

TSAR 2024

**The Third Workshop on Text Simplification, Accessibility and
Readability**

Proceedings of the Workshop

November 15, 2024

©2024 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-176-6

Introduction

The organisers are pleased to present the proceedings of the 3rd edition of the Workshop on Text Simplification, Accessibility and Readability (TSAR), hosted at The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), in Miami, Florida, USA.

This year the workshop was organised around two key tracks. The main track was of general interest to the audience and covered topics surrounding empirical research on text simplification, accessibility and readability. The special track encouraged participants to focus on the evaluation of systems with relevance to the workshop. In total, the workshop received twenty-eight submissions split between twenty submissions for the main track and eight submissions for the special track; special track submissions are indicated by * in the table of contents. These submissions covered a variety of current topics of interest to the TSAR community. In the main track, four papers considered simplification at the full text and lexical level, three papers presented work to identify various aspects of complex vocabulary and one paper considered accessibility through image generation. In the special track on evaluation two submissions focussed on the introduction of new evaluation methods, whereas three papers presented work with a particular focus on the evaluation of text simplification systems. All papers are listed below.

Main Track

- MultiLS: An End-to-End Lexical Simplification Framework
- OtoBERT: Identifying Suffixed Verbal Forms in Modern Hebrew Literature
- CompLex-ZH: A New Dataset for Lexical Complexity Prediction in Mandarin and Cantonese
- Images Speak Volumes: User-Centric Assessment of Image Generation for Accessible Communication
- Cochrane-auto: An Aligned Dataset for the Simplification of Biomedical Abstracts
- Considering Human Interaction and Variability in Automatic Text Simplification
- Society of Medical Simplifiers
- Difficult for Whom? A Study of Japanese Lexical Complexity

Special Track (*)

- Lexical Complexity Prediction and Lexical Simplification for Catalan and Spanish: Resource Creation, Quality Assessment, and Ethical Considerations
- SciGisPy: a Novel Metric for Biomedical Text Simplification via Gist Inference Score
- EASSE-DE & EASSE-multi: Easier Automatic Sentence Simplification Evaluation for German & Multiple Languages
- Evaluating the Simplification of Brazilian Legal Rulings in LLMs Using Readability Scores as a Target
- Measuring and Modifying the Readability of English Texts with GPT-4

All submissions were peer-reviewed by the members of the program committee which includes distinguished specialists in text simplification, accessibility, and readability. Out of the twenty-eight submissions to the workshop, two were desk-rejected, thirteen were rejected after review, eleven were accepted unconditionally and two were accepted subject to improvements in line with reviewer feedback. Out of

thirteen accepted papers, five were selected to be presented orally and eight as posters, which were also presented during a lightning-talk session. We additionally invited three non-archival poster presentations from relevant EMNLP Findings papers.

The workshop is held in-person, with online attendance for authors who were unable to attend due to constraints beyond the organisers control. The program encompasses: two invited talks, first by Dr. Iria Da Cunha, National Distance Education University (Spain) and secondly by Walburga Fröhlich, CEO and Co-Founder of Capito; two oral sessions, comprising five presentations; a round of lightning talks to introduce the poster presentations; and a hosted discussion session on current issues and trends in text simplification, accessibility and readability research.

We would like to thank the members of the program committee for their timely help in reviewing the submissions and all the authors for submitting their papers to the workshop. We also thank the EMNLP 2024 workshop chairs for their kind support in delivering the workshop and producing these proceedings.

TSAR Organizing Committee

Matthew Shardlow,

Fernando Alva-Manchego,

Kai North,

Regina Stodden,

Sanja Štajner,

Marcos Zampieri,

Horacio Saggion

Organizing Committee

Organizing Committee

Matthew Shardlow, Manchester Metropolitan University, UK

Horacio Saggion, Universitat Pompeu Fabra, Spain

Fernando Alva-Manchego, Cardiff University, UK

Marcos Zampieri, George Mason University, USA

Kai North, Cambium Assessment, USA

Sanja Štajner, Karlsruhe, Germany

Regina Stodden, Heinrich Heine University Düsseldorf, Germany

Program Committee

Program Committee

Rémi Cardon, UC Louvain
Richard Evans, University of Wolverhampton
Joseph Marvin, University of Bath
Yingqiang Gao, University of Zurich
Mounica Maddela, Bloomberg
Jan Trienes, University of Marburg
Freya Hewett, University of Potsdam
Liana Ermakova, Université de Bretagne Occidentale
Andreas Säuberli, University of Zurich
Piotr Przybyła, University Pompeu Fabra
Sowmya Vajjala, National Research Council, Canada
Emad Alghamdi, King Abdulaziz University
Silvana Deilen, University of Mainz
Sandaru Seneviratne, The Australian National University
Tadashi Nomoto, National Institute of Japanese Literature
Gustavo Paetzold, Federal University of Technology - Paraná
Tomoyuki Kajiwara, Ehime University
Claire Gardent, CNRS/LORIA
Sweta Agrawal, Instituto de Telecomunicações
Jaap Kamps, University of Amsterdam
Zihao LI, University of Manchester
Raquel Hervas, University Complutense of Madrid
Michael Gille, HAW Hamburg
Sarah Ebling, University of Zurich
Tannon Kew, University of Zurich
Giulia Venturi, Istituto di Linguistica Computazionale “Antonio Zampolli”
Natalia Grabar, University of Lille
Anna Dmitrieva, University of Helsinki
Laura Vásquez-Rodríguez, IDIAP
David Kauchak, Pomona College
Margot Madina, Darmstadt University of Applied Sciences
Jipeng Qiang, Yangzhou University
Lourdes Moreno, Universidad Carlos III de Madrid
Jasper Degraeuwe, Ghent University
Michael Ryan, Georgia Tech
Itziar Gonzalez-Dios, University of the Basque Country
Reno Kriz, John Hopkins
Yannick Parmentier, LORIA
Akio Hayakawa, University Pompeu Fabra
Oliver Alonzo, Rochester Institute of Technology
Patrick Haller, University of Zurich
Thomas François, UC Louvain
Susana Bautista, Universidad Francisco de Vitoria
Christina Niklaus, University of StGallen
Núria Gala, Aix Marseille Université
Maja Popović, Dublin City University

