

---

# Absorbed in Goodreads. A Computational Approach for the Study of Online Social Reading

Simone Rebora<sup>\*1,2</sup>, Piroska Lendvai<sup>1</sup>, and Moniek Kuijpers<sup>1</sup>

<sup>1</sup>University of Basel – Switzerland

<sup>2</sup>University of Verona – Italy

## Abstract

Our project proposes an expansion of empirical methods in reader experience research (Peer et al., 2012), by focusing on the growing phenomenon of "online social reading" (Cordon Garcia et al., 2013). Using software from the field of natural language processing, we will match the 18 statements from the Story World Absorption Scale (SWAS, cf. Kuijpers et al., 2014) – a questionnaire used to identify absorbing experiences such as narrative transportation – to reader reviews posted on the online platform Goodreads. The aims of this project are twofold: 1) validating the SWAS, and 2) enabling comparative analyses of absorption across different books, genres, and readers' groups.

We performed a manual analysis on 180 Goodreads reviews of three contemporary blockbuster novels, confirming that, in many cases, SWAS statements and particular sentences in Goodreads reviews overlap substantially. A reviewer writes: "I'm so absorbed in the world Martin produced out of his wits" (a sentence that matches with SWAS statement A3: "I felt absorbed in the story"); another reviewer expresses her identification with the main character: "I went through all the emotional ups and downs right along with her" (matching with EE4: "I felt how the main character was feeling"). A total of 132 matching sentences were identified.

In order to extend the analysis to the entire Goodreads corpus, which collects about 80 million book reviews, we tested two technologies: textual entailment detection software, i.e., EOP (Magnini et al., 2014) and text reuse detection software, i.e., TRACER (Büchler et al., 2017). However, preliminary experiments (Rebora et al., 2018) show that both tools need adaptation and training for this specific task (best "out of the box" recall score: 0.28).

The *Mining Goodreads* project (funded by SNSF and running until August 2020) was conceived for this specific goal. With five annotators working in parallel on the *brat* platform, we plan to produce a ground truth corpus to be used as training data for machine learning algorithms. At the DARIAH-DESIR Workshop, we will present the first results of the annotation process, the conceptual framework and annotation guidelines that we developed to optimize the workflow, together with the future steps and possible extensions of the project.

## Work cited:

Büchler, M., Franzini, G., Franzini, E. and Bulert, K. (2017). TRACER – a multilevel framework for historical text reuse detection, *Journal of Data Mining and Digital Humanities – Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages*.

---

\*Speaker

Cordon Garcia, J.A., Alonso Arevalo, J., Gomez Diaz, R. and Linder D. (2013). *Social Reading: Platforms, Applications, Clouds and Tags*. Oxford: Chandos Publishing.

Kuijpers, M., Hakemulder, F., Tan, E.E. and Doicaru, M.M. (2014). Exploring absorbing reading experiences. Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature*, 4(1): 89–122.

Magnini, B., Zanolini, R., Dagan, I., Eichler, K., Neumann, G., Noh, T. G. and Levy, O. (2014). The Excitement Open Platform for textual inferences. In *Proceedings of ACL Demo Session*. Baltimore: ACL, 43–48.

Peer, W. van, Hakemulder, F. and Zyngier, S. (2012). *Scientific Methods for the Humanities*. Amsterdam/Philadelphia: John Benjamins.

Rebora, S., Lendvai, P. and Kuijpers M. (2018). Reader experience labeling automatized: Text similarity classification of user-generated book reviews. In *EADH 2018 Book of Abstracts*, <https://eadh2018.exordo.com/programme/presentation/90>.