

Relating the Unread

Modellierungen der Literaturgeschichte

Judith Brottrager, TU Darmstadt

DHd, Doctoral Consortium, 09.03.2022

Überblick

- Forschungsfragen & Ziele
- Kontextreiche Metadatensätze für literaturgeschichtliche Fragestellungen
- Operationalisierung literaturwissenschaftlicher Konzepte für quantitative Analysen
- Ausblick: Erste Ergebnisse & weitere Schritte

Forschungsfragen

Wie verhält sich das Korpus zum Kanon? Inwiefern kann „The Great Unread“ abgebildet werden?

Wie können literaturwissenschaftliche Kategorien für quantitative Analysen operationalisiert werden?

Wie muss ein Datensatz aufgebaut werden, um nachhaltig „Data-Rich Literary History“ (Bode 2018) zu ermöglichen?

Diachrone & synchrone Vergleiche
von Kanonisierungs- und
Wertungsprozessen in englisch- und
deutschsprachiger
Literaturgeschichte (1688–1914)

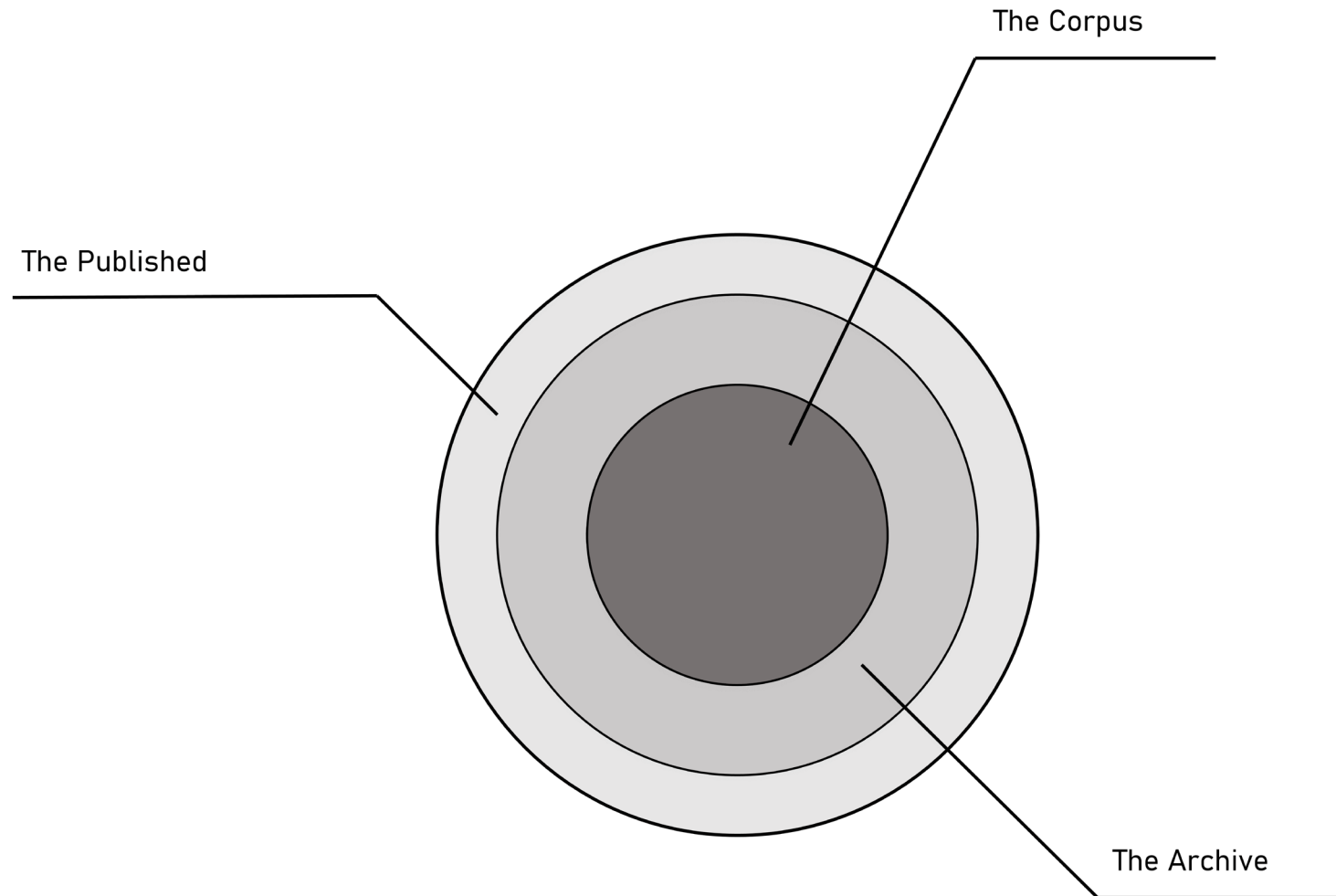
Quantitative Überprüfung etablierter
literaturhistorischer Narrative:
Kanonisierungsprozesse und
Autorengender (cf. Watt 1957; Raven
1987; Raven & Forster 2000; Tuchman
& Fortin 2012)

Modellierung von Textähnlichkeiten:
Gibt es Textmerkmale, die mit
Kanonisierungs- und
Wertungsprozessen in Verbindung
gebracht werden können?

