

Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristics of Narrative.

Project Statement

Thomas Bögel¹, Michael Gertz¹, Evelyn Gius², Janina Jacke², Jan Christoph Meister², Marco Petris², Jannik Strötgen¹

¹ Heidelberg University

² University of Hamburg

Introduction	2
1. heureCLÉA as a "Computational Narratology" Project	3
2. Conceptual and Methodological Premises of the heureCLÉA Project	3
2.1. Annotation and Text Analysis in the Humanities	3
2.1.1. Markup, Annotation, Interpretation as Knowledge Productive Humanistic Practices	3
2.1.2. Annotation in CATMA and heureCLÉA	10
2.2. Relevant Perspectives on Time Phenomena	12
2.2.1. Time Phenomena in a Narratological Perspective: from Surface Markers to Inference-based Constructs	12
2.2.2. Time Phenomena in a Linguistic and Computational Perspective: Extraction and Annotation of Time-related Expressions in Documents	14
2.3. The Role of Standards	15
3. Design and Current Outcomes of the heureCLÉA Project	17
3.1. Narratological Analysis of Temporal Phenomena: Current Outcomes of the Manual Annotation with CATMA	19
3.1.1. Concept Vagueness Resolved by Additional Surface Markers	21
3.1.2. Concept Vagueness resulting from Basic Narratological Problems	22
3.1.3. Methodological Consequences for the heureCLÉA Project	23
3.2. Automation of Complex Time Annotation: Current Outcomes	24
3.2.1. Temporal Signals	25
3.3. Implementation and Integration of the Heuristic heureCLÉA Module into CATMA: Current Outcomes	26
4. Conclusion	27
References	29
Standards	34
heureCLÉA Corpus	35

Introduction

This paper is about heureCLÉA¹, an interdisciplinary project for the development of a "digital heuristic" that can support research in narratives by automatically identifying narratologically salient features in textual narratives.² This heuristic will be integrated in the CATMA (*Computer Aided Textual Markup and Analysis*)³ working environment as a program module and will provide an automatic annotation functionality that complements the manual annotation functionality already provided by CATMA. While the current project focuses on a specific field of application - i.e., narratological text analysis and markup - heureCLÉA's methodological approach aims at bridging manual and automated procedures in a more general sense, thus making it relevant to other digital humanities oriented markup projects.

In heureCLÉA we are currently analyzing and annotating a corpus of 21 short fictional narratives in German language that were written mostly around 1900.⁴ These texts are investigated in a plain text format and were selected because of their—relatively—clear structure in terms of narrative levels. Most of the sample texts originate from the TextGrid repository.⁵

The work within heureCLÉA is an interdisciplinary collaboration between the University of Hamburg (Literary Studies) and Heidelberg University (Computer Science) and is supported by the German Federal Ministry of Education and Research (BMBF)⁶. The project team consists of two project leaders, five research assistants and five to ten student assistants. heureCLÉA started in February 2013 and will run until January 2016.

In this project statement we aim at presenting the more conceptual and methodological aspects of heureCLÉA as well as the practical aspects of the methods we use, including their outcomes at the current stage.

In section 1 we contextualize our project within the field of Computational Narratology. Section 2 discusses the conceptual premises of heureCLÉA in terms of annotation in the humanities, perspectives on time phenomena, and the role of standards. The discussion of annotation in the humanities is subdivided into the two aspects of (i) annotation in humanistic text analysis and (ii) its implementation in CATMA and heureCLÉA. The second part of this section is dedicated to time phenomena as they are discussed in the disciplines concerned, i.e., time phenomena in (i) a narratological perspective and in (ii) a linguistic and computational perspective. The third part of section 2 discusses the role of standards in our

¹ cf. the project page <http://www.heureclea.de> (accessed 14.08.2014).

² Narratology is a branch of narrative theory rooted in formalist and structuralist theories of literature. The so-called 'classical' variants of narratology focus on the taxonomic description of those structural and discursive features that are particular to the narrative encoding of information. Narratology's methodological orientation is therefore heuristic rather than hermeneutic; its intent is to render inter-subjectively verifiable object descriptions rather than holistic interpretations of narratives. - For an overview of the discipline of narratology, see Meister (2014).

³ For CATMA, see <http://www.catma.de> (accessed 14.08.2014). CATMA was initially developed as a reimplementations of TACT (Textual Analysis Computing Tools), a 1980s DOS-based suite of markup and text analysis tools programmed by John Bradley—some of them interactive, others operating in batch mode—which many DH researchers consider a pioneering feat. For more information on TACT, see <http://projects.chass.utoronto.ca/tact/TACT/tact0.html> (accessed 14.08.2014).

⁴ For a list of the texts used in heureCLÉA, see "heureCLÉA Corpus" at the end of this paper.

⁵ <http://www.textgridrep.de/repository.html> (accessed 14.08.2014).

⁶ BMBF grant no. 01UG1352A, "Verbundprojekt heureCLÉA"

project. Section 3 describes the overall design of the project and presents the current outcomes of (i) the manual narratological analysis of time phenomena, (ii) the automation of complex time annotation and (iii) the integration of the heuristic module into CATMA. In the concluding section 4 we reflect on the relevance of our outcomes in the digital humanities context.

1. heureCLÉA as a "Computational Narratology" Project

In his recent definition of *Computational Narratology* as research in "the algorithmic processes involved in creating and interpreting narratives, modeling narrative structure in terms of formal, computable representations", Mani (2013, para. 1) lists three main areas of related activities: (a) "approaches to storytelling in artificial intelligence systems and computer (and video) games", (b) "automatic interpretation and generation of stories", and (c) "exploration and testing of literary hypotheses through mining of narrative structure from corpora". heureCLÉA falls mainly into the latter category. However, since we regard "markup" and "annotation" to be essentially interpretive (and not just declarative) procedures, it also overlaps with the aim of an "automatic interpretation of narrative". In this vein the combination of manual and automated procedures is the distinguishing feature of heureCLÉA and its software components.

Our approach is bottom-up and based on a case study, and the practical research domain that serves as the exemplary environment is that of narrative studies. This inductive approach has two major consequences. First, the specific needs and questions arising in narrative studies and narratology are the actual drivers of the heureCLÉA project. Second, the hypotheses and models generated in the context of traditional narratological research are implicitly put to the test in heureCLÉA. Our starting point is therefore the "real life" narratological taxonomy developed in the structuralist tradition, which is today widely used in narrative studies; however, our conceptual goal is the validation of narratological theorems and concepts. Developing, testing, and optimizing CATMA as a tool and computational environment is what happens in between: heureCLÉA's ultimate focus is therefore not technological, but methodological.

2. Conceptual and Methodological Premises of the heureCLÉA Project

2.1. Annotation and Text Analysis in the Humanities

2.1.1. Markup, Annotation, Interpretation as Knowledge Productive Humanistic Practices

Beyond offering pragmatic affordances, the expansion from the traditional towards a computationally supported paradigm of research tools and practices in the humanities has triggered some methodological transformations that affect the humanistic disciplines in fundamental and methodological ways.⁷ In our view, thus far the most fundamental

⁷ See McCarty (1996) on the "promise of the new", which is often thought to be purely quantitative ("the new thing will do the old job faster, more efficiently, and more cheaply"), whereas in the long run

methodological shifts caused by the evolving "digital humanities" are the following four:

(1) A gradual *epistemological reconceptualization* of the objects that fall in the respective domain of interest. Traditionally, such objects and domains were conceptualized and interpreted by the humanities in a predominantly phenomenological fashion, that is: from an anthropocentric, historically-hermeneutic point of view. However, as a result of the digital turn, these objects may now equally well be approached in a more analytical and less hermeneutically inspired approach. They can also be modeled in a more objective manner as highly complex, multi-factorial and dynamic constructs.

This first shift is mainly owed to the digitization of representational media that has rendered hitherto opaque and holistically experienced humanistic objects of study in a more readily quantifiable and, hence, in a computable format.

(2) The *propagation of an empiricist humanistic research practice*. In this regard not all humanities disciplines have been affected to the same degree: the traditional practices of hitherto more speculatively oriented disciplines, such as, e.g., literary studies, have been more significantly put into question by the digital turn than that of, say, corpus linguistics. Yet despite this caveat, "big data" and "distant reading" based approaches have clearly extended the humanities' methodological repertoire and focus from the exemplary to the statistically relevant, from the normative singular case to the representative sample.

(3) A *renewed interest in the sociological dimension of interpretive practices*. Whereas the traditional hermeneutic paradigm tended to reify in particular aesthetic objects in order to study and explicate them as a 'given' in terms of what they are and mean, digital studies in the humanities nowadays increasingly focuses on what users—i.e. the members of the cultural context in which these artefacts become relevant—do with these objects, that is: how they process and interpret them. One of the methodological impacts of the digital humanities on literary studies in particular is thus to reactivate the 1970's paradigmatic shift from classical philological exegesis towards an aesthetics and empirics of reception.⁸

(4) An increasing awareness of the *social dimension of the humanities research practice itself*. Here, a theoretical and a pragmatic necessity go hand in hand. Unless restricted to the mere and faithful application of tried and tested digital technologies, digital humanities projects require us to cooperate with the disciplines that focus on the mathematical and engineering aspects of the computational paradigm. Finding an interested and competent partner in computer science has thus become a *sine qua non* for most advanced digital humanities projects. These partnerships require openness to interdisciplinary exchange as well as the ability to communicate one's own perspective on the relevant objects and processes in the partner's language, or more abstractly put: it requires a conceptual grasp of an inherently foreign, yet

the "new tool is not just a bigger lever and more secure fulcrum, rather a new way of conceptualizing the world, e.g. as something that can be levered." (1996, para. 1)

⁸ In this regard see Andrew Goldstone's provocative thesis "Let DH be sociological", presented at the DH 2014 in Lausanne (<http://dharchive.org/paper/DH2014/Paper-756.xml>, accessed 14.08.2014)

complementing approach and methodology.

In addition, the mere scope and complexity of "big data" centered digital humanities projects nowadays makes it increasingly impossible to conduct research as an individualistic endeavor. Defining an individual humanities scholar's research interests and agenda in terms of an interdisciplinary, team-oriented research practice poses a considerable challenge to a field where outstanding intellectual and academic achievements have traditionally been associated with individuals. This forth aspect affects not only researchers, but also the entire disciplinary and institutional infrastructure that supports them.

In text studies, as in any other domain of the humanities, digital environments and approaches no longer just emulate traditional humanities practices; they have also begun to transform it in an essential way, as Gradmann and Meister (2008) argue. For example, text annotation represents one of the oldest forms of object explication documented in cultural history. More astoundingly, despite the change in media, even the very early witnesses of this practice—such as interlinear annotations in monastic manuscripts—have maintained their explicative functionality. One reason for this longevity is the physical proximity between source and metatext in these original annotation techniques. Ranging from the mere underlining and highlighting of individual words to the complex and fully textualized commentaries in the margins, all of the original variants of annotation possess a decisive functional quality: an extremely stable and non-ambiguous way of referencing the *explicandum*. This is a functional characteristic which secondary and meta-texts cannot match.⁹

Today we emulate this practice by adding markup of various complexity and functionality to the digital representation of an object, or to the born digital object itself. However, in contrast to traditional philological practice we now find ourselves faced with two competing technological solutions that both claim to have replaced the old manual technique: embedded (or in-line) markup, which preserves the physical connection between object and annotation, and external stand-off markup, which trades the former's automatic and unambiguous object reference for a higher degree of flexibility. This flexibility is potentially of far reaching consequence: not only can a stand-off markup approach accommodate overlap more easily, but it also frees us from the constraint of disambiguation as conflicting annotations on the same source document might be captured in separate stand-off annotations.

