

LREC-COLING 2024

**3rd Workshop on Tools and Resources
for People with REAding Difficulties
(READI 2024) @LREC-COLING-2024**

Workshop Proceedings

Editors

Rodrigo Wilkens, Rémi Cardon, Amalia Todirascu and
Núria Gala

20 May, 2024
Torino, Italia

**Proceedings of the 3rd Workshop on Tools and Resources for People with
REAding Difficulties (READI 2024) @LREC-COLING-2024**

Copyright ELRA Language Resources Association (ELRA), 2024
These proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-34-0
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Message from the General Chair

Recent studies show that the number of children and adults facing difficulties in reading and understanding written texts is steadily growing. Reading challenges can show up early on and may include reading accuracy, speed, or comprehension to the extent that the impairment interferes with academic achievement or activities of daily life. Various technologies (text customization, text simplification, text-to-speech devices, and screening for readers through games and web applications, to name a few) have been developed to help poor readers to get better access to information as well as to support reading development. Among those technologies, text adaptations are a powerful way to leverage document accessibility by using NLP techniques.

The 3rd Workshop on Tools and Resources for READING Difficulties (READI), collocated with the International 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING), aims at presenting current state-of-the-art techniques and achievements for text adaptations together with existing reading aids and resources for lifelong learning. The materials can be addressed to children struggling with difficulties in learning to read, to the community of teachers, speech-language pathologists and parents seeking solutions, but also to professionals involved with adults struggling with reading (illiterates, aphasic readers, low vision readers, etc.).

18 propositions have been submitted at this third edition, from which 9 were accepted, i.e., a 50% acceptance rate. This acceptance rate is lower than in the first and second editions, which had an acceptance rate of 66% and 71%, because the third edition takes place on a half-day, unlike the previous ones, which were full-day events. The accepted papers come from 41 different authors from 11 different countries (United Kingdom 6, Spain 4, Belgium 4, Germany 4, France 3, Japan 3, United States 3, Switzerland 2, Iceland 1, Italy 1, and Russian Federation 1). READI also features one invited speaker, Giulia Venturi from the Istituto di Linguistica Computazionale Antonio Zampolli (CNR, Pisa, Italy).

We are thankful to the authors who submitted their work to this workshop, to our Program Committee members, the reviewers and the additional reviewers who did a thorough job evaluating submissions, to Giulia Venturi who kindly accepted to be our invited speaker, and to the LREC committee for including this workshop into their program.

The workshop has been funded by Centre de traitement automatique du langage (CENTAL, IL&C, Université catholique de Louvain), the Laboratoire Parole et Langage (LPL, CNRS UMR 7309 & Aix Marseille Université) and the Institute Language, Communication and the Brain (ILCB, Aix Marseille Université), the French National Agency for Research (ANR-16-CONV-0002), the Excellence Initiative of Aix Marseille University A*MIDEX (ANR-11-IDEX-0001-02), and the Research Network on Language and Communication (University of Strasbourg).

Organizing Committee

Organizers

Rémi Cardon, Université catholique de Louvain, Belgium
Thomas François, Université catholique de Louvain, Belgium
Núria Gala, Aix Marseille Université, France
Amalia Todirascu, Université de Strasbourg, France
Rodrigo Wilkens, Université catholique de Louvain, Belgium

Program Committee:

Delphine Bernhard, Lilpa, Université de Strasbourg, France
Dominique Brunato, Institute of Computational Linguistics “A. Zampolli” (ILC-CNR), Pisa, Italy
Rémi Cardon, CENTAL, ILC, Université Catholique de Louvain, Belgium
Mireia Farrus, Universitat de Barcelona, Spain
Thomas François, UCLouvain, CENTAL, Belgium
Núria Gala, LPL-CNRS, Aix Marseille Université, France
Lingyun Gao, UCLouvain, Belgium
Itziar Gonzalez-Dios, HiTZ Basque Center for Language Technologies - Ixa, University of the Basque Country UPV/EHU, Spain
Arne Jonsson, Linköping University, Sweden
Euan McGill, Universitat Pompeu Fabra, Spain
Didier Schwab, Univ. Grenoble Alpes, France
Matthew Shardlow, Manchester Metropolitan University, United Kingdom
Kim Cheng Sheang, Universitat Pompeu Fabra, Spain
Anaïs Tack, KU Leuven; imec; UCLouvain, Belgium
Amalia Todirascu, University of Strasbourg, France
Vincent Vandeghinste, Instituut voor de Nederlandse Taal, Netherlands
Giulia Venturi, Institute of Computational Linguistics “Antonio Zampolli” (ILC-CNR), Italy
Elena Volodina, University of Gothenburg, Sweden
Rodrigo Wilkens, CENTAL, ILC, Université Catholique de Louvain, Belgium