Kontextreiche Metadatensätze für literaturgeschichtliche Fragestellungen

The Published The Archive The Corpus

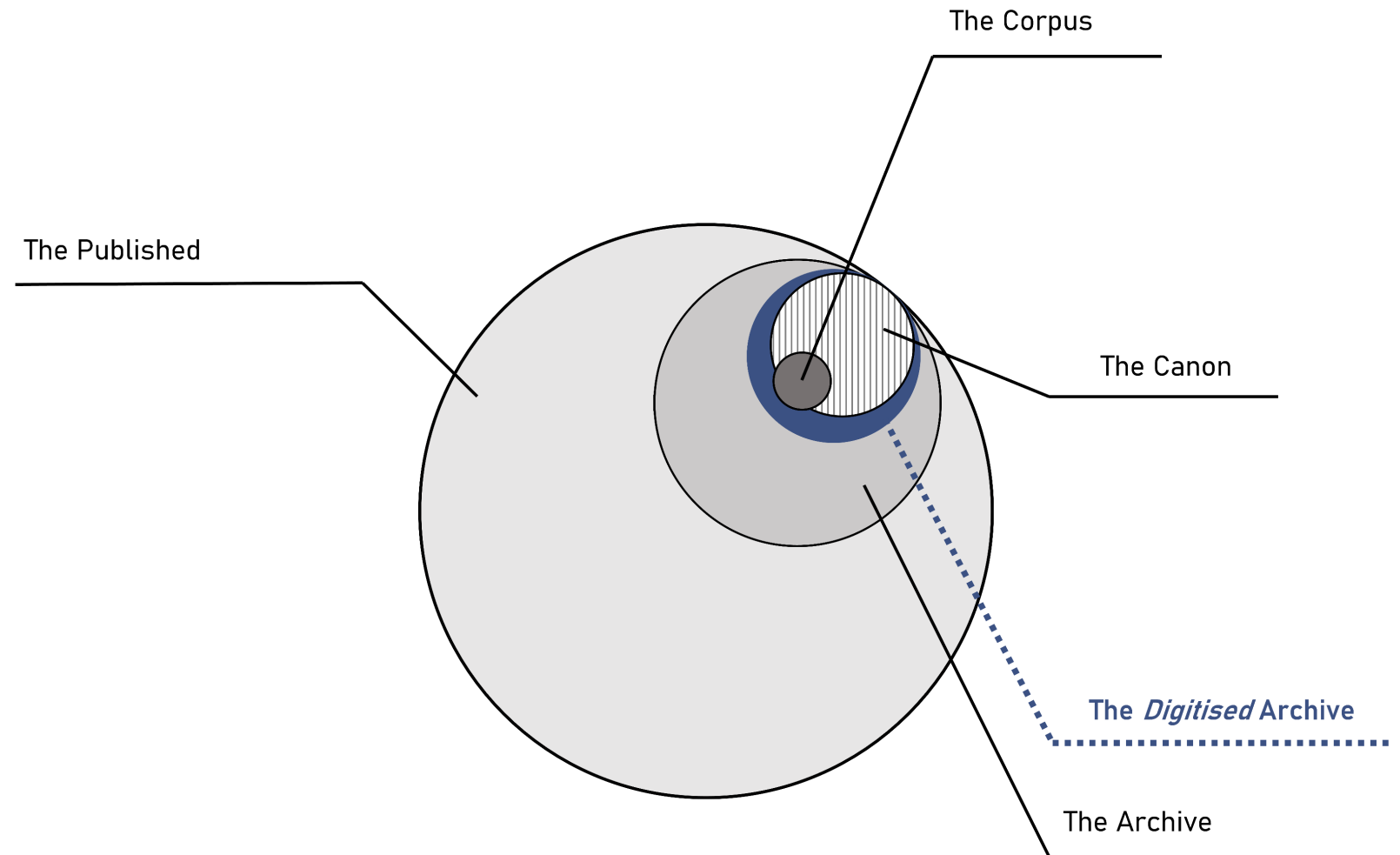
Algee-Hewitt et al. (2016)



The Published The Archive The Corpus

Algee-Hewitt et al. (2016)

Bode (2018)