But why would one want to allow for ambiguity in the first place? The answer is crucial: allowing for ambiguity is a *sine qua non* for an adequate conceptualization of the notion of "object" common to the humanities. Unlike the natural sciences, and unlike mathematics or pure logic, the humanistic disciplines necessarily conceive of their objects of study as historical phenomena which can only partly be "described" in a fixed and uncontroversial manner. The humanist's real task lies in the object's eventual hermeneutic interpretation—and this cannot be static or normatively declared because interpretation per se is always historical and, as such, dynamic. No matter how consistent an interpretation might seem to

⁹ See also Pliny (<http://pliny.cch.kcl.ac.uk/index.html>, accessed 14.08.2014), a work of John Bradley for another approach on emulating human practice of taking notes while reading a text or viewing a picture.

us today, it comes with an inherent expiry date defined by its conceptual context. Therefore, in a humanities perspective, the choice between embedded vs. stand-off markup is anything but technical: rather, the two competing technological solutions must be understood as representing two methodological opposites, one upholding the claim of objective and timeless ontologic¹⁰ universality, the other one supporting the idea of the historical and dynamic transience of object explication.

The debate on what markup is, can or should be has been ongoing for over two decades now.¹¹ In their 1987 article on "Markup Systems and the Future of Scholarly Text Processing", Coombs, DeRose, and Renear (1987, pp. 933-947) already argued that the only type of markup adequate to the intellectual goals of academic text studies ought to be *descriptive* rather than *procedural*. This distinction between a type of markup that instructs a machine how to process a string of characters regardless of its intrinsic semiotic or rhetorical function, and another type of markup that informs a human reader about the correct categorization of that string in reference to some external system or purpose (a grammar, a lexicon, a pragmatic context etc.) proved fundamental. Coombs et al. presented two main arguments in favor of descriptive markup: one, it is more flexible than procedural markup; two, only descriptive markup implies an accurate model of what a text actually is from the point of view of the human reader: namely a means of communication.¹²

An even more radical philosophical critique was formulated by Buzzetti (2002, pp. 61-88) in his reflections on '*Digital Representation and the Text Model*'. Buzzetti argued that strongly embedded markup implies a data model which can only capture the expression plane of literary texts, but is inherently inadequate to model the content plane. While the symbolic material that represents the expression is organized in a linear, sequential fashion closely emulated by embedded markup, the semantic content is organized as a non-sequential and multidimensional continuum that takes on the form of a matrix; a characteristic which calls for "a weakly embedded markup system and a non-linear data model" (Buzzetti, 2002, p. 76). However, the key criterion for an adequate markup of literary texts according to Buzzetti is the ability to model and define not only the sequential representational material from which texts are made, but, more importantly, the *one-to-many-relationships* that are characteristic of texts both in terms of textual pragmatics (construction, distribution, editing, critique, interpretation etc.) and textual semiotics. In short, Buzzetti strongly advocates variance in markup: "Various encoded textual portions, generated by different interpretations of one single expression, may be considered synonymous expressions of one single content" (Buzzetti, 2002, p. 83).

Despite these critical reflections, declarative embedded markup as defined by SGML and XML or implemented with HTML established itself as de facto technological standard—the result of a predominant engineering perspective on markup as a means of process control in generating print or screen output, as Schmidt (2010) shows in the latest critical overview on

¹⁰ Understood in the sense of: universality of the underlying concept ontology.

¹¹ The following three paragraphs partially rephrase and partially quote verbatim our summary presented in Meister (2012, pp. 114-117).

¹² On this, Renear later clarified: "The objects indicated by descriptive markup have an intrinsic direct connection with the intellectual content of the text; they are the underlying "logical" objects, components that get their identity directly from their role in carrying out and organizing communicative intention" (Renear, 2004, chap. 2.7).

the history of markup development.¹³ Schmidt himself challenges the adequacy of embedded markup with particular reference to cultural heritage texts, arguing that a number of technical deficiencies make embedded markup particularly inapt for such documents. For him the main problems with embedded markup lie in its inability to deal with overlapping structures, in the fact that it inadvertently encourages a manual encoding of information that could equally well, if not even more efficiently and consistently, be computed automatically, and most fundamentally: in the risk of inscribing potentially obsolescent technology and interpretations into the source text.

We, too, consider embedding markup (let alone technology) within an interpreted object inappropriate. But for us the decisive arguments are neither technological (e.g., the overlap problem) nor pragmatic (e.g., the counterproductive preservation of "obsolete" markup and interpretations)—problems which Schmidt discusses in depth.¹⁴ Beyond these theoretical and pragmatic concerns lie points which are of a conceptual, methodological and social nature: How can the traditional modes of annotation across their various levels of complexity (i.e. from base-level declarative markup of textual elements right up to that of holistic and context sensitive text interpretation) be *conceptualized as a practice of knowledge production*? And how can we *go beyond merely emulating traditional forms* of this practice in the digital domain?

Let us begin by looking at how this type of knowledge production has evolved in the modern non-digital humanities. From the 18th century onward, annotation and interpretation of primary objects overcame the pre-enlightenment model of canonical exegesis; the philologies replaced it with a critical, discursive hermeneutic practice. The failure to account for this self-reflexive quality of annotation conceptually is therefore arguably the single biggest shortcoming of embedded markup. For the historically aware humanist it smacks of an anachronistic methodological paradigm: how can one still try to enforce normative, canonical descriptions of texts?

However, to be realistic, let us acknowledge that different types of markup are required for different purposes: embedded markup remains, in all probability, the adequate and most efficient approach when it comes to base-level declarative (or procedural, output oriented) operations performed on a text and against the backdrop of a widely accepted taxonomy or syntax. But it is at the same time inadequate and conceptually outdated when higher-level hermeneutic operations are at stake.¹⁵ Hermeneutic practice in the contemporary humanities no longer follows the normative exegetic procedures that were in force before Schleiermacher, nor can it adequately be reflected upon and thus "modeled" without considering the fundamental scientification of hermeneutics that was brought about by the works of, among other, Dilthey and Gadamer. To conceive of markup and to devise a markup model without taking into account the social and discursive dimension of annotation

¹³ For a competing account that openly defends the OHCO (Ordered Hierarchy of Content Objects) data model of "text"—which Schmidt strongly opposes—as the underlying premise of most functional markup schemata, see Renear (2004).

¹⁴ Schmidt's list of monita includes: the legacy of an "output command orientation" in markup languages; the reference to a hierarchical text model as exemplified by the OHCO approach; the versioning problem, which is of particular concern to him and which he proposes to solve by his own Multi-Version Document Model.

¹⁵ Here we are consciously ignoring the aspect of procedural markup, for which a separate case can be made.

as a historical cultural practice, or to ignore the knowledge productive nature of disagreement and hermeneutic circularity for the humanities reveals an oversimplified model of its "world". It is this fundamental philosophical problematic from which the discontent of many traditional humanists with the emerging paradigm of the digital humanities stems: by contrast, the pragmatic inadequacies of the concretized digital tools and models, and in particular those encountered in specific markup technologies are only surface phenomena and as thus of a transient and symptomatic order.

One of the recent contributors to the markup debate who pays explicit tribute to the modern humanistic perspective is Piez (2010).¹⁶ He outlines an architecture aimed at supporting what he terms "hermeneutic markup":

By "hermeneutic" markup I mean markup that is deliberately interpretive. It is not limited to describing aspects or features of a text that can be formally defined and objectively verified. Instead, it is devoted to recording a scholar's or analyst's observations and conjectures in an open-ended way. As markup, it is capable of automated and semiautomated processing, so that it can be processed at scale and transformed into different representations. By means of a markup regimen perhaps peculiar to itself, a text would be exposed to further processing such as text analysis, visualization or rendition. Texts subjected to consistent interpretive methodologies, or different interpretive methodologies applied to the same text, can be compared. Rather than being devoted primarily to supporting data interchange and reuse—although these benefits would not be excluded—hermeneutic markup is focused on the presentation and explication of the interpretation it expresses. (Piez, 2010, p. 202)

Piez's notion of "hermeneutic markup" comes close to what a contemporary literary scholar would regard as a non-trivial (i.e., hermeneutically functional) form of annotation. At the same time, it would be false to conflate the concepts of "annotation" and "interpretation". In a similar vein, Renear (2000, pp. 411-420) had already warned against the conflation of "statements about documents" and statements about "domain (the kinds of objects named in those statements)". To him this risk stemmed from the inadequacy of the "procedural vs. declarative" perspective onto markup. Renear concluded that an "adequate markup taxonomy must, among other things, incorporate distinctions such as those developed in contemporary 'speech-act theory'." Against this background, Meister (2012) has recently suggested to replace the various dichotomous models by

...a scalar model that extends between the ideal-types of performative versus hermeneutic markup. The former type represents all markup practices intended to define the legitimate modes and procedures by which textual elements and textual objects on a whole may be *processed*; the latter stands for approaches to markup that are primarily motivated by the aim to define how these textual elements and objects ought to be *interpreted*. (Meister, 2012, pp. 115-116)

Note that the two ideal-types are theoretical abstractions: in reality no markup is probably ever absolutely and exclusively performative or hermeneutic.

¹⁶ Piez argues in a similar vein in a subsequent panel discussion held at the DH 2012 (see Baumann et al., 2012).

Finally, an important social and pragmatic point concerns markup as a type of practice. Again, it seems that the computer science and engineering quest for automated parsing and tagging routines has resulted in a somewhat biased and solipsist conceptualization of text declaration and text understanding. But, unlike a computer, no human being ever processes a text in isolation from his or her own social and cognitive context. This is where the vision of *collaborative markup* comes into play which Christian Wittern—without using the specific term—already outlined in a 1999 HUMANIST posting in which he stated that sharing of source documents should not be the only perspective of collaboration which digital technology can support:

What we need to develop is also some protocol through which distributed layered portions of markup, which might be located on entirely different physical locations, can be used to generate a view of a text. We might also want to think of "open workgroups", where markup can be added remotely to texts located somewhere in cyberspace. (Wittern, 1999, para. 4)

These principal considerations on what markup is and how we must conceptualize it from a theoretical, a practical and a social perspective in order to pay tribute to its role in and for humanities research form the conceptual backdrop to the heureCLÉA project. At the same time, they also help to identify our logical point of departure. Most of the established markup schemes that might be of relevance in terms of annotating narrative features were developed to support declarative and taxonomy driven annotation. We cannot simply build on them. For example, during the first project phase, heureCLÉA focuses on the identification of time-related expressions. Hence having on the one end a scheme like TimeML¹⁷, clearly declarative and ultimately aimed for facilitating non-ambiguous, automated markup, and on the other end manual, multi perspective and possibly ambiguous markup, we decided to start from both ends, enhancing the former with the latter.

Let us summarize our conceptual reflections in the form of two guiding questions:

- One, is it possible to support the generation of semantic markup that can subsequently be made functional for more traditional hermeneutic operations—i.e., for text interpretation—by way of a machine learning approach based on the analysis of manual text annotations? In particular, is it possible to use this approach not only for the computational modeling of low-level narratological surface phenomena in narrative texts, but also for more complex phenomena that require interpretive inferences? Note that in this context, "support" does not necessarily mean generating semantic markup per se, but can also take the form of generating sufficiently plausible markup suggestions that will then be presented to the human user for validation.
- Two, can the knowledge-productive openness of the original humanistic practice of annotation—a practice in which taxonomic declaration and heuristic exploration go hand in hand—be modeled computationally? Is it possible to build a digital heuristic that will not only generate answers, but which perhaps, more importantly, can also help to identify the as yet undefinable and thereby help to formulate new research

¹⁷ A XML-based standard for the encoding and annotation of temporal expressions and events; for further details see Pustejovsky et al. (2004), Pustejovsky (2005), Pustejovsky et al. (2010) and section [“2.1.3. The Role of Standards”](#) below.

questions?

In order to be able to answer "Yes" to both questions, we will need to achieve the mission-critical goals outlined by Piez (2010, p. 203).¹⁸

2.1.2. Annotation in CATMA and heureCLÉA

CATMA (*Computer Aided Textual Markup and Analysis*) is a text annotation and analysis tool developed at the University of Hamburg since 2008 and currently available in version 4.¹⁹ The first three versions of CATMA used a file based database in TEI format. The latest release is still JAVA-based but with the usual amount of Javascript, HTML and CSS, and it is backed by a SQL relational database.

