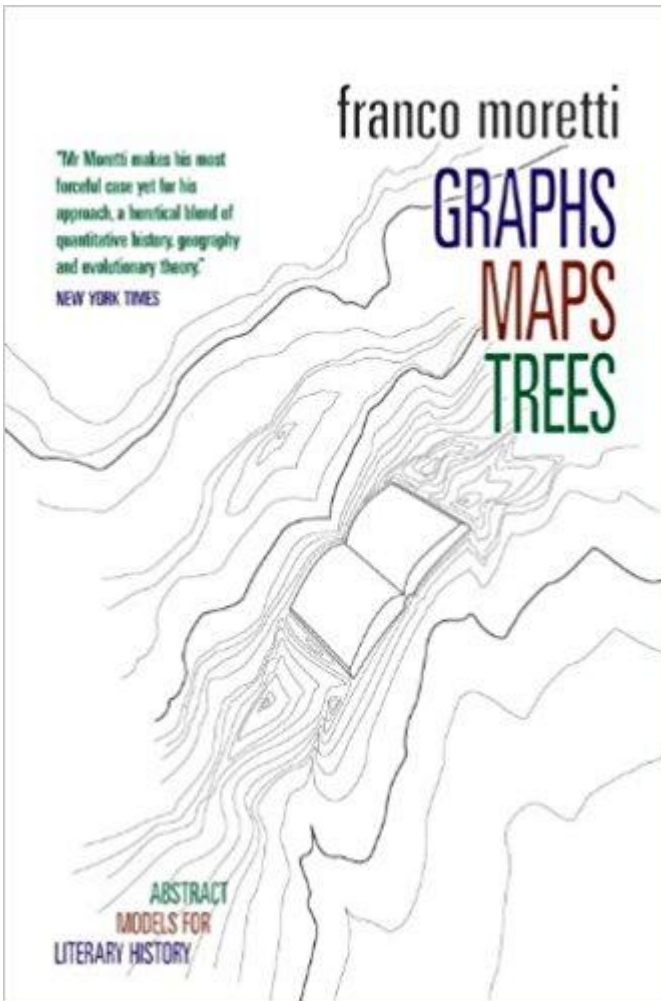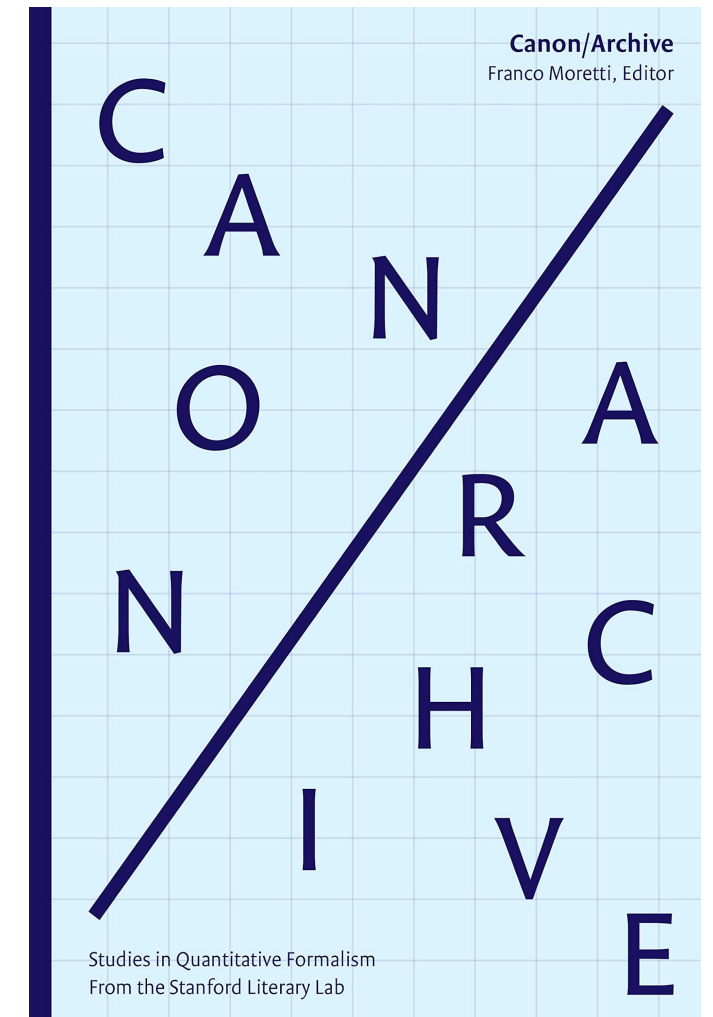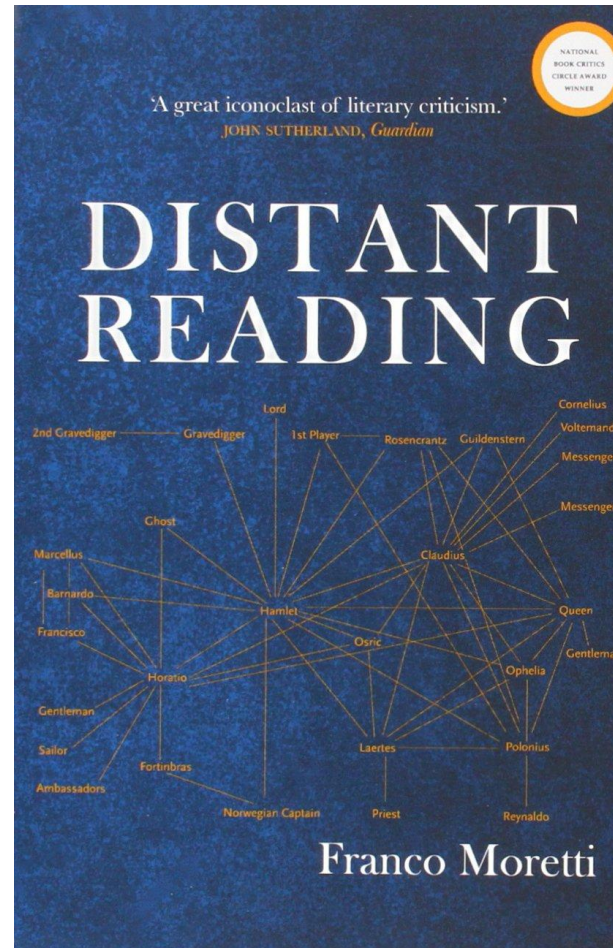# Why Distant Reading?

