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

Giorgio Maria Di Nunzio[1], Federica Vezzani[2] and Tiziana Sicilia[3,4]

[1]*Department of Information Engineering, University of Padua, Italy*
[2]*Department of Linguistic and Literary Studies, University of Padua, Italy*
[3]*European Association for Technical Communication, Tekom Europe*
[4]*COM&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)"[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,[2] proposed a controlled French language[3] in 1999, "le Français Rationalisé" (FR), allowing easy translation of technical documentation into Simplified English, whilst improving readability [7]. The "Italiano Tecnico Semplificato" (ITS) has been developed by COM&TEC[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,[5] and automatically harvest rule for lexical [8] and syntactic simplifications [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 [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:[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.

---

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

References

[1] S. Al-Thanyyan, A. M. Azmi, Automated text simplification: A survey, ACM Comput. Surv. 54 (2021) 43:1–43:36. URL: https://doi.org/10.1145/3442695. doi:10.1145/3442695.

[2] S. Stajner, Automatic text simplification for social good: Progress and challenges, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online, 2021, pp. 2637–2652. URL: https://aclanthology.org/2021.findings-acl.233. doi:10.18653/v1/2021.findings-acl.233.

[3] H. Saggion, Automatic Text Simplification, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2017. URL: https://doi.org/10.2200/S00700ED1V01Y201602HLT032. doi:10.2200/S00700ED1V01Y201602HLT032.

[4] L. Ermakova, P. Bellot, J. Kamps, D. Nurbakova, I. Ovchinnikova, E. SanJuan, É. Mathurin, S. Araújo, R. Hannachi, S. Huet, N. Poinsu, Automatic simplification of scientific texts: Simpletext lab at CLEF-2022, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval - 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10-14, 2022, Proceedings, Part II, volume 13186 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 364–373. URL: https://doi.org/10.1007/978-3-030-99739-7_46. doi:10.1007/978-3-030-99739-7_46.

[5] W. E. Britton, What is technical writing?, College Composition and Communication 16 (1965) 113–116. URL: http://www.jstor.org/stable/354886.

[6] G. Adriaens, D. Schreurs, From cogram to alcogram: Toward A controlled english grammar checker, in: 14th International Conference on Computational Linguistics, COLING 1992, Nantes, France, August 23-28, 1992, 1992, pp. 595–601. URL: https://aclanthology.org/C92-2090/.

[7] K. Barthe, C. Juaneda, D. Leseigneur, J.-C. Loquet, C. Morin, J. Escande, A. Vayrette, Gifas rationalized french: A controlled language for aerospace documentation in french, Technical Communication 46 (1999) 220–229. URL: https://www.ingentaconnect.com/content/stc/tc/1999/00000046/00000002/art00009.

[8] J. Qiang, Y. Li, Y. Zhu, Y. Yuan, Y. Shi, X. Wu, Lsbert: Lexical simplification based on bert, IEEE/ACM Transactions on Audio, Speech, and Language Processing 29 (2021) 3064–3076. doi:10.1109/TASLP.2021.3111589.

[9] A. Siddharthan, A. Mandya, Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules, in: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Gothenburg, Sweden, 2014, pp. 722–731. URL: https://aclanthology.org/E14-1076. doi:10.3115/v1/E14-1076.

[10] M. X. Chen, B. N. Lee, G. Bansal, Y. Cao, S. Zhang, J. Lu, J. Tsay, Y. Wang, A. M. Dai, Z. Chen, T. Sohn, Y. Wu, Gmail smart compose: Real-time assisted writing, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 2287–2295. URL: https://doi.org/10.1145/3292500.3330723. doi:10.1145/3292500.3330723.