Because of its low-threshold design, CATMA allows traditional humanities scholars to annotate digital texts in a very flexible, yet XML/TEI-compliant format without having to bother about technical issues or standards: users can either import a pre-defined tagset²⁰, or build their own tags in an *on-the-fly*-mode, or apply any combination of these two procedures. Moreover, they can also make use of a suite of integrated analytical as well as of some basic visualization functions.²¹

¹⁸ Piez outlines in his architectural sketch:

- ¹⁸A data model supporting arbitrary overlap.
- ¹⁸Interfaces, including a markup syntax, that facilitate the creation, editing and analysis of texts using this data model, with the capability of defining ad hoc elements and properties (attributes) on the fly.
- ¹⁸A transformation technology supporting (in addition to data transformations) analytical tools applicable to the markup as such (not just the raw text), with the capability of managing elements and their properties in sets, locating them, listing them by type, sorting, visualizing and comparing them.
- ¹⁸Schema-inferencing capabilities for describing the structural relations within either an entire marked-up corpus, or within identifiable segments, sections or profiles of it.
- ¹⁸In connection this, a schema technology that supports partial and modular validation.

¹⁸ Piez, 2010, p. 203

¹⁸CATMA has already made first steps to satisfy these goals by

- ¹⁸working with a character range based data model where (even discontinuous) chunks of text can be referenced by any number of annotations,
- ¹⁸providing a graphical user interface to apply annotations by mark and click actions and to create new tag and attribute definitions either conceptually in advance and separated from the text but also ad hoc during the tagging of a text,
- ¹⁸providing a powerful analysis module that works directly on the text and its annotations plus various export formats to connect to other analysis tools,
- ¹⁸having implemented the repository on top of a SQL relational database which provides at least very good structural partial and modular validation.

¹⁹ The source code is released under the GNU general public license v3 and can be accessed at GitHub: <https://github.com/mpetris/catma> (accessed 14.08.2014).

²⁰ In CATMA, "tagset" means a set of tags where each tag is a named collection of properties with a unique identifier (UUID). In the simplest case, a tag is just the name and the UUID without any other properties. Tags can form a hierarchy within their tagset by establishing an "is-a" relationship between one parent tag and one or more child tags. Therefore, a tag can span a subtagset that contains all its children. Tags are heavily inspired by Feature Structures as defined in http://www.iso.org/iso/catalogue_detail.htm?csnumber=37324 (accessed 14.08.2014) and can be modeled as such.

²¹ The analytical functions most notably include a graphical query interface to build concordances from the text and the annotations. The graphical interface is backed by a very powerful query language which provides even more possibilities.

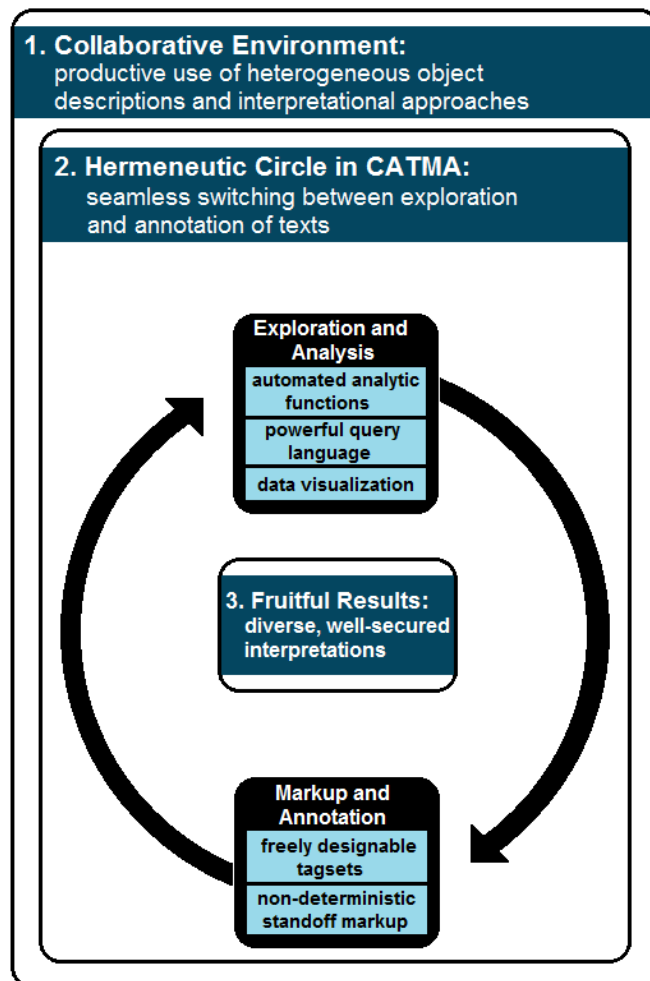


Fig. 1: Exploration and Annotation in CATMA

The overall goal in the development of CATMA was to create a working environment in which literary scholars can seamlessly switch back and forth between text annotation and text analysis (cf. fig. 1 “Exploration and Annotation in CATMA”). This integrated approach towards markup and analysis aims to emulate and support what actually happens in the real world of literary scholars: the continuous turning of the “hermeneutic wheel” (Burnard, 2001), that is, the back and forth between formal text analysis and declaration on the one hand, and the generation of semantically oriented interpretative hypotheses on the other.

Moreover, version 4 of CATMA—named CLÉA (*Collaborative Literature Exploration and Annotation*)²²—represents not only a technological change from the previous desktop to a web-based solution; more importantly, it implements the conceptual shift towards a practice of collaborative markup (Meister, 2012).²³ Markup as well as the analysis of texts and text corpora can now be undertaken jointly: using CATMA, researchers can share, reuse, amend

²² The inapt *accent deguis* is a bit of a private joke: it marks the diacritical concerns of non-Anglo-American philology which we wanted to highlight.

²³ The CLÉA development phase of CATMA was generously supported by two Google Digital Humanities Awards (2010, 2011). For further details see <http://www.catma.de/clea> (accessed 14.08.2014).

and dispute each other's markup *ad libitum* and effectively in real-time.²⁴

In this development line, the outcome of heureCLÉA will mark the latest step towards providing literary scholars with a working environment that acknowledges the full spectrum of text analytical and hermeneutic procedures of which their traditional practice has always consisted. In this regard, the prefix *heure* points to a crucial methodological aspect in the text analytical and interpretive workflow so far widely ignored in the digital humanities: the procedural component in exploration and reasoning commonly referred to as *heuristics*. A heuristic is a theory-based taxonomy and methodology whose primary function is not to validate an already formulated hypothesis or theory; rather, its purpose is of an exploratory and preparatory character. A heuristic is a "problem finding tool", an intellectual device and method designed to inspect an object or an object domain in order to identify potential theory-based inroads on which the subsequent generation of explanatory hypotheses and eventually complex theory development may be based. Our concrete goal in the heureCLÉA project is to add such a tool to the web-based version of CATMA: a computational heuristic module.

2.2. Relevant Perspectives on Time Phenomena

In order to constrain the tasks in heureCLÉA to a manageable amount in terms of the project's scope, we restrict our focus on one particular narratological domain: time. More domains may follow after the completion of heureCLÉA. The starting point is our corpus of narratives in German language from around 1900 in which time related narrative features are identified by way of manual markup. This markup functions as the base for the automation process. In order to understand the complexity of this task, we review how time phenomena are standardly theorized in narratology and computational linguistics.

2.2.1. Time Phenomena in a Narratological Perspective: from Surface Markers to Inference-based Constructs

Following Fludernik (2005, p. 608), narrative theory has taken three main approaches in its attempts to clarify the relation between time and narrative texts:

On the textual micro level, time appears in the form of grammatical and morphological elements, so-called time-markers. They manifest themselves as use of tense, such as the "epic preterite" (Hamburger, 1957) or of present and future tense (Margolin, 1999, pp. 142-166); also by way of relative and non-relative temporal expressions (de Toro, 1986), deictic expressions and so-called "temporal operators" (Meister, 2005a, p. 112). All of these elements contribute towards the construction of a temporal model of the referenced (fictional or real) "world".

Second, narrative texts also possess the "double temporality" (Lahn & Meister, 2013, p. 136) of the levels of story and discourse, that is, the time of the narrating act and the time of the narrated events (Genette, 1972, Chatman, 1978; as precursor: Müller, 1948). From the relation among these two levels stem three fundamental forms of time arrangement and modification: order (the sequence in which events are portrayed—asked for by the question

²⁴ Users are free to share primary texts, standalone libraries of tagsets and/or collections of annotations along with the corresponding tag definitions. We are currently developing a use-policy for future use of CATMA.

“In which order are the events narrated?”), duration (the expanse of time taken up by the event in the referenced “world” in relation to the time taken up by its narrative representation within a text—asked for by the question “How long does the act of narrating take?”) and frequency (the number of occurrences of a referenced event in relation to the number of times that it is being narrated—asked for by the question “How many times is an event narrated?”).²⁵

Thirdly, in a philosophical perspective, time is considered, since Kant, as a fundamental category that organizes the epistemic practice of human cognition (Markosian, 2008). In this respect, a distinction is made between the measurable and objective, linear time of physics, and the psychological and subjective time experience of humans (McTaggart, 1927; Dowden, 2011). In contemporary literature the latter form of time, which often appears as a highly complex and contradictory phenomenon, plays a central role: modernist literature in particular focuses on the question of how we as humans can organize and sustain our subjective memory and identity.

The modeling of complex time-related narrative phenomena remains a big challenge to the digital humanities (Meister, 2005b, 2012). Structural phenomena in particular cannot simply be analyzed on the basis of words or tokens in a computer assisted text analysis (Meister, 2011). heureCLÉA’s perspective on the domain of narrative time construction and management therefore needs to combine three perspectives. Starting with the tagging of pre-defined time markers in a digitized corpus of texts, one can in a first approach isolate, collect and classify temporal information that resides at the textual surface (Adolphs, 2006). In a next phase, one can then identify the more complex, partially metaphorically encoded or inference-based categories of time construction and management in a semi-manual fashion, as for example phenomena of character or utterance-position specific temporal perspective. Following a step-by-step approach, the overall time structure can eventually be outlined and, using analytical and statistical algorithms, correlations between time markers and text-inherent time phenomena may finally be identified. Against the background of the text’s or corpus’s specific time structure and time semantics one may now draw inferences on the overarching philosophical concepts of time and memory, which in turn may serve as an essential component in the overall interpretation of the text or corpus at hand.

Some of these features might be of a reasonably uncontroversial nature and can thus be expected to allow for an automatic description in terms of the *de facto* canonic narratological terminology developed in the structuralist tradition of Todorov, Genette, Schmid and others.²⁶ Others will be of a more interpretive and contextualist order; here, the digital heuristic’s task will in all probability be restricted to generate markup suggestions that are ranked in terms of statistical probability, but will in any event require an interactive validation by the user. A third category can be expected to yield neither of these options, but rather to point out potentially interesting phenomena for which none of the existing taxonomic categories seems to fit—and therefore support the quintessentially heuristic function of identifying potentially fruitful new research questions. Finally, a fourth category will comprise what, at least from the narratological perspective, is either a conceptual outlier or simply

²⁵ For a more detailed discussion of these and other narratological concepts for temporal analysis, see [section 3.1.](#) below.

²⁶ The narratological taxonomy of heureCLÉA is based on the comparative and summary account of structuralist narratology presented in Lahn and Meister (2008).

"white noise".

2.2.2. Time Phenomena in a Linguistic and Computational Perspective: Extraction and Annotation of Time-related Expressions in Documents

Extraction of time information from texts plays an important role in computational linguistics. An overview on current trends in information retrieval is presented in Alonso et al. (2011). They include the grouping of search results on the basis of temporal information embedded in documents (Alonso et al., 2009, pp. 97-106), the support of time related queries in search engines, or applications for the temporal exploration of single documents. Temporal information is also crucial in topic detection and tracking (Allan, 2002). Here, individual texts within a news stream can be classified according to whether they introduce a new topic, or extend an already existing one.

Regardless of the specific application context, the fundamental precondition for all of these techniques is that temporal expressions in texts can be automatically identified, extracted, and normalized. Programs performing these tasks are called "temporal taggers". While the identification and extraction of time expressions is of limited complexity, their correct normalization poses a big challenge, the reason being that temporal expressions do not just occur explicitly, but also in a relative or in an underspecified form (Schilder & Habel, 2001, pp. 65-72). In these instances, the correct reference time as well as the temporal relation to it must be determined in order to allow for the normalization of the expression.

