

# Reading Machines: promoting reading with computational text analysis

**Ciotti, Fabio**

fabio.ciotti@uniroma2.it

Università di Roma Tor Vergata, Italy

**Baldi, Alberto**

alberto.baldi@unifi.it

Università di Roma Tor Vergata, Italy

## Introduction

This paper presents the *Macchine per leggere* (Reading Machines) project, born out of the collaboration between the Department of Literary, Philosophical and Art History Studies at the University of Rome “Tor Vergata” and the Center for Books and Reading of the Italian Ministry of Culture.

Computational analysis of literary texts is a branch of the digital humanities that has developed and found various fields of experimentation in recent years, thanks to IT advances in machine learning and the steady increase of textual databases accessible online (Moretti, 2013; Underwood, 2019; Jockers, 2013). However, especially because of the high computing expertise required, these techniques have seldom crossed the niche of digital humanists and have remained confined to a highly specialized field. Therefore, in addition to not having spread to the scholarly community as a whole, they have not even been able to produce particular spin-offs on literature teaching, especially in high school, and on the promotion of reading. Instead, in our opinion, they could be used as a way for proposing to young born-digital students a different and alternative perspective for approaching the great books of Italian literature, in the apparently paradoxical hypothesis that by learning to “not read” – according to Franco Moretti’s formula (Moretti, 2000) – they can get closer to reading our literary heritage.

## State of the art

As of now, there are very few examples of the adoption of text analysis and mining methods to foster reading skills. In Italy we can cite the experiment carried out in a class at the Liceo Scientifico Galileo Galilei high school in Trento (Valitutti and Dalla Torre, 2021), in which students were guided in analyzing the sentiment of the novel *Io non ho paura* by Niccolò Ammaniti, and Alessandro Iannella’s (Iannella, 2019) proposal to create with Google Dialogflow a chatbot simulating a dialogue with the poetess Sappho. The idea of a virtual assistant to help young people in their approach to reading has precedents, such as, for example, the *Sobek* tool, developed “to support educational applications [...] from assisting teachers to review student’s work to helping kids in reading and writing activities” by two researchers at the Federal University of Rio Grande do Sul (Reategui et al., 2020) and aimed

at extracting a network of keywords from a text, and *Readerbench*, a digital platform (or “Personal Learning Environment”) created by the Polytechnic University of Bucharest (Dascalu et al., 2013) that targets students and teachers by providing a set of tools to analyze text corpora (in English or French), performing analyses such as keyword extraction, calculating the readability index of a text, and sentiment analysis.

## The *Macchine per leggere* project

Building on these few experiments and our knowledge in the field of text analysis, our project involved the creation of a digital environment (desktop and mobile) to introduce secondary school students to the knowledge and use of text analysis techniques. Ten novels from the nineteenth-century canon of Italian literature (including *I promessi sposi*, *I Malavoglia*, *Le avventure di Pinocchio*...) are presented on a web site ([macchineperleggere.it](http://macchineperleggere.it)). The main section of the site, entitled “Digital Analysis”, consists of a series of pages (one per novel) in which the outputs of a selection of some of the main text mining techniques are provided: linguistic-descriptive statistical analysis, topic modeling, sentiment analysis, network analysis. In detail, we propose:

- A selection of tools for linguistic-descriptive statistical analysis from the VoyantTools suite: Cirrus, for the Wordcloud; Terms, to generate a frequency index of terms in the corpus; Contexts, a concordance compiler; Microsearch, to graphically represent the distribution of terms in the text; Phrases, which identifies the most recurrent syntagmas.
- An example of topic modeling by proposing in five graphs some outputs of an analysis accomplished with BERTopic by Marten Grootendorst (Grootendorst, 2022);
- An example of sentiment analysis using the R library Syuzhet, created by Matthew Jockers (Jockers, 2015), which allows to reconstruct the plot of a novel by analyzing its sentiment;
- An example of network analysis, combining resources of two Python libraries: Spacy for Named Entity Recognition and NetworkX for node graph creation.

Each tool is introduced by a paragraph that introduces the tool and contextualizes the type of analysis performed. In later stages of the project, it is also planned to integrate a module for georeferencing novels, to connect the reading experience to users’ daily experience.