Dr. Simone Rebora, Johannes Gutenberg-Universität Mainz
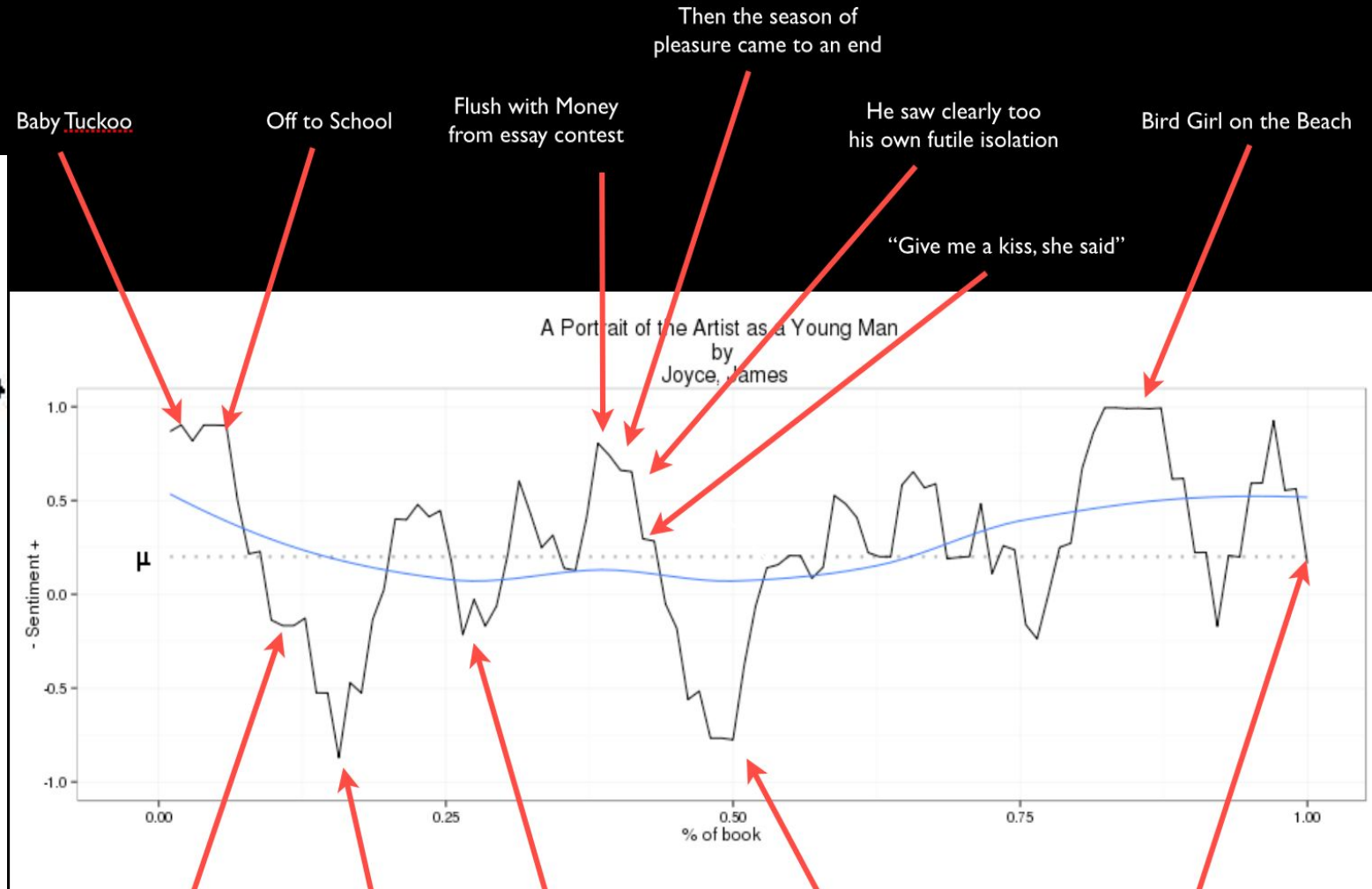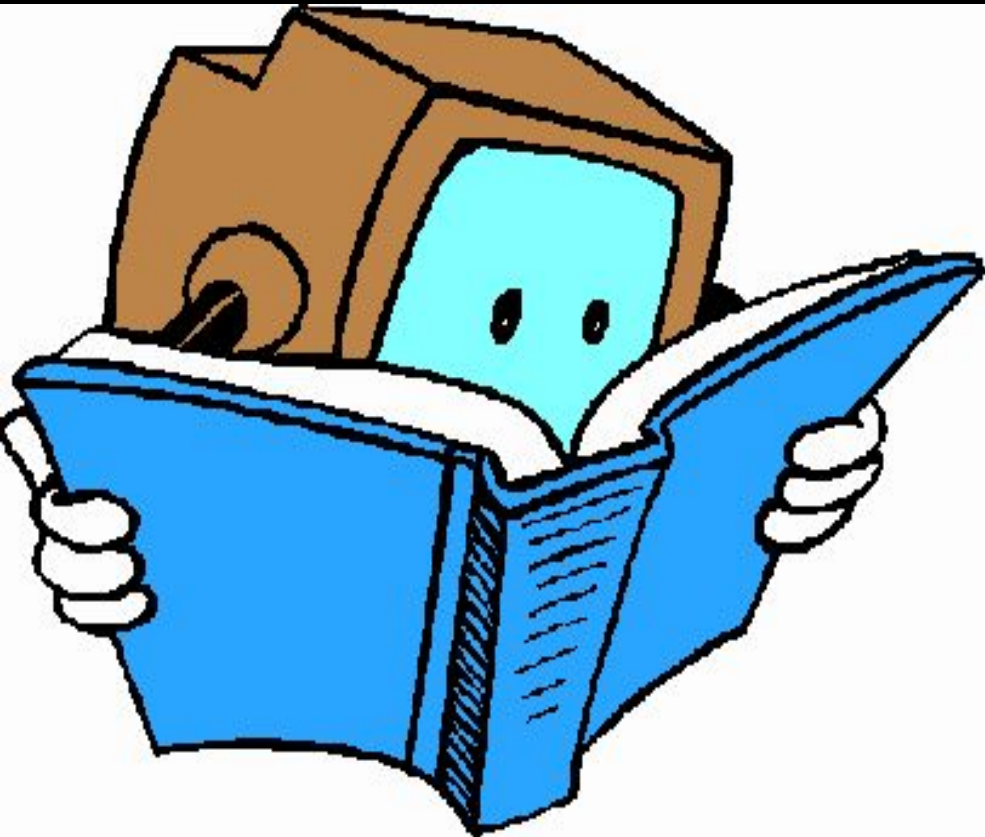Dr. Giulia Grisot, Universität Bielefeld