Because of the importance of temporal information in many applications, and because of the difficulties of normalizing time expressions, automatic temporal annotation is a very active field of research. Apart from numerous prototypes (Mani & Wilson, 2000, pp. 69-76; Mazur & Dale, 2009, pp. 245-257) this interest is also reflected in scientific competitions among research groups. For example, at the TempEval Workshops 2010 and 2013 (Verhagen et al., 2010, pp. 57-62; UzZaman et al., 2013, 1-9) the identification and normalization of temporal expressions in texts was one of the tasks for measuring the performance of automated systems. These developments in automated temporal annotation have led to annotation standards as well as to annotated corpora, such as the TimeBank Corpus, or the training and evaluation corpora ACE Tern 2004, 2005 and 2007²⁷, as well as the TempEval-2 and TempEval-3 data sets.²⁸

Despite the popularity of the research field almost all of the works in temporal annotation concentrate on news texts or similar genres. Texts taken from this expository domain normally come with a timestamp indicating the date of publication—which in turn bears a strong relation to the date of text production—that may serve as a fixed reference for subsequent normalization routines. In addition, these texts are normally quite short, making their temporal as well as their discourse structure fairly simple. By comparison, long "narrative" documents generally have a significantly more complex discourse structure, whereas the date of publication plays hardly a role for the normalization of time expressions contained within them. Mazur and Dale (2010, pp. 913-922) have compiled a temporally

²⁷ TIDES Standard for the Annotation of Temporal Expressions: <http://projects ldc.upenn.edu/ace/data/> (accessed 14.08.2014).

²⁸ TempEval-2: Evaluating Events, Time Expressions, and Temporal Relations, <http://www.timeml.org/tempeval2/> (accessed 14.08.2014) and <http://www.cs.york.ac.uk/semEval-2013/task1> (accessed 14.08.2014).

annotated corpus of Wikipedia articles on historically important wars. They show that the application of the usual strategies for normalization of temporal expressions in these documents will lead to many misinterpretations, and they argue that other normalization strategies ought to be developed in order to arrive at acceptable results for this text domain. Against this background, it can be expected that taking the next step up in the level of text complexity, i.e., that from the level of factual narrative accounts to the level of literary narratives, will result in even more complex challenges.

In the heureCLÉA project, we use our rule-based temporal tagger HeidelbergTime as the second main software component besides CATMA. Its goal is to detect and normalize temporal expressions with TimeML's TIMEX3 annotations. As previous temporal taggers, HeidelbergTime was originally developed for processing English news documents (Strötgen & Gertz, 2010, pp. 321-324) but was then extended to become the first multilingual, domain-sensitive temporal tagger (Strötgen & Gertz, 2013, pp. 269-298). In particular, the domain-sensitivity is an important aspect for the heureCLÉA project. In a cross-domain evaluation, we demonstrated the necessity for domain-sensitive temporal tagging: when processing (factual) narratives instead of the news-style documents with the (standard) news strategy, the normalization quality drops about 20 percentage points. However, running HeidelbergTime in its mode for processing narratives, this loss is fully compensated (Strötgen & Gertz, 2012, pp. 3746-3753). Although literary narratives result in further challenges compared to factual narratives, HeidelbergTime is designed in such a way that further domain-sensitive strategies can be added. Currently, HeidelbergTime supports 10 languages (Strötgen et al., 2014a, pp. 1-21) among which German is of particular interest in the heureCLÉA project. In addition, it is publicly available as standalone version and as UIMA²⁹ component.³⁰ Another important aspect for users from the digital humanities is that it was recently extended to also deal with temporal expressions referring to historic dates (Strötgen et al., 2014b, pp. 2390-2397).

2.3. The Role of Standards

As already evident from the description of our project and our tools, various standards are of relevance in the heureCLÉA context: CATMA is based on a TEI/XML compliant format, the HeidelbergTime temporal tagger³¹ uses TimeML, and the concept of annotation we are working with is in general terms based on the W3C recommendations. But why and to which extent do we consider standards and standard orientation important in a digital humanities project like ours in the first place?

From a computational point of view standards guarantee the interoperability of systems and research results. Moreover, adherence to standards helps to protect against 'reinventing the wheel' and the development of isolated applications which cannot be sustained in the long run. However, from a humanities point of view theoretically innovative projects cannot really rely on existing standards if they want to explore new conceptual aspects - the humanities' object domain is highly dynamic, and so are its conceptualizations of this domain. For this reason theoretically ambitious humanistic research has a tendency to constantly re-design and extend its models and taxonomies. In fact, the notion of a "standard description" of an

²⁹ <http://uima.apache.org/> (accessed 14.08.2014).

³⁰ <http://code.google.com/p/heideltime/> (accessed 14.08.2014).

³¹ cf. section "[2.2.2. Time Phenomena in a Linguistic and Computational Perspective: Extraction and Annotation of Time-related Expressions in Documents](#)" below for details about HeidelbergTime.

object is *per se* somewhat foreign to the contemporary ‘liberal arts’; even those disciplines which have developed more stringent and elaborate taxonomies, such as linguistics, will generally tolerate, if not even integrate competing descriptive approaches and theoretical models.

While the humanist may welcome and defend this flexibility as knowledge productive, even a minor conceptual redesign of a model or a theory can, from a computational point of view, nevertheless have far-reaching counter productive consequences; at worst it might even necessitate a complete re-encoding of research data in terms of a substantially different concept ontology. So which approach is favorable?

One way to resolve this question is to distinguish among two types of interoperability and to approach them separately. The first one is the *physical* interoperability that allows for technical interchange and re-use of information in a broad sense. Physical interoperability can be realized by the implementation of standards like XML or TimeML, and in heureCLÉA we strive to achieve this first type of interoperability by various means:

- Since CATMA allows for overlap of annotations, they cannot be organized hierarchically as requested by TEI/XML. In order to still make annotations TEI/XML compliant, they are stored as feature structure representations³², a solution specified in ISO 24610-1:2006.³³ More specifically, annotations are represented by a tag definition consisting of an identifier, a color code, optional properties and their possible values (optional). Tag definitions can be organized hierarchically. The annotations, so called “tag instances”, reference one or more text sequences and may contain additional values for their optional properties.³⁴
- HeidelTime annotates temporal expressions following the temporal markup language TimeML (Pustejovsky et al., 2004, pp. 28-34; Pustejovsky et al., 2005, pp. 123-164), in which the normalized information of temporal expressions is defined according to the ISO 8601 standard for temporal information with some extension.³⁵ Furthermore, TimeML itself is also specified as an ISO standard, the ISO-TimeML.³⁶
- We use UIMA (Lally et al., 2006, pp. 17-18) as our processing framework for automatic text annotations, which internally represents annotations and meta data of

³² For feature structures cf. Lee et al. (2004) and <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html> (accessed 14.08.2014). (“P5: Guidelines for Electronic Text Encoding and Interchange. 18 Feature Structures”).

³³ http://www.iso.org/iso/catalogue_detail.htm?csnumber=37324 (accessed 14.08.2014). (“ISO 24610-1:2006. Language resource management -- Feature structures -- Part 1: Feature structure representation”).

³⁴ The TEI/XML based implementation of the annotations is realized with segment elements (<seg>) that reference the annotations coded in feature structures with the @ana attribute. Tag definitions are implemented with the element <fsDecl> and tag instances with <fs> (= feature structure). The @type attribute in <fs> is used for tag names, property names are implemented in the @name attribute of the element <f> (= feature). Both attributes are of the type xsd:Name.

³⁵ <http://www.iso.org/iso/home/standards/iso8601.htm> (accessed 14.08.2014). (“ISO 8601 – Time and date format”).

³⁶ http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37331 (accessed 14.08.2014). (“ISO 24617-1:2012. Language resource management – Semantic annotation framework (SemAF) -- Part 1: Time and events (SemAF-Time, ISO-TimeML)”), see also Pustejovsky et al. (2010, pp. 394-397).

a document as an XML document using the established XML Metadata Interchange standard (XMI).³⁷ UIMA also provides the possibility to export annotations encoded in XMI.

The second type of interoperability is that of *logical* interoperability, which means interoperability of ontologies and concepts. Achieving logical interoperability, especially in projects relating to theoretical concepts from humanities fields, is a much more complex task than guaranteeing mere physical interoperability.

This holds for heureCLÉA, too: the project's aim is to automate the recognition of narratological concepts regarding time phenomena, which obviously requires a taxonomy of the narratological concepts we apply. In order to achieve the highest possible agreement among narratologists about these concepts, they must be as established, secured and well-tested as possible. As our starting point we therefore defined a set of compatible narratological categories and taxonomies of time phenomena that originated in the structuralist tradition of Todorov, Genette, Prince and others. From this well-established basis we intend to work towards a possibly more comprehensive narratological taxonomy of time phenomena. But rather than introducing fundamentally new concepts or concept ontologies *ad libitum*, our preferred approach will be reductive—though without precluding in principle a theoretically substantiated extension, where required.³⁸

The first step towards such a more comprehensive taxonomy is the already mentioned consensus oriented approach to narratological analysis.³⁹ Additionally, we focus on a preliminary stage of logical interoperability on the implementation side of our project: our approach to annotation ensures the extensibility of the data we produce. Our text annotations can therefore be complemented with additive annotations of various kinds as well as aggregated in order to build more complex annotations.

3. Design and Current Outcomes of the heureCLÉA Project

From a design point of view the heureCLÉA project spans three interrelated dimensions that need to be addressed not necessarily in sequential fashion, but rather in a more point-by-point interchange across domains (cf. fig. 2 “heureCLÉA Project Dimensions”).

The first dimension concerns the annotation of time phenomena. Our pragmatic starting

³⁷ <http://www.omg.org/spec/XMI/ISO/19509/PDF/> (accessed 14.08.2014) (“ISO/IEC 19509 -- Information technology -- Object Management Group XML Metadata Interchange (XMI)”)

³⁸ This ‘traditionalist’ orientation also precludes us from considering the fundamentally different, non-hierarchically structured new concept ontology for narrative information proposed by Zarri (2009).

³⁸ Our taxonomy is already being used—and extended—for other narratological research projects, the conceptually most advanced of which are three recent PhD theses at the University of Hamburg: (1) A broad narratological taxonomy including the time concepts used in heureCLÉA has been applied to the analysis of narrations about conflicts (Gius, 2013); (2) a sub-set of terms for the markup of so-called ‘eventfulness’ was elaborated and applied in a narratological analysis of song lyrics in order to measure their level of narrativity (Schüch, in preparation), and the taxonomy is also being applied in (3) an analysis of the representation of consciousness in fictional and factual narratives (Lagoni, in preparation).

³⁹ Its underlying concept of annotation and the role of ambiguity have already been discussed in section 2.1.1., the narratological aspects have been described in section 2.2.1. and will be expanded a bit more in section 3.1.

point and practical research problem is the identification of the temporal structure of events in our corpus of narrative texts. Here, two concurrent approaches are employed: manual collaborative annotation with CATMA, and automated temporal tagging with the temporal tagger HeidelTime (Strötgen and Gertz, 2013).

By using the method of collaborative tagging, the narrative representations of temporal phenomena in the corpus are manually provided with markup. For the decoding of temporal phenomena in the fields of *discours/histoire*, we draw upon the narratological taxonomy of Genette, which is supplemented by a taxonomy suited to capture phenomena connected to action and event segmentation. The markup itself is then generated by using CATMA to tag the corpus texts with the relevant taxonomic terms.

Methodologically the focal point of this project dimension lies in the field of the humanities, with an emphasis on the narratological analysis of temporal phenomena; an additional practical task is the implementation of further collaborative functionalities in CATMA. The third task in the dimension of the annotation of time phenomena, however, is the adaptation of existing automated procedures with a view to generating more complex time annotations. Here we focus on a subset of the markup operations which so far could only be manually conducted in CATMA.

