Workshop on Statistical Machine Translation 2006

(WMT 06)

Held at HLT/NAACL 2006

New York, New York, USA
8-9 June 2006

ISBN: 978-1-61738-493-6

Printed from e-media with permission by:

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

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

Copyright© (2006) 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

Table of Contents

Morpho-syntactic Information for Automatic Error Analysis of Statistical Machine Translation Output
Maja Popovic, Adrià de Gispert, Deepa Gupta, Patrik Lambert, Hermann Ney, José B. Mariño, Marcello Federico and Rafael Banchs 1

Initial Explorations in English to Turkish Statistical Machine Translation
ilknur Durgar El-Kahlout and Kemal Oflazer 7

Morpho-syntactic Arabic Preprocessing for Arabic to English Statistical Machine Translation
Anas El Isbihani, Shahram Khadivi, Oliver Bender and Hermann Ney 15

Quasi-Synchronous Grammars: Alignment by Soft Projection of Syntactic Dependencies
David Smith and Jason Eisner 23

Why Generative Phrase Models Underperform Surface Heuristics
John DeNero, Dan Gillick, James Zhang and Dan Klein 31

Phrase-Based SMT with Shallow Tree-Phrases
Philippe Langlais and Fabrizio Gotti 39

Searching for alignments in SMT. A novel approach based on an Estimation of Distribution Algorithm
Luis Rodríguez, Ismael García-Varea and Jose A. Gámez 47

Discriminative Reordering Models for Statistical Machine Translation
Richard Zens and Hermann Ney 55

Generalized Stack Decoding Algorithms for Statistical Machine Translation
Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta 64

N-Gram Posterior Probabilities for Statistical Machine Translation
Richard Zens and Hermann Ney 72

Partitioning Parallel Documents Using Binary Segmentation
Jia Xu, Richard Zens and Hermann Ney 78

Contextual Bitext-Derived Paraphrases in Automatic MT Evaluation
Karolina Owczarzak, Declan Groves, Josef Van Genabith and Andy Way 86

How Many Bits Are Needed To Store Probabilities for Phrase-Based Translation?
Marcello Federico and Nicola Bertoldi 94

Manual and Automatic Evaluation of Machine Translation between European Languages
Philipp Koehn and Christof Monz 102

NTT System Description for the WMT2006 Shared Task
Taro Watanabe, Hajime Tsukada and Hideki Isozaki 122

Mood at work: Ramses versus Pharaoh
Alexandre Patry, Fabrizio Gotti and Philippe Langlais 126

Stochastic Inversion Transduction Grammars for Obtaining Word Phrases for Phrase-based Statistical Machine Translation
Joan Andreu Sánchez and José Miguel Benedí 130

PORTAGE: with Smoothed Phrase Tables and Segment Choice Models
Howard Johnson, Fatiha Sadat, George Foster, Roland Kuhn, Michel Simard, Eric Joanis and Samuel Larkin 134

Syntax Augmented Machine Translation via Chart Parsing
Andreas Zollmann and Ashish Venugopal 138

TALP Phrase-based statistical translation system for European language pairs
Marta R. Costa-jussà, Josep M. Crego, Adrià de Gispert, Patrik Lambert, Maxim Khalilov, José B. Mariño, José A. R. Fonollosa and Rafael Banchs 142

Phramer - An Open Source Statistical Phrase-Based Translator
Marian Olteanu, Chris Davis, Ionut Volosen and Dan Moldovan 146

Language Models and Reranking for Machine Translation
Marian Olteanu, Pasin Suriyentrakorn and Dan Moldovan 150

Constraining the Phrase-Based, Joint Probability Statistical Translation Model
Alexandra Birch, Chris Callison-Burch, Miles Osborne and Philipp Koehn 154

Microsoft Research Treelet Translation System: NAACL 2006 Europarl Evaluation
Arul Menezes, Kristina Toutanova and Chris Quirk 158

N-gram-based SMT System Enhanced with Reordering Patterns
Josep M. Crego, Adrià de Gispert, Patrik Lambert, Marta R. Costa-jussà, Maxim Khalilov, Rafael Banchs, José B. Mariño and José A. R. Fonollosa 162

The LDV-COMBO system for SMT
Jesús Giménez and Lluís Màrquez 166