

# **6th Workshop on Statistical Machine Translation 2011**

**(WMT 2011)**

**Held at EMNLP 2011**

**Edinburgh, Scotland, UK  
30-31 July 2011**

**ISBN: 978-1-61839-255-8**

**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© (2011) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2011)

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 Grain of Salt for the WMT Manual Evaluation</i>	
Ondřej Bojar, Miloš Ercegovčević, Martin Popel and Omar Zaidan .....	1
<i>A Lightweight Evaluation Framework for Machine Translation Reordering</i>	
David Talbot, Hideto Kazawa, Hiroshi Ichikawa, Jason Katz-Brown, Masakazu Seno and Franz Och .....	12
<i>Findings of the 2011 Workshop on Statistical Machine Translation</i>	
Chris Callison-Burch, Philipp Koehn, Christof Monz and Omar Zaidan .....	22
<i>Evaluate with Confidence Estimation: Machine ranking of translation outputs using grammatical features</i>	
Eleftherios Avramidis, Maja Popović, David Vilar and Aljoscha Burchardt .....	65
<i>AMBER: A Modified BLEU, Enhanced Ranking Metric</i>	
Boxing Chen and Roland Kuhn .....	71
<i>TESLA at WMT 2011: Translation Evaluation and Tunable Metric</i>	
Daniel Dahlmeier, Chang Liu and Hwee Tou Ng .....	78
<i>Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems</i>	
Michael Denkowski and Alon Lavie .....	85
<i>Approximating a Deep-Syntactic Metric for MT Evaluation and Tuning</i>	
Matouš Macháček and Ondřej Bojar .....	92
<i>Evaluation without references: IBM1 scores as evaluation metrics</i>	
Maja Popović, David Vilar, Eleftherios Avramidis and Aljoscha Burchardt .....	99
<i>Morphemes and POS tags for n-gram based evaluation metrics</i>	
Maja Popović .....	104
<i>E-rating Machine Translation</i>	
Kristen Parton, Joel Tetreault, Nitin Madnani and Martin Chodorow .....	108
<i>TINE: A Metric to Assess MT Adequacy</i>	
Miguel Rios, Wilker Aziz and Lucia Specia .....	116
<i>Regression and Ranking based Optimisation for Sentence Level MT Evaluation</i>	
Xingyi Song and Trevor Cohn .....	123
<i>MAISE: A Flexible, Configurable, Extensible Open Source Package for Mass AI System Evaluation</i>	
Omar Zaidan .....	130

<i>MANY improvements for WMT'11</i>	
Loïc Barrault .....	135
<i>The UPV-PRHLT combination system for WMT 2011</i>	
Jesús González-Rubio and Francisco Casacuberta .....	140
<i>CMU System Combination in WMT 2011</i>	
Kenneth Heafield and Alon Lavie .....	145
<i>The RWTH System Combination System for WMT 2011</i>	
Gregor Leusch, Markus Freitag and Hermann Ney .....	152
<i>Expected BLEU Training for Graphs: BBN System Description for WMT11 System Combination Task</i>	
Antti-Veikko Rosti, Bing Zhang, Spyros Matsoukas and Richard Schwartz .....	159
<i>The UZH System Combination System for WMT 2011</i>	
Rico Sennrich .....	166
<i>Description of the JHU System Combination Scheme for WMT 2011</i>	
Daguang Xu, Yuan Cao and Damianos Karakos .....	171
<i>Multiple-stream Language Models for Statistical Machine Translation</i>	
Abby Levenberg, Miles Osborne and David Matthews .....	177
<i>KenLM: Faster and Smaller Language Model Queries</i>	
Kenneth Heafield .....	187
<i>Wider Context by Using Bilingual Language Models in Machine Translation</i>	
Jan Niehues, Teresa Herrmann, Stephan Vogel and Alex Waibel .....	198
<i>A Minimally Supervised Approach for Detecting and Ranking Document Translation Pairs</i>	
Kriste Krstovski and David A. Smith .....	207
<i>Agreement Constraints for Statistical Machine Translation into German</i>	
Philip Williams and Philipp Koehn .....	217
<i>Fuzzy Syntactic Reordering for Phrase-based Statistical Machine Translation</i>	
Jacob Andreas, Nizar Habash and Owen Rambow .....	227
<i>Filtering Antonymous, Trend-Contrasting, and Polarity-Dissimilar Distributional Paraphrases for Improving Statistical Machine Translation</i>	
Yuval Marton, Ahmed El Kholy and Nizar Habash .....	237
<i>Productive Generation of Compound Words in Statistical Machine Translation</i>	
Sara Stymne and Nicola Cancedda .....	250
<i>SampleRank Training for Phrase-Based Machine Translation</i>	
Barry Haddow, Abhishek Arun and Philipp Koehn .....	261

