COLING 2014

**Automatic Text Simplification
Methods and Applications in the Multilingual Society**

Proceedings of the Workshop

August 24th, 2014
Dublin, Ireland

Introduction

The remarkable development of language technology tools in recent years in terms of robustness, computational speed and volume of processed data, together with the increasing number of languages covered, made possible their usage not only for specific research applications, but also for real world applications which prove useful in everyday life. Automatic correction of text, machine translation, extraction of important information and interaction with devices using speech are just some of these applications. Language technology now has the maturity to be used for addressing societal challenges such as helping people with disabilities, the elderly and migrants.

However, due to the ambiguity and complexity of natural language, its automatic processing is still very challenging and benefits from processing shorter and less ambiguous information. The same is true for people who have difficulties understanding text due to disabilities, or who have to read texts in a language they do not have a good command of. In all these cases, automatic text simplification can prove to be very useful.

In contrast to controlled languages, which practically create a sublanguage by imposing constraints on the grammar rules, discourse style, number of words in a sentence etc., text simplification eliminates or replaces parts of sentences or paragraphs, or even reformulates them according to specific requirements of the target user groups. Among the most frequent techniques are: lexical substitution, verb forms replacement (for morphologically rich languages), word order adjustments, deletion of subordinate clauses, replacement of anaphoric pronouns by their reference, usage of synonym expressions with higher frequency as well as compound splitting.

This workshop intends to bring together scientists working in a variety of fields in which text simplification can be applied, computational linguists interested in the research problems of text simplification and of course users who can benefit from the simplified texts.

The innovative aspect of this workshop is its focus on text simplification from two perspectives: On the one hand, how computational linguistics applications which simplify texts can be used by people in real world situation, and on the other hand, how to simplify the input for other NLP-based applications in order to improve their accuracy.

We are happy we could include in the workshop programme contributions dealing with all aforementioned issues.

Iustin Dornescu, Richard Evans and Constantin Orăsan report in *Relative clause extraction for syntactic simplification* about their results on syntactic text simplification method which focuses on extracting embedded clauses from structurally complex sentences and rephrasing them without affecting the original meaning.

In the paper *Making Biographical Data in Wikipedia Readable: A Pattern-based Multilingual Approach* the authors (Itziar Gonzalez-Dios, María Jesús Aranzabe and Arantza Díaz de Ilarraza) present Biographix a tool meant to create simple readable and accessible sentences in Wikipedia articles related to biographies. The tools is originally designed for Basque and then adapted for five European languages.

The contribution *Exploring the effects of Sentence Simplification on Hindi to English Machine Translation System* by Kshitij Mishra, Ankush Soni, Rahul Sharma and Dipti Sharma shows how text simplification can be used for bringing forward research in Machine Translation.

Ruslav Mitkov and Sanja Štajner introduce in their paper *The Fewer the better? A Contrastive Study about Ways to Simplify* a minimal set of rules which ensure a readability close to that one obtained by applying a long list of more fine-grained rules.

In *Automatic Text Simplification For Handling Intellectual Property (The Case of Multiple Patent Claims)* Svetlana Sheremetyeva presents an on-going project on a multi-level text simplification to assist experts who work on handling intellectural property in patent claims.

A User-View on adequate language resources to be used for txt simplifications is presented in the paper *Assessing Conformance of Manually Simplified Corpora with User Requirements: the Case of Autistic Readers* by Sanja Štajner, Richard Evans and Iustin Dornescu

In *Making historical Texts accessible for the crowd*, the authors explain which kind of simplification and adaptation historical texts may go through in order to be accessible to researchers and broad public not familiar with languages of previous centuries.

We hope that the workshop will contribute to the development of a roadmap of activities, tools and resources on text simplification from a multilingual perspective, roadmap which we think to be absolutely necessary for ensuring advances in this intriguing research field.

The organising committee would like to thank to the Programme Committee which contributed with very fast but substantial reviews to the workshop programme

<div align="right">Constantin Orăsan, Petya Osenova and Cristina Vertan</div>

Organizers:

Constantin Orăsan, University of Wolverhampton, UK

Petya Osenova, Sofia University "St. Kl. Ohridski", Bulgaria

Cristina Vertan, University of Hamburg, Germany

Program Committee:

Eric Atwell, Leeds University, UK
Eduard Barbu, University of Jaen, Spain
Ann Copestake, University of Cambridge, UK
Iustin Dornescu, University of Wolverhampton, UK
Richard Evans, University of Wolverhampton, UK
Thomas François, University of Louvain, Belgium
David Gil, Deletrea, Spain
Vesna Jordanova, Imperial College London, UK
Walther v. Hahn, University of Hamburg, Germany
Veronique Hoste, University College Gent, Belgium
Elena Lloret, University of Alicante, Spain
Annie Louis, University of Edinburgh, UK
Maite Martin Valdivia, University of Jaen, Spain
Paloma Moreda, University of Alicante, Spain
Hitoshi Nishikawa, NTT, Japan
Maciej Ogrodniczuk, Polish Academy of Sciences, Poland
Pavel Pecina, Charles University Prague, Czech Republic
Gabor Proszeky, Morphologic, Hungary
Horacio Saggion, Universitat Pompeu Fabra, Spain
Advaith Siddharthan, University of Aberdeen, UK
Lucia Specia, Sheffield University, UK
Sara Tonelli, FBK, Italy
Hristo Tanev, JRC, Italy
Dan Tufis, Romanian Academy, Romania
Dusko Vitas, University of Belgrade, Serbia

Invited Speaker:

Advaith Siddharthan, University of Aberdeen, UK

Table of Contents