Jan Horstmann
Undogmatic Literary Annotation with CATMA

Manual, Semi-Automatic and Automated

Abstract: How can traditional literary scholars be introduced to digital methods? And how can the development of annotation tools be specifically oriented to theories and methods of Literary Studies like hermeneutics or scalable reading? Manual annotation is the easiest way to get started: adding comments, markings, underlining and links in the course of close reading may have a counterpart in the digital environment that comes with a number of advantages, such as collaborative work or the sustainability of annotation. The article introduces the manifold annotation modes of the web-based tool CATMA (Computer Assisted Text Markup and Analysis), which has been developed in Hamburg since 2008 against the background of hermeneutic-circular methods of text research and the method of scalable reading. With its 'undogmatic', stand-off-markup-based approach, CATMA offers all the freedom of traditional manual-analogue annotation and allows for multiple, overlapping and even taxonomically contradictory annotations by one or more users. CATMA's markup taxonomies (tagsets) are not limited to binary yes/no, right/wrong oppositions, but can also support the operationalization of semantically challenging literary concepts. In developing a tool for digital text research such as annotation, goals should include providing for an easy, low-threshold introduction to the method, supporting the unstructured and exploratory bottom-up approach characteristic of first-time text encounters and motivating users to apply functions unique to the digital environment. The users should be guided through a continuum of methods in digital text research which range from computer-supported, interactive-manual and 'close' to algorithm-based 'distant' reading.

Keywords: Levels of Annotation, Role of Interpretation, Vagueness, Commenting, Describing, Highlighting, Moderating, Questioning, Tagging, Ontologies, Tool, Literature, Methodology

Jan Horstmann, Leiter Digitales Labor, Forschungsverbund Marbach Weimar Wolfenbüttel, jan.horstmann@mww-forschung.de

1 Annotation as an introduction to the methods of digital text research

Methods developed within the framework of Digital Humanities or adapted from other disciplines for digitally supported investigation of (literary) texts are frequently characterized by their scarcely limited scope. Moretti (2000, 57) coined the term "distant reading" for the analysis of large and very large text corpora with the support of computational algorithms and digital tools. Conceptually, the greater distance to these text corpora and the quantifying perspective on text data enable new and paradigmatically different research questions than the ones common in traditional Literary Studies. Huge text corpora and quantifying perspectives on the one hand, small amounts of text or individual texts and the analytical proximity of close reading on the other characterize the broad field of contemporary Literary Studies. It is therefore not uncommon to ask whether the more traditional scholarship and Digital Literary Studies actually still belong to the same discipline, and if so, how this seemingly large gap could be bridged (cf. Trilcke and Fischer 2018 from a praxeological perspective).

One way of tackling this problem is to manually annotate individual texts or smaller text corpora in a digital working environment. Annotation is understood here as the manual or automatic addition of supplementary information usually in written form to a primary text or text corpus. In a broad use of the term, italicization or bold type can already be described as annotations, whereby in this form the appearance of the primary text itself is changed in order to open up new semantic dimensions. In methodological-praxeological terms, annotation is one of the most traditional ways of working in text research (cf. e.g. Moulin 2010). Ideally, in addition to manual annotation functions (close reading), the digital working environment also offers the possibility of taking a greater distance, i.e. integrating a quantifying and relational perspective, in order to smoothen and thus facilitate the methodological exchange between traditional and Digital Literary Studies within one and the same environment. The span between close and distant reading, which only supposedly requires a methodological paradigm shift, thus becomes recognizable as a continuum that can be explored in the course of a scalable reading (cf. Weitin and Werber 2017). In Digital Literary Studies, this not only poses new questions, but also allows old questions to be dealt with and new answers to these questions to be found (cf. Horstmann and Kleymann 2019).

The web-based tool CATMA[1] (https://catma.de [27.08.2020]) offers such a possibility from a conceptual point of view by providing a multitude of functions. Two of these (analysis and visualization) will in the following be addressed only in passing. Instead, I will focus on three distinct, yet related concepts which in that are fundamental to CATMA's three markup variants: "Highlight", "Comment" and "Annotation". The basic steps of creating tagsets, which can also be described as concept ontologies, and dimensions of collaborative annotation are also important to discuss in this context. The reflection of the possibilities for semi-automatic and automated annotation of texts in CATMA concludes the article.

2 Dissemination of digital routines, resources and tools in forTEXT

The DFG-funded research project forTEXT (https://fortext.net [27.08.2020]) at the Universität Hamburg is developing beginner-friendly method descriptions, self-learning units and teaching modules for university and school teaching. In addition, existing digital text collections and tools – from digitization and digital annotation to digitally supported interpretation and visualization of literature – are critically discussed. Tutorial videos, literary case analyses, further editable tagsets, libraries for secondary literature, a social media strategy (cf. Horstmann and Schumacher 2019), as well as a comprehensive glossary of basic terms from the field of Digital Humanities round off the dissemination model for routines, resources and tools for digitally supporting research and teaching projects. The main focus in forTEXT, however, is on the two most important activities of hermeneutic text indexing: text annotation and text analysis. forTEXT aims to open up direct, low-threshold and functional access to digital methods, especially for traditional Humanities scholars. Perhaps the biggest problem here is that computer-assisted methods have so far often only been superficially adapted to the needs and paradigms of Humanities work. This can easily give Humanities scholars the impression that the use of computer-assisted working methods necessarily demands a conceptual paradigm shift that calls into question the self-definition of, for example, Literary Studies as a hermeneutic discipline. forTEXT tries to counter this prejudice.

