

Distant Reading for European Literary History. A COST Action

Christof Schöch,¹ Maciej Eder², Carolin Odebrecht³, Mike Kestemont⁴, Antonija Primorac⁵, Justin Tonra⁶, Katja Mihurko Poniz⁷, Catherine Kanellopoulou⁸

¹ Department for Computational Linguistics and Digital Humanities, University of Trier, 54286 Trier
schoech@uni-trier.de

² Institute of Polish Language, Polish Academy of Sciences, al. Mickiewicza 31, 31-120 Krakow,
Poland
maciej.eder@ijp.pan.pl

³ Faculty of Language, Literature and Humanities, Humboldt University of Berlin, Dorotheenstraße
24, D- 10117 Berlin
carolin.odebrecht@hu-berlin.de

⁴ University of Antwerpen, Department of Literature, Grote Kauwenberg 18, S.D.115, 2000
Antwerpen, Belgium
mike.kestemont@uantwerpen.be

⁵ Faculty of Humanities and Social Sciences, University of Rijeka, Sveučilišna avenija 4, HR-51000
Rijeka, Croatia.
Antonija.Primorac@ffri.uniri.hr

⁶ School of Humanities, National University of Ireland Galway, University Road, Galway, H91 TK33,
Ireland
justin.tonra@nuigalway.ie

⁷ School of Humanities, University of Nova Gorica, Vipavska 13, 5000 Nova Gorica
katja.mihurko.poniz@ung.si

⁸ Ionian University Korfou, Department of Audio & Visual Arts 7, Tsirigoti Square, 49100 Corfu
catherine.kanellopoulou@gmail.com

1 Introduction

This poster aims to stimulate awareness of the existence of the newly-established COST Action on “Distant Reading for European Literary History” (2017-2021). In the context of this networking project, “distant reading” is understood as an umbrella term for recent computational, and particularly quantitative, approaches to the study of large collections of texts. This paradigm is here applied to the multilingual literary traditions of Europe in the long nineteenth century.

2 What is a COST Action?

COST (www.cost.eu) stands for ‘European Cooperation in Science and Technology’: COST Actions are essentially networking initiatives focused on a particular, timely and innovative research topic, aiming to bring together a critical mass of researchers from Europe and beyond. COST Actions coordinate their activities through working group meetings and offer Training Schools and opportunities for scientific exchange. Examples of previous COST Actions in Digital Humanities include Interedition (<http://www.interedition.eu/>, 2008-2012) and e-Lexicography (<http://www.elexicography.eu/>, 2013-2017).

2.1 Aims of the “Distant Reading” Action

The contribution of the Distant Reading paradigm to Literary Studies continues to be a matter of intense debate. In our view, recent, quantitative approaches clearly provide an important methodological perspective that usefully complements, and at times challenges, more established approaches to literary history and theory in areas like authorship attribution, genre analysis, periodization, canonization and intertextuality.

We aim to create a vibrant and diverse network of researchers jointly developing the resources and methods necessary to change the way European literary history is written. Fostering insight into crossnational, large-scale patterns and evolutions across European literary traditions, we will facilitate the creation of a broader, more inclusive and better-grounded account of European literary history and cultural identity. We will foster distributed research, the systematic exchange of expertise, and the visibility of all participants, activities and resources.

In terms of scientific objectives, we will coordinate the creation of a multilingual European Literary Text Collection (ELTeC). We will use the ELTeC to establish best practices and develop innovative methods of Distant Reading for the multiple European literary traditions. Furthermore, we will engage in an investigation into the theoretical consequences of Distant Reading approaches for literary history and literary theory. We also aim to foster the acquisition of state-of-the-art methods related to data curation, standards, best practices and quantitative analysis in workshops and training schools. Last but not least, we aim to address the current gender imbalance among practitioners of Distant Reading research.

2.2 The network

Our network of members is currently comprised of researchers in Corpus Linguistics, Computational Linguistics, (Digital) Literary History and Literary Theory from 26 different countries and more than 40 cities across Europe and beyond (Figure 1).