Distant Reading

"Instead of reading texts in the traditional way – so-called close reading –, he invites to count, to graph and to map or, in other words, to visualize them" **(Jänicke et al., 2015)**

Distant Reading

Distant Reading

"[…] a canon of two hundred novels, for example, sounds very large for nineteenth-century Britain (and is much larger than the current one), but is still less than one per cent of the novels that were actually published: twenty thousand, thirty, more, no one really knows—and close reading won't help here, a novel a day every day of the year would take a century or so"
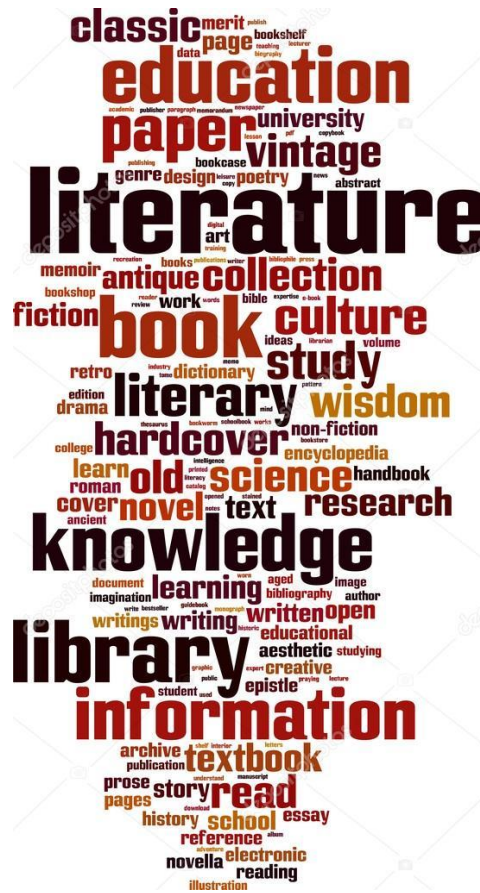**(Moretti, 2005)**

