

Budapest Training School – Canonization in Distant Reading Research

Vasilije Milnović
milnovic@unilib.rs
University Library
“Svetozar Marković”
Belgrade, Serbia

PAPER SUBMITTED: 10 November 2021

PAPER ACCEPTED: 13 November 2021

As a part of the COST action “Distant Reading for European Literary History” (COST Action CA16204), the third Training School was organized at the Centre for Digital Humanities – Eötvös Loránd University in Budapest from September 23 to 25. Within this Training School, participants could choose to attend one of three parallel tracks: Corpus design and text contribution for ELTeC, Natural Language Processing for Distant Reading and Canonization in Distant Reading Research.

Since my scientific expertise is related to the study of the relationship between tradition and avant-garde and a certain re-examination of canons in the traditional sense, I chose to attend the last track. I did not have real experience in using track-related digital tools. That was the reason why this kind of training was of great use to me to get acquainted with methodologies I might use for my future scientific work.

Lecturers within the chosen topic were:

- Marijan Dović (Slovenia, Literature History), who gave an introductory lecture on the topic “Literary Canon: From Traditional to Contemporary Approaches”,
- Maciej Eder (Poland, Linguistics), who gave guidelines for the application of Digital Humanities and Distance Reading in literary language research,
- Karina van Dalen-Oskam (Netherlands, Digital Humanities), who gave a presentation “From the Riddle of Literary Quality to the Riddle of the Literary Canon: a meta-perspective” and
- Christof Schoch (Germany, Digital Humanities), the Coordinator of the entire COST Action, with the topic of technical details of Distant Reading research “Operationalization, Formalization, Modeling: Measuring Degrees of Canonization for Distant Reading Research”.

After the importance of a kind of meta-perspective of Distant Reading and the possibility of applying such a tool to research the canonicity of

literary works was pointed out, participants were introduced to the many possibilities of digital research of literary text. Regardless of whether the research is conducted according to keywords, or digital tools are used in terms of some type of synthesis, digital humanities provide various possibilities that can also serve as a form of re-examination of existing canons. Distant Reading can be used for a new approach to interdisciplinary analysis of literary texts from the past, including those from the Serbian language collection, for the simple reason that the software can detect various aspects of language in specific texts, which cannot be observed by simple close reading.

In this sense, Distant Reading uses technology to get a bird's eye view of a corpus, and represents one of the most important resources for research in the field of linguistics and related language disciplines. It is used in all linguistic disciplines as a tool for literary and artistic texts. In such research, a special place is occupied by the statistical analysis of language, which can provide conclusions that cannot be drawn on the basis of close reading. As an illustration of distant reading methods, some of the opportunities regarding the use of certain words in the novels of Charles Dickens and Jane Austen, and their connection with existing language corpora, were presented to the participants of the training school.

Within all lectures, participants were presented with basic information about the traditional canon (literary and cultural), as well as the application of the philosophy of distant reading to study the canonicity of literary works. While reaching conclusions about both distant and close reading, an important methodological question for the digital humanities was raised during the discussion: what function might (canonical) close reading fulfil in the digital analysis of large corpora? Is there a way of combining close and distant reading – which many scholars have argued for – without seemingly undermining the necessity of one or the other?

To answer this question, participating in the COST action was very beneficial, because the Serbian ELTeC sub-collection now contains 100 novels, from the period 1840-1920, with more in its extension, including many works that were not part of the official national cultural canon. Both types of research can now be done on this sample - close or distant reading - and based on everything I have learned in this workshop – I am sure that it is possible to apply them both in studying the canonicity of a work. This possibility is especially well expressed when it comes to the avant-garde, the main topic of my scientific research, because even in the so-called developed or large cultures it is part of an alternative or underground canon. Therefore, I am

very interested in spreading my knowledge on the topic of distant reading and its possible application in scientific research.