


## Annotation: A Uniting, but Multifaceted Practice. A Review of Nantke and Schlupkothen (2020)

Melanie Andresen <melanie\_dot\_andresen\_at\_ims\_dot\_uni-stuttgart\_dot\_de>, Universität Stuttgart 

### Abstract

The volume *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*, edited by Julia Nantke and Frederik Schlupkothen, assembles research papers that are united in their focus on annotation, but display a broad variety of possible understandings of and approaches to annotation.

Annotations are ubiquitous in research in the digital humanities and beyond and have been declared a “scholarly primitive” by Unsworth. By this, he refers to basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation [Unsworth 2000]. Julia Nantke and Frederik Schlupkothen have set out to further deepen our understanding of annotation and foster interdisciplinary exchange on what annotation is and can be. To that end, they organized an interdisciplinary conference held at the University of Wuppertal, Germany, in February 2019. The conference was called *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization* and its results were published in a volume by the same name [Nantke and Schlupkothen 2020].

1

The introduction to the volume by Nantke/Schlupkothen is in itself worth mentioning: to give a comprehensive overview of the 16 articles without reducing them to one aspect only, they provide several visualizations of the volume’s content using the articles’ keywords. A word cloud shows which keywords are the most popular, a similarity matrix displays which articles combine to thematic clusters and an edge bundling visualization gives easy access to information on which articles share which keywords. The latter unfolds its full usefulness only in the interactive version that can be found on the publisher’s website (De Gruyter): <https://cloud.newsletter.degruyter.com/annotations/>. This review, in contrast, must force the articles back into the linearity of the medium of text. The guiding principle of the following is to question both what the authors mean by the multifaceted term “annotation” and whether they approach it from a primarily theoretical or applied perspective.

2

First, some articles focus on analysis of annotations that have already been done by someone else and are now the object of study. Freedman extends the understanding of annotations to footnotes in academic texts and traces that type of textual expression back to its origins in the 17th century. For Bamert, annotations are handwritten marginal notes in books for which he coins the term *pen traces* (“Stiftspuren”). That term allows one to include non-verbal forms like underlinings and stresses the materiality of such annotations. Through the example of pen traces made by Thomas Mann found in books in his estate, Bamert shows that four types of knowledge need to be included in the analysis of such annotations. Most evident is the importance of knowledge of the text and the reader: readers might underline or comment on passages where they learn something from the text that they did not know before. However, the notes by Thomas Mann also document his critical stance towards the author of the text, who seems to willingly modify his interpretations to conform with National Socialist ideology. Therefore, Bamert argues, the knowledge of the author as well as of the historical context need to be involved in the interpretation of such annotations.

3

While in these examples, the annotations analyzed have been done previous to the research, most articles in the volume deal with annotations that are produced at some point in the research process. One focus, as indicated in the title, is on annotations in the context of digital editions. In a digital edition, many different types of annotations can arise.

4

Schlupkoth/Schmidt suggest a systematic distinction between commentaries (“Kommentare”) and explanatory notes (“Erläuterungen”), two terms that have often been used synonymously. While commentaries refer to the form of the text and editorial decisions related to it, explanatory notes offer additional information regarding the text’s content and thus ensure understandability. Fanta presents the edition of the estate of the author Robert Musil, which holds special challenges for the editors: the high number of documents are interrelated in many ways that echo Musil’s writing process. Such process is reflected in a high number of comments, revisions, and explicit references to other documents by the author himself, making it possible to trace how different preliminary works lead to later, more elaborate documents. Fanta shows how the edition attempts to capture all these relations in TEI/XML. Sciuto shares lessons learned from the digital edition of the work of the French philosopher d’Holbach. He discusses the problem of how to target different audiences with one edition and proposes a two-level annotation system for this purpose. Koolen/Boot approach the topic from a more technical and infrastructural angle: for many editions, it would be desirable to allow for annotation by third-party users on the edition’s website and, where useful, publish them for other users. They discuss difficulties that result from the HTML presentation of the edition and propose a solution based on semantic web technology.