Invited Speaker:

Giulia Venturi, ILC-CNR, Pisa, Italy

Table of Contents

<i>Evaluating Document Simplification: On the Importance of Separately Assessing Simplicity and Meaning Preservation</i> Liam Cripwell, Joël Legrand and Claire Gardent	1
<i>Malmon: A Crowd-Sourcing Platform for Simple Language</i> Helgi Björn Hjartarson and Steinunn Rut Friðriksdóttir	15
<i>Automatic Generation and Evaluation of Reading Comprehension Test Items with Large Language Models</i> Andreas Säuberli and Simon Clematide	22
<i>An Extensible Massively Multilingual Lexical Simplification Pipeline Dataset using the MultiLS Framework</i> Matthew Shardlow, Fernando Alva-Manchego, Riza Batista-Navarro, Stefan Bott, Saul Calderon Ramirez, Rémi Cardon, Thomas François, Akio Hayakawa, Andrea Horbach, Anna Huelsing, Yusuke Ide, Joseph Marvin Imperial, Adam Nohejl, Kai North, Laura Occhipinti, Nelson Pérez Rojas, Nishat Raihan, Tharindu Ranasinghe, Martin Solis Salazar, Marcos Zampieri and Horacio Saggion	38
<i>SIERA: An Evaluation Metric for Text Simplification using the Ranking Model and Data Augmentation by Edit Operations</i> Hikaru Yamanaka and Takenobu Tokunaga	47
<i>Transfer Learning for Russian Legal Text Simplification</i> Mark Athugodage, Olga Mitrofanove and Vadim Gudkov	59
<i>Accessible Communication: a systematic review and comparative analysis of official English Easy-to-Understand (E2U) language guidelines</i> Andreea Maria Deleanu, Constantin Orasan and Sabine Braun	70
<i>LanguageTool as a CAT tool for Easy-to-Read in Spanish</i> Margot Madina, Itziar Gonzalez-Dios and Melanie Siegel	93
<i>Paying attention to the words: explaining readability prediction for French as a foreign language</i> Rodrigo Wilkens, Patrick Watrin and Thomas François	102

Whorshop Program

Monday, May 20, 2024

9:10–10:00 Invited speaker

9:10–10:00 *Putting the right book into the hands of the right reader" in the era of Large Language Models*
Giulia Venturi

**10:00–
10:30 Oral Section presentation**

10:00–
10:30 *Evaluating Document Simplification: On the Importance of Separately Assessing Simplicity and Meaning Preservation*
Liam Crippwell, Joël Legrand and Claire Gardent

**10:30–
11:00 coffee break**

**11:00–
12:00 Poster Section presentation**

11:00–
12:00 *Malmon: A Crowd-Sourcing Platform for Simple Language*
Helgi Björn Hjartarson and Steinunn Rut Fririksdóttir

11:00–
12:00 *Automatic Generation and Evaluation of Reading Comprehension Test Items with Large Language Models*
Andreas Säuberli and Simon Clematide

11:00–
12:00 *An Extensible Massively Multilingual Lexical Simplification Pipeline Dataset using the MultiLS Framework*
Matthew Shardlow, Fernando Alva-Manchego, Riza Batista-Navarro, Stefan Bott, Saul Calderon Ramirez, Rémi Cardon, Thomas François, Akio Hayakawa, Andrea Horbach, Anna Huelsing, Yusuke Ide, Joseph Marvin Imperial, Adam Nohejl, Kai North, Laura Occhipinti, Nelson Pérez Rojas, Md Nishat Raihan, Tharindu Ranasinghe, Martin Solis Salazar, Marcos Zampieri and Horacio Saggion

11:00–
12:00 *SIERA: An Evaluation Metric for Text Simplification using the Ranking Model and Data Augmentation by Edit Operations*
Hikaru Yamanaka and Takenobu Tokunaga

11:00–
12:00 *Transfer Learning for Russian Legal Text Simplification*
Mark Athugodage, Olga Mitrofanove and Vadim Gudkov

Monday, May 20, 2024 (continued)

11:00– *Accessible Communication: a systematic review and comparative analysis*
12:00 *of official English Easy-to-Understand (E2U) language guidelines*
Andreea Maria Deleanu, Constantin Orasan and Sabine Braun

12:00– Oral Section presentation
13:00

12:00– *LanguageTool as a CAT tool for Easy-to-Read in Spanish*
13:00
Margot Madina, Itziar Gonzalez-Dios and Melanie Siegel

12:00– *Paying attention to the words: explaining readability prediction for French as*
13:00 *a foreign language*
Rodrigo Wilkens, Watrin Patrick and Thomas François