Distant Reading Methodologies

Stylometry
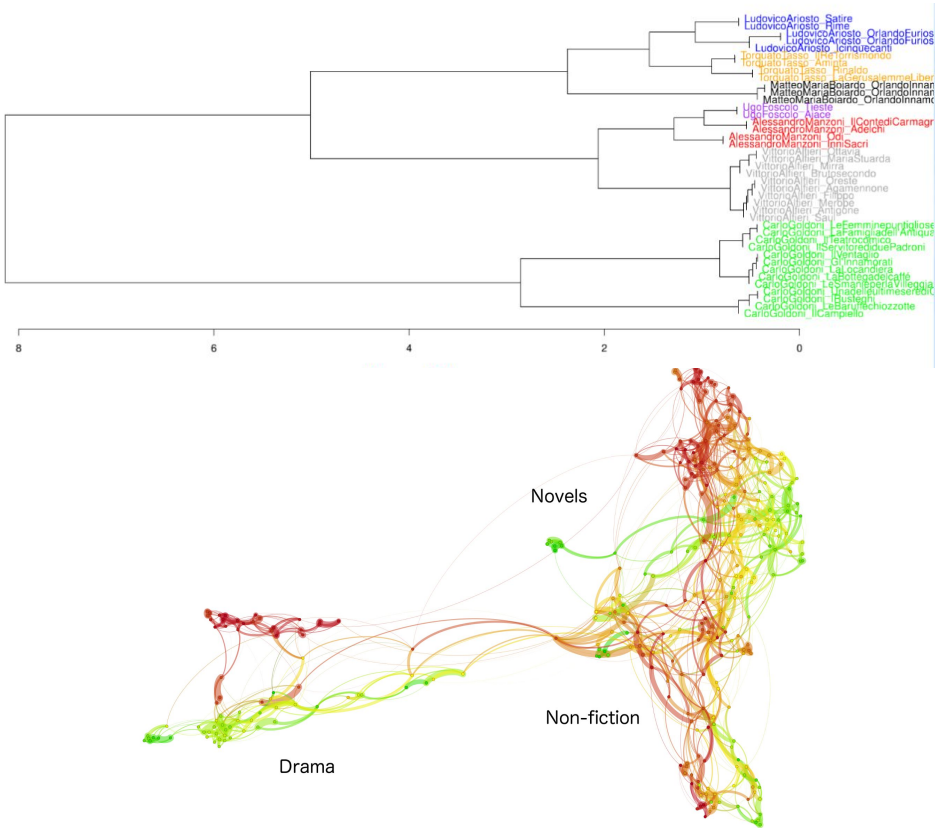
Topic Modeling

Sentiment Analysis

Networks

Maps

Stylometry Case Studies
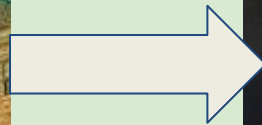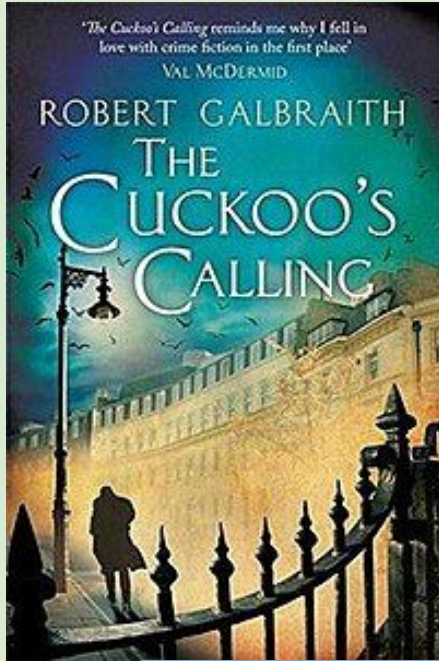
J. K. Rowling
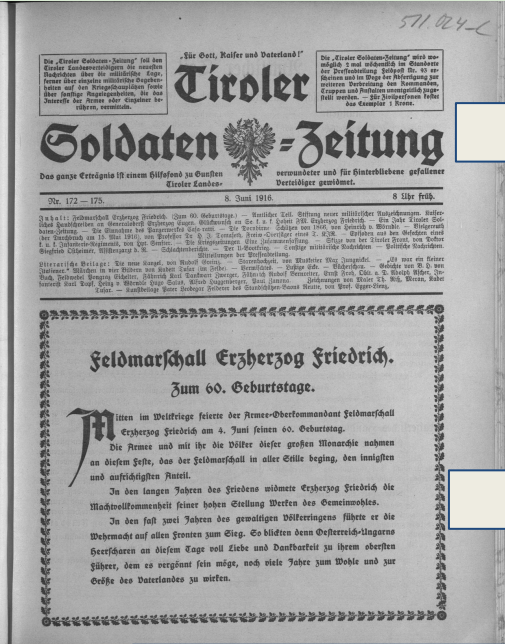
(Juola, 2015)
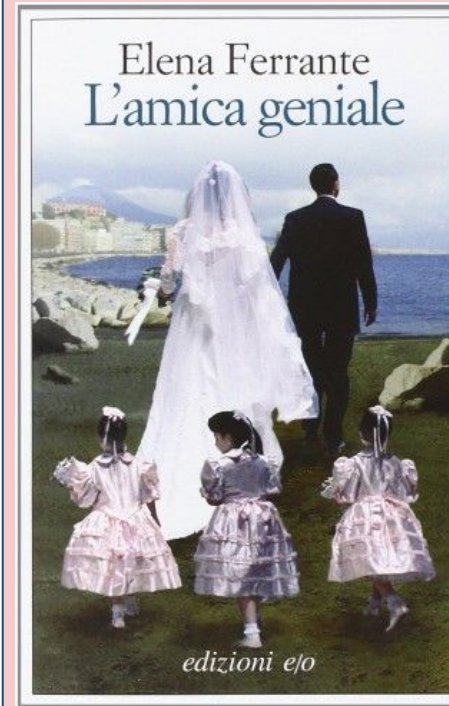
Domenico Starnone

(Tuzzi and Cortelazzo, 2018)

Stylometry Case Studies

Albert Ritter

