

# **A Shared Task for the Digital Humanities: Annotating Narrative Levels\***

Special Issue of *Cultural Analytics*

Evelyn Gius, Nils Reiter, Marcus Willand (editors)

\*This PDF was compiled using the articles provided by *Cultural Analytics* to allow citation and convenient browsing of the entire issue. Please cite individual articles as indicated on their respective title page. Please cite the entire issue as follows:

Evelyn Gius, Nils Reiter, and Marcus Willand, eds. (2019). *Cultural Analytics: A Shared Task for the Digital Humanities: Annotating Narrative Levels*. Special Issue. DOI.

# Contents

<i>Evelyn Gius, Nils Reiter, Marcus Willand</i>		
1	Foreword . . . . .	4
<b>I</b>	<b>Introduction</b>	<b>8</b>
<i>Nils Reiter, Marcus Willand, Evelyn Gius</i>		
2	Introduction to Annotation, Narrative Levels and Shared Tasks . . . . .	9
<i>Evelyn Gius, Nils Reiter, Marcus Willand</i>		
3	Evaluating Annotation Guidelines . . . . .	33
<i>Marcus Willand, Evelyn Gius, Nils Reiter</i>		
4	Description of Submitted Guidelines and Final Evaluation Results . . . . .	44
<b>II</b>	<b>Annotation Guidelines and Reviews</b>	<b>59</b>
<i>Joshua Eisenberg, Mark Finlayson</i>		
5	Annotation Guideline 1 . . . . .	60
<i>Meredith A. Martin</i>		
6	Review of Guideline 1 . . . . .	81
<i>Edward Kearns</i>		
7	Annotation Guideline 2 . . . . .	84
<i>Tilmann Köppe</i>		
8	Review of Guideline 2 . . . . .	96
<i>Nora Ketschik, Benjamin Krautter, Sandra Murr, Yvonne Zimmermann</i>		
9	Annotation Guideline 4 . . . . .	100
<i>J. Berenike Herrmann</i>		
10	Review of Guideline 4 . . . . .	113

## Contents

*Florian Barth*

11 Annotation Guideline 5 . . . . . 116

*Jan Horstman*

12 Review of Guideline 5 . . . . . 128

*Matthias Bauer, Miriam Lahrsow*

13 Annotation Guideline 6 . . . . . 131

*Natalie M. Houston*

14 Review of Guideline 6 . . . . . 145

*Mats Wirén, Adam Ek and Anna Kasaty*

15 Annotation Guideline 7 . . . . . 148

*Gunther Martens*

16 Review of Guideline 7 . . . . . 170

*Adam Hammond*

17 Annotation Guideline 8 . . . . . 176

*Tom McEnaney*

18 Review of Guideline 8 . . . . . 184

# Foreword to the Special Issue “A Shared Task for the Digital Humanities: Annotating Narrative Levels”

Evelyn Gius, Nils Reiter, Marcus Willand

08.20.19

*Article DOI: 10.22148/16.047*

*Journal ISSN: 2371-4549*

*Cite: Evelyn Gius, Nils Reiter, and Marcus Willand, “Foreword to the Special Issue ‘A Shared Task for the Digital Humanities: Annotating Narrative Levels,’” Journal of Cultural Analytics. November 4, 2019. doi: 10.22148/16.047*

This volume is the first of two, and it documents activities that we have been conducting in the past years. They are best described as “organizing shared tasks with/in/for the digital humanities” and have evolved significantly since we started.

Research in digital humanities entails a number of unique challenges, some of which are caused by the collaboration model that digital humanities projects often work in. This collaboration creates a lot of friction, but comes with huge potential: Different collaboration partners can *specialize* in different aspects of the shared goal. While a shared goal and a common language is still needed, each party does not have to know everything about the other party’s focus area. A (proto-)typical division of labor lends itself to the digital humanities: Computer scientists work on the technical aspects, while humanities scholars focus on the content side. Still, it turns out that knowing what each party is working on does not suffice, because content and technology need to be re-integrated at some point. How exactly this integration takes place depends on the specific project design and is a matter of *interface*.

While our initiative is concerned with narrative levels in concrete terms, it does, more abstractly, also establish such an interface: when it comes to transporting knowledge about the research subject itself, annotated data serves as an interface between humanists and computer scientists. In addition to annotated data as interface, the shared task format itself can already be seen as an interface for scholars and researchers to interact. This interaction does not have to be direct and it does not have to take place in the same project, the same country, or even in the same decade.