Five further articles consider annotation in a different sense as the application of analytical categories to an object of study by researchers. Lang reports on the semi-automatic annotation of alchemical code names (“Decknamen”) with a thesaurus. Alchemical texts often refer to substances via their code names which are usually unfamiliar to modern readers and should therefore be accompanied by additional information. Lück aims at an analysis of philosophical texts with respect to the examples used and, as a first step, needs to identify examples in the texts. He exploits the fact that many examples are explicitly marked as such (with text strings like *e.g.* or *for instance*) and uses these as a starting point for the heuristic identification of further candidates. The contribution by Reiter/Willand/Gius is about the annotation of narrative levels in literature. Noteworthy is that they approached this task by adapting the format of a Shared Task from computational linguistics and asked the research community to contribute annotation guidelines, test them empirically in a competitive setting, and engage in discussion about them. In a second part of the Shared Task, the aim is the development of a system for the automatic detection of narrative levels. The article by Drummond/Wildfeuer stands out by the fact that they tackle multimodal annotation. They investigate contemporary American TV series with respect to gender differences in representation of characters. Annotation categories capture aspects like camera perspective, sound design, how active the characters are and many more. Overall, female characters are displayed as in weaker positions.

Another group of articles focuses on annotation as a process and tools for its support. McCarty shares his thoughts on general note making in the research process and considers them a cognitive tool not only for writing down thoughts but for developing them in the first place. He stresses how the choice for a specific tool – be it digital or analogue – impacts our thinking and calls for a thorough study of annotation practices and their embedding in context prior to software development. Lange contributes empirical data on the annotation of research articles by scholars. To this end, he analyses public annotations by researchers in the online journal eLifeScience.com, which allows readers to annotate with the help of the Hypothes.is plug-in. In comparison to what we know about private note making, he finds that most annotations were clearly written (or rewritten) with publication in mind. Horstmann follows up on questions of tool development by presenting the annotation tool CATMA (Computer Assisted Text Markup and Analysis). CATMA supports close reading practices like highlighting and manual annotation as well as quantitative analysis of its annotations and automatic annotation of some phenomena. This way, Horstmann argues, the tool can contribute to building bridges between traditional literary scholars and digital humanists.

Finally, some articles primarily address theoretical aspects of annotation. Hinzmann revisits the concept of the hermeneutic circle. Despite the term’s popularity, it is not precisely defined and used for many different understandings of this concept. Hinzmann suggests a differentiation of several circular relationships that are involved in annotation and interpretation: 1) The interplay of part and whole of a text or corpus in the sense that interpretation of a part is influenced by knowledge of the whole and vice versa, 2) the interplay of individual annotation categories, the category system as a whole, and its theoretical foundation, 3) the interplay of the object’s historicity and systematic categories of modern research, and 4) the interplay of induction (bottom-up) and deduction (top-down) in research processes. That

5

6

7

suggestion is to be welcomed and communication in and beyond the digital humanities could profit greatly from more explicitness as to which understanding of hermeneutic circle is being referred to. Franken/Koch/Zinsmeister look at the different functions that annotations can fulfill in a research process. To this end, they compare typical uses of annotations in computational linguistics and cultural anthropology. One major difference is that computational linguistics predominantly works with pre-defined category systems which are then applied to data while cultural anthropology follows methods of grounded theory and develops the annotation categories in the annotation process itself. Rehm offers various reflections on annotation from the perspective of computational linguistics and artificial intelligence. Among others, he provides a list of “dimensions” by which annotations can be systematically described (e. g. research question, annotators, guidelines, complexity and evaluation).

