i>GHKM Rule Extraction and Scope-3 Parsing in Moses</i> Philip Williams and Philipp Koehn	388
<i>Data Issues of the Multilingual Translation Matrix</i> Daniel Zeman	395
<i>Constructing Parallel Corpora for Six Indian Languages via Crowdsourcing</i> Matt Post, Chris Callison-Burch and Miles Osborne	401
<i>Twitter Translation using Translation-Based Cross-Lingual Retrieval</i> Laura Jehl, Felix Hieber and Stefan Riezler	410
<i>Analysing the Effect of Out-of-Domain Data on SMT Systems</i> Barry Haddow and Philipp Koehn	422
<i>Evaluating the Learning Curve of Domain Adaptive Statistical Machine Translation Systems</i> Nicola Bertoldi, Mauro Cettolo, Marcello Federico and Christian Buck	433
<i>The Trouble with SMT Consistency</i> Marine Carpuat and Michel Simard	442
<i>Phrase Model Training for Statistical Machine Translation with Word Lattices of Preprocessing Alternatives</i> Joern Wuebker and Hermann Ney	450
<i>Leave-One-Out Phrase Model Training for Large-Scale Deployment</i> Joern Wuebker, Mei-Yuh Hwang and Chris Quirk	460
<i>Direct Error Rate Minimization for Statistical Machine Translation</i> Tagyoung Chung and Michel Galley	468
<i>Optimization Strategies for Online Large-Margin Learning in Machine Translation</i> Vladimir Eidelman	480