Korpuserstellung

Algee-Hewitt & McGurl (2015): Repräsentatives Korpus für die englischsprachige Literatur des 20. Jahrhunderts

normativ-exklusiven Kanon

populäre Texte

**von Expert:innen für Postkoloniale und
Feministische Literaturwissenschaft
vorgeschlagene Werke**

exklusive Literaturgeschichten & Anthologien

spezialisierte Sekundärliteratur

**erweiterter akademischer Kanon & spezialisierte
Sekundärliteratur**

Korpuserstellung

- 1.200 englisch- und deutschsprachige Prosatexte
 - Gängige Textsammlungen
 - Webscraping Repositorien
 - Umwandlung von EPUBs/PDFs
 - Retrodigitalisierung (ABBYY)
- Bedingungen:
 - Gemeinfrei
 - Keine Übersetzungen
 - Fiktionalität (z.B. keine Autobiografien)

Kontextreiche Datensätze

- Klassische Metadaten (Autor:in, VIAF, Publikationsjahr, etc.)
- Erwähnungen in Sekundärtexten (pro Ebene)
- Ausgaben: Studienausgaben, Gesamtausgaben
 - Oxford Classics | Reclam Universalbibliothek
 - Scraping BNB | DNB
- Zeitgenössische Rezensionen
 - OCR4all
- Einträge in Leihbibliothekskatalogen
 - OCR4all

Operationalisierung literaturwissenschaftlicher Konzepte

Operationalisierungen

Nutzung für quantitative Analysen

- Formalisierung als numerischer Score
- Literaturwissenschaftlich plausibilisierte Operationalisierung
- Vergleichbarkeit im jeweiligen Korpus
- Vergleichbarkeit über beide Korpora hinweg

