

PROJECT DESCRIPTION – HERMA: AUTOMATED MODELLING OF HERMENEUTIC PROCESSES

Uta Gaidys, Evelyn Gius, Margarete Jarchow, Gertraud Koch, Wolfgang Menzel, Dominik Orth, Heike Zinsmeister

Project description

In the project »Automated modelling of hermeneutic processes – The use of annotation for analyses in the health domain in social research and the humanities (hermA)«¹, the automation of annotations plays a key role in the context of hermeneutic text analyses. The project addresses the need for improvement of automated approaches that is aroused by the constantly growing number of digitally available texts and text formats. Key thesis and starting point of ›hermA‹ is that an advancement of knowledge in the automation of annotation is only possible in relation to epistemological and methodological questions. These methodological questions can best be studied in a differentiation of research methodologies into deductive, inductive and abductive approaches. With a collaboration across disciplines of the humanities and social sciences with computational linguistics and computer science, the aim of the project is therefore to identify and specify potential improvements of automated modelling for further research. Simultaneously, the options for a collaboration in the field of digital humanities across higher education institutions in Hamburg as a location for science are explored in ›hermA‹. The domain of health serves as an example.

Hermeneutic analyses are a central approach in the epistemologies of the social sciences and the humanities. One important challenge is that they are guided by different understandings of what hermeneutics means in detail and thus define the range and scope of its use in the epistemological processes differently². Annotation – depending on the research context and approach, often named differently such as for example coding or tagging – is a central work technique in all hermeneutic processes for signifying meaning or making references to historical and contemporary social or literary realities and by doing so making them accessible for scholarly analyses. In general, research in the humanities is done by using annotations for commenting, explicating, categorizing, interpreting and enriching texts with references.

1 hermA is funded by Hamburg's Ministry for Science and Research (Behörde für Wissenschaft, Forschung und Gleichstellung) from May 2017 to October 2020.

2 e.g. *Wilhelm Dilthey*: Die Entstehung der Hermeneutik. Stuttgart 1990; *Hans-Georg Gadamer*: Wahrheit und Methode. Grundzüge einer Philosophischen Hermeneutik. Tübingen 1972; *Hans-Georg Soeffner*: Auslegung des Alltags – der Alltag der Auslegung. Zur wissenschaftlichen Konzeption einer Sozialwissenschaftlichen Hermeneutik. Frankfurt am Main 1989; see for an overview *Kurt Eberhard*: Einführung in die Erkenntnis- und Wissenschaftstheorie: Geschichte und Praxis der konkurrierenden Erkenntniswege. Stuttgart 1999.

Annotations can be designed as standardized and unambiguous object categories³ or as a less unambiguous object interpretation⁴. Categories of annotation can be built iteratively in the context of annotation processes⁵ as for example in Grounded Theory⁶ or from already existing tag sets, taxonomies or ontologies. The concept of annotation denotes both the process of interpretative tagging as well as the representations emerging from this process⁷.

Texts are used as research data throughout all academic disciplines and are also the starting point of this research project. We have a broad spectrum of understandings of what a text is, also including non-standard languages in transcripts of spoken language, or social media data from the internet. This variety of texts poses some challenges for annotation and its automatization.

In hermeneutic processes, scholars analyse and annotate texts in a circular or iterative fashion, thus gaining an increasingly deeper understanding of meanings as well as of the process of analysis (see ill. 1)⁸. It is a strength of hermeneutic analysis that it can deal with ambiguity and vagueness in the text while most automatizations start with unambiguous, rule-governed annotations. Ambiguity and vagueness thus present a challenge to automated analysis that rather uses standardization as a means for the identification of (canonical) meaning. The affinity of manual annotations to ambiguity of meaning on the one hand and of automatized annotations to deterministic interpretations of meaning on the other hand indicate a relation that is relevant also for research processes: The more structured research methodologies are, the better automatized approaches of annotation will be applicable; and vice versa: the less standardized a research methodology is, the more difficult it is to apply standardized approaches for annotation. Therefore, automatization seems to be better applicable in deductive approaches with their orientation towards the application of a defined theory and their interest in the verification or evaluation of rules. Inductive research approaches, by contrast, require an openness for meanings created in a research group/

3 e.g. in linguistic corpora cf. *Geoffrey Leech/Andrew Wilson: EAGLES Guidelines: Recommendations for the Morphosyntactic Annotation of Corpora*. 1996 URL: <http://www.ilc.cnr.it/EAGLES/annotate/annotate.html> (date: 13. 11. 2017).