Therefore, an important aspect of the work in forTEXT has been the development of the sixth version of CATMA. This web-based collaborative annotation

---

1 CATMA stands for Computer Assisted Text Markup and Analysis. The tool is open access and open source.
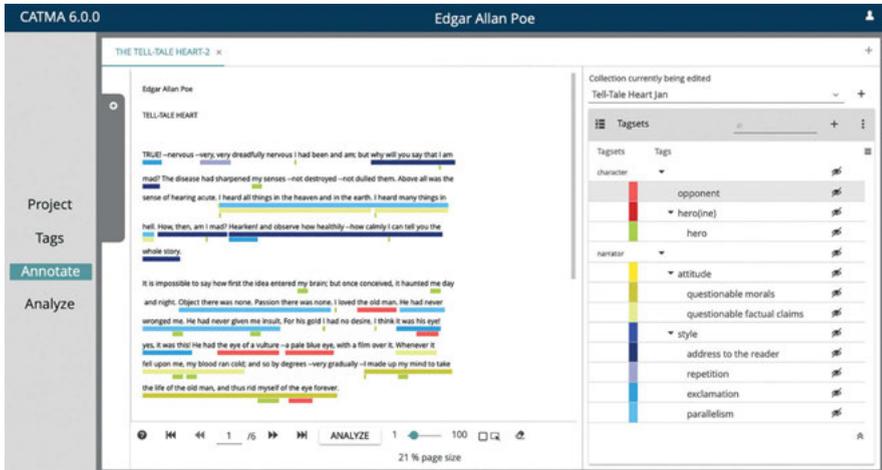
Fig. 1: The Annotate module in CATMA 6

and analysis tool has been developed at the Universität Hamburg since 2008 and currently has over 13,000 user accounts worldwide.[2] It is designed to support annotation and analysis routines that meet the specific requirements of literary-hermeneutical text research. The integration of text annotation, analysis and visualization within a web-based working environment against the background of a conceptual connection to theories of ('undogmatic') hermeneutic text research (more on this in sections 3 and 4) is unique in the field of DH tools (cf. Meister forthcoming). CATMA 6 provides a series of new functionalities[3] and a modernized, even more intuitive interface (cf. Fig. 1) based on Google's widely used and thus familiar Material Design. The backend system architecture of the tool was fundamentally re-designed as well and now offers – among other things – versioning of all the imported and created data, as well as a project-centered workflow. A CATMA

---

2 Of the current 13,033 accounts, 3030 were only used once and 1876 were guest accounts, so that 8127 users can be assumed (status: December 2019).

3 This includes, among other things, facilitating the sharing of documents, tagsets and annotations with other users, the versioning of documents, tagsets and annotations as well as the integration of a visualization concept based on the Humanities-oriented Vega Visualization Grammar according to the criteria of "Dynamic Data Visualization and Exploration for Digital Humanities Research" formulated in the project 3DH (http://threedh.net [27.08.2020]). The visualization language VEGA enables advanced users to parameterize and (re-)code their own data visualizations according to their needs and is based on the generic *Grammar of Graphics* developed by Wilkinson (2005). For epistemic dimensions of (digital) visualizations in humanistic work processes see Horstmann (2020).

project is a workspace in which users with different roles and permissions create and exchange text documents, tagsets and annotation collections. On loading a project this entire configuration becomes available at once.

3 Mixed methods, scalable reading and the hermeneutic circle

Since the mid-1980s, quantitative and qualitative methods in social and behavioral science have been joined by mixed-method approaches, which not only combine the two methods, but also go beyond them, because "the complexity of our research problems demands answers that contain more than just numbers in the quantitative sense and words in the qualitative sense" (Kuckartz 2014, 17).[4] Mixed-Method approaches have since become increasingly transdisciplinary and have also found their way into Digital Literary Studies, for example in the project "Reading at Scale. Mixed Methods in der literaturwissenschaftlichen Korpusanalyse"[5] at the Technische Universität Darmstadt.

The theory of scalable reading,[6] which is based on practical experiments, deals with the question of how individual text reading can be profitably combined with the analysis of larger and large amounts of text" (Weitin 2017, 1).[7] Underlying this metaphorical concept of scalability, which is based on semantic zooming, as interactive maps often allow, is the idea of a connection between distant and close reading in relation to a collection of texts. Weitin (2017, 2) points out, however, that the assumption of being able to switch continuously between the different approaches of qualitative analysis and quantitative reading or micro- and macroanalysis is misleading. The fact is, that most researchers were generally very familiar with the corpus they were investigating in quantifying regards (i.e. they also knew it on a micro level). Frequently, the results of a distant reading can only be made fruitful for an interpretation against the background of this well-founded knowledge of the texts. Therefore, if digital tools want to offer the possibility of interpretation and thus use the cognitive potential of scalable reading, they face concrete require-

---

4 In the original: "[d]ie Komplexität unserer Forschungsprobleme verlangt nach Antworten, die mehr als nur Zahlen im quantitativen und Worte im qualitativen Sinne beinhalten".

5 Cf. https://www.digitalhumanitiescooperation.de/projects/reading-at-scale/ (02.10.2019).

6 The term was originally coined by the classical philologist Martin Mueller (2012) in the form of a blog entry, where it was already firmly associated with Moretti's "distant reading".

7 In the original: "Frage, wie sich die Einzeltextlektüre mit der Analyse größerer und großer Textmengen gewinnbringend kombinieren lässt".

ments. On the one hand they should be able to generate or determine qualitative and quantitative data for heuristic and/or interpretational purposes. On the other hand, they should offer the possibility to put these data in a fruitful – and indeed scalable – exchange.