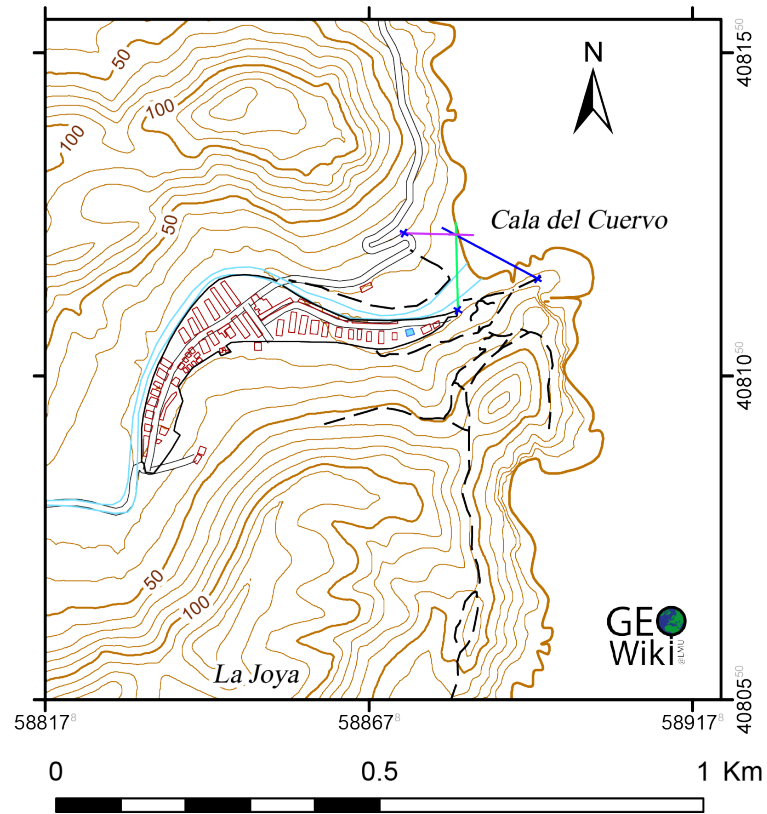
Kanonisierungsscore

Bisherige Forschungsansätze

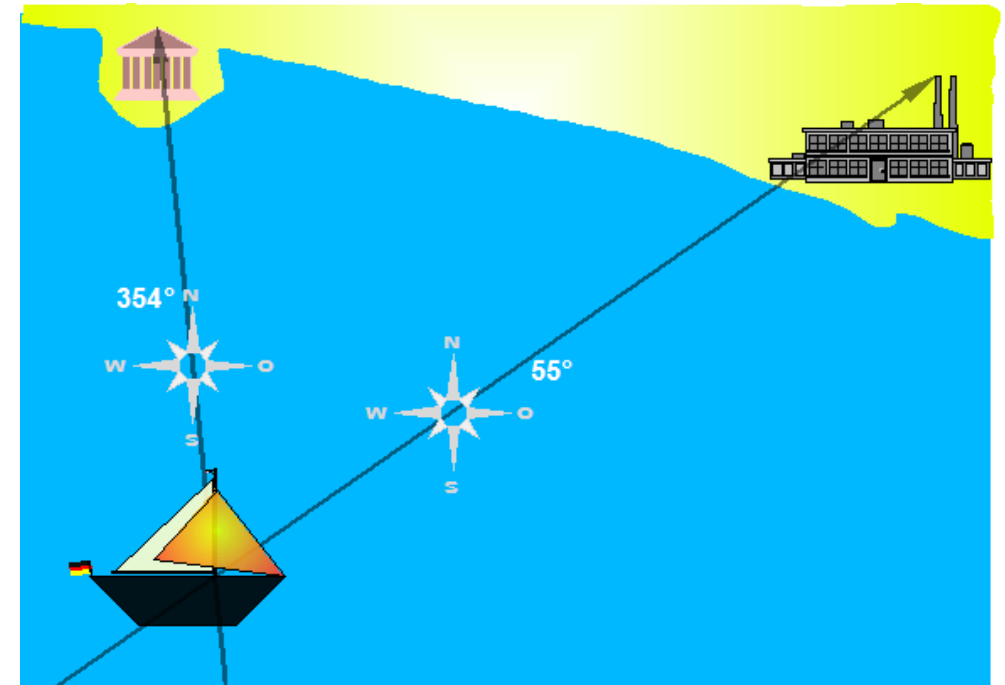
- Literarizität (Cranenburgh et al. 2019)
- Prestige und Popularität (Porter 2018, Underwood & Sellers 2016)

Fokus auf Prozesshaftigkeit (Kanon vs. Kanonisierung)

Analogie: Drei-Punkt-Peilung / Kreuzpeilung



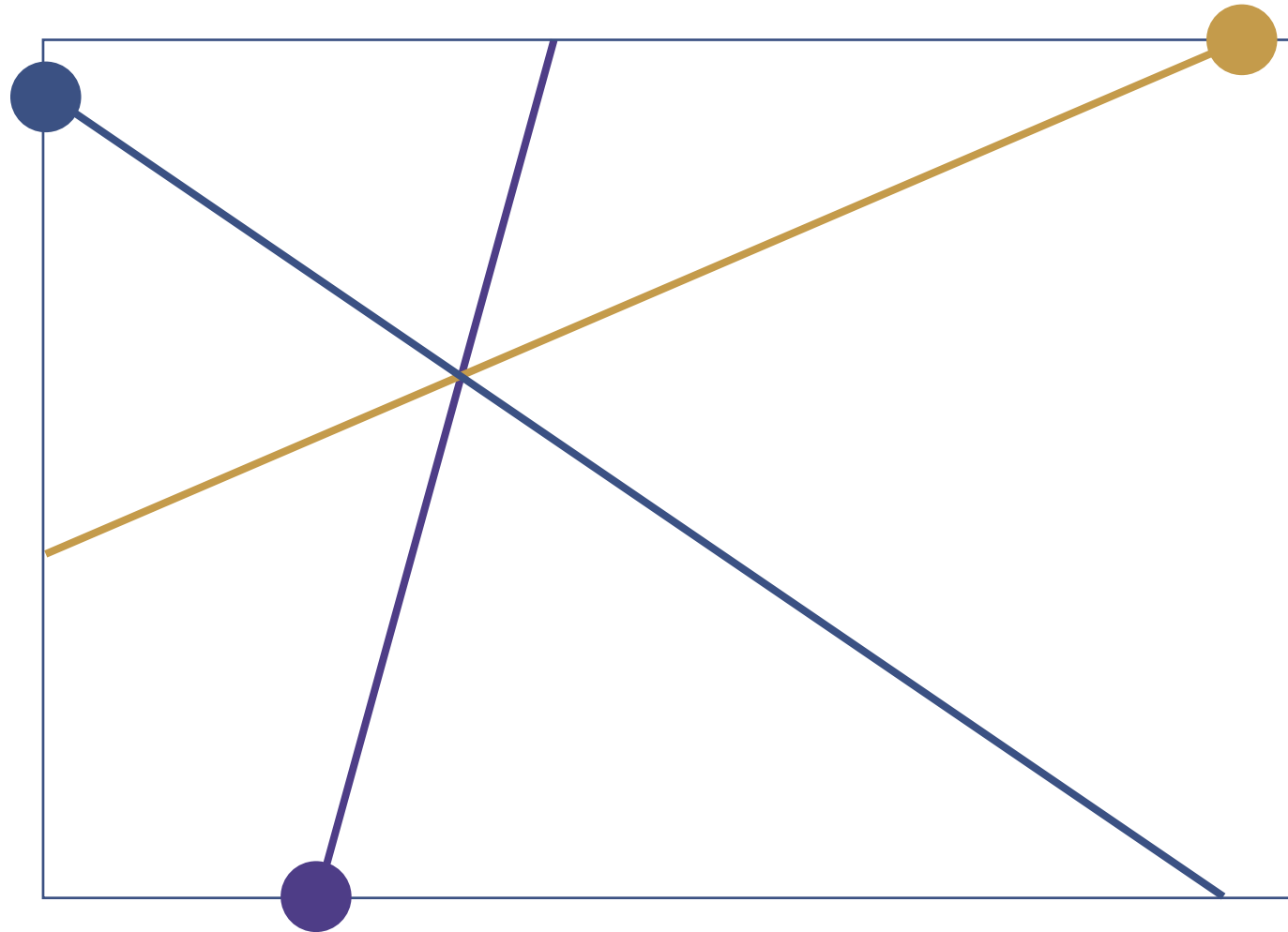
https://geowiki.geo.lmu.de/wiki/Datei:3PM_4.png/media/File:3PM_4.png



