Invited talk

The Interface between Readability and Automatic Text Simplification: Identifying Difficulties to Support Simple Writing

Thomas François

Assistant Professor at CENTAL, IL&C (UCLouvain)

For nearly a century, readability formulas have focused on the complex task of outputing a single numerical value consisting in an estimate of the difficulty of a text for a given population of readers. Although this synthetic approach has virtues in certain contexts, its main limitation is that it analyses how dozens or even hundreds of linguistic characteristics of a text affect the reading process, but lets the user know about this process only through this single numerical value. Automatic text simplification (ATS), for its part, aims to identify complex features in a text (words, syntactic structures, numbers, etc.) and automatically simplify them. Despite being a finer-grained approach, due to the lack of theoretical and empirical data, ATS still struggles to identify all linguistic characteristics that should be simplified.

In this talk, we will first set out our view of both fields and their current limitations in more details. In a second step, we will present several projects that are located at the interface between text readability and ATS, including the CEFRLex project (`http://cental.uclouvain.be/cefrlex/`), which is a set of lexical resources that can be used for readability and ATS purposes, the AMesure project (`http://cental.uclouvain.be/amesure/`), a platform to support simple writing of administrative texts, and ReSyf (`https://cental.uclouvain.be/resyf/`), which is a disambiguated and graded resource with synonyms. These projects will illustrate how automatically detecting complex segments of texts using readability techniques can inform semi-supervised or unsupervised simplification systems.

Bio: Thomas François is an Assistant Professor at UCLouvain (`http://cental.fltr.ucl.ac.be/team/tfrancois/`) in Applied Linguistics. His research focuses on readability, text simplification, automatic complex word identification, and efficient communication in professional contexts. He completed his Ph.D. at the Centre for Natural Language Processing (CENTAL, UCLouvain) and has received the best Ph.D. Thesis award by the ATALA in 2012. He spent a one-year research stay at IRCS (University of Pennsylvania) as a B.A.E.F. and Fulbright Fellow. As a follow up, he returned to UCLouvain and benefited from several post-doctoral research scholarships at CENTAL (founded by the FNRS and several regional projects such as iMediate and SPORTIC), before becoming a member of the UCLouvain academic staff. He has led projects such as CEFRLex (`http://cental.uclouvain.be/cefrlex/`), a CEFR-graded lexicon for foreign language learning or AMesure (`http://cental.uclouvain.be/amesure/`), a platform to support simple writing. He has also organized the CL4LC workshop, and has been invited to review for several NLP conferences (ACL, Coling, NAACL), journals (Computational Linguistics, ELRA), or book series (Synthesis Lectures on Human Technologies).

Proceedings of the 1st Workshop on Automatic Text Adaptation (ATA), pages 1–2,
Tilburg, The Netherlands, November 8 2018.