The second dimension of heureCLÉA concerns the learning of new rules for automated annotation from the manually generated markup. Here the heureCLÉA project takes on a highly dynamic nature: since users of the collaborative environment constantly generate new markup, the data basis for the machine learning analysis keeps growing. As the employed machine learning algorithms are based on descriptive and inductive statistics, their performance depends on the size of available annotated data. The constantly growing and improving amounts of manually created markup data is used to optimize the machine learning models incrementally. This active learning strategy improves the predictive power of the component that will eventually be programmed and integrated into the CATMA platform.

Depending on our outcomes we envisage to use methods for the derivation of rules—for example rules capturing combinations (co-occurrences) of temporal expressions. The derived rules will then be implemented in HeidelTime and corresponding tools in order to facilitate the automation of more complex annotations. Larger quantities of markup (and user feedback) are expected to facilitate more complex modeling strategies based on distributional approaches (such as Latent Semantic Analysis, LSA) and will further increase the accuracy of automatic annotations. Finally, patterns representing typical temporal sequences will be extracted (Sequence Mining) in order to serve as a basis for the visualization and exploration of the detected temporal structures in the documents as well as in the related markup.

At current the heureCLÉA team is concentrating on the first two dimensions—first insights and their relevance for the next project phase will be reported below.

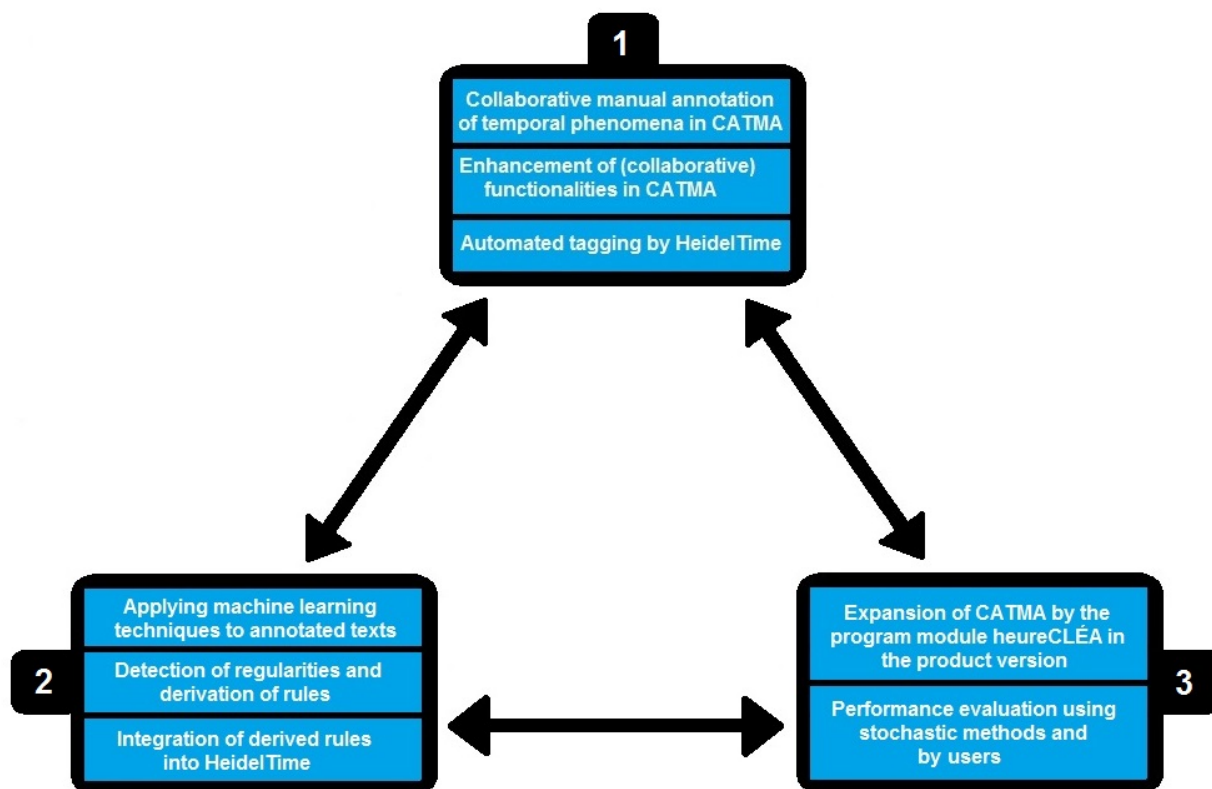


Fig. 2: heureCLÉA Project Dimensions

The third project dimension concerns the implementation and evaluation of the heuristic which we aim to develop. As soon as tests achieve a functional threshold—defined by the parameters of (a) reliability of the automated detection of temporal references of low complexity (i.e., of phenomena that are directly traceable on the textual surface level), and (b) performance and robustness—the annotation environment CATMA will be augmented by the program module heureCLÉA in its productive version. This module will provide a 'digital heuristics' for the partially automated, partially interactive generation of temporal markup. The main component of the module will be HeidelTime, which—based on a constantly expanded rule base—will be used to automatically specify temporal markup in texts. The upgraded functionality of the collaborative markup environment will then be tested and evaluated by users in a specific application context in the field of humanities. In principle, this individual operation serves the evaluation of the adequateness of the automatically generated temporal markup. In parallel, the functionality will also be evaluated using stochastic methods. The focal point of this project area is empirical—in the field of humanities as well as in the field of information theory.

In the remaining sections of this chapter we will report current outcomes regarding these dimensions of our project.

3.1. Narratological Analysis of Temporal Phenomena: Current Outcomes of the Manual Annotation with CATMA

At the moment, five experts in the field of narratology are using CATMA to annotate the corpus using a tagset that represents the most important canonical narratological categories for analyzing temporal phenomena in narrative texts. Currently, we concentrate on the annotation of temporal phenomena in three distinct classes: (1) temporal expressions

signifying time points and time spans (represented by the subtagset <dates>), (2) the temporal relation between the act of narrating and the narrated events (represented by the subtagset <relation_narrator–event_time>), and (3) the temporal relation between *discours* and *histoire*, i. e., between the way of presenting a story and the story itself (represented by Genette's three fields <order>, <frequency>, and <duration> in our subtagset <timerelation_discours–histoire> (Genette 1972)). The choice to annotate these particular temporal phenomena as a start was made as we expected these phenomena to be rather easy to detect on the surface level of a text, making a high degree of overlap with automatically generated markup probable. For example, some parts of the subtagset <dates> designate textual phenomena that can already be detected by the temporal tagger HeidelTime. Accordingly, the manually generated annotations serve, on the one hand, the alignment of the already available functionalities of HeidelTime with the specific requirements of narratological textual analysis, as well as the rapid expansion of HeidelTime's functionality towards that of a practical digital heuristics. More complex temporal phenomena that cannot be detected on the surface level of a text like temporal perspective (i. e. the distance between the initial perception of events and subsequent acts of perception and narration) will be tackled in a later phase of the project.

A rather unexpected preliminary finding concerns the social aspect of our crowdsourcing approach in the collection of manual annotation data. Since we are explicitly aiming for the implementation of a 'digital heuristic' that will also support the analysis and interpretation of complex narrative phenomena, we assumed that not all of the narratological categories—represented by the tags of our tagset—can be deemed as text-descriptive, and that their application on a text may turn out a matter of subjective interpretation at an early stage. In fact, the generation of varied manual markup suggestions is a pre-requisite in order for the machine learning analysis of the metadata to be able to arrive at more than just a fixed annotation rule, but rather at competing heuristic markup suggestions that can be ranked in terms of probability or likelihood.

However, our annotators themselves preferred to settle most controversies concerning the temporal tagging of the texts through reasoned argumentation—instead of merely regarding the categories as genuinely interpretative. As it turned out, many cases of dissent could indeed be resolved by either pointing to objective surface phenomena of the texts, or by referring to details of the definition of the category in question. Even the cases that could not be settled as easily provided interesting insights: every persistent incongruity proved to trace back to a deficiency in the definition of the particular narratological category. These shortcomings can be divided in two groups, namely

- a conceptual incompleteness that can easily be eliminated by complementing the application of a category through functional decisions that consider additional, hitherto overlooked surface level markers, or
- a theoretical incompleteness that traces back to more fundamental problems in narratological models and theories.

To understand how these two classes differ in detail it is necessary to take a brief look at what narratological categories are and what exact purpose they serve. Generally speaking, narratological categories are concepts that serve the analysis and interpretation of narrative texts. Normally, a narratological category designates textual features that are, on the one hand, deemed to be specific to narrative texts and, on the other hand, considered interesting and fruitful to express the (narrative) characteristics of a textual representation as such.

Many of these categories serve the denotation of structural phenomena that are mainly accessible on the surface level of a text—such as most of the categories for the analysis of explicitly marked temporal phenomena.⁴⁰ An example for such a category is <*analepse*>, which helps analyze the order of the narration and—roughly—designates a flashback in time: expressions that mark such a flashback in time may be very easy to identify as surface markers, e.g., the discursive time information 'two years ago' that instructs the reader to jump back in 'action time' when building a mental image of the story world.

Let us now return to the distinction between surface-level conceptual problems and basic theoretical narratological problems: when it comes to the application of a narratological taxonomy, it might become apparent that the taxonomy fails to make entirely clear which exact features of a text are to be captured with the help of a certain category—its account of the category may simply be too vague. This problem can sometimes be settled by making decisions that *only* affect the category in question. For example, the account of <*analepse*> as a flashback in time may be too vague to really determine which textual features are to be captured by it. Accordingly, the expression 'flashback in time' would have to be clarified and its definition expanded. This may be done by simply pointing out more specific textual surface features that should be called <*analepse*>. Such problems can be deemed surface-level problems in two respects: (a) Their solution often requires the more specific identification of textual surface phenomena that should be denoted by a category, and (b) the decisions that are made in order to solve these problems only affect the problematic category itself—hence the decision that is required can be deemed to lie on a surface level instead of deeply inside the theoretical narratological underpinning, where a decision would affect many different individual surface phenomena. However, a category may also be underspecified for more fundamental theoretical reasons whose scope is of such an order that it actually affects a number of categories. For example, many structuralist narratological concepts are based on what has been termed the "model of narrative communication" (Schmid, 2008). Though powerful, this model cannot conceptually account for narrative phenomena that fall into the cognitivist realm, or can only be theorized on the basis of a possible worlds semantics. We will call this type of problem a 'basic narratological problem'.⁴¹

In the following subsections, we will illustrate these two types of problems by using examples from some of the literary texts we are currently working with. We will then briefly discuss the theoretical implications of these practical experiences as well as the consequences for our further conceptual approach to heureCLÉA.

3.1.1. Concept Vagueness Resolved by Additional Surface Markers

The first example of a tagging disagreement concerned the tag <*prolepse*>, which—serving

⁴⁰ Still none of the phenomena that are relevant in the field of narratology are *merely* formal textual features in a strict sense, because the meaning of words and sentences is always crucial for their detection. Thus, though these phenomena may be deemed accessible on the textual surface, their identification may still be subject to interpretation—at least in a broader sense of the word.

⁴¹ The distinction between surface-level problems and basic narratological problems must be seen as an 'ideal type' distinction, though: The classification of individual problems according to these two types may in some cases be a matter of interpretation or point of view on a problem. Still, this distinction—or: the presentation of our interpretation of individual problems—helps understand our handling of these problems.

the analysis of the temporal order of a narrative—designates a flash forward in time, i. e. the disruption of a chronological report in favor of a section that anticipates future events. In the short story *Matteo* by Friedrich Hebbel, there is a brief section—an utterance by one of the characters—that became subject to discussion with regard to its potential status as a prolepsis:

(example 1) "Sieh, morgen feire ich meine Hochzeit; zum Zeichen, daß du mir nicht mehr böse bist, kommst du auch, meine Mutter wird dich gern sehen."⁴²
Hebbel, 1963, para. 4⁴³

In order to understand how the tagging of this passage as <prolepse> is controversial, it must be added that the addressed character does not attend the speaker's wedding on the following day. As a result, some annotators tagged the whole passage as <prolepse>, others only tagged the first sentence, arguing that the rest cannot be seen as an actual flash forward in time, since its content does not eventuate in the fictional world. This prompted a discussion of a narrower, consistent and easily applicable definition of <prolepse>, which resulted in the following account:

(new account of <prolepse>) A prolepsis is an actual anticipation of the plot; plans that are shown not to be implemented and prophecies that are shown not to come true do *not* count as prolepses.