https://commons.wikimedia.org/wiki/File:Rueckwaertseinschneiden_a6.png

Anhaltende
Pflege in den
literatur-
vermittelnden
Institutionen

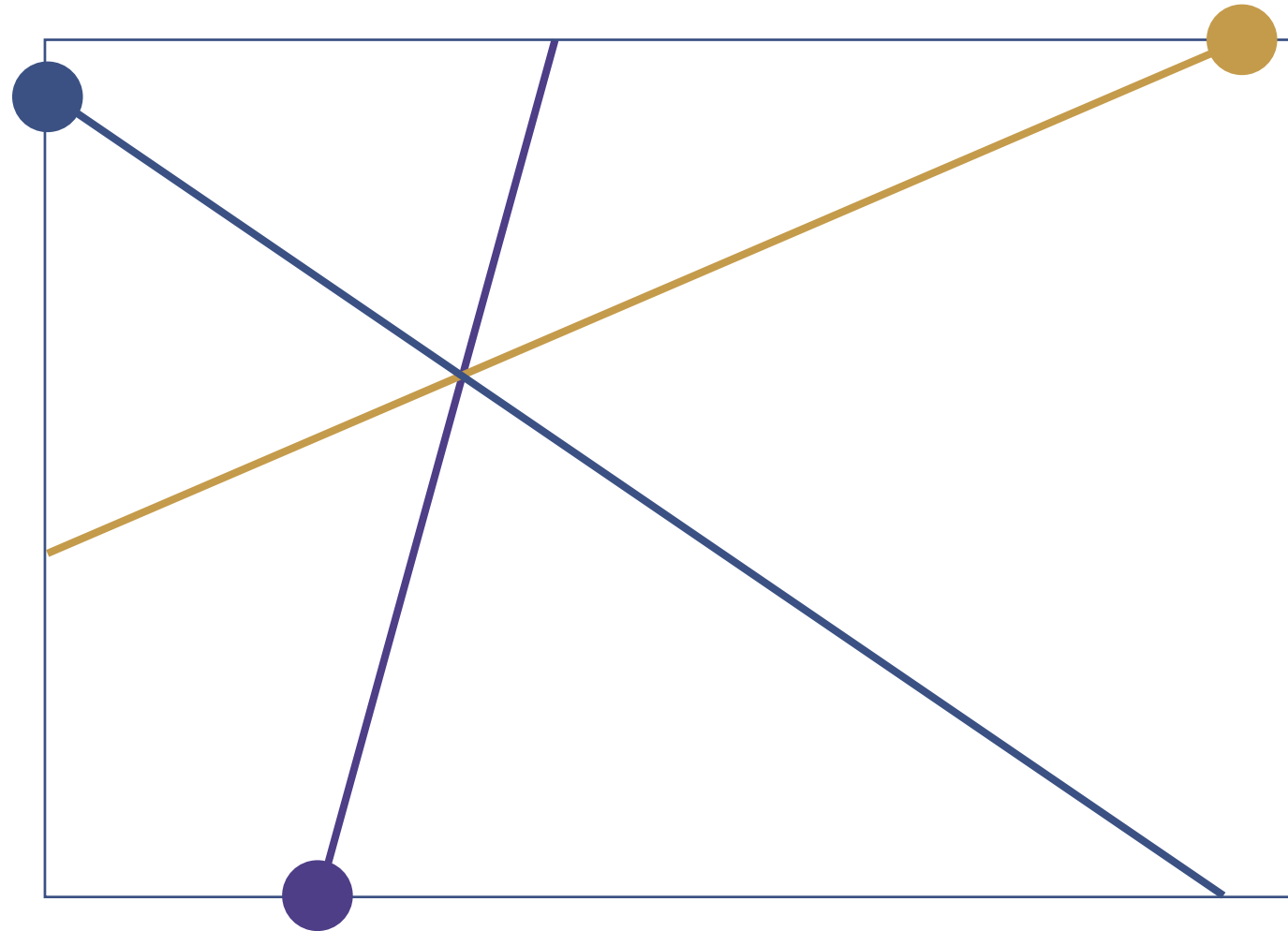
Regelmäßige und
ausführliche
Behandlung in
Literaturgeschichten,
Lexika u. a.