These constellations of integrated qualitative and quantitative – and one might add: relational – perspectives on a text or a collection of texts in the course of a process of understanding (in Literary Studies) can be mapped and traced back to a much older method, which forms one, if not *the* nexus of Literary Studies: hermeneutics. Hermeneutic text analysis and interpretation are frequently described as a circular movement. The so-called *hermeneutic* circle, which occupies a central position in the methodological research discourse of Literary Studies, is often traced back to a passage in Friedrich Ast's *Grundlinien der Grammatik, Hermeneutik und Kritik* (1808) in which he writes: "The foundational law of all understanding and knowledge is to find the spirit of the whole through the individual, and through the whole to grasp the individual; this the analytical, that the synthetic method of understanding" (Ast 1808, § 75, 178).[8] It was Schleiermacher (1838, 36–37) who described this integration of the individual and the whole in the cognitive process as a "circle", a metaphor that cannot yet be found in Ast, according to Danneberg (1995). While theoretical approaches such as those of Ast, Schleiermacher, Gadamer etc. outline hermeneutics as understanding and explaining, the term can be methodologically reduced to the understanding of texts in the practice of literary text analysis and be interpreted as an iterative "three-stage examination of texts in the form of description, analysis and interpretation", as Gius (forthcoming) does.[9] This constellation, which is fundamental for understanding texts, seems to correspond to the concept of scalable reading and to the metaphor of zooming in and out of texts or text corpora, on the basis of which an interpretation can then be formulated. Thus, both the centuries-old tradition of hermeneutics and the more recent approaches of mixed methods and scalable reading can be used to derive interrelated requirements for a digital working environment for literary scholars. These requirements should be met if one wishes to prevent a methodological paradigm shift and facilitate access to digital work – understood as supporting and expanding Literary Studies research. Annotation is a very good starting point for this process. While manual annotation is best

---

8 In the original: "Das Grundgesetz alles Verstehens und Erkennens ist, aus dem Einzelnen den Geist des Ganzen zu finden, und durch das Ganze das Einzelne zu begreifen; jenes die analytische, dieses die synthetische Methode der Erkenntniss".

9 In the original: "dreistufige Auseinandersetzung mit Texten in Form von Beschreibung, Analyse und Interpretation".

suited to guide the close reading, other forms of annotation (semi-automatic and automated) can productively support the process of zooming out.

4 Manual digital annotation in CATMA

4.1 Why undogmatic? – Special features of literary annotation

The traditional practice of annotating in the Humanities can now more or less be methodologically structured and thus related to the principles of hermeneutic text comprehension.[10] The tradition of digital annotation in Computational Linguistics is older than in Literary Studies. The reason for this seems to lie in different requirements when performing markup on literary texts. The concept of the 'undogmatic' can be used to sum up these requirements terminologically: in a literary annotation process one often does not want to make 'dogmatic', i.e. rigid, inflexible either-or decisions. Rather, it is at times a matter of acknowledging vagueness, polyvalence and uncertainty in the metadata in order to be able to represent the plausibility of an annotation as an interpretation of the text.[11] Determining ground truths, securing inter-annotator agreement or arriving at gold standards is not necessarily the prime objective in this field of practice, which is why a (literary) digital annotation tool needs to offer greater flexibility. In particular it should enable users to express and record contradiction and variance in annotations and stimulate a discourse about the respective literary artifact. There is no such thing as the one correct interpretation of a literary text, but the interpretation process itself is in principle incomplete: Literary Studies are a discursive discipline. This evokes a plurality of possible approaches to literary texts that can be classified in more detail with regard to at least three aspects.

1. Methodological variance: literary texts can be investigated in many ways, such as structurally, content-wise or content-transcending (cf. Shusterman 1978; Folde 2015[12]). Both the focus on content and the method of textual research often differ depending on the respective literary approach (cf. Danneberg 1999; Bühler 2003; Kindt and Köppe 2008).

---

10 For a general introduction to the method of manual digital annotation see Jacke (2018b).

11 For the concept of plausibility as a literary evaluation criterion see Winko (2015).

12 Relevant here is the distinction between content-specifying and content-transcending interpretations: interpretations can aim to identify non-explicit content elements (i.e. answer non-trivial questions about what is true in a fiction) or they can aim to highlight meaning beyond the content, e.g. transferred meanings or aesthetic functions.

2. Conceptual polyvalence: due to their ambiguity, literary texts can be understood differently even within one and the same approach (cf. Jannidis 2003; Bauer et al. 2010).[13]
3. Epistemic iteration: depending on the individual researcher, the workflow of text investigation can run cyclically through different methodological phases (cf. Delabar 2009, 63–83; Kocher and Krehl 2008, 99–137; Nünning and Nünning 2010, 10–21; Puhl et al. 2015, 42–46).

4.2 Annotation without category? – "Highlight" and "Comment"

The further development of CATMA toward its sixth version focuses on meeting these three procedural requirements of methodological variance, conceptual polyvalence and epistemic iteration in the Humanities: for instance, a comment function and the possibility of category-free annotation are being prepared for the new version, which enable unsystematized annotation that frequently appears with the first reading of a text. The orientation towards questions and workflows in the Humanities not only enables fruitful use in research, but also in both DH and Literary Studies teaching. Students not only get to know the method of digital annotation in practice, but also start to discuss concepts and taxonomies, which underlie the students' concrete research questions or even Literary Studies itself.

Many annotation tools (including CATMA 5 and earlier versions) only allow annotation with the aid of tagsets. In order to annotate a text in this way, however, researchers must already have at hand a formalized and structured category system with which they want to investigate the text. However, the digital annotation method should also be usable for unstructured text exploration. In CATMA 6, the two annotation modes "Highlight" and "Comment" are implemented for this purpose (cf. Fig. 2 on the next page).