Robert Musil

(Rebora et al., 2019)

Josefine Mutzenbacher
oder
Die Geschichte einer Wienerischen Dirne
von ihr selbst erzählt

Anonymous, 1906

Felix Salten

(Rebora and Salgaro, 2021)

Sentiment Analysis

What is SA?

- SA is commonly used to refer to the examination and identification of:

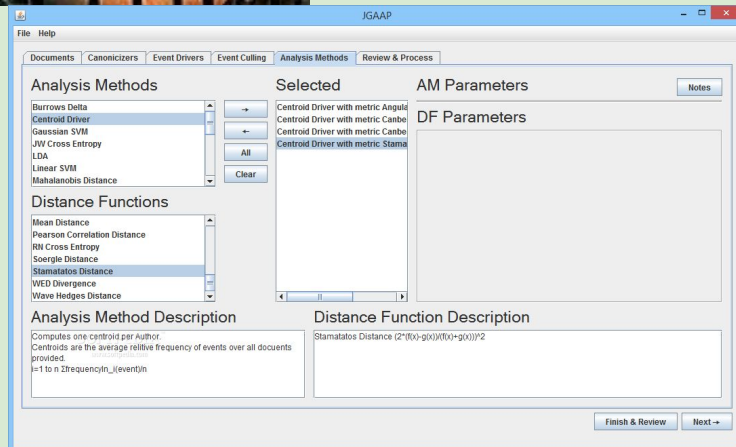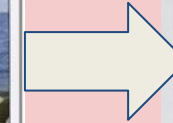   1. The **polarity** (or valence) of a piece of text (i.e. neg vs. pos)

