



LREC 2026

**Joint Workshop on Readability and Text Simplification
(READIxTSAR) @ LREC 2026**

Workshop Proceedings

Editors

**Matthew Shardlow, Thomas François, Raquel Amaro,
Jorge Baptista, Rémi Cardon, Eugénio Ribeiro,
Horacio Saggion, Regina Stodden, Amalia Todirascu,
and Rodrigo Wilkens**

11 May 2026

Proceedings of the Joint Workshop on Readability and Text Simplification (READIxTSAR) @
LREC 2026

©ELRA Language Resources Association (ELRA), 2026
These proceedings are licensed under a Creative Commons Attribution-
NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-91-3

Preface

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.

Thus, the growth of educational and assistive technologies for reading, aimed at enhancing the performance of individuals with reading difficulties, intellectual disabilities, or reading in a second language, provides an important setting for automated text simplification (ATS) research, which is one of the key strategies for enhancing text accessibility. Nevertheless, ATS research often remains insufficiently grounded in the practical realities of education and disability contexts, thereby reducing the number of evaluations conducted with these target groups and constraining the integration of theoretical principles from education and accessibility research into ATS systems. This READIXTSAR workshop aims to present state-of-the-art applications and approaches in technology-enhanced reading along with innovations in text accessibility with the aim of bringing together these two strands of research. The workshop addresses specialized technology, tools, and resources, their impact on learning to read and comprehension, and innovative works spanning research to fieldwork, particularly in light of recent AI advances.

READIXTSAR is co-located with LREC 2026, the 15th edition of the Language Resources and Evaluation Conference, held in Palau de Congressos de Palma, Palma de Mallorca (Spain), on 11-16 May 2026. READIXTSAR is a joint initiative between two previous workshops of mutual interest: Tools and Resources for REAding DIfficulties (READI, hosted at LREC 2020,22,24) and Text Simplification, Accessibility and Readability (TSAR, EMNLP 2022,24,25, RANLP 2023). This year at LREC, the organising committee and programme committee of the two events have merged to deliver a joint event, uniting accessibility research communities under a common umbrella. Whereas the TSAR community is more oriented towards text simplification technology and the production of NLP methods for advancing the state of the art in text simplification, the READI community is oriented towards the development of novel tools and resources for improving the understandability of written documents. By uniting these two communities for a collaborative event, we hope to share mutual experiences and best working practices in our fields of interest fostering new collaborations and developing capacity for future research avenues of mutual benefit.

We received 24 submissions to the workshop, of which none were desk rejected. All 24 submissions were submitted for peer review, with each submission receiving two or three peer reviews from the programme committee, with papers aligned to reviewer expertise. Additionally, each paper was assigned to an expert member of the organising committee to coordinate reviews and make a provisional decision. All decisions were ratified by the full organising committee, with a total of 17 papers accepted of which 6 papers were accepted as oral presentation and 11 as poster presentations. This gives an acceptance rate of 70.83%, which is in line with previous editions of READI and TSAR. The organising committee aimed to offer developmental feedback based on reviewer suggestions to ensure that camera ready versions of all papers were of high quality when presented to the workshop audience.

The authorship of the accepted works are geographically diverse, with authors from across

Europe and beyond (Switzerland, France, Belgium, Portugal, UK, Spain, The Netherlands, Italy, Germany, Japan, UAE, USA). Topics of the accepted papers ranged from: novel methods and evaluation metrics for aspects of text simplification (6 papers); novel studies on the nature of readability including LLM-mediated texts (6 papers); studies at the intersection of simplification and readability investigating the interplay between these related phenomena (2 papers); and studies on representational aspects of complexity in models and embeddings (3 papers). The papers also exhibited linguistic diversity in the texts studied and evaluated. Whereas English was the primary language of study for 9 accepted papers, submissions also were accepted which studied French (3), Portuguese (2), Arabic (1), German (1) and Galician/Spanish (1).

The organisers extend their thanks to the authors who kindly submitted their work to our workshop, and to our program committee members, the reviewers and the additional reviewers who did a thorough job evaluating submissions. We would also like to acknowledge the participants of the workshop who attended both the oral and poster sessions. We would like to thank the committee of the DeTermIt workshop, with whom we co-ordinated our programme to ensure that participants could benefit from a full day of presentation of simplification and readability related presentations.

