

# The impact of text simplification training on comprehensibility: Combining automated metrics and human evaluation

Alessandra Rossetti<sup>1</sup>, Luuk Van Waes<sup>1</sup>

<sup>1</sup>*Department of Management (Universiteit Antwerpen), Prinsstraat 13, 2000 Antwerpen, Belgium*

## Abstract

The need to communicate in simple language is growing, and so is the need to train writers in the production of reader-friendly texts. We trained university students (with English as a second language) to revise English business texts by carefully taking into account the needs and preferences of the target readers. We examined the impact of our training on the achieved comprehensibility (and lack thereof) of the texts produced. To this end, we combined automated metrics of text difficulty provided by Coh-Metrix with human evaluations obtained through comparative judgments in Comproved. We report the preliminary results and conclusions of our analysis.

## Keywords

text comprehensibility, text-simplification training, revision, automatic and human evaluation

## 1. Background

Text revision is a complex task which relies on the reviser's ability to simultaneously consider the intentions of the author, the features of the text, and the target audience [1]. In particular, revising texts to make them easier to read and more engaging requires adaptation to the reading skills, prior knowledge, and expectations of the reader [2]. Years of practice and levels of language proficiency influence how effectively revision tasks are carried out, with novice/low-proficiency revisers struggling to approach the text as a whole (e.g. by examining its cohesion) and to consider readership [3] [4]. Different types of training – in different contexts and domains – have been developed and tested, with mixed results [5] [6].

## 2. Research goal

Against this background, this study investigates the impact of text simplification training on the characteristics of the revised texts produced by university students in a second language. The training was delivered online and included theoretical and practical components. The content

---

*Words/Machines 4: Text simplification and adaptation, March 25, 2022, Brest, France*

✉ [alessandra.rossetti@uantwerpen.be](mailto:alessandra.rossetti@uantwerpen.be) (A. Rossetti); [luuk.vanwaes@uantwerpen.be](mailto:luuk.vanwaes@uantwerpen.be) (L. V. Waes)


🌐 <https://www.uantwerpen.be/nl/personeel/alessandra-rossetti/> (A. Rossetti);

<https://www.uantwerpen.be/nl/personeel/luuk-vanwaes/> (L. V. Waes)

🆔 0000-0002-2162-9639 (A. Rossetti); 0000-0002-3642-9533 (L. V. Waes)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

of the texts (i.e. corporate social responsibility [CSR]) and the fact that the students revised already existing texts in a second language represent elements of novelty of this study.

### **3. Method**

We recruited 42 Master's students (native speakers of Dutch with English as a second language) and we randomly assigned them to a control and to an experimental group. In a pre-test session, each student: 1) revised a business text on CSR to make it easier to read and more engaging; and 2) took part in online training. The experimental group received training on accessible, engaging communication applied to CSR content, while the control group received training exclusively on the topic of CSR. In a post-test session, the students revised another business text on CSR using what they had learnt from their respective modules. We pre-analysed and manipulated the features of the texts to be revised to make sure that they had similar readability features. Both the modules and the texts to be revised were in English. In total, the students produced 84 texts (42 in the pre-test and 42 in the post-test).

The texts revised by the students were evaluated automatically through Coh-Metrix [7], as well as by 25 human evaluators carrying out comparative judgments in Comproved [8]. Coh-Metrix is a computational tool that automatically analyses texts and provides scores of multiple linguistic features related to textual cohesion and text difficulty [9]. Human comparative judgment involves comparing one text with another (rather than evaluating each text independently) and deciding which text demonstrates the highest level of a pre-defined competence [10]. This method has shown promising results in terms of reliability and validity [8]. Most of our evaluators (N=21) were native speakers of a language other than English, however their level of English proficiency was very high – 17 evaluators reported having studied English for more than 10 years and 16 of them had a doctorate degree. Each evaluator read and compared 5 pairs of texts. Each text was assessed by 3 evaluators on average.