Highlight: The highlight is an annotation that initially serves exclusively to draw special attention to an interesting text passage. Users can mark the annotated passage as relevant, even if they only have a vague or no interpretation hypothesis at this point.[14] By using the analysis functions provided in CATMA, highlighted passages can be searched for and displayed as a list. From here they can for example

---

13 However, Jannidis argues against a strong concept of polyvalence that excludes the possibility of arguing against the plausibility of interpretations.
14 The modelling of vagueness and uncertainty, which is important for applications in the Humanities, can also be graph-based, cf. also the discussions in Kuczera, Wübbena and Kollatz 2019.
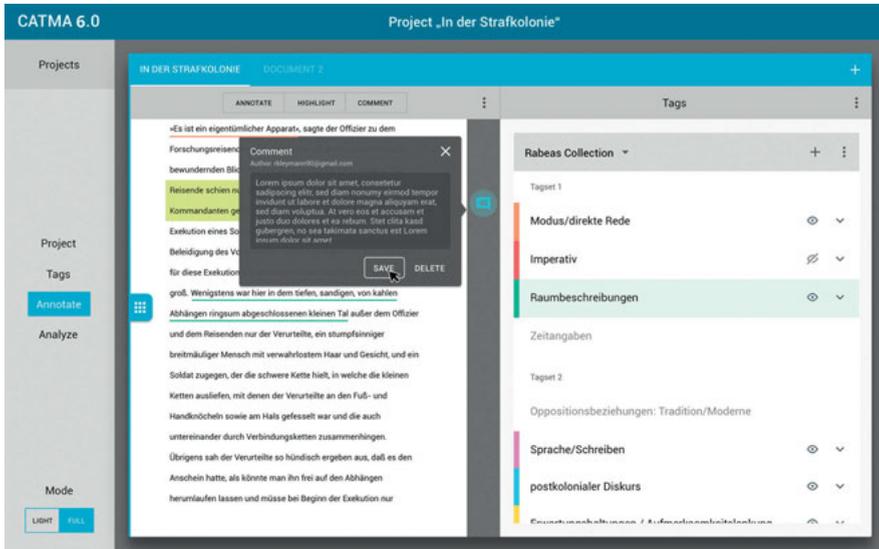
Fig. 2: The "Comment" function in CATMA 6 (wireframe)

also be annotated further with other annotation forms, once text analysis and interpretation are more advanced.

Comment: In comment mode, text passages can be commented freely. This annotation mode makes it possible to record thoughts on a text passage without using a structured concept repertoire – i.e. it also enables the modeling of vagueness. In future CATMA also wants to make it possible to simplify the creation of tagsets based on comments, for example by analyzing the comments (partially automated) and thus using them as hints to latent tag categories.

4.3 Taxonomy-based annotation – tagset creation and concept ontologies

The third and technically as well as conceptually most comprehensive annotation mode in CATMA is called *Annotation*. This means tag-based markup in CATMA, in which text passages are annotated using hierarchically structured concept ontologies. The creation of tagsets in CATMA is also undogmatic, i.e. categories can be freely selected and freely combined in their relationship to each other (see Fig. 3 on the following page). It is therefore not necessarily a prerequisite to make annotation categories hierarchically dependent on each other, if this for example contradicts one's own theoretical approach.
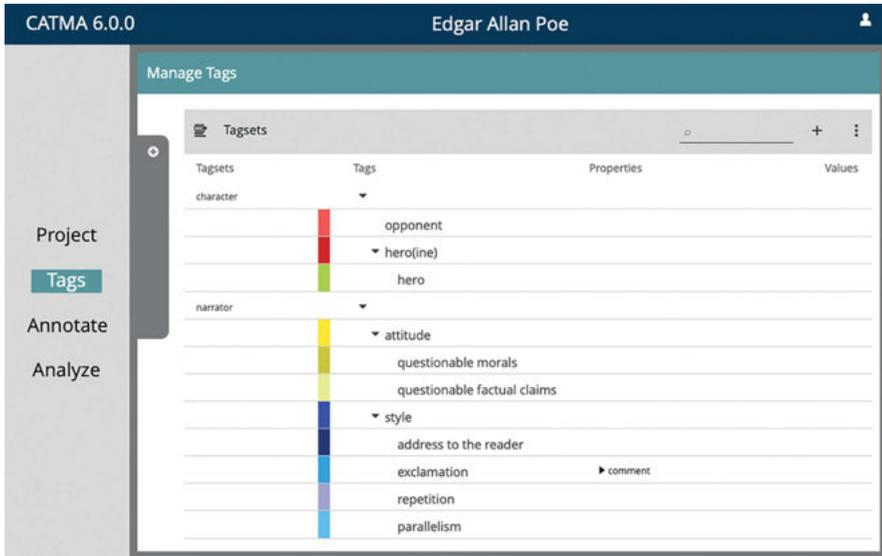
Fig. 3: Sample tagset in CATMA

The three annotation modes can interlock iteratively and thus also represent the so-called 'hermeneutic circle' of text investigation on this level. The tag-based form of annotation requires the most structuring and formalization. Tagsets created in CATMA can be changed continuously during the annotation process; the creation of such a concept ontology leads to very text-oriented work and requires the productive reflection of literary theories and methods.

To what extent is such tag-based annotation compatible with the dynamics of text investigation in Literary Studies outlined above? In order for structured annotation with tagsets to be used not only in the context of heuristic text *description*, but also to depict the scope of text *interpretation*,[15] a number of conditions must be met.

