

COLING 2012

**24th International Conference on
Computational Linguistics**

**Proceedings of the
Workshop on Reordering for Statistical
Machine Translation**

Workshop chairs:

**Karthik Visweswariah, Ananthakrishnan Ramanathan
and Mitesh M. Khapra**

9 December 2012

Mumbai, India

Diamond sponsors

Tata Consultancy Services
Linguistic Data Consortium for Indian Languages (LDC-IL)

Gold Sponsors

Microsoft Research
Beijing Baidu Netcon Science Technology Co. Ltd.

Silver sponsors

IBM, India Private Limited
Crimson Interactive Pvt. Ltd.
Yahoo
Easy Transcription & Software Pvt. Ltd.

Proceedings of the Workshop on Reordering for Statistical Machine Translation
Karthik Visweswariah, Ananthakrishnan Ramanathan and Mitesh M. Khapra
(eds.)
Revised preprint edition, 2012

Published by The COLING 2012 Organizing Committee
Indian Institute of Technology Bombay,
Powai,
Mumbai-400076
India
Phone: 91-22-25764729
Fax: 91-22-2572 0022
Email: pb@cse.iitb.ac.in

This volume © 2012 The COLING 2012 Organizing Committee.
Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Nonported* license.
<http://creativecommons.org/licenses/by-nc-sa/3.0/>
Some rights reserved.

Contributed content copyright the contributing authors.
Used with permission.

Also available online in the ACL Anthology at <http://aclweb.org>

Preface

Statistical Machine Translation (SMT) is currently a very active area of research. Top NLP conferences typically include many papers on SMT, and the past decade has also seen many workshops on this topic (e.g., WMT, SSST). Results reported in papers on SMT include the influence of various components and processes, one of the most important of these being reordering (except for structurally similar language pairs like English-French). Traditional phrase-based SMT systems, which have been the state-of-the-art in the previous decade, do not handle reordering in a satisfactory manner, and various new and more sophisticated methods for reordering have been introduced in the past couple of years. However, unlike most other potential components of MT systems, such as transliteration, WSD, and anaphora resolution, reordering has not had a dedicated forum for its evaluation. The proposed workshop will be a platform to bring together different MT systems and compare how they tackle this crucial subtask.

A shared task on "learning reordering from word-alignments" will be at the heart of this workshop. Parser-based reordering has been a popular method, but many languages do not have parsers (e.g., no Indian language has a publicly available parser), and using alignments to learn parsers (and thereafter reordering) or to learn reordering models directly is an important new idea in MT. The task is to develop a system for reordering a source sentence to best match the order of the corresponding target sentence. For example, the English (SVO language) sentence "Ram drinks water" is translated into Hindi (SOV language) as "Ram paanii piitaa hai (Ram water drinks)". Thus, the correct reordering of this English sentence which matches the target (Hindi) order is "Ram water drinks".

We released high-quality word-alignments (annotated by hand) between English and 3 languages (Farsi, Italian and Urdu), and described one or two baseline techniques for reordering based on publicly available tools (such as GIZA++, Moses). We also made available part-of-speech tags for this data to enable participants to experiment with these additional features easily. The participants have to reorder the English sentences to match the order of the target language. Participants can choose either to come up with their own reordering models or tweak the baseline system to improve performance.

Workshop Chairs

Mitesh M. Khapra, IBM Research India
Ananthakrishnan Ramanathan, IBM Research India
Karthik Visweswariah, IBM Research India

Organizers:

Mitesh M. Khapra (IBM Research India)
Ananthakrishnan Ramanathan (IBM Research India)
Karthik Visweswariah (IBM Research India)

Program Committee:

Abe Ittycheriah (IBM Watson Research Lab)
John DeNero (Google)
Pushpak Bhattacharyya (IIT Bombay)
Philipp Koehn (University of Edinburgh)
Phil Blunsom (Oxford University)

Table of Contents

<i>Whitepaper for Shared Task on Learning Reordering from Word Alignments at RSMT 2012</i> Mitesh M. Khapra, Ananthakrishnan Ramanathan and Karthik Visweswariah	1
<i>Report of the Shared Task on Learning Reordering from Word Alignments at RSMT 2012</i> Mitesh M. Khapra, Ananthakrishnan Ramanathan and Karthik Visweswariah	9
<i>A Tagging-style Reordering Model for Phrase-based SMT</i> Minwei Feng and Hermann Ney	17
<i>Building a reordering system using tree-to-string hierarchical model</i> Jacob Dlougach and Irina Galinskaya	27
<i>Learning Improved Reordering Models for Urdu, Farsi and Italian using SMT</i> Rohit Gupta, Raj Nath Patel and Ritesh Shah	37
<i>Partially modelling word reordering as a sequence labelling problem</i> Anoop Kunchukuttan and Pushpak Bhattacharyya	47

Workshop on Reordering for Statistical Machine Translation

Program

Sunday, 9 December 2012

- 14:30–15:30 **Invited talk:**
A survey of reordering for MT with a focus on pre-ordering
Karthik Visweswariah, IBM Research India
- 15:30–15:40 Introduction to the shared task
Whitepaper for Shared Task on Learning Reordering from Word Alignments at RSMT 2012
Report of the Shared Task on Learning Reordering from Word Alignments at RSMT 2012
- 15:40–16:10 *A Tagging-style Reordering Model for Phrase-based SMT*
Minwei Feng and Hermann Ney
- 16:10–16:30 *Building a reordering system using tree-to-string hierarchical model*
Jacob Dlougach and Irina Galinskaya
- 16:30–17:00 Tea break
- 17:00–17:20 *Learning Improved Reordering Models for Urdu, Farsi and Italian using SMT*
Rohit Gupta, Raj Nath Patel and Ritesh Shah
- 17:20–17:40 *Partially modelling word reordering as a sequence labelling problem*
Anoop Kunchukuttan and Pushpak Bhattacharyya
- 17:40–18:00 Concluding remarks by workshop chairs

