

Tenth Workshop on Building and Using Comparable Corpora 2017

Held at ACL 2017

Vancouver, Canada
3 August 2017

ISBN: 978-1-5108-4575-6

Printed from e-media with permission by:

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



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

Copyright© (2017) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Users and Data: The Two Neglected Children of Bilingual Natural Language Processing Research</i> Phillippe Langlais	1
<i>Deep Investigation of Cross-Language Plagiarism Detection Methods</i> Jérémy Ferrero, Laurent Besacier, Didier Schwab and Frédéric Agnès	6
<i>Sentence Alignment using Unfolding Recursive Autoencoders</i> Jeenu Grover and Pabitra Mitra	16
<i>Acquisition of Translation Lexicons for Historically Unwritten Languages via Bridging Loanwords</i> Michael Bloodgood and Benjamin Strauss	21
<i>Toward a Comparable Corpus of Latvian, Russian and English Tweets</i> Dmitrijs Milajevs	26
<i>Automatic Extraction of Parallel Speech Corpora from Dubbed Movies</i> Alp Öktem, Mireia Farrús and Leo Wanner	31
<i>A parallel collection of clinical trials in Portuguese and English</i> Mariana Neves	36
<i>Weighted Set-Theoretic Alignment of Comparable Sentences</i> Andoni Azpeitia, Thierry Etchegoyhen and Eva Martínez Garcia	41
<i>BUCC 2017 Shared Task: a First Attempt Toward a Deep Learning Framework for Identifying Parallel Sentences in Comparable Corpora</i> Francis Grégoire and Philippe Langlais	46
<i>zNLP: Identifying Parallel Sentences in Chinese-English Comparable Corpora</i> Zheng Zhang and Pierre Zweigenbaum	51
<i>BUCC2017: A Hybrid Approach for Identifying Parallel Sentences in Comparable Corpora</i> Sainik Mahata, Dipankar Das and Sivaji Bandyopadhyay	56
<i>Overview of the Second BUCC Shared Task: Spotting Parallel Sentences in Comparable Corpora</i> Pierre Zweigenbaum, Serge Sharoff and Reinhard Rapp	60