Towards Textual Studies’ High Tide? Studying Research Epistemologies in Digital Humanities after the Machine Learning Turn

Dr. J. Berenike Herrmann (Basel)
Dr. des. Thomas Messerli (Basel)
Prof. Noah Bubenhofer (Winterthur)
Daniel Knuchel (Winterthur)

Digitization
- Vast amount of digital data & powerful text mining tools for Humanities scholars, e.g., sentiment analysis.
- Many tools were developed within engineering and commercial application.
- Machine Learning and Deep Learning can be effective even in the absence of any expert knowledge about the make-up of text.
- Need for reflection of Digital Humanities practices.

Overarching Research Questions
- How is digitization transforming the humanities and social sciences?
- What is the application benefit of digital tools in the textual humanities?
- Do they answer pertinent questions and offer new perspectives?
- Do they aid—or possibly obstruct—a reflective Humanities scholarship?

Research epistemologies (»Forschungslogiken«)
- Guiding schemata for scientific/scholarly action
- Mutual determination with regard to reality & humanity
- Showing in research practice
- Types of research questions, argumentation, interpretation, methodology
- Thought styles (»Denkstile«; Fleck, 1912, first 1935)
- Epistemic cultures (»Wissenskulturen«; Knorr-Cetina, 1999)
  - see also Bubenhofer & Dressen (2018)

Research Design

Epistemological research  Two Web 2.0 Case Studies

Sentiment Analysis

Theoretical Premises: What understanding of language and communication motivate the different approaches? What assumptions are made about the world, reality and humanity?

Operationalization: What, if any, linguistic and literary categories and theories are being operationalized? What covert and overt decisions inform existing analytical practices?

Discussion: What rationale is used to discuss and interpret results? How do interpretations of results tie in with the respective theoretical frameworks? What open questions are foregrounded? What role and importance are given to the interpretation of data within the overall research projects? How is subjectivity discussed?

Research Practice: What explicit and implicit rationale motivates the employment of particular research processes?

Evaluation Practices
- NLP and Psychological Sentiment Analysis (e.g., Buechel et al., 2017; Egloff et al., 2018; Jakobs, 2018; Liu, 2012)
- (Digital) Linguistics (e.g., Benamara et al., 2012; Biber & Zhang, 2008; Hunston & Thompson, 2000; Sandig, 1979)
- Aesthetic judgement (Bourdieu, 1984; Cassigner, 1932; Kant, 1790; Wolff, 1732)
- Literary Criticism (e.g., Hayderbrand & Winko, 1996; Reif, 2017)
- Distant Reading (e.g., Archer & Jockers, 2016; Piper & Portelance, 2016)
- Web2.0 Digital Lives: Writer-Reader (“prosumer”) (e.g., Kaplan & Haenlein, 2010; Social Digital Reading (e.g., Piper, 2012; Stein, 2010)

Methods
- Harvesting and preprocessing of two corpora
- Literary and linguistic methods: Corpus linguistics, distant reading, hermeneutic
- NLP: Sentiment and opinion mining tools
- Machine learning: Statistical and neuronal
- Cultural anthropology: Participatory observation, Meta-annotation (e.g., Giuss & Jacke, 2017), logbooks (Glaser, 2008)

Perspectives & Potential
Applied reflection of Digital Humanities practices, with new insights on:
- Research Epistemologies in two communities of textual DH
- Sentiment Analysis across textual DH
- The Role of Machine Learning in textual DH
- (E)valuative practices of Web 2.0 users
- Digital transformation of Humanities: Perils and potentials