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

# Table of Contents

*Repetition and Language Models and Comparable Corpora*

Ken Church ................................................................. 1

*Extracting Lay Paraphrases of Specialized Expressions from Monolingual Comparable Medical Corpora*

Louise Deléger and Pierre Zweigenbaum ........................................... 2

*An Extensible Crosslinguistic Readability Framework*

Jesse Kirchner, Justin Nuger and Yi Zhang ....................................... 11

*An Analysis of the Calque Phenomena Based on Comparable Corpora*

Marie Garnier and Patrick Saint-Dizier ............................................... 19

*Active Learning of Extractive Reference Summaries for Lecture Speech Summarization*

Jian Zhang and Pascale Fung .......................................................... 23

*Train the Machine with What It Can Learn—Corpus Selection for SMT*

Xiwu Han, Hanzhang Li and Tiejun Zhao ............................................ 27

*Mining Name Translations from Comparable Corpora by Creating Bilingual Information Networks*

Heng Ji .......................................................... 34

*Chinese-Uyghur Sentence Alignment: An Approach Based on Anchor Sentences*

Samat Mamitimin and Min Hou ......................................................... 38

*Exploiting Comparable Corpora with TER and TERp*

Sadaf Abdul Rauf and Holger Schwenk ............................................. 46

*Compilation of Specialized Comparable Corpora in French and Japanese*

Lorraine Goeuriot, Emmanuel Morin and Béatrice Daille ....................... 55

*Toward Categorization of Sign Language Corpora*

Jérémie Segouat and Annelies Braffort ........................................... 64