

EACL 2014

**14th Conference of the European Chapter of the
Association for Computational Linguistics**



**Proceedings of the 3rd Workshop on Predicting and
Improving Text Readability for Target Reader Populations
(PITR)**

April 27, 2014
Gothenburg, Sweden

©2014 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-91-6

Introduction

Welcome to the Third International Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR).

The last few years have seen a resurgence of work on text simplification and readability. Examples include learning lexical and syntactic simplification operations from Simple English Wikipedia revision histories, exploring more complex lexico-syntactic simplification operations requiring morphological changes as well as constituent reordering, simplifying mathematical form, applications for target users such as dyslexics, deaf students, second language learners and low literacy adults, and fresh attempts at predicting readability.

The PITR 2014 workshop has been organised to provide a cross-disciplinary forum for discussing key issues related to predicting and improving text readability for target users. It will be held on April 27, 2014 in conjunction with the 14th Conference of the European Association for Computational Linguistics in Gothenburg, Sweden, and is sponsored by the ACL Special Interest Group on Speech and Language Processing for Assistive Technologies (SIG-SLPAT).

These proceedings include fifteen papers that cover various perspectives on the topic: simplification in specific domains such as medicine and patents, simplification for specific languages, tailoring text for specific users (e.g., dyslexia and autism), development of new corpora, automatic system evaluation, analyses of human simplifications, studies of human reading, and predicting the reading level of text in general and for particular genres.

We hope this volume is a valuable addition to the literature, and look forward to an exciting Workshop.

Sandra Williams
Advaith Siddharthan
Ani Nenkova

Organizers:

Sandra Williams, The Open University, UK.
Advaith Siddharthan, University of Aberdeen, UK.
Ani Nenkova, University of Pennsylvania, USA.

Programme Committee:

Stefan Bott, Universitat Pompeu Fabra, Spain
Kevyn Collins-Thompson, University of Michigan, USA
Siobhan Devlin, University of Sunderland, UK
Micha Elsner, Ohio State University, USA
Richard Evans, University of Wolverhampton, UK
Oliver Ferschke, Technische Universität Darmstadt, Germany
Thomas Francois, University of Louvain, Belgium
Caroline Gasperin, SwiftKey, UK
Albert Gatt, University of Malta, Malta
Raquel Hervas, Universidad Complutense de Madrid, Spain
Veronique Hoste, University College Ghent, Belgium
Matt Huenerfauth, The City University of New York (CUNY), USA
David Kauchak, Middlebury College, USA
Annie Louis, University of Edinburgh, UK
Ruslan Mitkov, University of Wolverhampton, UK
Hitoshi Nishikawa, NTT, Japan
Ehud Reiter, University of Aberdeen, UK
Matthew Shardlow, Uni of Manchester, UK
Lucia Specia, University of Sheffield, UK
Ivelina Stoyanova, BAS, Bulgaria
Irina Temnikova, Qatar Computing Research Institute, Qatar
Sowmya Vajjala, Uni Tuebingen, Germany
Ielka van der Sluis, University of Groningen, The Netherlands
Jennifer Williams, MIT, USA
Kristian Woodsend, University of Edinburgh, UK