   2. Someone's **attitude** towards a particular target or topic

   3. **Discrete emotions** encoded in a piece of text (fear, joy, etc.)

   (Mohammad 2017)

- SA aims to discover the
   - entity,
   - aspect,
   - sentiment,
   - holder,
   - time

   of a document

   (additionally also the reason and qualifier for the sentiments)

   (Liu 2015)

Sentiment Analysis
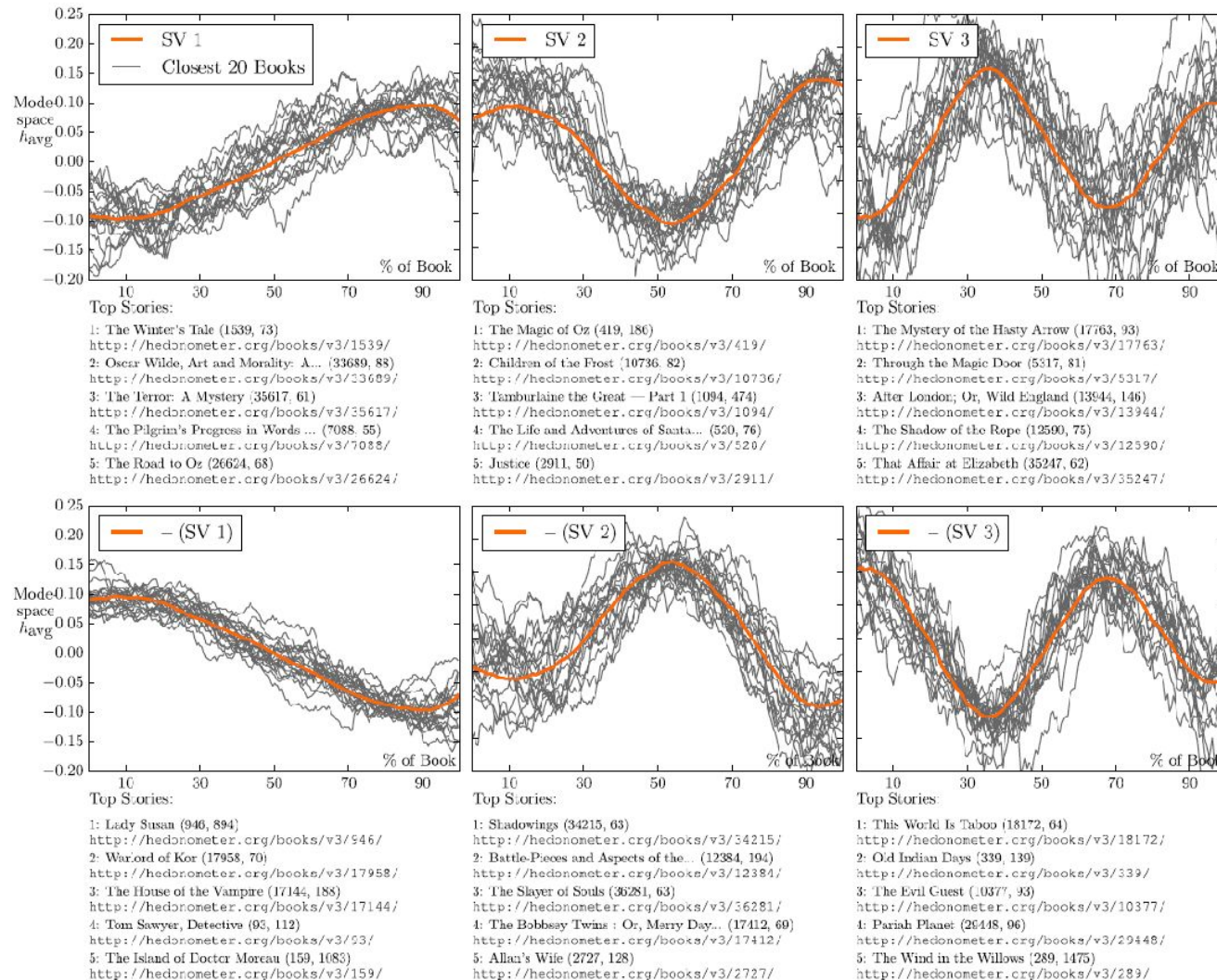
- Sentiment analysis can be applied to a large number of books to find the archetypal "shapes" of stories
- Reagan et al., 2016

