

15th Workshop on Building and Using Comparable Corpora (BUCC 2022)

Held at LREC 2022

Marseille, France
25 June 2022

Editors:

**Reinhard Rapp
Pierre Zweigenbaum
Serge Sharoff**

ISBN: 978-1-7138-6113-3

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© (2022) by the Association for Computational Linguistics
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-ND. (<https://creativecommons.org/licenses/by-nd/4.0/>)

Printed with permission by Curran Associates, Inc. (2022)

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>Evaluating Monolingual and Crosslingual Embeddings on Datasets of Word Association Norms</i> Trina Kwong, Emmanuele Chersoni and Rong Xiang	1
<i>About Evaluating Bilingual Lexicon Induction</i> Martin Laville, Emmanuel Morin and Phillippe Langlais	8
<i>Don't Forget Cheap Training Signals Before Building Unsupervised Bilingual Word Embeddings</i> Silvia Severini, Viktor Hangya, Masoud Jalili Sabet, Alexander Fraser and Hinrich Schütze	15
<i>Building Domain-specific Corpora from the Web: the Case of European Digital Service Infrastructures</i> Rik van Noord, Cristian García-Romero, Miquel Esplà-Gomis, Leopoldo Pla Sempere and Antonio Toral	23
<i>Multilingual Comparative Analysis of Deep-Learning Dependency Parsing Results Using Parallel Corpora</i> Diego Alves, Marko Tadić and Božo Bekavac	33
<i>CUNI Submission to the BUCC 2022 Shared Task on Bilingual Term Alignment</i> Borek Požár, Klára Tauchmanová, Kristýna Neumannová, Ivana Kvapilíková and Ondřej Bojar	43
<i>Challenges of Building Domain-Specific Parallel Corpora from Public Administration Documents</i> Filip Klubička, Lorena Kasunić, Danijel Blazsetin and Petra Bago	50
<i>Setting Up Bilingual Comparable Corpora with Non-Contemporary Languages</i> Helena Bermudez Sabel, Francesca Dell'Oro, Cyrielle Montrichard and Corinne Rossari	56
<i>Fusion of linguistic, neural and sentence-transformer features for improved term alignment</i> Andraz Repar, Senja Pollak, Matej Ulčar and Boshko Koloski	61