Table of Contents

<i>One Step Closer to Automatic Evaluation of Text Simplification Systems</i>	
Sanja Štajner, Ruslan Mitkov and Horacio Saggion	1
<i>Automatic diagnosis of understanding of medical words</i>	
Natalia Grabar, Thierry Hamon and Dany Amiot	11
<i>Exploring Measures of “Readability” for Spoken Language: Analyzing linguistic features of subtitles to identify age-specific TV programs</i>	
Sowmya Vajjala and Detmar Meurers	21
<i>Keyword Highlighting Improves Comprehension for People with Dyslexia</i>	
Luz Rello, Horacio Saggion and Ricardo Baeza-Yates	30
<i>An eye-tracking evaluation of some parser complexity metrics</i>	
Matthew J. Green	38
<i>Syntactic Sentence Simplification for French</i>	
Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat and Thomas Francois	47
<i>Medical text simplification using synonym replacement: Adapting assessment of word difficulty to a compounding language</i>	
Emil Abrahamsson, Timothy Forni, Maria Skeppstedt and Maria Kvist	57
<i>Segmentation of patent claims for improving their readability</i>	
Gabriela Ferraro, Hanna Suominen and Jaume Nualart	66
<i>Improving Readability of Swedish Electronic Health Records through Lexical Simplification: First Results</i>	
Gintare Grigonyte, Maria Kvist, Sumithra Velupillai and Mats Wirén	74
<i>An Open Corpus of Everyday Documents for Simplification Tasks</i>	
David Pellow and Maxine Eskenazi	84
<i>EACL - Expansion of Abbreviations in Clinical text</i>	
Lisa Tengstrand, Beáta Megyesi, Aron Henriksson, Martin Duneld and Maria Kvist	94
<i>A Quantitative Insight into the Impact of Translation on Readability</i>	
Alina Maria Ciobanu and Liviu Dinu	104
<i>Classifying easy-to-read texts without parsing</i>	
Johan Falkenjack and Arne Jonsson	114
<i>An Analysis of Crowdsourced Text Simplifications</i>	
Marcelo Amancio and Lucia Specia	123
<i>An evaluation of syntactic simplification rules for people with autism</i>	
Richard Evans, Constantin Orasan and Iustin Dornescu	131

Workshop Program

Sunday April 27, 2014

(09:00) Session 1 - Keynote

- 09:00 Welcome and opening remarks
09:10 Keynote: Ehud Reiter *Choosing Appropriate Words in Generated Texts for Low-Skill Readers*
10:30 Coffee break

(11:00) Session 2 - Papers

- 11:00 *One Step Closer to Automatic Evaluation of Text Simplification Systems*
Sanja Štajner, Ruslan Mitkov and Horacio Saggion
11:20 *Automatic diagnosis of understanding of medical words*
Natalia Grabar, Thierry Hamon and Dany Amiot
11:40 *Exploring Measures of “Readability” for Spoken Language: Analyzing linguistic features of subtitles to identify age-specific TV programs*
Sowmya Vajjala and Detmar Meurers
12:00 *Keyword Highlighting Improves Comprehension for People with Dyslexia*
Luz Rello, Horacio Saggion and Ricardo Baeza-Yates
12:20 Discussion
12:30 Lunch break

(14:00) Session 3 - Papers

- 14:00 *An eye-tracking evaluation of some parser complexity metrics*
Matthew J. Green
14:20 *Syntactic Sentence Simplification for French*
Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat and Thomas Francois
14:40 Panel
15:30 Coffee break

(16:00) Session 4 - Posters

- Medical text simplification using synonym replacement: Adapting assessment of word difficulty to a compounding language*
Emil Abrahamsson, Timothy Forni, Maria Skeppstedt and Maria Kvist

Sunday April 27, 2014 continued

Segmentation of patent claims for improving their readability

Gabriela Ferraro, Hanna Suominen and Jaume Nualart

Improving Readability of Swedish Electronic Health Records through Lexical Simplification: First Results

Gintare Grigonyte, Maria Kvist, Sumithra Velupillai and Mats Wirén

An Open Corpus of Everyday Documents for Simplification Tasks

David Pellow and Maxine Eskenazi

EACL - Expansion of Abbreviations in Clinical text

Lisa Tengstrand, Beáta Megyesi, Aron Henriksson, Martin Duneld and Maria Kvist

A Quantitative Insight into the Impact of Translation on Readability

Alina Maria Ciobanu and Liviu Dinu

Classifying easy-to-read texts without parsing

Johan Falkenjack and Arne Jonsson

An Analysis of Crowdsourced Text Simplifications

Marcelo Amancio and Lucia Specia

An evaluation of syntactic simplification rules for people with autism

Richard Evans, Constantin Orasan and Iustin Dornescu

- + Guest Poster, a preview of an EACL Main Session Paper: *Assessing the Relative Reading Level of Sentence Pairs for Text Simplification*
Sowmya Vajjala and Detmar Meurers

- 16:50 Closing Remarks