Tomas Goldsack, University of Sheffield
Diana Galvan-Sosa, University of Cambridge
Daniele Schicchi, CNR-ITD
Dennis Aumiller, Cohere

Keynote Talk

Artificial Intelligence and Plain Language

Iria da Cunha

National Distance Education University

November 15, 2024 –

Abstract: Natural Language Processing (NLP), a branch of artificial intelligence, has evolved significantly over recent decades. Broadly speaking, three main paradigms have been employed: rule-based systems, machine learning, and, more recently, language models. Simultaneously, the international plain language movement emerged, advocating for specialized texts addressed to the general public to be written in a simpler, more accessible manner. Over the past decade, a synergy has developed between these two fields—NLP and plain language—which has led to the creation of technological tools aimed at producing clearer texts. These tools follow the aforementioned NLP paradigms and can be classified, according to their function, into clarity testers, writing assistants, and clear text generators.

This presentation will offer an overview of the available tools, outlining their functionalities as well as the advantages and disadvantages of each NLP paradigm. Particular attention will be given to tools developed for the Spanish language, with a specific focus on the arText system (<https://sistema-artext.com/>), the first writing assistant for Spanish designed to help public administration staff draft texts in plain language addressed to citizens. This system has been developed with funding obtained through various competitive grants, in collaboration with different Spanish governments, and is available online free of charge.

Bio: Iria da Cunha is a lecturer at the National Distance Education University (Spain). She holds a PhD in Applied Linguistics from the Universitat Pompeu Fabra. Her research lines are Plain Language, Natural Language Processing, Terminology, and Specialized Discourse. She is the director of the arText team.

Keynote Talk

Easy-to-Understand Writing with AI Assistance

Walburga Fröhlich
CEO and Co-Founder of Capito
November 15, 2024 –

Abstract: Easy-to-understand language is not only needed by people with learning difficulties and disabilities. International studies show that many more people have difficulties understanding serious information from public authorities or companies.

In this presentation we will introduce rules and criteria for easy-to-understand language and show some good examples from companies and public authorities.

Since most people who are responsible for writing and providing information do not know how to write easy-to-understand texts, AI-based writing assistance can be very useful.

In the talk, we will show how we can analyze and simplify complicated information very easily with an AI-based writing assistance.

Bio: Walburga Fröhlich is Co-Founder and CEO of “capito”. For 30 years, she has been concerned with the question of how the potential of disabled people can be made visible and developed in our society and what accessible communication does for our society. Together with her team, she has developed AI-based digital tools for easier comprehensible information.

Walburga was born in 1966 in Austria, she studied first social work and has in addition a Masters degree in Social Management. She has already received many awards for her work, including the European Woman Innovator Prize or the European Innovation Council “Seal for Excellence”.

Table of Contents

<i>MultiLS: An End-to-End Lexical Simplification Framework</i> Kai North, Tharindu Ranasinghe, Matthew Shardlow and Marcos Zampieri	1
<i>OtoBERT: Identifying Suffixed Verbal Forms in Modern Hebrew Literature</i> Avi Shmidman and Shaltiel Shmidman	12
<i>CompLex-ZH: A New Dataset for Lexical Complexity Prediction in Mandarin and Cantonese</i> Le Qiu, Shanyue Guo, Tak-Sum Wong, Emmanuele Chersoni, John Lee and Chu-Ren Huang .	20
<i>Images Speak Volumes: User-Centric Assessment of Image Generation for Accessible Communication</i> Miriam Anshütz, Tringa Sylaj and Georg Groh	27
<i>Cochrane-auto: An Aligned Dataset for the Simplification of Biomedical Abstracts</i> Jan Bakker and Jaap Kamps	41
<i>Considering Human Interaction and Variability in Automatic Text Simplification</i> Jenia Kim, Stefan Leijnen and Lisa Beinborn	52
<i>Society of Medical Simplifiers</i> Chen Lyu and Gabriele Pergola	61
<i>Difficult for Whom? A Study of Japanese Lexical Complexity</i> Adam Nohejl, Akio Hayakawa, Yusuke Ide and Taro Watanabe	69
<i>*Lexical Complexity Prediction and Lexical Simplification for Catalan and Spanish: Resource Creation, Quality Assessment, and Ethical Considerations</i> Horacio Saggion, Stefan Bott, Sandra Szasz, Nelson Pérez, Saúl Calderón and Martín Solís . . .	82
<i>*SciGisPy: a Novel Metric for Biomedical Text Simplification via Gist Inference Score</i> Chen Lyu and Gabriele Pergola	95
<i>*EASSE-DE & EASSE-multi: Easier Automatic Sentence Simplification Evaluation for German & Multiple Languages</i> Regina Stodden	107
<i>*Evaluating the Simplification of Brazilian Legal Rulings in LLMs Using Readability Scores as a Target</i> Antonio Flavio Paula and Celso Camilo-Junior	117
<i>*Measuring and Modifying the Readability of English Texts with GPT-4</i> Sean Trott and Pamela Rivière	126