Gesamtausgabe(n)

cf. Heydebrand & Winko (1996)

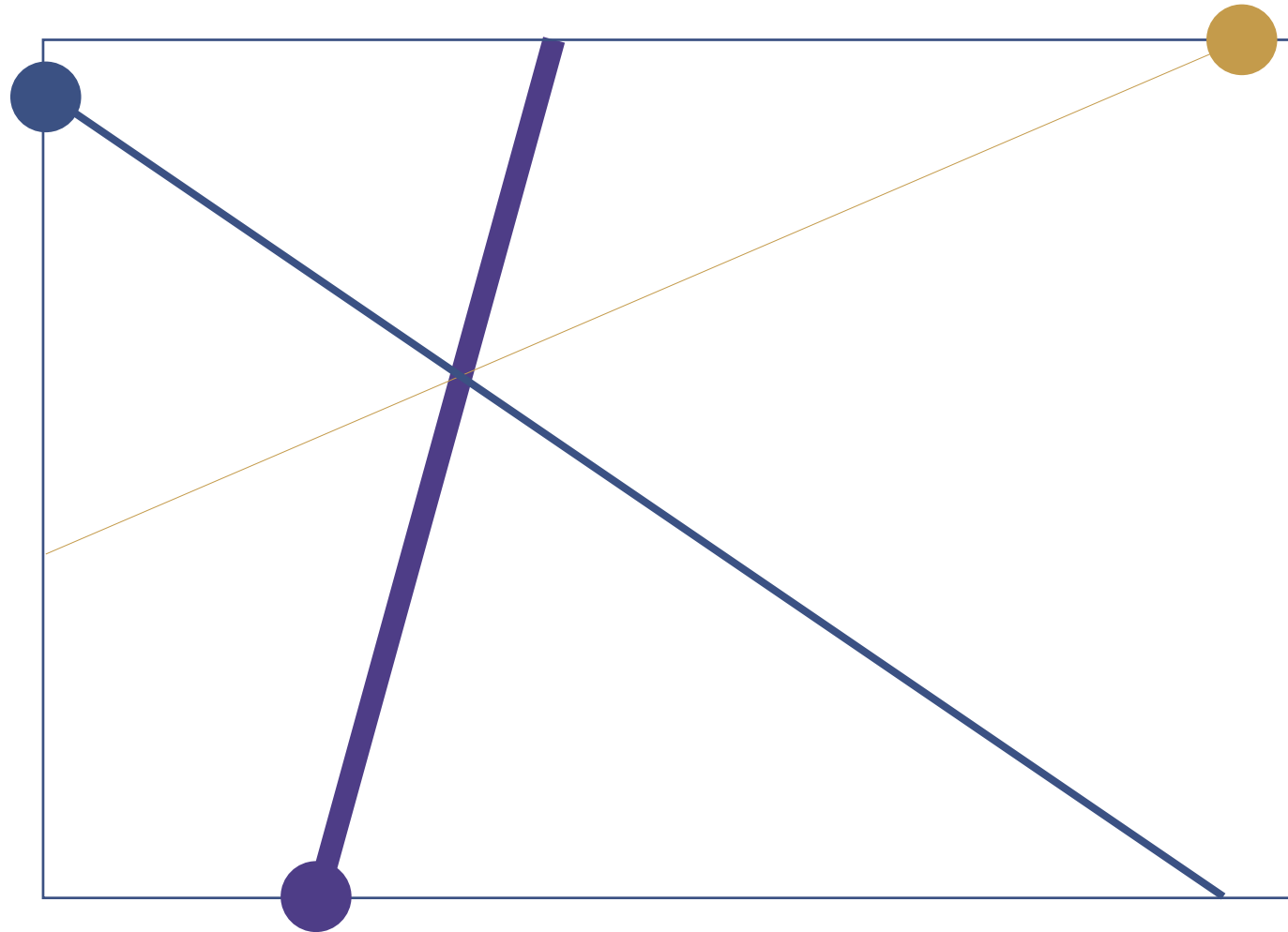
Studien-
ausgaben



Regelmäßige und
ausführliche
Behandlung in
Literaturgeschichten,
Lexika u. a.

Gesamtausgabe(n)

Studien-
ausgaben



Regelmäßige und
ausführliche
Behandlung in
Literaturgeschichten,
Lexika u. a.