<i>Instance Selection for Machine Translation using Feature Decay Algorithms</i> Ergun Biciçi and Deniz Yuret .....	272
<i>Investigations on Translation Model Adaptation Using Monolingual Data</i> Patrik Lambert, Holger Schwenk, Christophe Servan and Sadaf Abdul-Rauf .....	284
<i>Topic Adaptation for Lecture Translation through Bilingual Latent Semantic Models</i> Nick Ruiz and Marcello Federico .....	294
<i>Personal Translator at WMT2011</i> Vera Aleksic and Gregor Thurmair .....	303
<i>LIMSI @ WMT11</i> Alexandre Allauzen, H�el�ene Bonneau-Maynard, Hai-Son Le, Aur�elien Max, Guillaume Wisniewski, Fran�ois Yvon, Gilles Adda, Josep Maria Crego, Adrien Lardilleux, Thomas Lavergne and Artem Sokolov .....	309
<i>Shallow Semantic Trees for SMT</i> Wilker Aziz, Miguel Rios and Lucia Specia .....	316
<i>RegMT System for Machine Translation, System Combination, and Evaluation</i> Ergun Biciçi and Deniz Yuret .....	323
<i>Improving Translation Model by Monolingual Data</i> Ondr�ej Bojar and Aleř Tamchyna .....	330
<i>The CMU-ARK German-English Translation System</i> Chris Dyer, Kevin Gimpel, Jonathan H. Clark and Noah A. Smith .....	337
<i>Noisy SMS Machine Translation in Low-Density Languages</i> Vladimir Eidelman, Kristy Hollingshead and Philip Resnik .....	344
<i>Stochastic Parse Tree Selection for an Existing RBMT System</i> Christian Federmann and Sabine Hunsicker .....	351
<i>Joint WMT Submission of the QUAERO Project</i> Markus Freitag, Gregor Leusch, Joern Wuebker, Stephan Peitz, Hermann Ney, Teresa Herrmann, Jan Niehues, Alex Waibel, Alexandre Allauzen, Gilles Adda, Josep Maria Crego, Bianka Buschbeck, Tonio Wandmacher and Jean Senellart .....	358
<i>CMU Syntax-Based Machine Translation at WMT 2011</i> Greg Hanneman and Alon Lavie .....	365
<i>The Uppsala-FBK systems at WMT 2011</i> Christian Hardmeier, J�org Tiedemann, Markus Saers, Marcello Federico and Mathur Prashant	372
<i>The Karlsruhe Institute of Technology Translation Systems for the WMT 2011</i> Teresa Herrmann, Mohammed Mediani, Jan Niehues and Alex Waibel .....	379

