

Presentation of the COST action Distant reading for European Literary History

Diana Santos

d.s.m.santos@ilos.uio.no

Distant  Reading

30 august 2018



COST action CA16204

Three doors

- http://www.cost.eu/COST_Actions/ca/CA16204
- <https://www.distant-reading.net/>
- <https://distantreading.github.io/>

This action will

*affect the way scholars in the Humanities do research, but also
the way institutions like libraries will make their holdings
available to researchers in the future*

Distant  Reading





- build a multilingual European Literary Text Collection (ELTeC), ultimately containing around 2,500 full-text novels in at least 10 different languages, permitting to test methods and compare results across national traditions;
- establish and share best practices and develop innovative computational methods of text analysis adapted to Europe's multilingual literary traditions;
- consider the consequences of such resources and methods for rethinking fundamental concepts in literary theory and history.

Distant  Reading

Wg1: European Literary Text Collection (ELTeC)

This work can be split into three distinct tasks:

- 1 defining selection criteria (corpus design)
- 2 developing basic encoding methods (both for data and for metadata);
- 3 defining a suitable workflow for preparation of the corpus

Already available:

https://distantreading.github.io/sampling_proposal.pdf

https://distantreading.github.io/encoding_proposal.pdf

More: <https://www.distant-reading.net/wg-1/>

Distant  Reading

WG2: Methods and tools

- ① stimulate the awareness of the possibilities of the many frameworks and libraries which are already out there, through the organization of hands-on training schools and through the development of dedicated online tutorials on specific topics.
- ② unify existing language and text technologies in Digital Humanities
- ③ raise awareness about the strengths and weaknesses of established modelling approaches in the field, for instance for topic modelling and text classification, including the standards and metrics that are necessary for their empirical evaluation

More: <https://www.distant-reading.net/wg-2/>

Distant  Reading

WG3: Literary Theory and History

Exploring theoretical concerns that stem from the application of Distant Reading methods to literary history

- ① provide a theoretical underpinning to the project's activities,
- ② use Distant Reading research for a re-evaluation and redefinition of key concepts for literary history, such as genre, style or authorship

Ultimately explore their implications for periodisation and canonisation in European literary history

More: <https://www.distant-reading.net/wg-3/>

Distant  Reading

- Digitization projects?
- Other projects?
- Networking
- Contact with observers around the world - Australian, Brazilian, ...
- Short term scientific missions (STSMs)
- Training schools (and ITC conference grants)

Distant Reading

What's in it for me?

- Participation in an interesting project
- Possibility of influencing the outcome or at least discuss the issues
- Help for developing smaller projects or project proposals
- Being well informed about distant reading in Europe
- Participate (as teacher or student) in specific courses or tutorials

First: join the mailing list(s)!

Distant Reading

Timeline proposal

- submitted first proposal March 2014
- resubmitted March 2015
- resubmitted April 2016 (not eligible because not anonymous enough)
- resubmitted December 2016
- approved June 2017
- Christian-Emil Ore and Diana Santos are approved by NFR (and before by ILN and ILOS respectively) and COST as members of the management committee, representing Norway, September 2017
- COST action begins, November 3rd 2017, first management committee in Brussels.

Distant  Reading

COST timeline

- 3 Nov 2017 COST action begins, first management committee in Brussels.
- Jan 2018 Working group meeting 4 (dissemination) in Amsterdam
- 12 Feb 2018 Second management committee in Prague
- 12-14 Feb 2018 Working group meeting 1 in Prague
- 8 Oct 2018 Third management committee in Antwerp
- 8-10 Oct 2018 Working group meeting 2 in Antwerp
- Jan 2019 Fourth management committee in Lisbon

Distant  Reading

- The novel
- The canon
- The purpose of the collection
- The annotation

Distant Reading

Examples

- Lusophone literature: are there groups and tendencies independent of nation, or Brazilian literature and Portuguese literature are two radically different objects?
- Do authors cite or refer to other European authors or same language authors?
- Can the themes from different literary schools be identified automatically? Clustering correlating with genre or period?
- Is author identity stronger than genre or school identity?

Distant Reading

Some research questions

- Literary influence across literatures
- Themes/tropes in fashion
- Character description
- Narrator types
- Maps according to time
- Intertextuality
- ...

Distant [📖] Reading

Limitations

- It is only 100 novels per national literature in the period ca. 1840-1919
- “Nations” are hard to define, “national literatures” even worse
- Translation not taken into consideration, although it arguably is the best way of influencing across borders
- Some computational applications require (such as authorship attribution) several works by the same authors...

Solution: extended collections! And connected projects.

Distant [📖] Reading

WG1: Testing the workflow with two novels for each language, level 0 and 1.

- TEI-header
- Everything on github

WG2: Preparing the meeting in October

- OsloMet group(s) will send some proposals
- Soon receiving green light for discussion

Distant [📖] Reading

Acknowledgments

I am grateful to Christian-Emil Ore for putting me in contact with the proposal writers.

I am grateful to Christof Schöh for the launching of the whole project, and running it beautifully.

I am grateful to all COST action members for interesting discussions.



COST is supported by the EU Framework Programme Horizon 2020

Distant [📖] Reading

- https://e-services.cost.eu/files/domain_files/CA/Action_CA16204/mou/CA16204-e.pdf
- Christof Schöch, Presentation of CA16204 in DARIAH 2018 event in Paris: <https://github.com/distantreading/slides-en>
- Santos, Diana, Cláudia Freitas & João Marques Lopes. "Comparando a literatura lusófona com outras literaturas: recursos para leitura a distância em português". Atas do *HDRio2018 – I Congresso Internacional em Humanidades Digitais no Rio de Janeiro* (9-13/4/2018)

Distant Reading

Obrigada!