Overall, the volume gives a very broad and inspiring insight into what can be considered annotation, what role in the research process annotation can play, and in what contexts and with which tools annotation as a practice takes place today. With regard to Unsworth’s [2000] “scholarly primitive,” the multitude of phenomena and approaches sometimes makes it difficult to see the primitive that they are all supposed to share.<sup>[1]</sup> An abstraction of all understandings of annotation could result in a formula like:

*A adds information B to object C in mode D with purpose E.*

This formula provides a grid for a structured discussion of different types of annotations, depending on how the variables are filled. As for A, the role of the annotator is either fulfilled by the researcher (as in most papers in the volume) or by some other scholar or author that is the object of study (academic writers in Freedman, Thomas Mann in Bamert). Fanta combines both types as the edition involves annotations by Robert Musil, but also their representation in the edition. For automatic annotation, the computer can be considered another type of annotator, even though its application is, of course, driven by a researcher. For the type of information (B) encoded in annotations, possibilities are essentially unlimited and as free as research itself, as the volume impressively illustrates. Of course, some categories can be more easily mapped to an annotation scheme or are even standardized, while others are more difficult to pin down at text surface (see Franken/Koch/Zinsmeister). Note that while the text is the most common object of annotation (C) in the book and beyond, annotation can be applied to other research objects as well, like movies in Drummond/Wildfeuer. Also, the annotation of “text” holds its ambiguities, e.g. whether the text is considered in its materiality or not. The mode of annotation (D), in a technical sense, can be divided in analogue and digital approaches, the former constituting the vast majority for annotations created today (with the exception of personal note-taking, as in McCarty). A further differentiation can be made among the digital approaches by looking at annotation formats like XML (Fanta and others), HTML (Koolen/Boot) and TCF (Lück) and tools like CATMA (Horstmann) or ELAN (Drummond/Wildfeuer). Finally, annotations can be described by their purpose (E) which is also dependent on the addressee of the annotations. Purposes of annotations in the volume range from personal notes the author addresses to themselves (Bamert, McCarty), analytical categories that are subsequently analyzed (Lang, Hinzmann, Reiter/Willand/Gius, and others), and annotations that are directed at a third-party reader like other researchers (Lange) or the reader of an edition (Schlupkothén/Schmidt, Sciuto). Some annotations are primarily addressed to the computer that maps the annotations to specific forms of visualization (typically) on the screen.

In highly interdisciplinary contexts like the digital humanities, communication can be challenging. On an abstract level, we will oftentimes agree that we all “do annotation” and thus regard it a scholarly primitive. However, concrete practices differ substantially and what a computational linguist and a historian have in mind when they say “annotation” might have less in common than both of them expect. A description model like the simple formula proposed here enables systematic discussions about concepts of annotation. Grounded in a solid understanding of commonalities and differences, multiple approaches to annotation can benefit greatly from confrontation with each other. In this way, annotation as a shared practice can foster interdisciplinary exchange. The volume by Julia Nantke and Frederik Schlupkothén is an impressive example of how the focus on an abstract practice like annotation allows various research endeavors to find parallels in their work.

## Notes

[1] The discussion about what scholarly primitives are is still in progress and the answer is highly dependent on the purpose of their compilation [Palmer, Teffeau, and Pirmann 2009] [Blanke 2013].

## Works Cited

- Blanke 2013** Blanke, T. and M. Hedges. "Scholarly Primitives: Building Institutional Infrastructure for Humanities e-Science", *Future Generation Computer Systems* 29.2 (2013): 654-661. <https://doi.org/10.1016/j.future.2011.06.006>.
- Nantke and Schlupkothen 2020** Nantke, J. and F. Schlupkothen. *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*. De Gruyter, Berlin/New York (2020). <https://doi.org/10.1515/9783110689112>.
- Palmer, Teffeau, and Pirmann 2009** Palmer, C., Teffeau, L., and C. Pirmann. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Report commissioned by OCLC Research (2009). <https://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf>.
- Unsworth 2000** Unsworth, J. *Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?*, in *Symposium on Humanities Computing: Formal Methods, Experimental Practice*. King's College, London (2000). <https://johnunsworth.name/Kings.5-00/primitives.html>.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.