T4TEC: A Prototype for Simplified Multilingual Technical Communication. [Abstract]

Giorgio Maria Di Nunzio¹, Federica Vezzani² and Tiziana Sicilia³,⁴

¹Department of Information Engineering, University of Padua, Italy
²Department of Linguistic and Literary Studies, University of Padua, Italy
³European Association for Technical Communication, Tekom Europe
⁴COM&TEC Italian Association for Technical Communication

Abstract
We present the design and implementation of a prototype to support the field of technical communication and in particular the multilingual technical writing, named Terminology for Technical Efficient Communication (T4TEC), jointly developed between the University of Padua and COM&TEC (Associazione Italiana per la Comunicazione Tecnica). T4TEC is an innovative prototype tool among the software solutions for technical writing and technical communication. T4TEC includes a hybrid approach (rule based + data driven) based on 1) the official hand-written rules proposed by the Simplified Technical English, the Français Rationalisé, and Italiano Tecnico Semplificato experts and users; and 2) automatically harvest rule for lexical and syntactic simplifications by means of neural models.

Keywords
Automatic text simplification, Technical writing, Technical communication, Hybrid systems, Terminology

1. Background
Automatic text simplification (ATS) is the process of reducing the linguistic complexity of a text to improve its understandability and readability, improving contents and meaning and providing safe, clear and unambiguous information. [1]. From a broader perspective, ATS has a significant impact on society: it may help people with low literacy levels or those who suffer from reading comprehension [2]. ATS has usually focused on two different tasks: lexical simplification and syntactic simplification [3]. From an IR perspective, text simplification can be seen as a potential pre-processing step for making texts ‘easier’ to handle for some tasks, such as information extraction [4]. Key questions in this research field are: 1) how to identify complexity in the text, and 2) how to simplify the text in order to match the reading ability of the target reader.

In this paper, we focus on a specific case of text simplification which involves technical writing and communication [5] and controlled languages [6]. Technical communication is: “the process of defining, creating and delivering information products for the safe, efficient
and effective use of products (technical systems, software, services)”\textsuperscript{1}. In this context, the AeroSpace and Defence Industries Association of Europe (ASD) has led since the 1980s the most important initiatives to develop a Simplified Technical English (ASD-STE100) to help users to unambiguously understand technical manuals written in English. Later, more and more organizations created their own controlled language to simplify different languages. For example, the French aerospace industry association, the GIFAS,\textsuperscript{2} proposed a controlled French language\textsuperscript{3} in 1999, “le Français Rationalisé” (FR), allowing easy translation of technical documentation into Simplified English, whilst improving readability \textsuperscript{[7]}. The “Italiano Tecnico Semplificato” (ITS) has been developed by COM&TEC\textsuperscript{4} to include methods for creating and writing, specific consultancy, technical training and certification for those who adopt and use it.

2. Challenges

In this abstract, we present the design and implementation of T4TEC, a prototype for the support of multilingual technical writing jointly developed between the University of Padua and COM&TEC. T4TEC is an innovative tool in the technical writing software panorama, because:

- it is a multilingual tool for writing and translating technical documentation in English, French, and Italian;
- it is a hybrid system (rule based + data driven) that uses 1) the official hand-written rules proposed by the STE, the FR, and ITS,\textsuperscript{5} and automatically harvest rule for lexical \textsuperscript{[8]} and syntactic simplifications \textsuperscript{[9]};
- it is a real-time suggestion tool; users receive suggestions while they are typing.

In addition, we are designing T4TEC to tackle different research challenges posed by the authors of the Google Smart Compose system \textsuperscript{[10]} — that generates interactive, real-time suggestions in Gmail that assists users in writing mails by reducing repetitive typing — but adapted to this specific use case for technical communicators and writers:\textsuperscript{6}

- Latency. Since suggestions appear as the user is typing, minimizing end-to-end latency is critical. Nevertheless, inference about suggestions needs not to be performed on every keystroke. The optimal amount of text needed to check rules (lexical and syntax) is to be found;
- Personalization. Users often have their own unique writing styles. The system needs not to capture the uniqueness of their personal style. The model needs to have enough capacity so that it is able to make tailored and high-quality suggestions for the company for which the technical document is redacted (and the targeted user);
- Fairness and Privacy. Make sure that our models never expose company private information. In this context, the user (technical writer) does not write for him/herself but for the company.

\textsuperscript{1}https://www.technical-communication.org/
\textsuperscript{2}https://www.gifas.fr/
\textsuperscript{3}https://journals.openedition.org/traduire/389
\textsuperscript{4}http://comtec-italia.org/
\textsuperscript{5}SimplifiedTechnicalItalian,https://www.italianotecnicosemplificato.it/
\textsuperscript{6}We cited the challenges in \textsuperscript{[10]} and added not to indicate that the challenge is different in this context.
References