Sentiment Analysis

Figure 3.2 The stability of fictional London, 1700-1900

1700-1749

1800-1849

1750-1799

1850-1899

- Or to investigate patterns that are otherwise very difficult to identify
- e.g. Heuser et al 2016 - *Mapping the Emotions of London*

..and many more methods and approaches

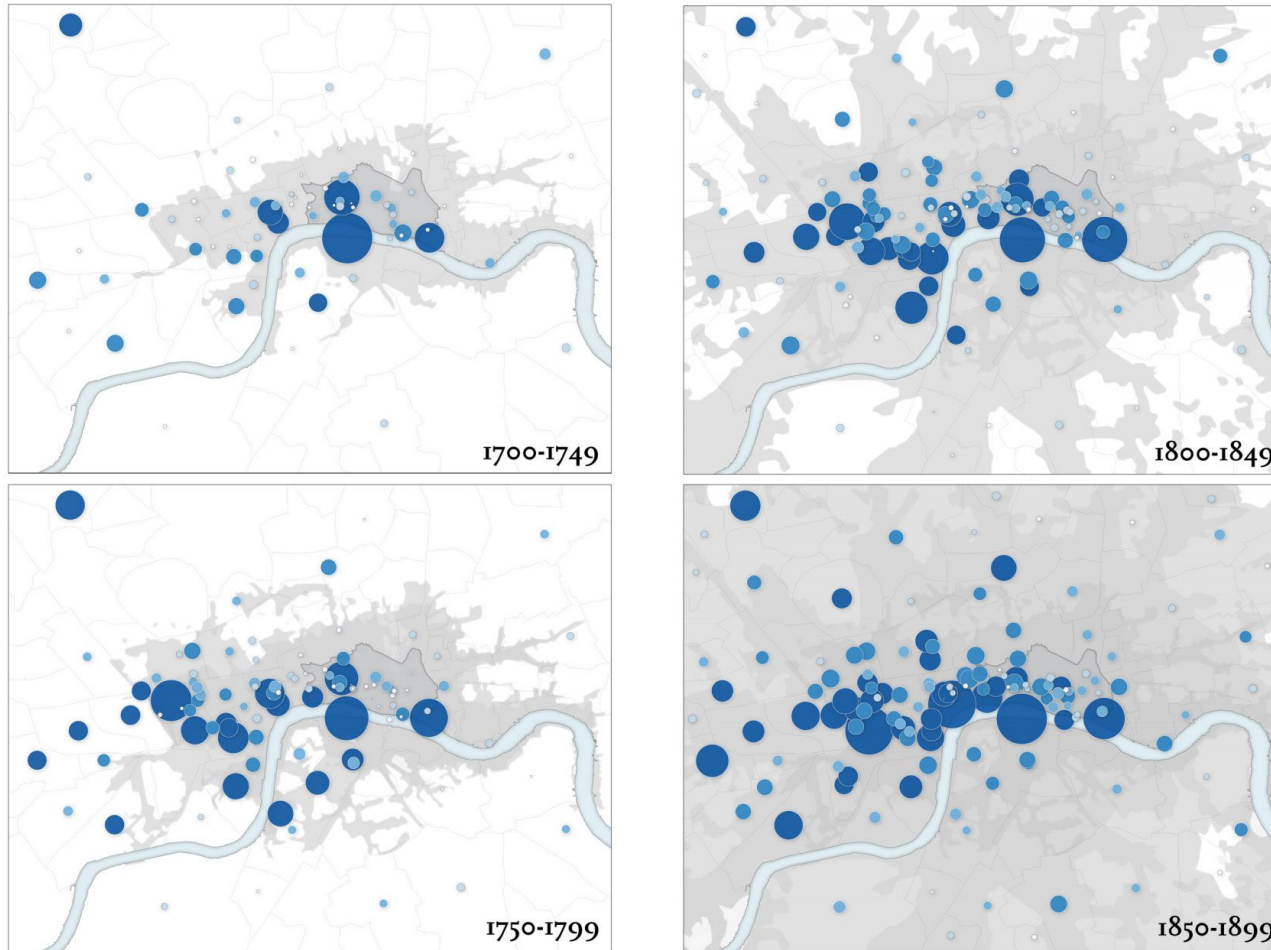## Corpus-Based Literary Studies

e.g. Herrmann & Lauer 2018

J. Berenike Herrmann & Gerhard Lauer

Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne

Ein Fach oder eine Teildisziplin ,Korpusliteraturwissenschaft' gibt es nicht, nicht so jedenfalls, wie es Korpuslinguistik als einen Bereich der Sprachwissenschaft gibt, der in systematisierender wie historischer Absicht natürlichsprachliche Äußerungen sammelt (Bubenhofer 2009; Biber/Conrad/Reppen 2006; Lemnitzer/Zinsmeister 2015). Auch spricht man in der Literaturwissenschaft und selbst in der Literaturgeschichte nur selten von einem Quellenkorpus, wie das in historischen Fächern üblich ist, wenn unter Quellen die je nach Fragestellung systematische Zusammenstellung möglichst breit angelegter Sammlungen von Zeugnissen verstanden wird (Maurer 2002). Außerdem fehlt in der Literaturwissenschaft weitgehend die Verbindung von qualitativen und quantitativen Methoden der Textanalyse, wie sie in der Linguistik zu finden ist (Kim 2012; Mei-ell 2011)

Das Fehlen ein...
in der Literaturwi...

Osnabrücker Beiträge zur Sprachtheorie, 92 (2018)

Machine Learning

e.g. Goval & Prakash 2022

Springer Link

Advances in Data and Information Sciences pp 297–305 | Cite as

Statistical and Deep Learning Approaches for Literary Genre Classification

Anshaj Goyal & V. Prem Prakash

Conference paper | First Online: 08 February 2022

Circuit is open

Part of the Lecture Notes in Networks and Systems book series (LNNS,volume 318)

Abstract

The problem of automatically classifying literary texts into their corresponding genres/subgenres occurs in Natural Language Processing (NLP) and has applications in library classification systems and semantic metadata extraction. Existing literature on this

Today's workshop

Methods

- Frequencies and Wordclouds
- Keyness analysis
- Stylometry
