

# **2nd Workshop on Building and Using Comparable Corpora: from Parallel to Non-Parallel Corpora 2009**

**(BUCC 2009)**

**Held at ACL-IJCNLP 2009**

**Singapore  
6 August 2009**

**ISBN: 978-1-61738-452-3**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571  
[www.proceedings.com](http://www.proceedings.com)

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

Copyright© (2009) 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](mailto:acl@aclweb.org)

## Table of Contents

<i>Repetition and Language Models and Comparable Corpora</i> Ken Church .....	1
<i>Extracting Lay Paraphrases of Specialized Expressions from Monolingual Comparable Medical Corpora</i> Louise Deléger and Pierre Zweigenbaum .....	2
<i>An Extensible Crosslinguistic Readability Framework</i> Jesse Kirchner, Justin Nuger and Yi Zhang .....	11
<i>An Analysis of the Calque Phenomena Based on Comparable Corpora</i> Marie Garnier and Patrick Saint-Dizier .....	19
<i>Active Learning of Extractive Reference Summaries for Lecture Speech Summarization</i> Jian Zhang and Pascale Fung .....	23
<i>Train the Machine with What It Can Learn—Corpus Selection for SMT</i> Xiwu Han, Hanzhang Li and Tiejun Zhao .....	27
<i>Mining Name Translations from Comparable Corpora by Creating Bilingual Information Networks</i> Heng Ji .....	34
<i>Chinese-Uyghur Sentence Alignment: An Approach Based on Anchor Sentences</i> Samat Mamitimin and Min Hou .....	38
<i>Exploiting Comparable Corpora with TER and TERp</i> Sadaf Abdul Rauf and Holger Schwenk .....	46
<i>Compilation of Specialized Comparable Corpora in French and Japanese</i> Lorraine Goeuriot, Emmanuel Morin and Béatrice Daille .....	55
<i>Toward Categorization of Sign Language Corpora</i> Jérémie Segouat and Annelies Braffort .....	64