The organisers also wish to acknowledge the participation of the iRead4Skills and iDEM projects within the workshop, both of which provided an invited speaker to discuss the accessibility focussed research arising from each project.

Finally, we wish to acknowledge the work of the LREC organising committee and particularly the workshop chairs in selecting and supporting our workshop in the LREC 2026 programme of events.

The READIxTSAR Organising Committee

Organising Committee

- Matthew Shardlow, Manchester Metropolitan University, UK
- Thomas François, UCLouvain, Belgium
- Raquel Amaro, NOVA University Lisbon, Portugal
- Jorge Baptista, Universidade do Algarve & INESC-ID Lisboa, Portugal
- Rémi Cardon, Universidad Carlos III de Madrid, Spain
- Eugénio Ribeiro, Iscte-IUL & INESC-ID Lisboa, Portugal
- Horacio Saggion, Universitat Pompeu Fabra, Spain
- Regina Stodden, University Bielefeld, Germany
- Amalia Todirascu, Université de Strasbourg, France
- Rodrigo Wilkens, University of Exeter, UK

Table of Contents

<i>Revisiting German Complex Word Identification: Contextualized LLMs and Feature Injection</i> Thorben Schomacker, Seid Muhie Yimam, Chris Biemann and Marina Tropmann-Frick . .	1
<i>Book Complexity Level Assignment in French and Portuguese</i> Jorge Baptista, David Antunes, Wafa Aissa, Julien Zakhia Doueihi, Hanh Trang Tran Pham, Eugénio Ribeiro, Thomas François and Raquel Amaro	12
<i>Taming CATS: Controllable Automatic Text Simplification through Instruction Fine-Tuning with Control Tokens</i> Hanna Hubarava and Yingqiang Gao	26
<i>PLABA-EVAL: A Multi-Dimensional, In-Context Sentence Readability Dataset for Medical Text</i> Kexin Bian, Su-Youn Yoon and Mamoru Komachi	49
<i>Automatic Extraction of Textual and Phonemic Complexity for French Cued Speech</i> Magali Norré, Brigitte Bigi, Núria Gala, Ludivine Javourey Drevet and Thomas François	61
<i>Can LLMs Control Readability? A Multi-Dimensional Evaluation Framework for CEFR-Controlled Arabic Generation</i> Nour Rabi, Chatrine Qwaider and Ted Briscoe	74
<i>Lexical Conditioning of Model's Distribution through Uncertainty-gated Soft-Mixing of Probabilities</i> Michele Papucci, Giulia Venturi and Felice Dell'Orletta	89
<i>A Comparative Study of Multilingual Fine-tuning and Prompting for Automatic Text Readability Classification in Galician</i> Sandra Rodríguez Rey and Marcos Garcia	101
<i>Plan-Guided Text Simplification with Extended Contexts</i> Pascal Mathas, Jan Bakker and Jaap Kamps	121
<i>LLM-Generated Stories for Students with Significant Cognitive Disabilities: Promise, Gaps, and Evaluation Framework</i> Pragati Maheshwary, Ananya Ganesh and Shamy Karumbaiah	130
<i>Evaluating Transformer Model Family Representations Through Automated Essay Scoring</i> Akchay Ozten and Rodrigo Wilkens	142
<i>Proficiency-Controlled Text Simplification in European Portuguese: A Preliminary Study using Prompting Approaches</i> Eugénio Ribeiro, David Antunes, Nuno Mamede and Jorge Baptista	151
<i>Automatic Text Simplification for French Medical Documents with LLMs: The Role of Target Audience and Genre</i> Rémi Cardon and A. Seza Doğruöz	164
<i>A Learner-Oriented Annotated Resource of French Multiword Expressions for Text Adaptation in Foreign Language Reading</i> Anna Kalinina, Thomas François, Hélène Vassiliadou and Amalia Todirascu	181

<i>A Meta-evaluation of Automatic Metrics for Elaborative Simplification</i>	
Abdullah Alshatti, Steven Schockaert and Fernando Alva-Manchego	193
<i>Readability Measures in Automatic Text Simplification: Is Simplification Quality a Coherent Construct?</i>	
Rémi Cardon and A. Seza Doğruöz	210
<i>Language Proficiency as a Recoverable Dimension in Multilingual LLM Embeddings</i>	
Rodrigo Wilkens	227

