



## Optimierung und Offene Publikation des Korpus der literarischen Moderne

Prof. Dr. Berenike Herrmann und Dr. Alan van Beek

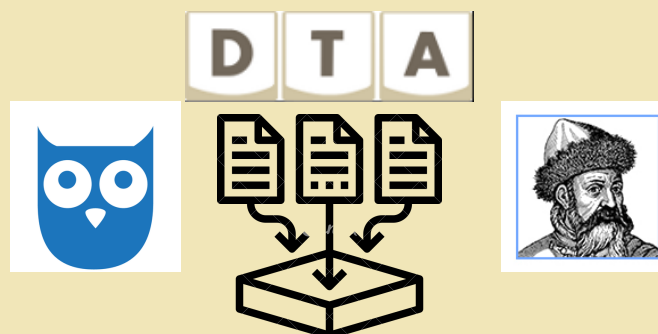


### EINFÜHRUNG

- **Desiderat:** eine Korpusliteraturwissenschaft, die das Potenzial der Data Science für die Literaturwissenschaft nutzbar macht. Dafür braucht es **interoperable Korpora**.
- bestehendes Korpus-Datenset (KOLIMO Beta)
- Update zu KOLIMO+:
- Aufbereitung für die wissenschaftliche Community der Computational Literary Studies nach neuesten philologischen und datenwissenschaftlichen Standards

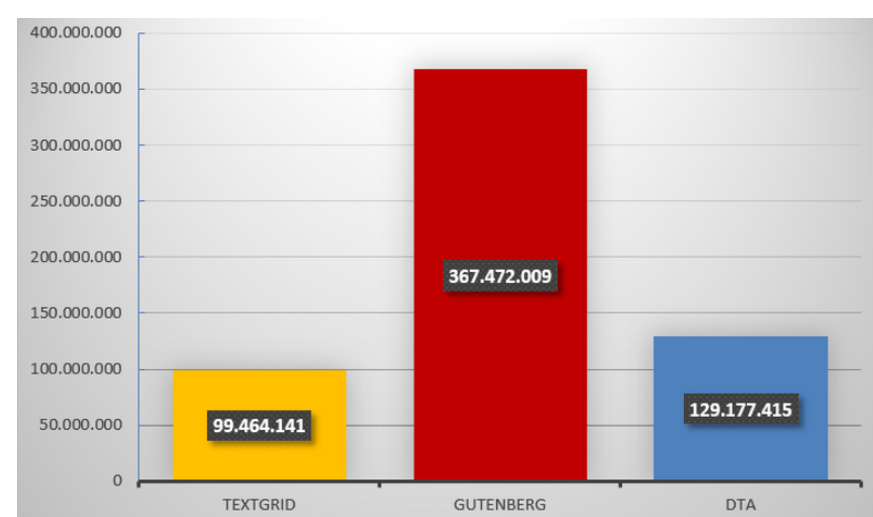
### DATEN

- Format: TEI-P5 (DTABf)
- Texte aus verschiedenen Repositorien
- ELTeC
- TextGrid Repository
- Deutsches Textarchiv
- Gutenberg-DE
- Einpflege neuer Metadaten und Primärtexte



### ZIELE

- Cleanup der Metadaten, XML-TEI-KOLIMO-Format (Version des DTA-Basisformats)
- Anreicherung mit Metadaten, z.B.: Raum, Zeitraum, Gattung, Gender, Literatur- vs. Gebrauchstexte uvm.
- Beta-Version des Korpus (eXist-Datenbank mit browserbasiertem GUI) soll zur Überführung in TextGrid aufbereitet und mit einem Webauftritt (mit GUI und API) versehen werden
- Interoperabilität durch Text+ und Reintegration in TextGrid
- FAIR-Prinzipien



KOLIMObeta umfasst 42.694 Texte von 2.029 verschiedenen Autor\*innen:

- 1.304 Texte aus dem DTA
- 27.410 Texte aus TextGrid und
- 2.400 Texte aus Gutenberg-DE

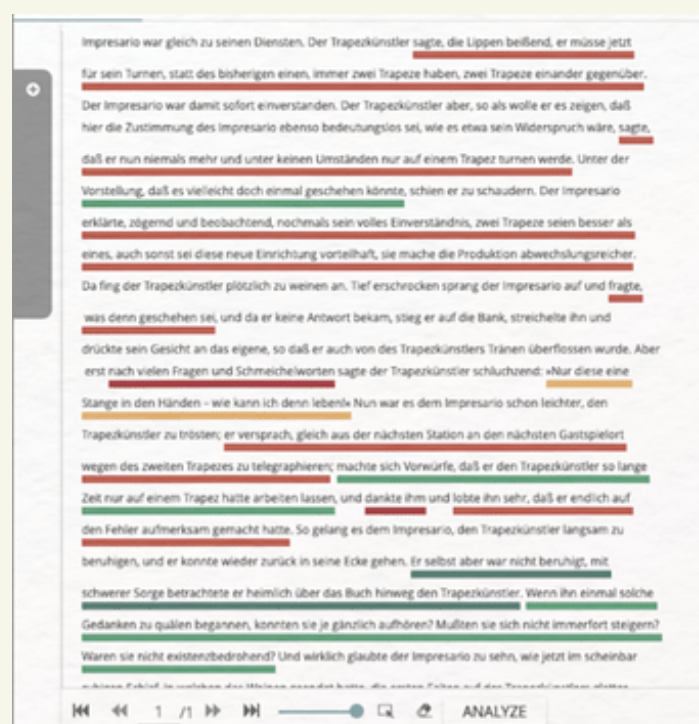
### FORSCHUNG



Wie kann man Texte an ihrer Oberfläche auf ihre **Literarizität** hin befragen? Wie **verändern** sich literarische Gattungen über die Zeit? Welche Merkmale **unterscheiden** fiktionale von nichtfiktionale Texten? Gibt es **Autor\*innenstil**, und wenn ja, wie viele? An welchen **Orten** spielen literarische Texte? Wie sind diese **emotional enkodiert**? Wie gut können **maschinelle Classifier** solche Kategorien unterscheiden?

### LEHRE

- curriculare Einbindung von Data Literacy in die interdisziplinäre Lehre
- Annotation und Analyse
- mixed methods
- Projekt: BiLinked Community of Practice *Data Literacy*
- Projekt: Curriculum 4.0. nrw



Beispielhafte Sentimentanalyse durch Annotation mit CATMA  
<https://catma.de/>

### KOOPERATION

- Universität Bielefeld
- SUB Göttingen
- 3 Workshops des Projekts KOLIMO+ der Uni Bielefeld und des Text+ Datenzentrums 2023
- Dissemination der Ressource über Text+ und TextGridRepository