In addition to free generation and iterative revision of tagsets, one condition for the use of tagset-based annotation as an interpretation-supporting method is the possibility of multiple annotation of the same text passage (whereby the annotations may be different or even contradictory). On the one hand, this takes account of the fact that a text can be examined from different perspectives: for example, one and the same text passage can contain intermedial references, address gender issues and allow conclusions to be drawn about the author's communicative in-

---

15 For the distinction between description and interpretation see Kindt and Müller (2003).

tentions. A multidimensional categorization of the text passage must therefore be possible. Technically, this can be achieved by standoff markup: When uploading a text in CATMA, all characters (letters, punctuation marks, etc.) are numbered consecutively. An annotation is then stored in a annotation collection, where all annotations included in this collection contain the respective start and end digits (character offsets) of the annotated text passage, so that clear referencing is possible. This allows users to create any number of annotations, including overlapping ones. From a technical point of view, annotations are modeled according to the Web Annotation Data Model[16] and have the class dataset as body type. The structure is a list of key/multi-value pairs and the tag of the annotation specifies the possible keys.

On the other hand, passages of literary texts are often open to interpretation, which is why different, sometimes contradictory interpretations can be equally valid. For example, (incompatible) assertions about who/what is to be embodied by a character appearing in the text may all be plausible.[17] In CATMA 6, free tag generation and multiple annotations of a text passage are possible by modeling text passages, annotations and tags as nodes in a graph structure that allows very flexible linking options.[18]

Since the scope of literary interpretation is not limitless and must be executed according to various rules (cf. e.g. Jannidis 2003), an annotation environment that supports taxonomy-based interpretation also needs to offer possibilities for classifying, explaining and negotiating interpretations. This role is fulfilled in CATMA by meta-annotations, which can be used taxonomy-based (as properties and values) or as meta commentary. Annotation categories that are used in a tagset can be provided with properties to which fixed or ad hoc assignable values can be assigned in order to qualify annotations more precisely. The same function is performed by free-text-based meta-annotations. Whether meta-annotations are used as free comments or on a taxonomical basis can depend on (a) the degree of theoretical elaboration of the heuristics of interpretation that are used, (b) the context of investigation or (c) the personally preferred way of working.

While meta-annotations can be used to add analytical categories to a tagset on a horizontal level,[19] they can also be used to classify interpretative decisions.

---

16 Cf. https://www.w3.org/TR/annotation-model/ (27.08.2020)
17 See, for example, Føllesdal (1979), who presents different interpretations of the foreign passenger in Ibsen's *Peer Gynt*, of which the last two (embodiment of Lord Byron or the devil) are convincing.
18 Cf. http://tinkerpop.apache.org/ (27.08.2020)
19 For example, in the tagset for irony developed by Horstmann and Kleymann (2019) each tag could be supplemented with the properties "subject" and "object", in which the values "author"

For example, researchers could indicate which literary or interpretation theory (e.g. reception aesthetics, poststructuralism or hermeneutic intentionalism, cf. Köppe and Winko 2013) they used to make a particular (potentially controversial) interpretative decision. Likewise, contextual information that influences the interpretation can be listed, or interpretations can be located on a certain-uncertain scale.[20] Such meta-annotations help to understand hermeneutic annotations in the context of theoretical and subjective embedding; they make it possible, at least in rudimentary form, to give arguments for interpretative decisions and create the conditions for a literary debate on the plausibility of interpretative hypotheses. Meta-annotations are particularly necessary when text passages actually contain several apparently contradictory annotations. This is especially the case in the context of collaborative annotation and interpretation.

4.4 Dimensions of collaborative work in a digital environment

In the Humanities, collaborative practices are primarily used in the production of editions, introductions, etc., but less in the actual annotation and analysis of primary texts (cf. Schönert 1993; Hoppe et al. 2016; Lange 2005). Collaborative annotation – understood here as the joint work of several researchers on the same text with the same question, which requires an increased degree of coordination of the annotation process itself – is an established method in linguistics to secure high-quality, consistent annotation decisions (cf. Wissler et al. 2014). Collaborative work is not yet well established in Literary Studies (cf. Röcke 2016); a more cautious approach is also advisable when it comes to verifying annotation and interpretation decisions.[21] It makes sense to develop a tagset that is used jointly by all those involved in the annotation process and to support this tagset with a guideline that contains definitions for the individual tag categories as well as examples for their use. In the guideline-supported collaborative annotation project heureCLÉA (https://heureclea.de [27.08.2020]) an iterative approach has proven to be appropriate and fruitful, in which collaboratively annotating researchers first create their own

---

or "narrator" could be specified to determine from which instance the irony emanates and which instance is ironized in the annotated passages.

20 Cf. Genette (1982) in relation to the knowledge of other texts in the case of intertextual approaches. For the modeling of uncertainties related to the field of text visualization see Drucker 2011.

21 For an introduction to the method of collaborative literary annotation, its tradition lines and a discussion see Jacke (2018a).

annotations and only then consider the annotations of the other annotators.[22] In the comparison they can discuss discrepantly annotated passages in order to highlight reasons for different decisions (cf. Gius and Jacke 2017). A thorough meta-annotation can slim down this complex workflow considerably. Depending on the reason, it can be considered whether the respective disagreement should be marked as legitimate. In this way, the scope for interpretation in collaborative annotation can be preserved and meaningfully limited at the same time.

While collaborative annotation in CATMA was already possible from version 4 on, CATMA 6 additionally offers a complex role and rights system that allows for a differentiated definition of the group structure even in the conception of a collaborative annotation project. Based on a GitLab backend, CATMA projects (GitLab groups)[23] can be created that can have several project members and are equipped with text documents, tagsets and annotation collections. Different project contexts (e.g. academic research projects with several project leaders, employees and assistants, seminar projects in university teaching or teaching projects in school teaching) require the definition of different rights for the collaborators. For this reason, the following roles can be assigned in CATMA 6 for each project: project owner, partner, assistant, observer and student.[24] The roles are provided with fixed rights configurations in the fields of project and member administration as well as the creation, editing and deletion of text documents, tagsets and annotation data. By individualizing the cooperation mode, it is possible to determine how much access each project member should have to the data. This functionality can therefore also be understood as a measure to carefully restrict a policy of anything goes in text interpretation by defining certain rules.

