

Digital Humanities and Distant Reading

Distant  *Reading*

Christof Schöch (Trier, Germany)
INTELE Webinar, June 2021

website: <http://www.distant-reading.net>

 **cost**
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY



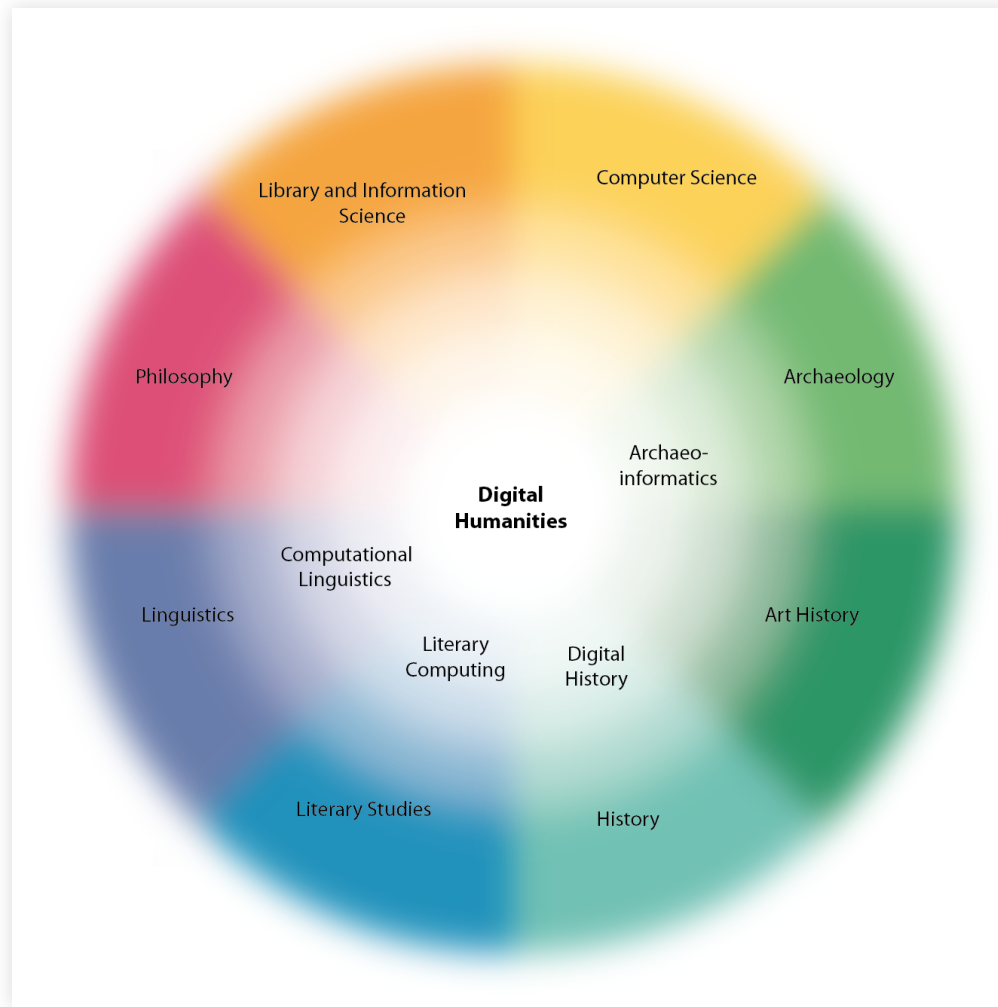
Funded by the Horizon 2020 Framework Programme
of the European Union

Overview

1. What is Digital Humanities?
2. What is Distant Reading?
3. What are COST Actions?
4. What is the Distant Reading COST Action about?
5. Conclusion

What is Digital Humanities?

Digital X vs. Core-DH



What is Distant Reading?

Did you say Distant Reading?

- Many alternative terms
 - Distant Reading
 - Stylometry
 - Digital Literary Stylistics
 - Computational Literary Studies
- A shared conceptual core
 - Research that uses digital text corpora
 - Focus on literary texts in a wide sense
 - Computational and statistical methods of analysis
 - Research questions from literary history and theory
 - Unique methodological research questions
(i.e. data modeling, adaptation of methods, evaluation)

2. What are COST Actions?

Networking Grants

- Support network and capacity building around specific research topics
- Networks of European countries and international partners
- Around 300 Actions are running in parallel (very few of them in the Humanities)

Some Actions relevant to DH

- [Interedition](#), 2008-2012
- [Women Writers In History](#), 2009-2013
- [European Network of e-Lexicography](#), 2013-2017
- [Reassembling the Republic of Letters](#), 2015-2018
- [NEP4DISSENT - Cultures of Dissent](#), 2017-2021
- [Nexus Linguarum](#), 2019-2023

Some key features

- Action duration is usually 4 years
- Structured into working groups
- No funding for staff, only for networking activities
- Various forms of "networking activities"

COST networking activities

- Working Group Meetings
- Training Schools
- Short Term Scientific Missions
- Conference Grants

2. What is *Distant Reading* for
European Literary History about?

Action: Research objectives

- Resources: Build a multilingual reference collection of European novels ("ELTeC")
- Methods: Explore, evaluate, adapt and share computational methods of text analysis for ELTeC
- Theory: Think through consequences of digital data and methods for literary history and theory