The initiative consists of two separate, but tightly linked shared tasks. The first one focuses on annotation guidelines for narrative levels and produces a balanced and consensual assessment of guidelines. The guideline that is best suited to the goal is then used for corpus annotation. The second shared task aims at automatizing the detection of narrative levels and will employ the annotated corpus in order to achieve this. The core benefit of this approach is that the resulting automatic detection systems incorporate the conceptual thinking that went into the guidelines in the first task. Decisions on the complexity and granularity of the concepts to be detected are made by the scholars who developed the annotation guidelines, and they do not have to make compromises for pragmatic or technical reasons. Consequently, fast success in terms of automatization is not guaranteed. The automatization task might be challenging for years to come, but at least the task definition is adequate for analyzing literature in the future.

Since this is a new format that has not been employed before in the digital humanities, this volume contains an extensive introduction covering the motivation and reasoning behind the first shared task in detail, a discussion of the evaluation setup and, finally, the guidelines as they were submitted, discussed and evaluated during a workshop we held.

Writing guidelines, however, is an iterative process. Therefore, this special issue will receive an update in the form of a second volume, which will contain the improved guidelines, based on the discussion and evaluation from the first shared task.

Shared tasks depend entirely on their participants. Initially, we could not be certain at all that this activity would attract a critical mass of interested researchers and scholars. As of now, we are very happy that such a diverse crowd participated in the first shared task, and we would like to emphasize their commitment and thank them sincerely for not only having discussed narrative levels with us in a remarkably intensive way but also, in some cases, taking on transatlantic flights to participate.

The participants of the first phase of the shared task are:<sup>1</sup>

- Matthias Bauer, guideline VI, English literature, Tübingen University, Germany
- Florian Barth, guideline V, digital humanities/literary studies, Stuttgart University, Germany
- Kristina Burghardt, guideline VI, English literature, Tübingen University, Germany
- Joshua Eisenberg, guideline I, natural language processing, Florida International University, Miami, U.S.A.
- Adam Ek, guideline VII, computational linguistics, Stockholm University, Sweden
- Mark Finlayson, guideline I, natural language processing, Florida International University, Miami, U.S.A.
- Adam Hammond, guideline VIII, English literature, University of Toronto, Canada
- Anna Kasaty, guideline VII, computational linguistics, Stockholm University, Sweden
- Edward Kearns, guideline II, English, National University of Ireland Galway, Ireland
- Nora Ketschik, guideline IV, literary studies/digital humanities, Stuttgart University, Germany
- Benjamin Krautter, guideline IV, (computational) literary studies, Stuttgart University, Germany
- Miriam Lahrsow, guideline VI, English literature, Tübingen University, Germany
- Sandra Murr, guideline IV, literary studies/digital humanities, Stuttgart University, Germany
- Ella Ujhelyi, guideline VI, English literature, Tübingen University, Germany
- Mats Wirén, guideline VII, computational linguistics, Stockholm University, Sweden
- Yvonne Zimmermann, guideline IV, literary studies, Stuttgart University, Germany

There are, next to the participants, a number of people that supported this initiative in various stages, for which we are very grateful. We thank Jannik Strötgen, who was involved in the initial steps of planning, but has left academia since.

---

<sup>1</sup>There were eight submissions, but Guideline III was withdrawn after the workshop. While it was still evaluated as the others, we therefore leave its authors anonymous.

We are also thankful to our advisory board, consisting of Janina Jacke, Fotis Jan-nidis, Jonas Kuhn, and Jan Christoph Meister. The shared task has been—and still is—supported by the Volkswagen foundation, which generously funded the workshop in Hamburg and subsequent work. The Centre for Reflected Text Ana-lytics (CRETA) at Stuttgart University provided the funding for the student anno-tators. We thank CRETA for the funding and Hanna Winter, Tanja Preuß, Nina Stark, and Linda Kessler for the annotation work. Katharina Krüger and Carla Sökefeld supported the realization of the workshop and did a lot of preparatory work, Carla Sökefeld and Felicitas Otte supported the writing of this introduc-tion. We would also like to thank the authors of the guideline reviews who will-ingly agreed to be part of this endeavour. And finally we would like to express our warmest thanks to Andrew Piper and the editorial board of Cultural Analyt-ics for their flexibility to publish this—at least up until now—unorthodox format as a special issue.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

**Part I.**

# **Introduction**



# A Shared Task for the Digital Humanities

## Chapter 1: Introduction to Annotation, Narrative Levels and Shared Tasks

Nils Reiter, Marcus Willand, Evelyn Gius

08.20.19

*Article DOI: 10.22148/16.048*

*Journal ISSN: 2371-4549*

*Cite: Nils Reiter, Marcus Willand, Evelyn Gius, "A Shared Task for the Digital Humanities Chapter 1: Introduction to Annotation, Narrative Levels and Shared Tasks," Journal of Cultural Analytics. November 4, 2019. doi: 10.22148/16.048*

Annotation guidelines for literary phenomena are a clear desideratum within the field of text-oriented digital humanities. Creating guidelines that are widely applicable, however, is almost only possible in large annotation projects, which are naturally expensive. Moreover, scholars interested in large-scale analyses of literary texts are required to perform a lot of tasks that are outside of their core expertise, while researchers from computer science interested in method development for literary texts are required to create annotated data by themselves. Shared tasks, a workshop and research format that is popular in natural language processing, are a way to address both issues at the same time. This volume documents the setup and the results of the first shared task conducted within the digital humanities. The shared task started in May 2018 and is the first one that has the development of annotation guidelines as its main goal.