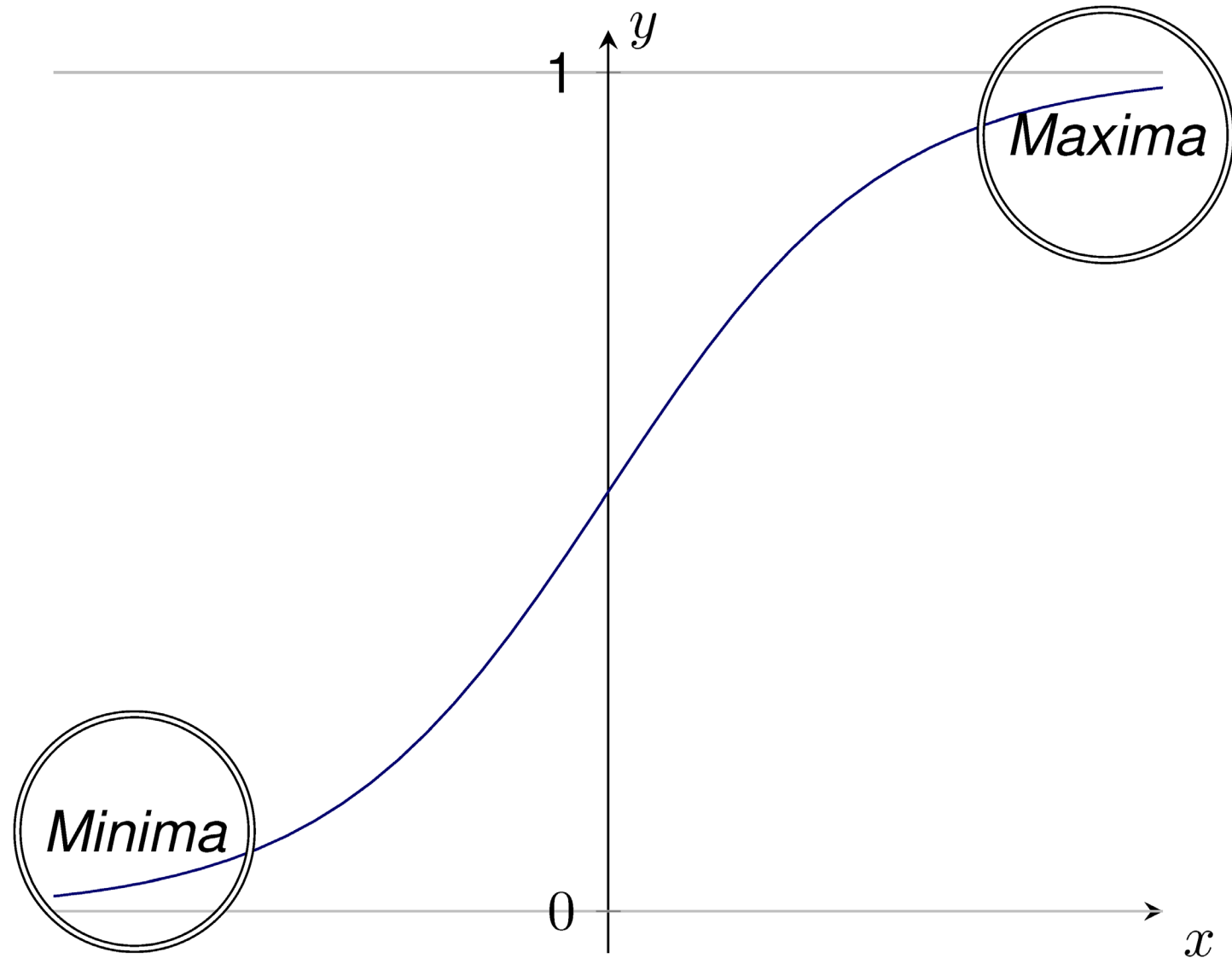
Gesamtausgabe(n)