5 Semi-automatic and automated annotation in CATMA

If the practice of annotation in the Humanities and thus close reading are successfully implemented in a digital working environment in the described way, traditional researchers and students can methodically proceed on the basis of the expertise unique to their disciplines. At the same time, however, they will also

---

22 Cf. the article of Gius, Reiter and Willand in the present volume for a collaborative annotation approach. In CATMA it is possible to switch annotation collections to "visible" or "invisible". Thus, even a collaboratively annotated document does not have to display all annotations at all times.

23 Cf. https://docs.gitlab.com/ee/user/group/ (27.08.2020)

24 The corresponding GitLab roles are owner, maintainer, assistant, reporter and guest.

Fig. 4: *KWIC* view in CATMA 6 for semi-automatic annotation (displaying sounds in Edgar Allan Poe's *The Tell-Tale Heart*)

become familiar with the machine logic of computers that do not – like fellow researchers or participants in a seminar – happily 'think along' and compensate for procedural gaps. This agnosticism of the program in relation to any semantics, conversely, demands the highest possible degree of accuracy on the part of the annotators: an effect that usually leads to productive frictions in terms of both content and method.

In CATMA the manual annotation of individual text passages can be supplemented by annotation procedures which, on the one hand, require a greater distance to the text and, on the other hand, also accelerate the annotation process – especially in the case of words or passages that are to be annotated repeatedly with the same category. Simulating zooming out of the text within the method of annotation lowers the threshold towards quantitatively oriented methods.

For this reason, semi-automatic annotation takes place within the *Analyze* module in CATMA. With one click, users can create a word list that displays all word types occurring in the text or text corpus sorted by frequency (see Fig. 4). From this word frequency list, keywords can now be collected that are to be annotated with a certain tag category (e.g. all verbs in the different past tenses can be annotated with the tag "past"). This KWIC list (keyword in context) can then be annotated with the selected tag or several tags at once. This way users can save a lot of time with repeated annotations.

According to the two-way screen postulate developed in the 3DH project (http://threedh.net [27.08.2020]) with regards to a dynamic data visualization and exploration for Digital Humanities research, CATMA 6 also offers the possibil-
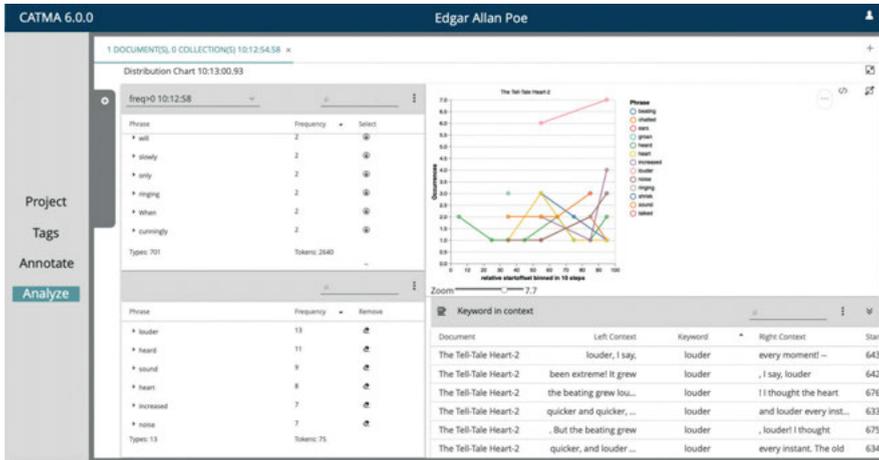
Fig. 5: An interactive distribution graph in CATMA for the support of semi-automatic annotation (displaying sounds in Edgar Allan Poe's *The Tell-Tale Heart*)

ity of using the implemented visualizations interactively. Thus, the visualizations allow jumping to concrete text passages and back or compiling KWIC lists for semi-automatic annotations (see Fig. 5). The annotation, analysis and visualization components in CATMA functionally interact and therefore conceptually reflect the discussed circular-hermeneutic work process of Humanities text research.

This type of a self-initiated semi-automatic annotation prepares the users for an even more distanced way of reading: fully automated annotation. At current CATMA supports three such routines: (1) the automatic annotation of grammatical tenses in German-language texts, (2) the automatic annotation of temporal signals in German-language texts and (3) the automatic annotation of part of speech also in German-language texts. The automated recognition of word types and grammatical tenses in particular are algorithms widely used in Computational Linguistics. It is planned to integrate more of these for other languages in CATMA in the near future. The automated annotation of time signals, however, is based on a more complex procedure. Guideline-based collaborative annotation in the already mentioned project heureCLÉA provided the conceptual framework for this: the (collaboratively) annotated medium-sized corpus of German-language short stories from around 1900 was used for a machine learning (cf. Bögel et al. 2015). The very dense annotation of the entire corpus – made intersubjectively comprehensible by meta-annotations in relation to (a) time signals and (b) the order categories by Gérard Genette (1972) – formed a comparatively small training corpus by machine learning standards. Nevertheless, the procedure led to unexpectedly

good results. As a consequence, it is now possible in CATMA to automatically annotate time signals in other German-speaking corpora. The implementation of the automated marking of order relationships is also planned. The more similar the texts are to German short stories from the nineteenth/twentieth century (i.e. the training corpus), the better the result will be. The algorithm was derived from a collaborative annotation process and a comprehensive machine learning routine, that was meticulously carried out and accompanied. Automation itself thus can take quite a lot of time. The big advantage, however, is that the algorithm can now be used very time-savingly, since 'only' annotations that the machine assigned incorrectly have to be manually corrected afterwards.