This special issue comes in two volumes. The first one is structured as follows: In this introduction (Chapter 1), we will cover the goals and underlying motivations of the project, describe basic assumptions on (this kind of) annotation, give background on narratological theory and on the role of narrative levels in text analysis and introduce our shared task procedure. Chapter 2 ("Evaluating Annotation

Guidelines”) explains how the submitted annotation guidelines have been evaluated. This also includes a description of the metric used for inter-annotator agreement. Chapter 3 (“Annotation Guidelines Overview and Evaluation Results”) provides a structured overview and comparison of the guidelines and presents the evaluation results. The remaining chapters document the annotation guidelines and contain an introductory rationale for each guideline and a review. The guidelines are published as they were submitted (besides layout and minor language editing). Thus, the evaluation results are based on the guidelines you find in this volume.

In sum, this first volume documents the preparatory work and the results of the workshop we held to complete the first shared task. Since the discussions and insights of this workshop gave rise to a large number of improvements to the guidelines, we decided to publish the revised guidelines as well. The improved guidelines, which document the final outcomes of the first shared task, will be published in the second volume of the special issue.

Please note that this shared task (called SANTA, for “Systematic Analysis of Narrative levels Through Annotation”) will be followed by a second one, with the goal of automatic detection of narrative levels.

## Motivation

This project addresses two issues prevalent in digital humanities and computational literary studies: The distribution of labor, competences, and tasks in the interdisciplinary research field of digital humanities and the inter-subjective manual and reliable automatic recognition of narrative levels in narrative texts.

### Distribution of Labor, Tasks and Competences

Given the current state of computational analysis of narrative texts<sup>1</sup> digital humanities projects that aim at analyzing content-related aspects of such texts *on a large scale* need to make technical-methodological progress in order to automatically detect the phenomena of interest. Therefore, many such projects are collaborative projects between researchers from computer science/natural language processing and literary or cultural studies. Although there is a growing number

---

<sup>1</sup>Performance on narrative texts is not systematically evaluated, but can be expected to be less than what is considered the state of the art: <https://nlpprogress.com>

of tutorials, how-tos and textbooks for various digital humanities topics,<sup>2</sup> the daily organization of such digital humanities projects remains challenging for a number of reasons:

Developing a shared language and common understanding of the subject at hand is one of the first tasks that new digital humanities projects often have to tackle. At times, computer scientists are only interested in the methodological part (without interpreting the results in reference to the texts under examination), while humanities scholars typically focus on conceptual issues or interpretation of the results. Thus, the individual goals of partners might be different even within the same project.

We believe that formats such as this adapted shared task offer unique opportunities to members of the digital humanities community with both backgrounds. In such a shared task, participants can focus on what they do best. Literary scholars can focus on the literary phenomenon that they are interested in and experienced with. Given their disciplinary routines and text experience, they are best qualified for exploring, defining, and exemplifying the narratological concepts without worrying about the implementability of their concepts or about making their findings automatable. Restricting oneself to simpler concepts just because one thinks they might be easier to detect automatically is a dead end for methodological innovation, as the limitations of computers are often only hearsay and constantly evolving. Moreover, if the conceptual complexity has been included in an annotation guideline that can be applied inter-subjectively, and a corpus with annotated concepts has been created, any computer scientist and/or machine learning expert can work on the automatic detection of the concepts, even if they are not experts in narratology (because a “ground truth” is available in the annotations). Similarly, as the shared task provides an empirical evaluation that can be trusted, machine learning models do not have to be transparent or explainable. When applying machine learning in a digital humanities scenario, there is often a trade-off between performance and transparency: Machine learning models that achieve better performance (e.g. neural networks) may be less transparent, while transparent models (e.g. decision trees) often lack in performance. In this case, because of the empirical evaluation, computer scientists can opt for the best performance.