This clarification of the former underspecified category <prolepse> results in much more coinciding annotation of prolepses, which suggests that the identification of a prolepsis might not be a matter of subjective interpretation.⁴⁴

3.1.2. Concept Vagueness resulting from Basic Narratological Problems

In contrast to the rather easy-to-solve problems as the one described above, the type of controversy that will be presented in the following paragraphs traces back to basic narratological problems. Their solution would necessarily entail consequences for many other singular problems.

The following example affects the subtagset <duration>, which serves the analysis of the speed of narration, i. e. the relation between the time of the narrated events and the time of the act of narration itself. Interestingly, some passages in the text *Der Tod* by Thomas Mann were tagged by some annotators as <summary> (i.e. less time is used for the act of

⁴² "Look, tomorrow I am celebrating my marriage; to show that you are not angry with me any more, you come, too. My mother will be happy to see you." (our translation)

⁴³ For reference, see "[heureCLÉA Corpus](#)" at the end of this paper.

⁴⁴ It must be remarked that it might be a matter of interpretation to characterize this issue as surface-level problem. We saw the decision that clarified the concept of <prolepse> as a singular decision. But our decision, by any chance, also coincides with a specific notion of a deeper rooted problem—namely *narrativity*—which will be discussed in the next section: to Schmid (2008), only strictly *narrative* elements of stories can be analyzed with narratological instruments; to be narrative in a strict sense, a text passage must contain the report of *events*; and to count as an event, an occurrence that is reported must hold the feature of *facticity*, meaning that it must *actually* happen. It is easy to see how this notion coincides with our new account of <prolepse>. But since this account was not developed with reference to Schmid's notion of *narrativity*, and since we did not deem our solution to entail consequences for other singular theoretical matters, we interpreted it as being located on a surface level.

narrating than it took for the narrated events to happen), by others as <pause> (extreme case of: more time is used for the act of narration than it took for the event to happen—since the narration of events is interrupted by something else, e.g., a description). One example passage is the following:

(example 2) Ich habe die ganze Nacht hinausgeblickt, und mich dünkte, so müsse der Tod sein oder das Nach dem Tode: dort drüben und draußen ein unendliches, dumpf brausendes Dunkel. Wird dort ein Gedanke, eine Ahnung von mir fortleben und -weben und ewig auf das unbegreifliche Brausen horchen?⁴⁵
Mann, 2004, p. 76⁴⁶

When we discussed the arguments for either decision, it turned out that the classification of the duration of this passage—or more precisely: the passage after the first comma—depends on the classification of the presentation of inner thoughts in terms of *narrativity* or *eventfulness*: depending on whether it is deemed necessary for a textual passage to contain the depiction of *events* for it to be narrative (Abbott, 2014), and whether it is considered necessary for an *event* to hold certain features such as *facticity* (Hühn, 2013), this passage either counts as narrative and can be categorized as a summary, since the reported thoughts presumably took longer than the few seconds that have been used for reporting them. Or the passage is not considered to be narrative, which makes it a pause, since the reported thoughts are seen as an inserted description that interrupts the narrating of events.

This type of annotation conflict concerning duration traces back to the complicated and well-discussed narratological topics of narrativity and eventfulness, for which no consensual theoretical account exists as yet. That makes it difficult to settle the matter of conflicting annotations pragmatically, as we did in the previously discussed cases. It is important to note that the disagreement does not necessarily emerge from a genuinely interpretative feature of the category in question: "duration" is not per se under-defined, rather it has different definitions in different fundamental theoretical perspectives.

Other problems of this type arose, again, while analyzing the speed of narration or while trying to define the main narration among different narrative levels:⁴⁷ both concepts turned out to be strongly dependent from the segmentation of narratives, another theoretically complex category.

3.1.3. Methodological Consequences for the heureCLÉA Project

As these examples demonstrate, many cases of conflicting annotation of temporal narrative phenomena are rooted in the fact that some of the narratological categories are insufficiently defined. Sometimes, this shortfall could only be overcome by settling basic narratological problems that are connected to the categories in question. The question is now how the latter kind of problem should be handled within the frameworks of the project heureCLÉA. We see two possibilities:

(1) Ambiguous markup will be admitted, just as intended in the initial concept of the project,

⁴⁵ 'I have gazed outside all night, and methought this was how death must be like, or the after-death: over there and outside an infinite, hollowly roaring darkness. Will a thought, a notion of mine linger and weave on there and eternally hark to the intangible roaring?' (our translation)

⁴⁶ For reference, see the section "[heureCLÉA Corpus](#)" at the end of this paper.

⁴⁷ For the concepts of narrative levels, see Pier (2014).

albeit for slightly different reasons: while the categories as such are not necessarily of an interpretative nature, their unambiguous definition would presuppose basic theoretical decisions which lie beyond the scope of our project. The practical consequences are the same, though: the machine learning processes will be applied to ambiguous markup, which will result in non-deterministic markup suggestions when it comes to the implementation of our digital heuristic.

There are two reasons why this approach may be considered appropriate: first, the theoretical discussions at hand are obviously not of the kind that could or should be cut off rashly for pragmatic purposes—after all, narratologists have had good reason to discuss these issues for many years without a consensual solution in sight. While the solving of basic theoretical problems is not our main focus, it would be counterproductive nevertheless to answer open questions of principle through ad hoc stipulations. On the contrary, it can be theoretically fruitful to admit ambiguous markup in this case, for this allows an extensive collection of intuitions and arguments that can be attributed to different positions in the theoretical debates. In the end, such conflicting data may eventually help to solve the underlying difficult questions more constructively.

The second reason for not trying to settle these problems is a practical one: the complete determination of very detailed phenomena may be simply too complex for the machine learning processes to detect utilizable data. For example, if we try to settle the problems relating to narrative levels, it may become necessary not only to annotate the different narrative levels, but also to diagnose all the temporal narrative phenomena relatively to the respective narrative levels. The metadata generated in this approach may then easily overburden the machine learning approach. Accordingly, it may be better to leave the decision of how to determine a shift in narrative level in abeyance, and to allow texts with one narrative level only to the corpus as far as possible.

(2) Though the first way of handling deeper-rooted problems was the one we have pursued in heureCLÉA in most instances up to this point, we recently started to try out a different strategy: we actually engaged in a discussion on the basic problem in question. As for the annotation problems connected to the subject of narrative levels, it finally showed that the neglect of the theoretical problems in question entailed the disregard of too many relevant narratological details. The results of this attempt, though, are still awaited.

3.2. Automation of Complex Time Annotation: Current Outcomes

As a starting point for tackling the task of extracting temporal structure from a perspective of computational linguistics and data mining, we applied our temporal tagger HeidelTime to narrative texts. This was self-evident, as most work in these research areas tackle the problem of extracting temporal structures from a viewpoint of explicit temporal markers and apply it to texts, e.g., to news texts in the TempEval challenges (Verhagen et al., 2009; Verhagen et al., 2010; UzZaman et al., 2013).

Temporal tagging in isolation, however, proved to be insufficient for a deep analysis in particular due to the low number of temporal expressions occurring in literary-style texts. In Bögel et al. (2014), we compare a selection of German Wikipedia texts with our heureCLÉA corpus and show that—while the general structure of the documents is very similar in terms of sentence length etc.—literary-style texts contain only a fraction of temporal expressions

compared to non-fictional texts.

Thus, in addition to temporal expressions, further temporal information has to be extracted. For this, we implemented a system to extract the tense for "temporal clusters", representing a contiguous sequence of tokens belonging to the same sub-sentence. We chose subsentences as our annotation target based on a discourse perspective: often, one subsentence only refers to one single event in a narrative text, while a sentence might contain references to multiple events. In the end, we want to be able to extract and order all events occurring in narrative texts based on temporal aspects (tense, among others) as described similarly in Mani (2012). Thus, tagging sub-sentences ensures that the temporal cluster refers to all elements of one specific event, even though most words do not contain explicit tense information.

In order to empirically evaluate our approach, a group of annotators with a background in narratology annotated the tense of temporal clusters according to very fine-grained annotation guidelines which resulted in a pairwise agreement score of $K = 0.897$ (Fleiss et al., 1981), where agreement is measured on the token level and holds if two tokens (within a sub-sentence) are assigned the same tense by two independent annotators.

We complement the manual annotation by establishing a UIMA pipeline that includes HeidelTime as a temporal tagger, the TreeTagger (Schmidt, 1994) as a POS tagger, and Morphisto (Zielinski et al., 2009) to perform a morphological analysis. HeidelTime uses its rule base for temporal tagging. Based on the information extracted by this preprocessing pipeline, we developed a rule set to predict the tense of a sub-sentence (Bögel et al., 2014) using various temporal markers (auxiliaries, participle components etc.). To handle discontinuities, i.e., sub-sentences that do not contain any tense information, we use the neighboring sentences to the left and right to transfer the tense annotation if the context both to the left and right has the same tense. If the tenses of the neighboring sentences differs, the discontinuity is not annotated with any tense.

The best setting correctly tags about 95% of all tokens/verbs in the preterite—comprising the majority of all annotated tokens with more than 77%—with the correct tense. Even for rare tenses in our corpus, such as the future tense with only 122 annotated tokens, about 90% of all tokens are annotated correctly. Overall, these promising results indicate a high potential for reducing manual annotation efforts. While tense annotations themselves are not of particularly high interest for narratological research, they will be used as one feature for the prediction of more complex phenomena (e.g., order, duration/speed of narration).

Besides the perspective of annotation prediction, we also used the predicted tense-clusters for visualization by presenting a bird's eye view of the narration with respect to tense. A coarse-grained visualization of tense patterns within the text can be used as one possible entry point for a further analysis of interesting text passages (e.g., shifts and discontinuities of tenses).

3.2.1. Temporal Signals

We are currently developing a machine learning based system to annotate explicit and

implicit temporal signals⁴⁸ to facilitate deeper narratological research and order events in narratives. While there is a clear resemblance between temporal signals in narratives and temporal expressions (as extracted by HeidelTime), we adopt a hybrid approach that combines rule-based heuristics with machine learning to benefit from our existing temporal tagger and eliminate the need for exhaustive training data—which would be required by a sequence labeling approach that aims at predicting temporal signals right from the start.

As temporal signals in narratives are much more broadly defined than temporal expressions in typical NLP settings, our goal is not to extract temporal signals in a strictly rule-based manner, but to develop HeidelTime rules for temporal signals and combine them with a machine learning approach for validating the rule-based suggestions. Thus, the goal of HeidelTime is a recall optimization. For this, we first extend the rule-set of HeidelTime to extract an extra type of expression in addition to TimeML's four TIMEX3 categories date, time, duration and set expressions. For this, we model patterns, which partially contain part-of-speech constraints, for temporal signals based on insights of narratologists. This yields a very high recall (meaning that most of the temporal signals in the corpus are found) but causes a low precision as the system predicts many more temporal signals than are actually present in the corpus. This is due to the often-occurring ambiguities of terms which, depending on the context, either carry a temporal meaning or not.

This is where the machine learning part of our hybrid approach comes in to help increase precision and remove erroneously predicted temporal signals: by letting manual annotators correct the output of the modified version of HeidelTime, we are currently training a machine learning classifier that learns whether an annotation is correct or not. Due to the early stage of development, we refrain from reporting performance results for the classifier at the moment. However, our first results look promising.

Overall, we think that this hybrid setting that combines heuristics with complex machine learning is a reasonable approach to counterbalance the effect of data sparsity for machine learning as it reduces the complexity of the machine learning setting by employing robust rules to decrease the decision space.

3.3. Implementation and Integration of the Heuristic heureCLÉA Module into CATMA: Current Outcomes

In order to work with annotated data in CATMA, we implemented a component that interfaces CATMA with our UIMA pipeline. As CATMA is a popular tool in the humanities, we developed the interface as a stand-alone component that can easily be used by others to combine the strengths of CATMA as an annotation framework with the analytical and predictive power of UIMA pipelines. The interface is geared to literary scientists with no knowledge of programming. To achieve a simple configuration, the user only has to specify mappings between CATMA and UIMA types in a single XML file. By providing an easy-to-use interface, we want to lower the bar for other projects in the humanities to employ simple yet effective NLP tools and thereby alleviate manual annotations.