Workshop Program

Monday, May 11, 2026

09:00–09:10 Welcome Session

09:10–10:00 Keynote Session

09:10–09:35 *The iRead4Skills Project*
Raquel Amaro

09:35–10:00 *The iDEM Project*
Horacio Saggion

10:00–10:30 Oral Presentations I

10:00–10:15 *Revisiting German Complex Word Identification: Contextualized LLMs and Feature Injection*
Thorben Schomacker, Seid Muhie Yimam, Chris Biemann and Marina Tropmann-Frick

10:15–10:30 *Book Complexity Level Assignment in French and Portuguese*
Jorge Baptista, David Antunes, Wafa Aissa, Julien Zakhia Doueihi, Hanh Trang Tran Pham, Eugénio Ribeiro, Thomas François and Raquel Amaro

10:30–11:00 Coffee Break

Monday, May 11, 2026 (continued)

**10:30–
11:30 Poster Session**

10:30– *Taming CATS: Controllable Automatic Text Simplification through Instruction*
11:30 *Fine-Tuning with Control Tokens*
Hanna Hubarava and Yingqiang Gao

10:30– *PLABA-EVAL: A Multi-Dimensional, In-Context Sentence Readability*
11:30 *Dataset for Medical Text*
Kexin Bian, Su-Youn Yoon and Mamoru Komachi

10:30– *Automatic Extraction of Textual and Phonemic Complexity for French Cued*
11:30 *Speech*
Magali Norré, Brigitte Bigi, Núria Gala, Ludivine Javourey Drevet and Thomas François

10:30– *Can LLMs Control Readability? A Multi-Dimensional Evaluation Framework*
11:30 *for CEFR-Controlled Arabic Generation*
Nour Rabih, Chatrine Qwaider and Ted Briscoe

10:30– *Lexical Conditioning of Model's Distribution through Uncertainty-gated Soft-*
11:30 *Mixing of Probabilities*
Michele Papucci, Giulia Venturi and Felice Dell'Orletta

10:30– *A Comparative Study of Multilingual Fine-tuning and Prompting for Automatic*
11:30 *Text Readability Classification in Galician*
Sandra Rodríguez Rey and Marcos Garcia

10:30– *Plan-Guided Text Simplification with Extended Contexts*
11:30
Pascal Mathas, Jan Bakker and Jaap Kamps

10:30– *LLM-Generated Stories for Students with Significant Cognitive Disabilities:*
11:30 *Promise, Gaps, and Evaluation Framework*
Pragati Maheshwary, Ananya Ganesh and Shamyia Karumbaiah

10:30– *Evaluating Transformer Model Family Representations Through Automated*
11:30 *Essay Scoring*
Akchay Ozten and Rodrigo Wilkens

10:30– *Proficiency-Controlled Text Simplification in European Portuguese: A Pre-*
11:30 *liminary Study using Prompting Approaches*
Eugénio Ribeiro, David Antunes, Nuno Mamede and Jorge Baptista

10:30– *Automatic Text Simplification for French Medical Documents with LLMs: The*
11:30 *Role of Target Audience and Genre*
Rémi Cardon and A. Seza Doğruöz

Monday, May 11, 2026 (continued)

**11:30–
13:00** **Oral Presentations II**

11:30–
11:50 *A Learner-Oriented Annotated Resource of French Multiword Expressions
for Text Adaptation in Foreign Language Reading*
Anna Kalinina, Thomas François, Hélène Vassiliadou and Amalia Todorascu

11:50–
12:10 *A Meta-evaluation of Automatic Metrics for Elaborative Simplification*
Abdullah Alshatti, Steven Schockaert and Fernando Alva-Manchego

12:10–
12:30 *Readability Measures in Automatic Text Simplification: Is Simplification Qual-
ity a Coherent Construct?*
Rémi Cardon and A. Seza Doğruöz

12:30–
12:50 *Language Proficiency as a Recoverable Dimension in Multilingual LLM Em-
beddings*
Rodrigo Wilkens

**12:50–
13:00** **Closing Session**

