ACL 2007

PRAGUE

ACL 2007

Proceedings of the Second Workshop on Statistical Machine Translation

June 23, 2007
Prague, Czech Republic

The Association for Computational Linguistics

Production and Manufacturing by
Omnipress
2600 Anderson Street
Madison, WI 53704
USA

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Preface

The ACL 2007 Workshop on Statistical Machine Translation (WMT-07) took place on Saturday, June 23 in Prague, Czech Republic, immediately preceding the annual meeting of the Association for Computational Linguistics, which was hosted by Charles University. This was the second time this workshop had been held, following the first workshop at the 2006 HLT-NAACL conference. But its ancestry can be traced back farther to the ACL 2005 Workshop on Building and Using Parallel Texts (when we started our evaluation campaign on European languages), and even the ACL 2001 Workshop on Data-Driven Machine Translation (which was the first ACL workshop mostly directed at statistical machine translation).

Over the last years, interest in statistical machine translation has been risen dramatically. We received an overwhelming number of full paper submission for a one-day workshop, 38 in total. Given our limited capacity, we were only able to accept 12 full papers for oral presentation and 9 papers for poster presentation, an acceptance rate of 55%. In a second poster session, 16 additional shared task papers were presented. The workshop also featured an invited talk by Jean Senellart of SYSTRAN Language Translation Technology, Paris.

Prior to the workshop, in addition to soliciting relevant papers for review and possible presentation we conducted a shared task that brought together machine translation systems for an evaluation on previously unseen data. This year's task resembled the shared tasks of previous years in many ways. Its focus was again the translation of European languages, using a relatively large training corpus. This year, we included a variety of manual evaluations of the MT systems' outputs, and a variety of automated evaluation metrics. Also, as a special challenge this year, we posed the problem of domain adaptation.

The results of the shared task were announced at the workshop, and these proceedings also include an overview paper for the shared task that summarizes the results, as well as provides information about the data used and any procedures that were followed in conducting or scoring the task. In addition, there are short papers from each participating team that describe their underlying system in some detail.

We would like to thank the members of the Program Committee for their timely reviews. We also would like to thank the participants of the shared task, the participants of the MT Marathon, which was organized by the University of Edinburgh in March this year, and all the other volunteers who helped with the manual evaluations. We also acknowledge financial support for the manual evaluation by the EuroMatrix project (funded by the European Commission under the Framework Programme 6).

Chris Callison-Burch, Philipp Koehn, Christof Monz, and Cameron Shaw Fordyce

Co-Organizers

Organizers

Chairs:

Chris Callison-Burch (Johns Hopkins University)
Philipp Koehn (University of Edinburgh)
Christof Monz (Queen Mary, University of London)
Cameron Shaw Fordyce (Center for the Evaluation of Language and Communication Technologies)

Invited Speaker:

Jean Senellart (SYSTRAN Language Translation Technology, Paris)

Program Committee:

Lars Ahrenberg (Linköping University)
Francisco Casacuberta (University of Valencia)
Colin Cherry (University of Alberta)
Stephen Clark (Oxford University)
Brooke Cowan (Massachusetts Institute of Technology)
Mona Diab (Columbia University)
Chris Dyer (University of Maryland)
Andreas Eisele (University Saarbrücken)
Marcello Federico (ITC-IRST)
George Foster (Canada National Research Council)
Alex Fraser (ISI/University of Southern California)
Ulrich Germann (University of Toronto)
Rebecca Hwa (University of Pittsburgh)
Kevin Knight (ISI/University of Southern California)
Philippe Langlais (University of Montreal)
Alon Lavie (Carnegie Melon University)
Lori Levin (Carnegie Mellon University)
Daniel Marcu (ISI/University of Southern California)
Bob Moore (Microsoft Research)
Miles Osborne (University of Edinburgh)
Michel Simard (Canada National Research Council)
Eiichiro Sumita (NICT/ATR)
Jörg Tiedemann (University of Groningen)
Christoph Tillmann (IBM Research)
Dan Tufiş (Romanian Academy)
Taro Watanabe (NTT)
Dekai Wu (HKUST)
Richard Zens (RWTH Aachen)

Additional Reviewers: Joshua Albrecht, Marine Carpuat, Hirofumi Yamamoto, and Keiji Yasuda.