In addition to the section dedicated to computational analysis, the site has a section designed to hosting the full texts of the novels, also offered in the Bionic Reading typographic format, a solution developed by Renato Casutt (<https://bionic-reading.com/>) that uses boldface combinations to aid concentration and make the most of the reading experience; a section with a guide intended for students that goes over the basic concepts of text analysis, tracing a short history, illustrating its theoretical assumptions and briefing on the state of the art of the various methodologies, combined with an essential glossary; and finally, a section in which the techniques demonstrated (with the exception of network analysis) are re-proposed in the form of web apps (realized, for BERTopic and Syuzhet, with Streamlit and Shiny), so that users can experiment independently and on other texts with the distant approach proposed within the project.

## Future developments

In early 2023, the project has moved to its experimental phase in some secondary school classrooms. We have prepared a set of video tutorials that illustrate the basic theoretical assumptions of distant reading and introduce the site, suggesting possible strategies – for example, reading paths or methods for analyzing outputs – to integrate the platform and its tools into traditional teaching. These educational supports have been presented to a group of High School teachers, in order to develop experimental activities that will take place during 2023/24 school year. Once the experimental period is over, feedback will be collected from teachers and students through questionnaires, evaluating the ergonomics of the platform, the effectiveness of its tools, and the possible impact in the teaching of literature and the enjoyment of the great classics by new generations.

## Bibliography

- Dascalu, M., Dessus, P., Trausan-Matu, Ș., Bianco, M. and Nardy, A.** (2013). ReaderBench, an Environment for Analyzing Text Complexity and Reading Strategies. In Lane, H. C., Yacef, K., Mostow, J. and Pavlik, P. (eds), *Artificial Intelligence in Education*, vol. 7926. (Lecture Notes in Computer Science). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 379–88 doi:10.1007/978-3-642-39112-5\_39. [http://link.springer.com/10.1007/978-3-642-39112-5\\_39](http://link.springer.com/10.1007/978-3-642-39112-5_39) (accessed 2 November 2022).
- Grootendorst, M.** (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv doi:10.48550/ARXIV.2203.05794. <https://arxiv.org/abs/2203.05794> (accessed 2 November 2022).
- Iannella, A.** (2019). ‘Ok Google, vorrei parlare con la poetessa Saffo’: Intelligenza Artificiale, assistenti virtuali e didattica della letteratura. *Thamyris, Nova Series: Revista de Didáctica de Cultura Clásica, Griego y Latín*, 10: 81–104.
- Jockers, M.** (2015). Revealing Sentiment and Plot Arcs with the Syuzhet Package <https://www.matthewjockers.net/>.
- Jockers, M. L.** (2013). *Macroanalysis: Digital Methods and Literary History*. 1st Edition edition. Urbana: University of Illinois Press.
- Moretti, F.** (2000). Conjectures on World Literature. *The New Left Review* <http://newleftreview.org/A2094>.
- Moretti, F.** (2013). *Distant Reading*. London: Verso.
- Reategui, E., Epstein, D., Bastiani, E. and Carniato, M.** (2020). Can Text Mining Support Reading Comprehension?. In Gennari, R., Vittorini, P., De la Prieta, F., Di Mascio, T., Temperini, M., Azambuja Silveira, R. and Ovalle Carranza, D. A. (eds), *Methodologies and Intelligent Systems for Technology Enhanced Learning, 9th International Conference*, vol. 1007. (Advances in Intelligent Systems and Computing). Cham: Springer International Publishing, pp. 37–44 doi:10.1007/978-3-030-23990-9\_5. [http://link.springer.com/10.1007/978-3-030-23990-9\\_5](http://link.springer.com/10.1007/978-3-030-23990-9_5) (accessed 2 November 2022).
- Underwood, T.** (2019). *Distant Horizons: Digital Evidence and Literary Change*. Chicago: The University of Chicago Press.
- Valitutti, A. and Dalla Torre, C.** (2021). ‘Io non ho paura’: Sentiment analysis nell’analisi di testi narrativi. *Proceedings of Didamatica-21*. AICA, pp. 166–69.