Logistische Regression

Trainingsdaten

Minima & Maxima

- Spezialisierte Sekundärliteratur
- Universitäre Leselisten

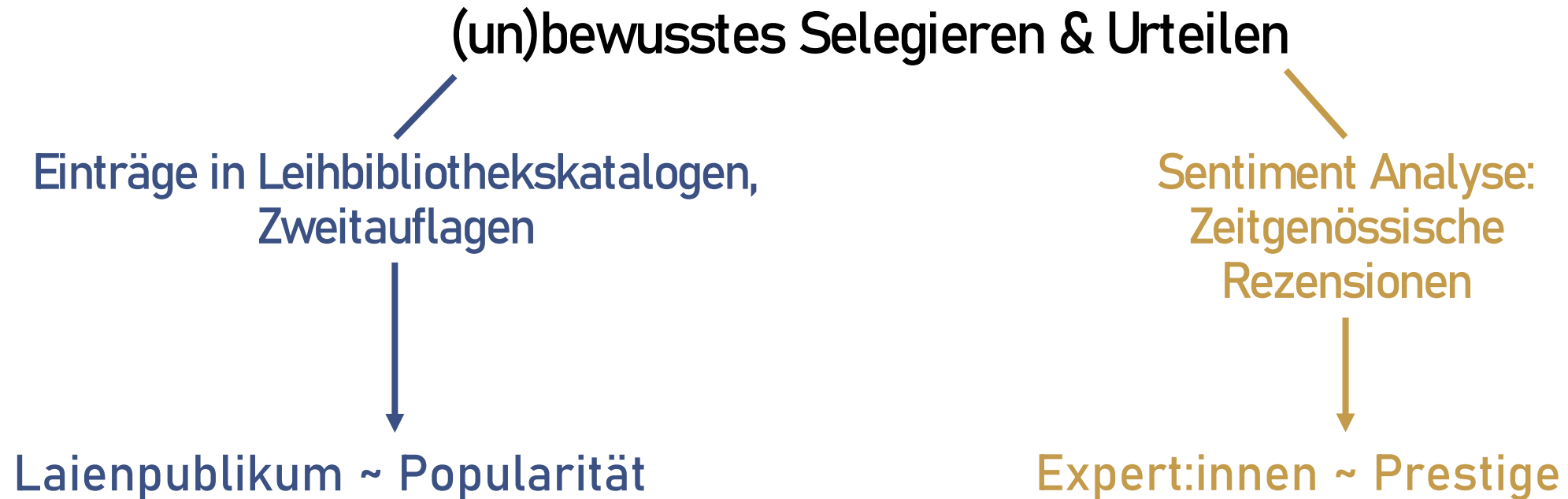


Kanonisierungsscore



Wertungsscore

Heydebrand & Winko (1996, 78–89): Werten als soziales Handeln



Ausblick: Erste Ergebnisse & weitere Schritte

Erste Ergebnisse

Kanonisierung

- Vergleich: Metadaten-basierter Kanonisierungsscore & Textmerkmale
- Korrelation
 - ENG: $r=0.242$
 - GER: $r=0.285$

Predicting Canonization: Comparing Canonization Scores Based on Text-Extrinsic and -Intrinsic Features

Judith Brottrager ¹ Annina Stahl ² Arda Arslan ²

¹Technical University of Darmstadt, Institute of Linguistics and Literary Studies

²ETH Zurich, Social Networks Lab

CHR2021
November 18, 2021

→ http://ceur-ws.org/Vol-2989/short_paper21.pdf

Erste Ergebnisse

Wertung

- Vergleich: Wertungsmarker (Rezensionen, Leihbibliothekskataloge) & Textmerkmale
- Sentiment Analyse: adaptierter SentiArt-Ansatz (Jacobs 2019)
- Genauigkeit
 - Rezensiert: ENG 0.69 | GER: 0.604
 - Leihbibliothekskataloge: ENG 0.71 | GER 0.61

Weitere Schritte

- Kohorten: Bildung von Subkorpora
- Datenbereinigung & nachhaltige Publikation
 - FDM
 - Nachnutzbarkeit
- Anhand Daten literaturwissenschaftliche Fragestellungen weiterentwickeln

Quellen

- Algee-Hewitt, Mark / Allison, Sarah / Gemma, Marissa / Heuser, Ryan / Moretti, Franco / Walser, Hannah (2016): "Canon/Archive. Large-scale dynamics in the literary field", Pamphlets of the Stanford Literary Lab 11. <https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf>
- Algee-Hewitt, Mark / McGurl, Mark (2015): "Between canon and corpus: Six perspectives on 20th-century novels", Pamphlets of the Stanford Literary Lab 8. <https://litlab.stanford.edu/LiteraryLabPamphlet8.pdf>
- Bode, Katherine (2018): A world of fiction: Digital collections and the future of literary history. Ann Arbor: University of Michigan Press.
- Cranenburgh, Andreas van / Dalen-Oskam, Karina van / Zundert, Joris (2019): "Vector space explorations of literary language", Language Resources and Evaluation 53(4), 625–650.
- Garside, Peter (2011): British Fiction 1800–1829: A Database of Production, Circulation & Reception. <http://www.british-fiction.cf.ac.uk/>
- Heydebrand, Renate von / Winko, Simone (1996): Einführung in die Wertung von Literatur: Systematik – Geschichte – Legitimation. Paderborn: Schöningh.
- Jacobs, Arthur (2019): „Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics", Frontiers in Robotics and AI 6(35), 4853.
- Porter, J. D. (2018): "Popularity/Prestige", Pamphlets of the Stanford Literary Lab 17. <https://litlab.stanford.edu/LiteraryLabPamphlet17.pdf>
- Reul, Christian / Christ, Dennis / Hartelt, Alexander / Balbach, Nico / Wehner, Maximilian / Springmann, Uwe / Wick, Christoph / Grundig, Christine / Büttner, Andreas / Puppe, Frank (2019): "OCR4all: An Open-Source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings", Applied Sciences 9(22)
- Underwood, Ted; Sellers, Jordan (2016). "The Longue Durée of Literary Prestige", Modern Language Quarterly 77(3), 321–344.

Kontakt

✉ judith.brottrager@tu-darmstadt.de

🐦 @JBrottrager



ETH zürich

linglit



Gefördert durch
DFG Deutsche
Forschungsgemeinschaft