<i>CMU Haitian Creole-English Translation System for WMT 2011</i>	
Sanjika Hewavitharana, Nguyen Bach, Qin Gao, Vamshi Ambati and Stephan Vogel . . . . .	386
<i>Experiments with word alignment, normalization and clause reordering for SMT between English and German</i>	
Maria Holmqvist, Sara Stymne and Lars Ahrenberg . . . . .	393
<i>The Value of Monolingual Crowdsourcing in a Real-World Translation Scenario: Simulation using Haitian Creole Emergency SMS Messages</i>	
Chang Hu, Philip Resnik, Yakov Kronrod, Vladimir Eidelman, Olivia Buzek and Benjamin B. Bederson . . . . .	399
<i>The RWTH Aachen Machine Translation System for WMT 2011</i>	
Matthias Huck, Joern Wuebker, Christoph Schmidt, Markus Freitag, Stephan Peitz, Daniel Stein, Arnaud Dagnelies, Saab Mansour, Gregor Leusch and Hermann Ney . . . . .	405
<i>ILLC-UvA translation system for EMNLP-WMT 2011</i>	
Maxim Khalilov and Khalil Sima'an . . . . .	413
<i>UPM system for the translation task</i>	
Verónica López-Ludeña and Rubén San-Segundo . . . . .	420
<i>Two-step translation with grammatical post-processing</i>	
David Mareček, Rudolf Rosa, Petra Galuščáková and Ondřej Bojar . . . . .	426
<i>Influence of Parser Choice on Dependency-Based MT</i>	
Martin Popel, David Mareček, Nathan Green and Zdenek Zabokrtsky . . . . .	433
<i>The LIGA (LIG/LIA) Machine Translation System for WMT 2011</i>	
Marion Potet, Raphaël Rubino, Benjamin Lecouteux, Stéphane Huet, Laurent Besacier, Hervé Blanchon and Fabrice Lefèvre . . . . .	440
<i>Factored Translation with Unsupervised Word Clusters</i>	
Christian Rishøj and Anders Søgaard . . . . .	447
<i>The BM-I2R Haitian-Créole-to-English translation system description for the WMT 2011 evaluation campaign</i>	
Marta R. Costa-jussà and Rafael E. Banchs . . . . .	452
<i>The Universitat d'Alacant hybrid machine translation system for WMT 2011</i>	
Víctor M. Sánchez-Cartagena, Felipe Sánchez-Martínez and Juan Antonio Pérez-Ortiz . . . . .	457
<i>LIUM's SMT Machine Translation Systems for WMT 2011</i>	
Holger Schwenk, Patrik Lambert, Loïc Barrault, Christophe Servan, Sadaf Abdul-Rauf, Haithem Afli and Kashif Shah . . . . .	464
<i>Spell Checking Techniques for Replacement of Unknown Words and Data Cleaning for Haitian Creole SMS Translation</i>	
Sara Stymne . . . . .	470

<i>Joshua 3.0: Syntax-based Machine Translation with the Thrax Grammar Extractor</i>	
Jonathan Weese, Juri Ganitkevitch, Chris Callison-Burch, Matt Post and Adam Lopez . . . . .	478
<i>DFKI Hybrid Machine Translation System for WMT 2011 - On the Integration of SMT and RBMT</i>	
Jia Xu, Hans Uszkoreit, Casey Kennington, David Vilar and Xiaojun Zhang . . . . .	485
<i>CEU-UPV English-Spanish system for WMT11</i>	
Francisco Zamora-Martinez and Maria Jose Castro-Bleda . . . . .	490
<i>Hierarchical Phrase-Based MT at the Charles University for the WMT 2011 Shared Task</i>	
Daniel Zeman . . . . .	496
<i>Crisis MT: Developing A Cookbook for MT in Crisis Situations</i>	
William Lewis, Robert Munro and Stephan Vogel . . . . .	501
<i>Generative Models of Monolingual and Bilingual Gappy Patterns</i>	
Kevin Gimpel and Noah A. Smith . . . . .	512
<i>Extraction Programs: A Unified Approach to Translation Rule Extraction</i>	
Mark Hopkins, Greg Langmead and Tai Vo . . . . .	523
<i>Bayesian Extraction of Minimal SCFG Rules for Hierarchical Phrase-based Translation</i>	
Baskaran Sankaran, Gholamreza Haffari and Anoop Sarkar . . . . .	533
<i>From n-gram-based to CRF-based Translation Models</i>	
Thomas Lavergne, Alexandre Allauzen, Josep Maria Crego and François Yvon . . . . .	542