Action: Networking objectives

- Bring together corpus linguists, computational linguists, digital literary scholars, literary historians and theorists
- Spread and share competencies in the three areas above among these groups
- Support relevant collaborative grant proposals on the national and European levels

Current Network

- 32 countries are involved
- 200+ scholars are participating
- several spin-off projects
- a follow-up project (CLS INFRA)

Key Output: ELTeC

Language	Last update	AUTHORSHIP				LENGTH			TIME SLOT					REPRINT COUNT		E5C		
		Texts	Words	Male	Female	1-title	3-title	Short	Medium	Long	1840-59	1860-79	1880-99	1900-20	range		Frequent	Rare
cze	2021-04-09	100	5621667	88	12	62	6	43	49	8	12	21	39	28	27	1	19	80.00
deu	2021-04-09	100	12738842	67	33	35	9	20	37	43	25	25	25	25	0	48	46	96.92
eng	2021-04-09	100	12227703	49	51	70	10	27	27	46	21	22	31	26	10	32	68	100.00
fra	2021-04-09	100	8712219	66	34	58	10	32	38	30	25	25	25	25	0	44	56	101.54
hrv	2021-03-22	21	1440018	21	0	4	0	6	12	3	6	12	2	1	11	1	0	23.08
hun	2021-04-09	100	6948590	79	21	71	9	47	31	22	22	21	27	30	9	32	67	100.00
ita	2019-11-21	34	3328244	32	2	19	3	13	10	11	5	12	10	7	7	12	0	55.97
lav	2020-12-20	2	106045	2	0	2	0	0	2	0	0	0	1	1	1	0	1	21.54
lit	2020-08-20	25	636132	18	7	16	1	19	3	2	5	3	3	14	11	6	18	55.38
nor	2021-04-09	53	3432676	38	15	18	11	26	18	9	4	2	30	17	28	32	21	67.69
pol	2021-04-09	100	8500172	58	42	1	33	33	35	32	8	11	35	46	38	39	61	80.00
por	2021-04-09	100	6799385	83	17	73	9	40	41	19	13	37	19	31	24	26	60	94.62
rom	2021-04-09	95	5558961	74	16	54	9	46	31	18	4	17	25	49	45	24	71	80.00
slv	2021-04-09	100	5682120	89	11	26	5	53	39	8	2	13	36	49	47	48	52	78.46
spa	2021-04-09	84	7147890	67	17	45	5	31	28	25	17	17	25	25	8	42	42	92.31
srp	2021-04-09	90	3961405	82	8	35	11	56	32	2	2	12	38	38	36	32	58	73.68
swe	2021-04-09	58	4960085	29	28	18	8	16	24	18	15	3	20	20	17	17	41	76.92
ukr	2021-04-09	50	1840062	37	13	23	7	34	13	3	5	10	11	24	19	30	20	70.77

Summary produced: 2021-04-09

- European Literary Text Collection (ELTeC)
- Comparable sets of novels for at least 10 European languages
- Each set: 100 novels published between 1840 and 1920
- Currently: over 1300 novels published!
- <https://distantreading.github.io/ELTeC/>
- ELTeC is a key output and basis for further work!

Time for questions!

To learn more: websites

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

To learn more: readings

- Burnard, Lou, Schoöch, and Odebrecht. "In Search of Comity: TEI for Distant Reading." November 1, 2019. <https://doi.org/10.5281/zenodo.3552489>.
- Byszuk, Joanna, Michał Woźniak, Mike Kestemont, Albert Leśniak, Wojciech Łukasik, Artjoms Šeļa, and Maciej Eder. "Detecting Direct Speech in Multilingual Collection of 19th Century Novels." In Proceedings of LT4HALA 2020-1st Workshop on Language Technologies for Historical and Ancient Languages, edited by Rachele Sprungoli and Marco Passarotti, <https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/LT4HALAbook.pdf:100-104>. Paris: European Language Resources Association (ELRA), 2020.
- Cinková, Silvie, et al. "Evaluation of Taggers for 19th-Century Fiction." DH_Budapest_2019, edited by Gábor Pálko, ELTE, 2019, http://elte-dh.hu/dh_budapest_2019-abstract-booklet/.
- Evert, Stefan, Fotis Jannidis, Thomas Proisl, Steffen Pielström, Thorsten Vitt, Christof Schöch, and Isabella Reger. "Understanding and Explaining Distance Measures for Authorship Attribution." Digital Scholarship in the Humanities, 2017. https://academic.oup.com/dsh/article-pdf/32/suppl_2/ii4/21298943/fqx023.pdf.
- Patras, Roxana, Ioana Galleron, Camelia GRĂDINARU, Ioana Lionte, and Lucreția Pascaru. "The Splendors and Mist(Eries) of Romanian Digital Literary Studies: A State-of-the-Art Just before Horizons 2020 Closes Off." Hermeneia 23 (2019): 207-22.
- Stankovic, Ranka, Francesca Frontini, Tomaž Erjavec, and Carmen Brando. "Named Entity Recognition for Distant Reading in Several European Literatures." In DH_Budapest_2019, edited by Gábor Pálko. Budapest: ELTE, 2019. http://elte-dh.hu/dh_budapest_2019-abstract-booklet/.
- Schöch, Christof; Roxana Patras; Tomaž Erjavec; Diana Santos: "Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives", submitted.

Danke! · Thank you! · Merci! · Ačiū!

CC-BY 4.0