With this step taken, users have travelled the path from the conceptually familiar manual annotation of text passages, as it corresponds to more traditional ways of researching, over semi-automatic annotation of selected keywords in the entire text or in the entire corpus, to the fully automated annotation of selected categories – all in one and the same annotation environment. They have thus used and linked both methods of close and distant reading in a hermeneutic manner and followed a scalable reading procedure. Supplementing text research with additional Digital Humanities methods from the distant reading repertoire will hopefully now meet fewer reservations on the humanist side than may have been the case before.

Bibliography

Ast, Friedrich. Grundlinien der Grammatik, Hermeneutik und Kritik. Thomann, Landshut. 1808. URN: urn:nbn:de:bvb:12-bsb10582791-6.

Bauer, Matthias, Joachim Knape, Peter Koch, and Susanne Winkler. Dimensionen der Ambiguität. In: Zeitschrift für Literaturwissenschaft und Linguistik 40 (158). Stuttgart, Weimar: Verlag J.B. Metzler. 2010, pp. 7–75.

Bögel, Thomas, Michael Gertz, Evelyn Gius, Janina Jacke, Jan Christoph Meister, Marco Petris, and Jannik Strötgen. Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative. In: DHCommons Journal 1 (July). 2015. DOI: 10.5281/zenodo. 3240591.

Bühler, Axel. Die Vielfalt des Interpretierens. In: Axel Bühler (Ed.), Hermeneutik. Basistexte zur Einführung in die wissenschaftstheoretischen Grundlagen von Verstehen und Interpretation. Heidelberg: Synchron Wissenschaftsverlag der Autoren. 2003, pp. 99–120.

Danneberg, Lutz. Die Historiographie des Hermeneutischen Zirkels. Fake und Fiction eines Behauptungsdiskurses. In: Zeitschrift für Germanistik, Neue Folge 5.3. Bern: Peter Lang AG. 1995, S. 611–624.

Danneberg, Lutz. Zum Autorkonstrukt und zu einem methodologischen Konzept der Autor-intention. In: Fotis Jannidis, Gerhard Lauer, Matías Martínez, and Simone Winko (Eds.),

Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs. Tübingen: Niemeyer. 1999, pp. 77–105.

Delabar, Walter. Literaturwissenschaftliche Arbeitstechniken. Eine Einführung. Darmstadt: Wiss. Buchges. 2009.

Drucker, Johanna. Humanities Approaches to Graphical Display. Digital Humanities Quarterly 5 (1). 2011. URL: http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html (04.10.2019).

Folde, Christian. Grounding Interpretation. In: British Journal of Aesthetics 55 (3). 2015, pp. 361–374. DOI: 10.1093/aesthj/ayv020.

Føllesdal, Dagfinn. Hermeneutics and the Hypothetico-Deductive Method. Dialectica 33. 1979, pp. 319–336. DOI: 10.1111/j.1746-8361.1979.tb00759.x.

Genette, Gérard. Discours du récit. Figures III. Paris: Editions du Seuil. 1972, pp. 67–282.

Genette, Gérard. Palimpsestes: La littérature au second degré. Paris: Editions du Seuil. 1982.

Gius, Evelyn. Digitale Hermeneutik: Computergestütztes Close Reading als Literaturwissenschaftliches Forschungsparadigma? In: Fotis Jannidis, Simone Winko, Andrea Rapp, Jan Christoph Meister, and Thomas Stäcker (Eds.), Digitale Literaturwissenschaft. DFG-Symposium Villa Vigoni. Berlin, New York: De Gruyter. Forthcoming.

Gius, Evelyn and Janina Jacke. The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Analysis. In: International Journal of Humanities and Arts Computing 11 (2). Edinburgh: Edinburgh University Press. 2017, pp. 233–254. DOI: 10.3366/ijhac.2017.0194.

Hoppe, Vinzenz, Marcel Lepper, and Stefanie Stockhorst (Eds.). Symphilologie. Formen der Kooperation in den Geisteswissenschaften. Göttingen: V & R unipress. 2016.

Horstmann, Jan. Textvisualisierung: Epistemik des Bildlichen im Digitalen. In: Martin Huber, Sybille Krämer, and Claus Pias (Eds.), Wovon sprechen wir, wenn wir von Digitalisierung sprechen? Gehalte und Revisionen zentraler Begriffe des Digitalen. Contributions to the symposion "Digitalität in den Geisteswissenschaften" at the Universität Bayreuth, February 2019. Frankfurt a. M.: Universitätsbibliothek Johann Christian Senckenberg. 2020. URN: urn:nbn:de:hebis:30:3-548619.

Horstmann, Jan and Rabea Kleymann. Alte Fragen, neue Methoden – Philologische und digitale Verfahren im Dialog. Ein Beitrag zum Forschungsdiskurs um Entsagung und Ironie bei Goethe. In: Zeitschrift für digitale Geisteswissenschaften. Wolfenbüttel. 2019. DOI: 10.17175/2019_007.

Horstmann, Jan and Mareike Schumacher. Social Media, YouTube und Co: Multimediale, multimodale und multicodierte Dissemination von Forschungsmethoden in forTEXT. In: Patrick Sahle (Ed.), DHd 2019. Digital Humanities: multimedial & multimodal. Konferenzabstracts. 2019, pp. 207–211. DOI: 10.5281/zenodo.2596095.