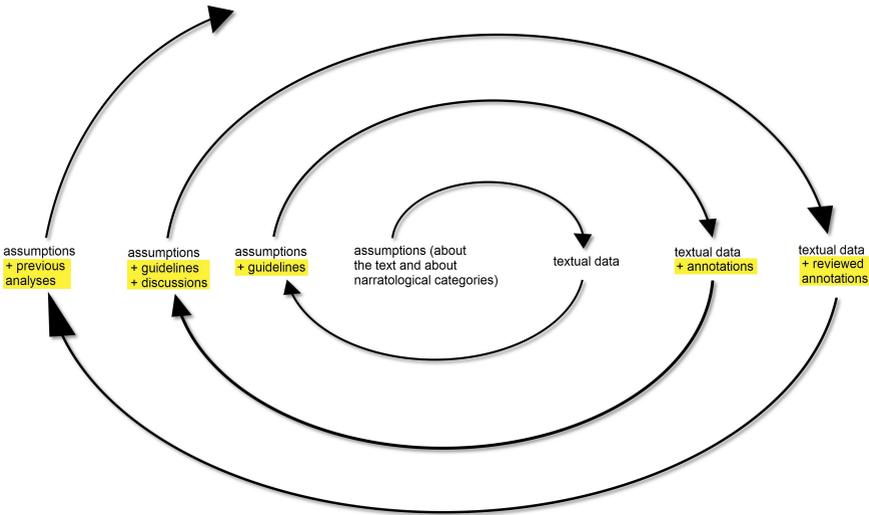
4 e.g. in social tagging cf. *Patricia Arnold/Lars Kilian/Anne Thillosen/Gerhard Zimmer* (eds.): *Handbuch E-Learning: Lehren und Lernen mit digitalen Medien*. Bielefeld 2013; *Herbert Frohner: Social Tagging: Grundlagen, Anwendungen, Auswirkungen auf Wissensorganisation und soziale Strukturen der User*. Boizenburg 2010.

5 *James Pustejovsky/Amber Stubbs: Natural Language Annotation for Machine Learning*. Beijing 2012, see p. 23–32).

6 *Barney G. Glaser/Anselm L. Strauss: The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York 1967.

7 *Geoffrey Leech: Introducing Corpus Annotation*. In: Roger Garside/Geoffrey Leech/Tony McEnery (ed.): *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London/New York 1997.

8 *Evelyn Gius/Janina Jacke: The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Text Analysis*. In: *International Journal of Humanities and Arts Computing* 11 (2017), no. 2, p. 233–54.



Ill. 1: Visualization of a hermeneutic circle for narratological analyses

research field and abductive research designs demand maximum flexibility in order to understand the emergence and constitution of new phenomena. Inductive and abductive research approaches thus show a smaller degree of standardization and are mostly guided by more than one theoretical reference. In our understanding, this has consequences for the potential of automatizing annotations. Open, only vaguely specified contexts of research have little affinity to standardized approaches of annotation. They demand alternative approaches, which have so far not been defined but need to be developed, for example in reference to existing methodologies such as learning ontologies from text⁹. However, even in domains with highly specified categories, the automatization of annotation is still a challenging endeavour with many open questions. There is a fundamental tension between manual hermeneutic annotation and automatized computational approaches in deductive, inductive, and abductive research designs that also has epistemological consequences. Both the tension and the consequences are hardly reflected¹⁰, which is another need that will be addressed in this project. The diversity of epistemological goals present in deductive, inductive, and abductive research designs demands a reflexive approach towards the potential

9 Mehrnoush Shamsfard/Ahmad Abdollahzadeh Barforoush: Learning Ontologies from Natural Language Texts. In: International Journal of Human-Computer Studies 60 (2004), no. 1, pp. 17–63.