The tight collaboration between both project partners made it possible to develop an API that

⁴⁸ Note that although "temporal signals" are also defined in TimeML, we deal with "temporal signals" according to the annotation standards developed by the narratologists.

allows instant access to all the user's documents and annotations stored in CATMA. We therefore improved the TEI/XML-based import/export interface of CATMA, and the UIMA pipeline can be terminated with a component that writes annotations in TEI/XML⁴⁹ format. This allows us to read annotations from CATMA, use them for our predictions (as described in Section 3.2.1) and feed the results back to CATMA so that they can be presented in the annotation interface instantly. This will allow us to implement a feedback loop for our machine learning approach that facilitates instant updates to the underlying model by taking into account user feedback.

4. Conclusion

At this stage of the heureCLÉA project, the most interesting results do not yet concern the actual goal of facilitating an automated detection of temporal phenomena. While both the development of the collaborative annotation platform CATMA and of the temporal tagger HeidelTime have been brought forward in significant aspects, and while the first comparisons between manual and automated annotations of time expressions indicate the overall feasibility of our approach, we obviously still have to go a long way before we can begin to implement, test, and integrate the anticipated heureCLÉA digital heuristics module into CATMA.

However, the practical cases discussed in [section 3.1](#) clearly point to the conceptual benefits derived from the methodological decision to tolerate and, indeed, make fruitful humanistic and hermeneutic fuzziness in a digital humanities context. Ambiguous markup may not just be a matter of interpretation—that is: of the inconsistent or idiosyncratic application of a descriptive taxonomy by annotators—but rather a logical consequence of the theoretical under-determination of foundational humanistic (and in this particular instance: narratological) categories which has hitherto gone unnoticed, or which the original discipline normally ignores if not resolves by way of conceptual workarounds. In a humanist perspective, a digital humanities project like heureCLÉA provides an empirical testbed that can bring fundamental problems of theories to light—provided that the digital humanities methods employed remain sensitive to the fundamentally hermeneutic orientation of the client domain. In this regard, the heureCLÉA project exemplifies the third variant of a "computational narratology" in that—to repeat the definition in Mani (2013)—it serves the purpose of an "exploration and testing of literary hypotheses through mining of narrative structure from corpora".

We believe that a strong application orientation as well as the connection to a specific humanistic research question are crucial to the success of heureCLÉA: these characteristics provide for a mutual basis for the project partners from the humanities and computer science. The joint project is interdisciplinary, fundamentally collaborative and generic in nature. By developing a complex narratological tagset in the frameworks of the project, it is also ensured that the applied techniques and concepts can both be re-used⁵⁰ and expanded

⁴⁹ This component is an implementation of a CasConsumer (in the terminology of UIMA), that outputs the generated annotations as TEI/XML. It extracts temporal expressions as defined by the temporal markup language TimeML, i.e., so-called date, time, duration, and set expressions.

⁵⁰ Further processing of the annotations is possible due to their TEI/XML-compliant format, for further

to other narratological markup categories (e.g. space, character, perspective, etc.).

Submitted version for DHCommons

processing of the analysis results there is an export (Excel and csv) available, too. CATMA separates the definitions of the tags from the application of the definitions to the text, i. e. creating the actual tag instances. This enables reusability of the tag definitions for other projects.

References

- Abbott, H. P.** (2014). Narrativity. *the living handbook of narratology*. <http://www.lhn.uni-hamburg.de/article/narrative-levels> (accessed 14.08.2014). First published 2011.
- Adolphs, S.** (2006). *Introducing Electronic Text Analysis*. New York: Taylor & Francis.
- Allan, J.** (ed.) (2002). *Topic Detection and Tracking: Event-based Information Organization*. Boston/Dordrecht/London: Kluwer Academic Publishers.
- Alonso, O., Gertz, M., and Baeza-Yates, R.** (2009). Clustering and Exploring Search Results using Timeline Constructions. *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*. Hong Kong, pp. 97-106.
- Alonso, O., Strötgen, J., Baeza-Yates, R., and Gertz, M.** (2011). Temporal Information Retrieval: Challenges and Opportunities. *Proceedings of the 1st International Temporal Web Analytics Workshop (TAWAW 2011)*. Hyderabad, pp. 1-8.
- Baumann, S., Hoover, D., Van Dalen-Oskam, K., and Piez, W.** (2012). Text Analysis Meets Text Encoding. In Meister, J. C. (ed.), *Digital Humanities 2012: Conference Abstracts*. Hamburg: Hamburg University Press, pp. 33-35.
- Burnard, L.** (2001). On the hermeneutic implications of text encoding. In Fiormonte, D. and Usher, J. (eds.), *New Media and the Humanities: Research and Applications*. Oxford: Humanities Computing Unit, pp. 31-38. Available online in the 1998 version at <http://users.ox.ac.uk/~lou/wip/herman.htm> (accessed 14.08.2014).
- Buzzetti, D.** (2002). Digital Representation and the Text Model. *New Literary History*, 33 (1): 61-88.
- Chatman, S.** (1978). *Story and Discourse: Narrative Structure in Fiction and Film*. Ithaca: Cornell University Press.
- Coombs, J. H., DeRose, S. J., and Renear, A. H.** (1987). Markup systems and the future of scholarly text processing. *Communications of the ACM*, 30 (11): 933-47.
- De Toro, A.** (1986). *Die Zeitstruktur im Gegenwartsroman: am Beispiel von G. Garca Márquez' Cien años de soledad, M. Vargas Llosas La casa verde und A. Robbe-Grillet's La maison de rendez-vous*. Tübingen: Narr.
- Dowden, B.** (2013). Time. *Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/time/> (accessed 14.08.2014).
- Finlayson, M. A., Fisseni, B., Löwe, B., Meister, J. C.** (2013). Preface. *2013 Workshop on Computational Models of Narrative (CMN 2013)*. Hamburg, pp. vii-viii. <http://drops.dagstuhl.de/portals/oasics/index.php?semnr=13008> (accessed 14.08.2014).
- Fludernik, M.** (2005). Time in Narrative. In Herman, D., Jahn, M., and Ryan, M.-L. (eds), *Routledge Encyclopedia of Narrative Theory*, London: Routledge, pp. 608-612.

- Genette, G.** (1972). Discours du récit. In id., *Figures III*. Paris: Editions Du Seuil, pp. 67-282.
- Gius, E.** (2013). Erzählen über Konflikte. Eine computergestützte narratologische Untersuchung von narrativen Interviews zu Arbeitskonflikten. Dissertation, University of Hamburg, August 2013.
- Goldstone, A.** (2014) "Let DH be sociological". *Digital Humanities 2014, Lausanne*. <http://dharchive.org/paper/DH2014/Paper-756.xml> (accessed 14.08.2014).
- Grabes, H.** (2014). Sequentiality. the living handbook of narratology. <http://www.lhn.uni-hamburg.de/article/sequentiality> (accessed 14.08.2014). First published 2013.
- Gradmann, S. and Meister, J. C.** (2008). Digital Document and Interpretation: Re-thinking "Text" and Scholarship in Electronic Settings, *Poiesis and Praxis: International Journal of Ethics of Science and Technology Assessment*, 5 (2): 139-153.
- Hamburger, K.** (1957). *Die Logik der Dichtung*. Stuttgart: Klett.
- Hühn, P.** (2013). Event and Eventfulness. *the living handbook of narratology*. <http://www.lhn.uni-hamburg.de/article/event-and-eventfulness> (accessed 14.08.2014). First published 2011.
- Lagoni, F.** (in preparation). Fiktional / faktual. Ein historisch-narratologischer Vergleich literarischer Erzählformen. Dissertation, University of Hamburg.
- Lahn, S. and Meister, J. C.** (2013). *Einführung in die Erzähltextanalyse*: 2nd, updated edition. Stuttgart: Metzler.
- Lally, A., Gruhl, D., Epstein, E., Schor, M., Murdock, J.W., Frenkiel, A., Brown, E.W., Hampp, T., Doganata, Y., Welty, C., Amini, L., Kofman, G., Kozakov, L., and Mass, Y.** (2006). Towards an interoperability standard for text and multi-modal analytics. *IBM Research Report*.
- Lee, K., Burnard, L., Romary, L., De La Clergerie, E., Declerck, T., Baumann, S., Bunt, H., Clement, L., Erjavec, T., Roussanaly, A. and Roux, C.** (2004). Towards an international standard on feature structures representation. *4th International Conference on Language Resources and Evaluation (2004)*. Lisbon, pp. 373-376. <http://hal.inria.fr/inria-00099855> (accessed 14.08.2014).
- Mani, I.** (2013). Computational Narratology. *the living handbook of narratology*. <http://www.lhn.uni-hamburg.de/article/computational-narratology> (accessed 14.08.2014).
- Mani, I. and Wilson, G.** (2000). Temporal Processing of News. *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL 2000)*. Morristown, pp. 69-76.
- Margolin, U.** (1999). Of What Is Past, Is Passing, or to Come: Temporality, Aspectuality, Modality and the Nature of Literary Narrative. In Herman, D. (ed.) *Narratologies. New Perspectives on Narrative Analysis*. Columbus: Ohio State University Press: pp. 142-66.
- Markosian, N.** (2014). Time. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/time/> (accessed 14.08.2014). First published 2002.

- Mazur, P. and Dale, R.** (2009). The DANTE Temporal Expression Tagger. *Proceedings of the 3rd Language and Technology Conference (2009)*. Poznan, pp. 245-257.
- Mazur, P. and Dale, R.** (2010). WikiWars: A New Corpus for Research on Temporal Expressions. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010)*. Massachusetts, pp. 913-922.
- McCarty, W.** (1996). Implicit Patterns in Ovid's *Metamorphoses*. *Centre for Computing in the Humanities (CHWP 1996)*. http://projects.chass.utoronto.ca/chwp/mccarty/mcc_8.html (accessed 14.08.2014). First published 1991.
- McTaggart, J. M. E.** (1927). The Unreality of Time. In id., *The Nature of Existence*, vol. 2. Cambridge: Cambridge University Press. First published 1908.
- Meister, J. C.** (2005a). Tagging Time in Prolog: The Temporality Effect. *Literary and Linguistic Computing*, 20: 107-124.
- Meister, J. C.** (2005b). Computational Approaches to Narrative. In Herman, D., Ryan, M.-L. (eds.), *Routledge Encyclopedia of Narratology*. London/New York, pp. 78-80.
- Meister, J. C.** (2011). The Temporality Effect: Towards a Process Model of Narrative Time Construction. In Meister, J. C. and Schernus, W. (eds), *Time. From Concept to Narrative Construct. A Reader*. Berlin/New York: de Gruyter, pp. 171-216.
- Meister, J. C.** (2012). Crowd sourcing "true meaning". A collaborative markup approach to textual interpretation. In McCarty, W. and Deegan, M. (eds.), *Collaborative Research in the Digital Humanities. Festschrift for Harold Short*. Ashgate Publishers: Farnham, Surrey/Burlington, pp. 105-122.
- Meister, J. C.** (2014). Narratology. *the living handbook of narratology*. <http://www.lhn.uni-hamburg.de/article/narratology> (accessed 13.08.2014). First published 2011.
- Müller, G.** (1948). Erzählzeit und erzählte Zeit. In *Festschrift für Paul Kluckhohn und Hermann Schneider. Gewidmet zu ihrem 60. Geburtstag, hrsg. von ihren Tübinger Studenten*. Tübingen: Mohr, pp. 195-212.
- Pier, J.** (2014). Narrative Levels. *the living handbook of narratology*. <http://www.lhn.uni-hamburg.de/article/narrative-levels-revised-version-uploaded-23-april-2014> (accessed 14.08.2014). First published 2011.
- Piez, W.** (2010): Towards Hermeneutic Markup: an Architectural Outline. *Digital Humanities 2010. Conference Abstracts*. London: Office for Humanities Communication, Centre for Computing in the Humanities, King's College London, pp. 202-205.
- Pustejovsky, J., Castano, J., Ingria, R., Sauri, R., Gaizauskas, R., Setzer, A., Katz, G., and Radev, D.** (2004). TimeML: Robust Specification of Event and Temporal Expressions in Text. In Maybury, M. T. (ed.), *New Directions in Question Answering*. Menlo Park, Calif.: AAAI Press/The MIT Press, pp. 28-34.
- Pustejovsky, J., Knippen, R., Littman, J., and Sauri, R.** (2005). Temporal and Event Information in Natural Language Text. *Language Resources and Evaluation* 39 (2-3): 123-164.