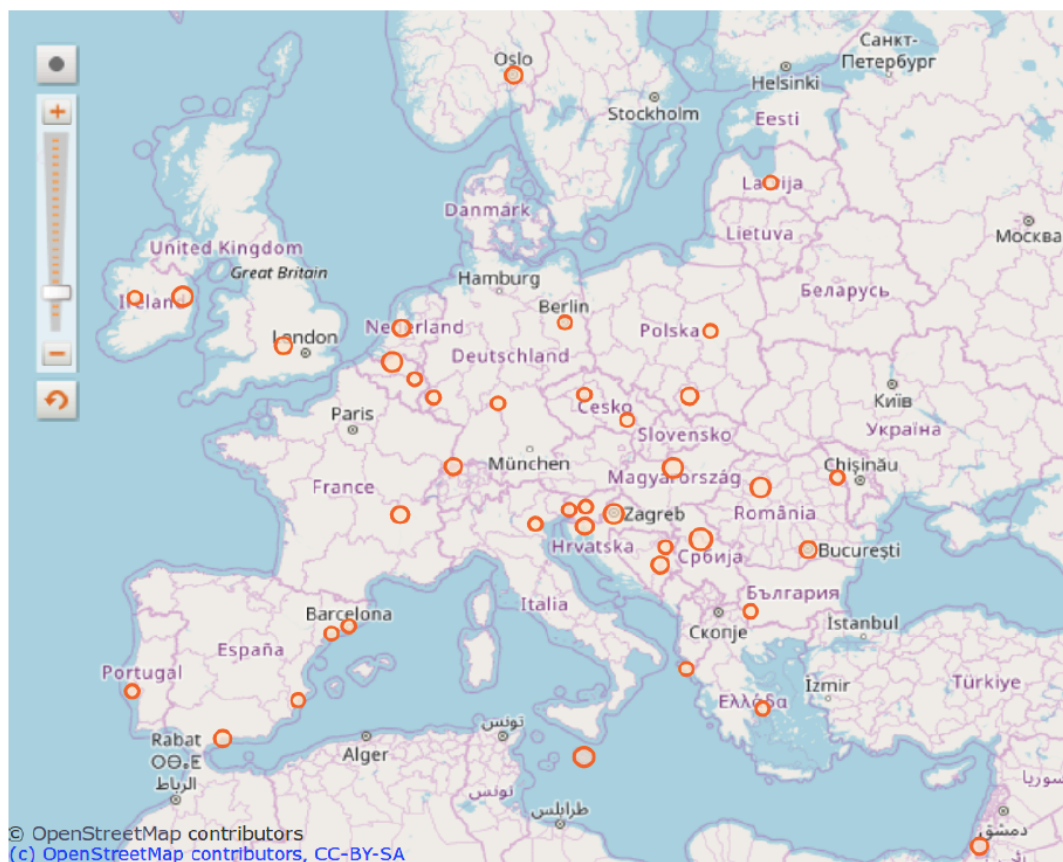


Figure 1: Map of Europe with the locations of Action members, created using the DARIAH GeoBrowser. Interactive version: <https://geobrowser.de.dariah.eu/?csv1=https://geobrowser.de.dariah.eu/storage/515798>.

2.3. Our key deliverable: the ELTeC

Our key deliverable is the European Literary Text Collection (ELTeC) that brings together comparable sets of nineteenth-century novels from at least 10 different European languages. Each set will comprise 100 different novels published in the late nineteenth century, with extensions covering the early nineteenth century or adding additional novels from the late nineteenth century. The purpose of the ELTeC is to serve as a benchmark corpus for the evaluation and development of annotation tools and distant reading methods across languages and as the basis for investigations into patterns and trends in literary history in multiple literary traditions. Text curation will happen on GitHub (with distributed access, issue tracking and version control). Experience with similar, national initiatives shows the importance of openness, standards and technical sustainability. As the Action progresses, linguistic annotation will be added to the texts (at least, concerning lemmata, part of speech and named entities). A shared format for the representation of document-level metadata and linguistic annotation across subcollections will be used, based on best practices in the field as recommended e.g. by DARIAH (Digital Research Infrastructure for the Arts and Humanities) and CLARIN (Common Language Resources and Technology Infrastructure).

The ELTeC will contain linguistic annotation in a cross-linguistically compatible manner by mapping each language-specific tagset onto a coarse-grained, shared tagset.

2.4. Research strands

Our activities are divided into three main research strands (organized in working groups):

- “Scholarly Resources”, focused on structuring, annotating and publishing the ELTeC;
- “Methods and Tools”, concerned with using, evaluating and developing methods for distant reading analysis;
- “Literary History and Theory”, dedicated to the theoretical consequences of distant reading methods for literary history and theory.

In addition, a working group on “Dissemination” provides infrastructure services, enables communication within the Action and gives visibility to the Action’s activities and results.

3. Learn more, learn how to join

To learn more, see the Action’s profile page www.cost.eu/COST_Actions/ca/CA16204 and the full proposal linked there (“Memorandum of Understanding”). The Action’s website is available at <https://www.distant-reading.net/>. Researchers from Computational Linguistics, (Digital) Literary Studies as well as Computer Scientists and Librarians are welcome to get involved!

4. References

- Andrew Goldstone. 2017. The Doxa of Reading. *PMLA*, 3(132):636–42.
- Franco Moretti. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. New York: Verso.
- Andrew Piper. 2017. Think Small: On Literary Modeling. *PMLA*, 3(132):651–58.
- Richard Jean So. 2017. ‘All Models Are Wrong’. *PMLA*, 3(132):668–673.
- Ted, Underwood. 2017. A Genealogy of Distant Reading. *Digital Humanities Quarterly*, 11, 2 <http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html>.