10 Stephen Ramsay: Algorithmic Criticism. In: Ray Siemens/Susan Schreibman: A Companion to Digital Literary Studies. Malden 2007, pp. 477–91, URL: <http://www.digitalhumanities.org/companionDLS/> (date: 13.11.2017).

gains and losses brought to the hermeneutic process by the automatization. With respect to the epistemologies of the academic production of knowledge on the large scale and also for the specific use cases in hermA, this includes questions on how, where in the research process and by which means (tools, approaches) an automatization of the diverse research approaches – in the breadth of its variety in the humanities – is helpful in terms of efficiency and quality, and what the new qualities emerging from the automatized parts in the research process are.

Project design

According to our hypotheses of diverging potentials of automatization in deductive, inductive and abductive approaches, the research design of the project is structured and applied in five sub-projects (see below). This structure is an initial way to differentiate the variety of hermeneutic approaches in social sciences and the humanities, and thus includes epistemological dimensions of the humanities in a more sophisticated way in the digital approaches. We see it as a starting point for further differentiation rather than claiming for this to be the one and only way to enhance approaches of automatization.

A central aim for the work in hermA is the identification of starting points for methodological reflection and enhancement of hermeneutic processes in the intersection of humanities and computational knowledge production. Moreover, the project explores options for future technological developments. Use cases form the basis for a realistic exploration of future options and thus guarantee an orientation towards the needs of humanities research and a collaborative specification of these needs in the sense of feasibility studies, which are preliminary for future work. A comprehensive automatization of manual hermeneutic processes is not intended; based on the current state of the art this would be an unrealistic intention. Instead, the intention is to work on the gap between established hermeneutic interpretation and automated approaches, in particular on the level of the decision process for one or the other, for either man-made or machine-made analyses, during the development of research designs. The interdisciplinary approach will facilitate a transfer of knowledge into the diverse disciplines and furthermore supports a broad understanding of the requirements from the humanities to the computer sciences.

The project subsumes the following five sub-projects:

- SP 1: Deductive Research Approach: Annotation and the exploration of genre patterns. Medical engineering in literary anti-utopias. Heads of project: Prof. Dr. Margarete Jarchow & Dr. Dominik Orth, Humanities, Hamburg University of Technology (TUHH)
- SP 2: Deductive Research Approach: Annotation of gender-specific presentation of diseases in literary works. Head of project: Dr. Evelyn Gius, Literary Studies, University of Hamburg (UHH)

- SP 3: Inductive Research Approach: Annotation of meaning representations in crisis-laden health care situations. Head of project: Prof. Dr. Uta Gaidys, Health Care Studies, Hamburg University of Applied Sciences (HAW)
- SP 4: Automatization potentials of hermeneutic processes in discourse ethnography dealing with acceptance challenges of telemedicine. Head of project: Prof. Dr. Gertraud Koch, European Ethnology/Cultural Anthropology, University of Hamburg (UHH)
- SP 5: Computational Linguistics Approach: Exploitation of imperfect manual and automated annotations for research in the social sciences and the humanities with a special focus on the health domain. Heads of project: Prof. Dr.-Ing. Wolfgang Menzel (Informatics) & Prof. Dr. Heike Zinsmeister (Computational Linguistics), University of Hamburg (UHH)

Co-speakers of the project are Prof. Dr. Gertraud Koch & Prof. Dr. Heike Zinsmeister



Website: <https://www.herma.uni-hamburg.de>

Contact:

Prof. Dr. Gertraud Koch
 c/o Institut für Volkskunde/
 Kulturanthropologie
 Universität Hamburg, Germany
 Edmund-Siemers-Allee 1 (West)
 20146 Hamburg
gertraud.koch@uni-hamburg.de
 @GertraudKoch

Prof. Dr. Heike Zinsmeister
 c/o Institut für Germanistik
 Universität Hamburg, Germany
 Überseering 35, Postfach #15
 22297 Hamburg
heike.zinsmeister@uni-hamburg.de