# Research in Nordic literary collections: What is possible and what is relevant?

Mads Rosendahl Thomsen, Mats Malm and Kristoffer Laigaard Nielbo
Aarhus University, Gothenburg University, University of Southern Denmark

There are a growing number of digital literary collections in the Nordic countries that make the literary heritage accessible and have great potential for research that takes advantage of machine readable texts. These collections range from very large collections such as the Norwegian Bokhylla, medium-sized collections such as the Swedish Litteraturbanken and the Danish Arkiv for Dansk Litteratur, to one-author collections, e.g. the collected works of N.F.S. Grundtvig. In this presentation we will discuss some of the obstacles for a more widespread use of these collections by literary scholars and present outcomes of a series of seminars – UCLA 2015, Aarhus 2016, UCLA 2017 – sponsored by the Fondation Maison des sciences de l'homme courtesy of a grant from the Andrew Carnegie Mellon Foundation.

We find that there are two important thresholds in the use of collections:
1) The technical obstacles for collecting the right corpora and applying the appropriate tools for analysis are too high for the majority of researchers working in literary studies. While much has been done to advance the access to works, differences in formats and metadata make it difficult to work across the collections. Our project has addressed this issue by creating a Nordic github repository for literary texts, CLEAR, which provides cleaned versions of Nordic literary works, as well as a suite of tools in Python.

2) The capacity to combine traditional hermeneutical approaches to literary studies with computational approaches is still in its infancy despite numerous good studies from the past years, e.g. by Stanford Literary Lab, Leonard and Tangherlini and Ted Underwood. We have worked to bring together scholars with great technical prowess and more traditionally trained literary scholars in a series of seminars, to generate projects that are technically feasible and scholarly relevant. The process of expanding the methodological vocabulary of literary studies is complicated and requires significant domain expertise to verify the outcome of computational analyses, and conversely, openness to work with results that cannot be verified by close readings.

In this paper, we will present a Nordic repository for literature and discuss the challenges and choices involved with this. We shall also present a test-case on how thematic variation and readability can provide new perspectives on Swedish and Danish literature, and how prototypes of a dashboard for accessing the repository and data on textual coherence can improve scholarship on literature.

**Literature**

Algree-Hewitt, Mark et al. 2016. "Canon/Archive. Large-scale Dynamics in the Literary Field." *Stanford Literary Lab* Pamphlet 11.

Heise, Ursula. 2017. "Comparative literature and computational criticism: A conversation with Franco Moretti." *Futures of Comparative Literature: ACLA State of the Discipline Report*. London: Routledge, 2017.

Leonard, Peter and Timothy R. Tangherlini. 2013. "Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research". *Poetics* 41(6): 725-749.

Thomsen, Mads Rosendahl et al. 2015. "No Future without Humanities." *Humanities* 1.

Underwood, Ted. 2013. *Why Literary Period Mattered*. Stanford: Stanford University Press.