- Sentiment Analysis

Approaches

- From research questions to data analysis

Corpus

We will use a corpus based on the German ELTeC collection (65 of the original 100 texts), produced by the "Distant Reading for European Literary History" COST Action.

Info: https://www.distant-reading.net/eltec/

Posit Cloud

Instructions:

1. Connect to https://posit.cloud/
2. Create a new account (Cloud Free Plan; Sign Up with Google)
3. Create a new project from Git Repository
4. Copy-paste the link: https://github.com/distant-reading-Bie22/distant-reading-Bie22.github.io.git (you can find it in the Website, under "resources")
5. Click on OK!

References

- Herrmann, J. B., & Lauer, G. (2018). Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne. Osnabrücker Beiträge zur Sprachtheorie, (92).
- Heuser, R., Moretti, F., & Steiner, E. B. (2016). The emotions of London. Literary Lab.
- Lei, L., & Liu, D. (2021). Conducting sentiment analysis. https://doi.org/10.1017/9781108909679
- Liu, A. (2013). The Meaning of the Digital Humanities. PMLA/Publications of the Modern Language Association of America, 128(2), 409–423. https://doi.org/10.1632/pmla.2013.128.2.409
- Liu, H., Qiu, Q., Wu, L., Li, W., Wang, B., & Zhou, Y. (2022). Few-shot learning for name entity recognition in geological text based on GeoBERT. Earth Science Informatics. https://doi.org/10.1007/s12145-022-00775-x
- Pianzola, F., Rebora, S., & Lauer, G. (2020). Wattpad as a resource for literary studies. Quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins. PloS one, 15(1), e0226708.
- Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. EPJ Data Science, 5(1), 1-12.
- Rebora, S., Herrmann, J. B., Lauer, G., & Salgaro, M. (2019). Robert Musil, a war journal, and stylometry: Tackling the issue of short texts in authorship attribution. Digital Scholarship in the Humanities, 34(3), 582-605.
- Underwood, T., Bamman, D., & Lee, S. (2018). The Transformation of Gender in English-Language Fiction. Journal of Cultural Analytics. https://doi.org/10.22148/16.019
- van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. Communication Methods and Measures, 15(2), 121–140. https://doi.org/10.1080/19312458.2020.1869198