Table of Contents

Conference Program

Saturday, June 23, 2007

8:30–8:35 Opening Remarks

Full Paper Session 1

8:35–8:55 *Using Dependency Order Templates to Improve Generality in Translation*
Arul Menezes and Chris Quirk

8:55–9:15 *CCG Supertags in Factored Statistical Machine Translation*
Alexandra Birch, Miles Osborne and Philipp Koehn

9:15–9:35 *Integration of an Arabic Transliteration Module into a Statistical Machine Translation System*
Mehdi M. Kashani, Eric Joanis, Roland Kuhn, George Foster and Fred Popowich

9:35–9:55 *Exploring Different Representational Units in English-to-Turkish Statistical Machine Translation*
Kemal Oflazer and Ilknur Durgar El-Kahlout

9:55–10:15 *Can We Translate Letters?*
David Vilar, Jan-Thorsten Peter and Hermann Ney

Poster Session for Full Papers

10:15–10:45 Boaster session for full paper posters

10:45–11:45 Coffee break + full paper poster session

A Dependency Treelet String Correspondence Model for Statistical Machine Translation
Deyi Xiong, Qun Liu and Shouxun Lin

Word Error Rates: Decomposition over POS classes and Applications for Error Analysis
Maja Popovic and Hermann Ney

Speech-Input Multi-Target Machine Translation
Alicia Pérez, M. Teresa González, M. Inés Torres and Francisco Casacuberta

Saturday, June 23, 2007 (continued)

Meta-Structure Transformation Model for Statistical Machine Translation
Jiadong Sun, Tiejun Zhao and Huashen Liang

Training Non-Parametric Features for Statistical Machine Translation
Patrick Nguyen, Milind Mahajan and Xiaodong He

Using Word-Dependent Transition Models in HMM-Based Word Alignment for Statistical Machine Translation
Xiaodong He

Efficient Handling of N-gram Language Models for Statistical Machine Translation
Marcello Federico and Mauro Cettolo

Human Evaluation of Machine Translation Through Binary System Comparisons
David Vilar, Gregor Leusch, Hermann Ney and Rafael E. Banchs

Labelled Dependencies in Machine Translation Evaluation
Karolina Owczarzak, Josef van Genabith and Andy Way

Invited Talk

11:45–12:30 Invited Talk by Jean Senellart, Systran

Lunch

Full Paper Session 2

14:00–14:20 *An Iteratively-Trained Segmentation-Free Phrase Translation Model for Statistical Machine Translation*
Robert Moore and Chris Quirk

14:20–14:40 *Using Paraphrases for Parameter Tuning in Statistical Machine Translation*
Nitin Madnani, Necip Fazil Ayan, Philip Resnik and Bonnie Dorr

14:40–15:00 *Mixture-Model Adaptation for SMT*
George Foster and Roland Kuhn

Saturday, June 23, 2007 (continued)

Shared Task

15:00–15:15 *(Meta-) Evaluation of Machine Translation*
Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz and Josh Schroeder

15:15–15:45 Boaster session for shared task posters

15:45–16:45 Coffee break + shared task poster session

Context-aware Discriminative Phrase Selection for Statistical Machine Translation
Jesús Giménez and Lluís Màrquez

Ngram-Based Statistical Machine Translation Enhanced with Multiple Weighted Reordering Hypotheses
Marta R. Costa-jussià, Josep M. Crego, Patrik Lambert, Maxim Khalilov, José A. R. Fonollosa, José B. Mario and Rafael E. Banchs

Analysis of Statistical and Morphological Classes to Generate Weigthed Reordering Hypotheses on a Statistical Machine Translation System
Marta R. Costa-jussià and José A. R. Fonollosa

Domain Adaptation in Statistical Machine Translation with Mixture Modelling
Jorge Civera and Alfons Juan

Getting to Know Moses: Initial Experiments on German-English Factored Translation
Maria Holmqvist, Sara Stymne and Lars Ahrenberg

NRC's PORTAGE System for WMT 2007
Nicola Ueffing, Michel Simard, Samuel Larkin and Howard Johnson

Building a Statistical Machine Translation System for French Using the Europarl Corpus
Holger Schwenk

Multi-Engine Machine Translation with an Open-Source SMT Decoder
Yu Chen, Andreas Eisele, Christian Federmann, Eva Hasler, Michael Jellinghaus and Silke Theison

The ISL Phrase-Based MT System for the 2007 ACL Workshop on Statistical Machine Translation
Matthias Paulik, Kay Rottmann, Jan Niehues, Silja Hildebrand and Stephan Vogel

Saturday, June 23, 2007 (continued)