

# Swe-Clarín Theme: Literary Text

Daniel Brodén,<sup>1</sup> Mats Dahllöf,<sup>3</sup> Adam Ek,<sup>2</sup> Mats Malm<sup>1</sup> & Mats Wirén<sup>2</sup>

<sup>1</sup>Department of Literature, History of Ideas, and Religion, University of Gothenburg

<sup>2</sup>Department of Linguistics, Stockholm University

<sup>3</sup>Department of Linguistics and Philology, Uppsala University

## Abstract

The aim of the Swe-Clarín theme Literary Texts is to find new methods for searching and analysing Sweden's literary heritage based on large-scale computational analysis, thereby supplementing the kind of close reading typically used in literary studies. The approach is interdisciplinary and draws upon theory and methodology from language technology, literary studies and digital humanities. The theme collaborates with international research institutions, including UCLA and Yale University Library, and participates in initiatives such as SANTA (Systematic Analysis of Narrative Texts through Annotation). Resource-wise, the theme is firmly based on the Swedish Literature Bank with its currently over 3,000 literary works. Below we outline our main activities contributing to the theme.

## University of Gothenburg

### Sub-Corpus Topic Modeling on Swedish Texts

The Centre for digital humanities, University of Gothenburg and the Swedish Literature Bank collaborate with the Scandinavian Dept., UCLA and Yale University Library in creating new ways to explore Sweden's literary heritage. Sub Corpus Topic Modeling was devised by Peter Leonard and Timothy Tangherlini as a way of directing studies through the researcher's domain expertise (2013). Through this collaboration it will be implemented at litteraturbanken.se, enabling exploration and pattern-finding over vast materials, but also making it possible to narrow down on single occurrences and their context. Thus enabling qualitative studies through quantitative methods and combining macro levels with micro levels, the project aims at expanding the ways into Swedish literary heritage and provide an open source tool. The image renders Leonard and Tangherlini's envisioning of the 'dashboard'.

### Project application to The Swedish Research Council

"Literary Heritage as Source Material for the Humanities and Social Sciences" will explore how literary cultural heritage can be used as a source material for a range of Humanities and Social Sciences disciplines. The ambition is to further the development of data-driven research through interdisciplinary collaboration and exchange of methods, new combinations of quantitative and qualitative approaches, and chronological perspectives for mapping transformations in society and culture. The work will be performed with an attention to possibilities as well as limitations and risks of different methods. The project will make a corpus of the Swedish literary works published separately 1800—1900 freely available, together with tools for exploring the material.

## Stockholm University

### Analysis of narrator's and characters' discourses in literary fiction

Literary fiction typically consists of passages alternating between two levels of narrative transmission: the narrator's telling of the story and the characters' speaking to each other in that story (as mediated by the narrator). Here is an example from Strindberg's *Röda rummet*:

*Olle gjorde med stor färdighet en påse av ett lakan och stoppade in alltihop under Lundells ivrigaste protester. Sedan påsen var färdig tog Olle den under armen, knäppte ihop den trasiga bonjouren för att bättre kunna dölja västens frånvaro och vandrade i väg åt staden. – Han ser ut som en tjuv, sade Sellén, där han stod i fönstret och tittade illmarigt ut åt vägen. – Får han bara gå i fred för polisen nu, så är det bra! – Raska på Olle! ropade han efter den bortgående! Köp sex franska bröd och två halva öl, om det blir något över på färgen!*

These narrative levels require different types of analysis. In a narrator's discourse, we are primarily interested in a sequence of events, the ordering of these, and how they form a plot. Although this is true also for characters' discourse, the fact that the narration is expressed through the speech of the characters makes the problem different. To begin with, we need to identify who the speakers and addressees are.

We are approaching these problems by developing methods for distinguishing narrator's and characters' discourses and for identifying speakers and addressees, assuming that these will be important ingredients in an environment for exploring Sweden's literary heritage.

## Uppsala University

### From Close Reading to Distant Reading: Digital Humanities and New Forms for Literary Analysis

A research programme funded by the University. Led by Johan Svedjedal.

Participants: Sociology of Literature (Department of Literature), Swedish (Department of Scandinavian Languages), Computational Linguistics (Department of Linguistics and Philology)

Ongoing studies (the two first have resulted in submissions):

–Prose rhythm in narrative fiction: the case of Karin Boye's *Kallockain*

–Topic modeling and gendered themes in classical and modern Swedish prose fiction

–A study on dialogue and quotation in fiction



## Workshop: Digital literary analysis

The Centre for Digital Humanities at the University of Gothenburg will host an international interdisciplinary workshop with the Danish Research Network, "Netværk for digital litteraturforskning" 18—19 December 2018. Participants will include Swedish and Danish researchers from a range of disciplines, including literary studies, language technology and digital humanities.