Jacke, Janina. Kollaboratives literaturwissenschaftliches Annotieren. forTEXT. Literatur digital erforschen. 2018a. URL: https://fortext.net/routinen/methoden/kollaboratives-literaturwissenschaftliches-annotieren (04.10.2019).

Jacke, Janina. Manuelle Annotation. forTEXT. Literatur digital erforschen. 2018b. URL: https://fortext.net/routinen/methoden/manuelle-annotation (04.10.2019).

Jannidis, Fotis. Polyvalenz – Konvention – Autonomie. In: Fotis Jannidis, Gerhard Lauer, Matías Martínez, and Simone Winko (Eds.), Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte. Berlin, New York: De Gruyter. 2003, pp. 305–328.

Kindt, Tom and Tilmann Köppe (Eds.). Moderne Interpretationstheorien. Ein Reader. Göttingen: V & R unipress. 2008.

Kocher, Ursula and Carolin Krehl. Literaturwissenschaft. Studium – Wissenschaft – Beruf. Berlin: Akademie-Verlag. 2008.

Köppe, Tilmann and Simone Winko. Neuere Literaturtheorien. Stuttgart: Metzler. 2013.

Kuckartz, Udo. Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren. Wiesbaden: Springer. 2014. DOI: 10.1007/978-3-531-93267-5.

Kuczera, Andreas, Thorsten Wübbena, and Thomas Kollatz (Eds.). Die Modellierung des Zweifels. Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. In: Zeitschrift für digitale Geisteswissenschaften, Sonderband 4. 2019. DOI: 10.17175/sb004.

Lange, Tanja. Vernetzte Wissenschaft? Zu Perspektiven computergestützter Kollaboration für Forschung und Lehre in den Geisteswissenschaften. In: Harro Segeberg and Simone Winko (Eds.), Digitalität und Literalität. Zur Zukunft der Literatur. Munich: Fink. 2005, pp. 271–294.

Meister, Jan Christoph. From TACT to CATMA or a Mindful Approach to Text Annotation and Analysis. In: Geoffrey Rockwell and Stéfan Sinclair (Eds.), Festschrift for John Bradley. Forthcoming.

Moretti, Franco. Conjectures on World Literature. In: New Left Review 1. 2000, pp. 54–68. URL: https://newleftreview.org/issues/II1/articles/franco-moretti-conjectures-on-world-literature (07.01.2020).

Moulin, Claudine. Am Rande der Blätter. Gebrauchsspuren, Glossen und Annotationen in Handschriften und Büchern aus kulturhistorischer Perspektive. In: Quarto. Zeitschrift des Schweizerischen Literaturarchivs 30/31. Genf: Slatkine. 2010, pp. 19–26.

Müller, Martin. Scalable Reading. Scalable Reading (blog). May 29, 2012. URL: https://scalablereading.northwestern.edu/?page_id=22 (02.10.2019).

Nünning, Vera and Ansgar Nünning (Eds.). Methoden der literatur- und kulturwissenschaftlichen Textanalyse: Ansätze – Grundlagen – Modellanalysen. Stuttgart, Weimar: Metzler. 2010.

Puhl, Johanna, Peter Andorfer, Mareike Höckendorff, Stefan Schmunk, Juliane Stiller, and Klaus Thoden. Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften. In: DARIAH-DE Working Papers No. 11. Göttingen: DARIAH-DE. 2015. URN: urn:nbn:de:gbv:7-dariah-2015-4-4.

Röcke, Werner. Geleitwort. In: Stefanie Stockhorst, Marcel Lepper, and Vinzenz Hoppe (Eds.), Symphilologie. Formen der Kooperation in den Geisteswissenschaften. Göttingen: V & R unipress. 2016, p. 7.

Schleiermacher, Friedrich. Hermeneutik und Kritik. Berlin: Reimer. 1838. URN: urn:nbn:de:kobv:b4-200905197348.

Schönert, Jörg. Konstellationen und Perspektiven kooperativer Forschung. In: Peter J. Brenner (Ed.), Geist, Geld und Wissenschaft. Frankfurt a. M.: Suhrkamp. 1993, pp. 384–408.

Shusterman, Richard. The Logic of Interpretation. In: The Philosophical Quarterly 28 (113). 1978, pp. 310–324. DOI: 10.2307/2219083.

Trilcke, Peer and Frank Fischer. Literaturwissenschaft als Hackathon. Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen. In: Martin Huber and Sybille Krämer (Eds.), Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden. Sonderband der Zeitschrift für digitale Geisteswissenschaften, 3. 2018. DOI: 10.17175/sb003_003.

Weitin, Thomas. Scalable Reading. In: Thomas Weitin and Niels Werber (Eds.), Scalable Reading, Zeitschrift für Literaturwissenschaft und Linguistik 47 (1). 2017, pp. 1–6. DOI: 10.1007/s41244-017-0048-4.

Weitin, Thomas and Niels Werber (Eds.). Scalable Reading, Zeitschrift für Literaturwissenschaft und Linguistik 47 (1). Stuttgart: Metzler. 2017.

Wilkinson, Leland. The Grammar of Graphics. Berlin, Heidelberg: Springer. 2005.

Winko, Simone. Zur Plausibilität als Beurteilungskriterium literaturwissenschaftlicher Interpretationen. In: Andrea Albrecht, Lutz Danneberg, Olav Krämer, and Carlos Spoerhase (Eds.), Theorien, Methoden und Praktiken des Interpretierens. Berlin, Boston: De Gruyter. 2015, pp. 483–511.

Wissler, Lars, Mohammed Almashraee, Dagmar Monett, and Adrian Paschke. The Gold Standard in Corpus Annotation. Proceedings of the 5th IEEE Germany Student Conference. 2014. DOI: 10.13140/2.1.4316.3523.