- Pustejovsky, J., Lee, K., Bunt, H., and Romary, L.** (2010). ISO-TimeML: An International Standard for Semantic Annotation. *Proceedings of the 7th Edition of the Language Resources and Evaluation Conference (LREC 2010)*. Valletta, pp. 394-397.
- Renear, A. H.** (2000). The Descriptive/Procedural Distinction is Flawed. *Markup Languages: Theory and Practice 2*: 411–20.
- Renear, A. H.** (2004). Text Encoding. In Schreibman, S., Siemens, R., and Unsworth, J. (eds), *A Companion to Digital Humanities*. Oxford: Blackwell, pp. 218–39.
<http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&doc.view=print&chunk.id=ss1-3-5&toc.depth=1&toc.id=0> (accessed 14.08.2014).
- Schilder, F. and Habel, C.** (2001). From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages. *Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing (2001)*. Toulouse, pp. 65-72.
- Schmid, W.** (2008). *Elemente der Narratologie*: 2nd, improved edition. Berlin: de Gruyter.
- Schmidt, D.** (2010). The Inadequacy of Embedded Markup for Cultural Heritage Texts. *Literary and Linguistic Computing 25*: 337-356.
- Schmidt, H.** (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*. Manchester.
- Schüch, L.** (in preparation). Die Narrativität kontemporärer deutscher und englischer Songtexte. Dissertation, University of Hamburg.
- Simon-Shoshan, M.** (2013). Narrativity and Textuality in the Study of Stories. *2013 Workshop on Computational Models of Narrative (CMN 2013)*. Hamburg, pp. 228-237.
<http://drops.dagstuhl.de/opus/volltexte/2013/4136/pdf/p228-simon-shoshan.pdf> (accessed 14.08.2014).
- Strötgen, J. and Gertz, M.** (2010). HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions. *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval 2010)*. Uppsala, pp. 321-324.
- Strötgen, J. and Gertz, M.** (2012). Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards. *Proceedings of the 8th Edition of the Language Resources and Evaluation Conference (LREC 2012)*. Istanbul, pp. 3746-3753.
- Strötgen, J. and Gertz, M.** (2013). Multilingual and Cross-domain Temporal Tagging. *Language Resources and Evaluation 47 (2)*: 269-298.
- Strötgen, J., Armiti, A., Tran, C.V., Zell, J., and Gertz, M.** (2014a). Time for More Languages: Temporal Tagging of Arabic, Italian, Spanish, and Vietnamese. *ACM Transactions on Asian Language Information Processing 13 (1)*: 1-21.
- Strötgen, J., Bögel, T., Zell, J., Armiti, A., Tran, C.V., and Gertz, M.** (2014b). Extending HeidelTime for Temporal Expressions Referring to Historic Dates. *Proceedings of the 9th Edition of the Language Resources and Evaluation Conference (LREC 2014)*. Reykjavik, pp.

2390-2397.

UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., and Pustejovsky, J. (2013). SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)* Atlanta, pp. 1–9.

Verhagen, M., Sauri, R., Caselli, T., and Pustejovsky, J. (2010). Semeval-2010 Task 13: TempEval-2. *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval 2010)* Uppsala, pp. 57–62.

Wheeler, M. (2011). Martin Heidegger. *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/spr2013/entries/heidegger/> (accessed 14.08.2014).

Wilensky, R. W. (1978). *Understanding Goal-Based Stories (Outstanding Dissertations in the Computer Sciences)*, New York: Garland Publishing.

Wittern, C. (1999). "RE: 13.0281 perfectability of texts". *Humanist Discussion Group no. 289*. <http://www.digitalhumanities.org/humanist/Archives/Virginia/v13/0289.html> (accessed 14.08.2014).

Zarri, G. P. (2009). *Representation and management of narrative information: theoretical principles and implementation*. Advanced information and knowledge processing. London: Springer.

Zielinski, A., Simon, C., and Wittl, T. (2009). Morphisto: Service-Oriented Open Source Morphology for German. *State of the Art in Computational Morphology: Workshop on Systems and Frameworks for Computational Morphology (SFCM 2009)*. Zürich, pp. 64-75.

Submitted version for PHO Commons

Standards

Feature Structures:

- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html> (accessed 14.08.2014). (“P5: Guidelines for Electronic Text Encoding and Interchange. 18 Feature Structures”).
- http://www.iso.org/iso/catalogue_detail.htm?csnumber=37324 (accessed 14.08.2014). (“ISO 24610-1:2006. Language resource management -- Feature structures -- Part 1: Feature structure representation”).

Temporal Information:

- http://www.iso.org/iso/catalogue_detail.htm?csnumber=37324 (accessed 14.08.2014). (“ISO 24610-1:2006. Language resource management -- Feature structures -- Part 1: Feature structure representation”).

TimeML:

- http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37331 (accessed 14.08.2014). (“ISO 24617-1:2012. Language resource management – Semantic annotation framework (SemAF) -- Part 1: Time and events (SemAF-Time, ISO-TimeML)”), see also Pustejovsky et al. (2010, pp. 394-397).

XML Metadata Interchange:

- <http://www.omg.org/spec/XMI/ISO/19509/PDF/> (accessed 14.08.2014) (“ISO/IEC 19509 -- Information technology -- Object Management Group XML Metadata Interchange (XMI)”).

Submitted version for JEL Commons

heureCLÉA Corpus

Auerbach, B. (1893). Die Kriegspfeife. In id., *Gesammelte Schriften, 2. Gesamtausgabe*, vol. 1, Stuttgart: Cotta. <http://www.textgridrep.de/browse.html?id=textgrid:k8dc.0> (accessed 13.08.2014).

Christen, A. (1964). Nachbar Krippelmacher. In id., *Das Haus zur Blauen Gans. Erzählungen und Gedichte*, ed. by Hanna-Heide Kraze, Berlin: Union Verlag. <http://www.textgridrep.de/browse.html?id=textgrid:kxd1.0> (accessed 13.08.2014). First published 1884.

Ebner-Eschenbach, M. v. (1956 - 1958). Krambambuli. In id., *Gesammelte Werke in drei Bänden*, ed. by Johannes Klein, München: Winkler. <http://www.textgridrep.de/browse.html?id=textgrid:mktb.0> (accessed 13.08.2014). First published 1883.

Eichendorff, J. v. (1970ff.). Auch ich war in Arkadien. In id., *Werke*, ed. by Ansgar Hillach, München: Winkler. <http://www.textgridrep.de/browse.html?id=textgrid:msfh.0> (accessed 13.08.2014). First published 1866.

Frapan, I. (1908). Die Liebe ist gerettet. In id., *Zwischen Elbe und Alster. Hamburger Novellen*, 3rd edition, Berlin: Gebrüder Paetel. <http://www.textgridrep.de/browse.html?id=textgrid:n27x.0> (accessed 13.08.2014).

Hebbel, F. (1963). Matteo. In id., *Werke*, ed. by Gerhard Fricke, Werner Keller and Karl Pörnbacher, München: Hanser. <http://www.textgridrep.de/browse.html?id=textgrid:pksg.0> (accessed 13.08.2014). First published 1841.

Hemingway, E. (1950). Die Mutter eines Schwulen. In id., *49 Stories*. Hamburg: Rowohlt. First published 1933.

Heym, G. (1960ff.). Ein Nachmittag. In id., *Dichtungen und Schriften. Gesamtausgabe*, ed. by Karl Ludwig Schneider, Hamburg, München: Ellermann. <http://www.textgridrep.de/browse.html?id=textgrid:q3jt.0> (accessed 13.08.2014).

Hofmannsthal, H. v. (1979). Reitergeschichte. In id., *Gesammelte Werke in zehn Einzelbänden. Erzählungen, Erfundene Gespräche und Briefe, Reisen*, ed. by Bernd Schoeller with Rudolf Hirsch, Frankfurt a.M.: S. Fischer. <http://www.textgridrep.de/browse.html?id=textgrid:qdbf.0> (accessed 13.08.2014). First published 1899.

Janitschek, M. (1897). Poverino. In id., *Kreuzfahrer*, Leipzig: Verlag Kreisende Ringe (Max Spohr). <http://www.textgridrep.de/browse.html?id=textgrid:gmpm.0> (accessed 13.08.2014).

Mann, T. (2004). Der Tod. In id., *Große kommentierte Frankfurter Ausgabe. Werke – Briefe – Tagebücher*, ed. by Heinrich Detering et al., vol. 2.1: *Frühe Erzählungen. 1893–1912*, ed. by Terence J. Reed. Frankfurt a. M.: S. Fischer. First published 1897.

Proelß, J. (1883). Lili. In id., *Katastrophen. Poetische Bilder aus unserer Zeit*, Stuttgart:

Adolph Bonz & Comp. <http://www.textgridrep.de/browse.html?id=textgrid:t4sp.0> (accessed 13.08.2014)

Gräfin zu Reventlow, F. (1980): Das polierte Männchen. In id., *Autobiographisches. Ellen Olestjerne. Novellen, Schriften, Selbstzeugnisse*, ed. by Else Reventlow. München: Langen Müller. <http://www.textgridrep.de/browse.html?id=textgrid:tbcm.0> (accessed 13.08.2014).

First published 1917.

Rilke, R. M. (1955 - 1966). Die Turnstunde. In id., *Sämtliche Werke* ed. by Rilke-Archiv with Ruth Sieber-Rilke, Wiesbaden and Frankfurt a.M.: Insel.

<http://www.textgridrep.de/browse.html?id=textgrid:tbz9.0> (accessed 13.08.2014). First published 1902.

Ringelnatz, J. (1994). Nervosipopel [1921]. In id., *Das Gesamtwerk in sieben Bänden*, ed. by Walter Pape, Zürich: Diogenes. <http://www.textgridrep.de/browse.html?id=textgrid:tf0b.0> (accessed 13.08.2014). First published 1921.

Schnitzler, A. (2004). Blumen. In id., *Ausgewählte Werke in acht Bänden*, ed. by Heinz Ludwig Arnold, vol. 1: *Leutnant Gustl. Erzählungen 1892–1907*. Frankfurt a. M.: S. Fischer. First published 1894.

Storm, T. (1967). Veronika. In id., *Sämtliche Werke in vier Bänden*, ed. by Peter Goldammer, Berlin and Weimar: Aufbau.

<http://www.textgridrep.de/browse.html?id=textgrid:vtkx.0> (accessed 13.08.2014). First published 1861.

Thoma, L. (1968). Meine erste Liebe. In id., *Gesammelte Werke in sechs Bänden. Erweiterte Neuauflage*, ed. by Albrecht Knaus, München: Piper.

<http://www.textgridrep.de/browse.html?id=textgrid:wc7q.0> (accessed 13.08.2014)

Tieck, L. (1963). Der Pokal. In id., *Werke in vier Bänden*, ed. by Marianne

Thalman, München: Winkler. <http://www.textgridrep.de/browse.html?id=textgrid:wdz7.0> (accessed 14.08.2014). First published 1812.

Trakl, G. (1972). Traumland. Eine Episode. In id., *Das dichterische Werk*, ed. by Walther Killy, Hans Szklénar and Friedrich Kur, München: Deutscher Taschenbuchverlag.

<http://www.textgridrep.de/browse.html?id=textgrid:wfkr.0> (accessed 13.08.2014). First published 1906.

Wedekind, F. (1969). Die Schutzimpfung. In id., *Werke in drei Bänden*, vol. 3: *Prosa*. Berlin and Weimar, Aufbau-Verlag, 1969. <http://www.textgridrep.de/browse.html?id=textgrid:x36s.0> (accessed 13.08.2014). First published 1903.