### **4. Preliminary results**

Automated results from Coh-Metrix comparing pre-test and post-test sessions showed a growth in the readability of the texts revised from both experimental and control group, possibly as a result of their respective training. In the pre-test, the texts revised by control and experimental group did not significantly differ along any of the readability dimensions investigated. However, in the post-test session (i.e. after having taken part in our training), the experimental group produced texts that contained fewer words, had shorter sentences, and were syntactically simpler, compared with the texts of the control group. This result is in line with previous studies showing the positive impact of technological support and feedback on syntactic simplicity [11]. Furthermore, the revised texts from the experimental group could be read and understood by people with fewer years of study, as indicated by the Flesch-Kincaid Grade Level. On the other hand, compared with the experimental group, the control group in the post-test produced revised texts with higher argument overlap between adjacent sentences, possibly as a result of their training centered around the topic of CSR and associated concepts. No significant between-group differences were found in relation to other text features (e.g. cohesion).

Our initial explorations showed that the human evaluations were partially in line with the automated scores. Both groups received higher scores in the post-test compared with the pre-test. In the pre-test, the texts revised by the control group received significantly lower scores in terms of accessibility and engagement ( $M=-.78$ ), compared with the texts revised by the experimental group ( $M=-.19$ ). These differences at the pre-test stage were not reflected in the automated scores. In the post-test, the differences between control group ( $M=.3$ ) and experimental group ( $M=.58$ ) were no longer significant. In other words, despite being shorter, syntactically simpler, and adapted to a lower grade level, the texts revised by the experimental group in the post-test were not rated as significantly more accessible and engaging by our human assessors. This result might be due to a lack of improvement of cohesion, a text feature that has been shown to facilitate reading comprehension [12].

## 5. Conclusions

Our training on revising texts for comprehensibility in a second language helped students make changes at the word and sentence level. However, the ability to revise global text features (e.g. cohesion) seems to require more practice, a finding which is in line with previous research. Students' inability to significantly improve these global text features might explain the lack of positive impact on the readers'/evaluators' experience, whose scores were only partially aligned with automated scores.

## Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 888918. We thank the students and the evaluators who participated in our study.

## References

- [1] R. T. Kellogg, Training writing skills: A cognitive developmental perspective, *Journal of writing research* 1 (2008).
- [2] K. A. Schriver, Teaching writers to anticipate readers' needs: A classroom-evaluated pedagogy, *Written communication* 9 (1992) 179–208.
- [3] K. Barkaoui, What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills, *The Modern Language Journal* 100 (2016) 320–340.
- [4] F. Zhang, H. B. Hashemi, R. Hwa, D. Litman, A corpus of annotated revisions for studying argumentative writing, in: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 1568–1578.
- [5] D. Janssen, Improving the quality of public documents, *Reading and Writing Public Documents: Problems, Solutions, and Characteristics* 1 (2001) 105.
- [6] M. Kuteeva, Wikis and academic writing: Changing the writer–reader relationship, *English for Specific Purposes* 30 (2011) 44–57.

- [7] D. S. McNamara, A. C. Graesser, P. M. McCarthy, Z. Cai, *Automated evaluation of text and discourse with Coh-Metrix*, Cambridge University Press, 2014.
- [8] T. van Daal, M. Lesterhuis, L. Coertjens, V. Donche, S. De Maeyer, *Validity of comparative judgement to assess academic writing: Examining implications of its holistic character and building on a shared consensus*, *Assessment in Education: Principles, Policy & Practice* 26 (2019) 59–74.
- [9] D. S. McNamara, M. M. Louwerse, P. M. McCarthy, A. C. Graesser, *Coh-metrix: Capturing linguistic features of cohesion*, *Discourse Processes* 47 (2010) 292–330.
- [10] A. V. Mortier, M. Lesterhuis, P. Vlerick, S. D. Maeyer, *Comparative judgment within online assessment: exploring students feedback reactions*, in: *International Computer Assisted Assessment Conference*, Springer, 2015, pp. 69–79.
- [11] A. Rossetti, *Intralingual translation and cascading crises*, *Translation in Cascading Crises* (2019).
- [12] S. E. Pickren, M. Stacy, S. N. Del Tufo, M. Spencer, L. E. Cutting, *The contribution of text characteristics to reading comprehension: Investigating the influence of text emotionality*, *Reading Research Quarterly* (2021).