

# **Second Workshop on Corpus Generation and Corpus Augmentation for Machine Translation (CoCo4MT 2023)**

Macau, China  
5 September 2023

ISBN: 978-1-7138-9038-6

**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© (2023) by the authors.  
All rights reserved.

These articles are licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Printed with permission by Curran Associates, Inc. (2024)

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)

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Do Not Discard – Extracting Useful Fragments from Low-Quality Parallel Data to Improve Machine Translation</i>	
Steinþór Steingrímsson, Pintu Lohar, Hrafn Loftsson and Andy Way .....	1
<i>Development of Urdu-English Religious Domain Parallel Corpus</i>	
Sadaf Abdul Rauf and Noor e Hira .....	14
<i>Findings of the CoCo4MT 2023 Shared Task on Corpus Construction for Machine Translation</i>	
Ananya Ganesh, Marine Carpuat, William Chen, Katharina Kann, Constantine Lignos, John E. Ortega, Jonne Saleva, Shabnam Tafreshi and Rodolfo Zevallos .....	22
<i>Williams College’s Submission for the CoCo4MT 2023 Shared Task</i>	
Alex Root and Mark Hopkins .....	28
<i>The AST Submission for the CoCo4MT 2023 Shared Task on Corpus Construction for Low-Resource Machine Translation</i>	
Steinþór Steingrímsson .....	33