---

<sup>2</sup>Among many others, cf. Susan Schreibman, Ray Siemens and John Unsworth, eds., *Companion to Digital Humanities* (Blackwell, 2004), <http://www.digitalhumanities.org/companion/>; Ray Siemens and Susan Schreibman, eds., *A Companion to Digital Literary Studies* (Oxford: Wiley-Blackwell, 2008); Matthew Lee Jockers, *Text Analysis with R for Students of Literature* (Cham: Springer, 2014); Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, eds. *Digital Humanities. Eine Einführung* (Stuttgart: Metzler, 2017); Anandi Silva Knuppel and Maria José Afanador Llach, eds., “Programming Historian,” accessed January 17, 2019, <https://programminghistorian.org>.

The two shared tasks we are organizing focus on two different sides of annotation. The first task, which consists of guideline development, forms the basis for an independent and reliable empirical evaluation of the automatic detection systems later on. Thus, a machine learning model that performed well in the second task may be safely used for new texts of the same kind as the test data (which is again transparent to scholars).

In addition to allowing everyone to focus on their field of expertise, this setup also renders a decoupling of the conceptual from the technical work possible. Scholars can focus on the development of annotation guidelines. This includes conceptual work, as well as a first step to operationalize scholarly concepts (to the extent of being applicable in an intersubjective manner). In this shared tasks model, the scholars do not have to be in the same project, at the same university or even on the same continent as the researchers developing the automatic detection tools (which includes technical work). This lowers entry barriers, as one does not have to work in a well-funded, interdisciplinary project in order to contribute to the overarching goals. Instead, scholars and researchers can contribute to the shared task at their own pace and integrate this single contribution more easily into their own research agendas. Moreover, this is possible without an augmentation of the workload in interdisciplinary collaborations.

### **Annotation Guidelines for Narrative Levels**

The detection of narrative levels and through that the identification of coherent text parts is required for the analysis of narrative texts to facilitate subsequent, content-related literary research based on the data obtained (about plot, characters, narrated world, etc.). While there is no exact statistics on this, narrative levels are such a common phenomenon that they are very often not even explicated in literary studies. Thus, automatically detecting narrative levels is a crucial contribution and groundwork research in the field of computational literary studies. Moreover, narrative levels can be a mediator connecting hermeneutic and automatic text analysis. Even though the complexity of narrative levels is considered comparably low from a literary studies point of view and comparably high from a natural language processing perspective, it is potentially relevant for text analysis of all sorts. Additionally, in comparison to other phenomena narrative levels are a rather little disputed phenomenon within literary studies. Finally, the definitions of narrative levels are usually based on textual and narrative features. For example, verbs of utterance and subsequent direct speech can be textual signals for narrative levels as well as the presence of a different story world that can be identified through the analysis of space or other narrative phenomena. Narrative

levels are therefore useful for the analysis of texts displaying a divergence between their textual structure and the structure of the narrated.

In sum, we consider narrative levels a good choice for a shared task. Its most important quality for our purposes is that it can bridge the gap between the theoretical discussion of a phenomenon and the application in text mining.

Most of today's text processing software is based on machine learning of various types. Given their interdependence with other phenomena as well as surface and content level characteristics, machine learning is the prime technique to automatically detect levels in texts. Machine learning, however, can only be successfully applied if training and testing data is available in high quantities. Such data needs to be annotated, i.e., it is a necessity to have texts in which narrative levels are already marked. These annotated texts can then be used to train models to detect levels in new, not previously annotated texts.

Annotation guidelines are needed not only to ensure the coherence of the annotations, but also to deal with unusual cases and to allow the annotation to be done by non-experts. Since annotation processes are expensive and time-consuming, it is unrealistic that different theoretical approaches to one concept (e.g. narrative levels) will be used as a basis for the annotation of this concept. Thus, a certain conceptual agreement within the community needs to be reached beforehand. Ideally, it should be one that leads to annotations which are useful for as many scholars as possible, even the ones with a different theoretical understanding of the concept. In the case of narrative levels, this can be shown by the question of whether simple or complex level concepts should be used. On the one hand, a less complex concept might reach higher inter-annotator agreement, but on the other hand, it might also lead to annotations that are less interesting for literary scholars to work with, as more differentiated concepts are thought to cover more complex literary phenomena.

In sum, annotation guidelines are a core ingredient towards automatic recognition of a concept. To ensure the scholarly usefulness of the resulting automatic recognition tools, experts in narratology need to be involved in the process of guideline creation.

## **Annotation**

The term "annotation" is used with different meanings within the digital humanities community. In our project, the term is used for the process of marking segments of a text to belong to a defined category. We also assume that such

categories are determined beforehand (ruling out exploratory or explanatory annotation) and that their detection is based on the contents of the text and not on structure or formatting (ruling out the annotation of e.g. text structure in TEI XML).<sup>3</sup> This also entails that detecting these categories is not trivial and requires text understanding and a certain level of text interpretation.

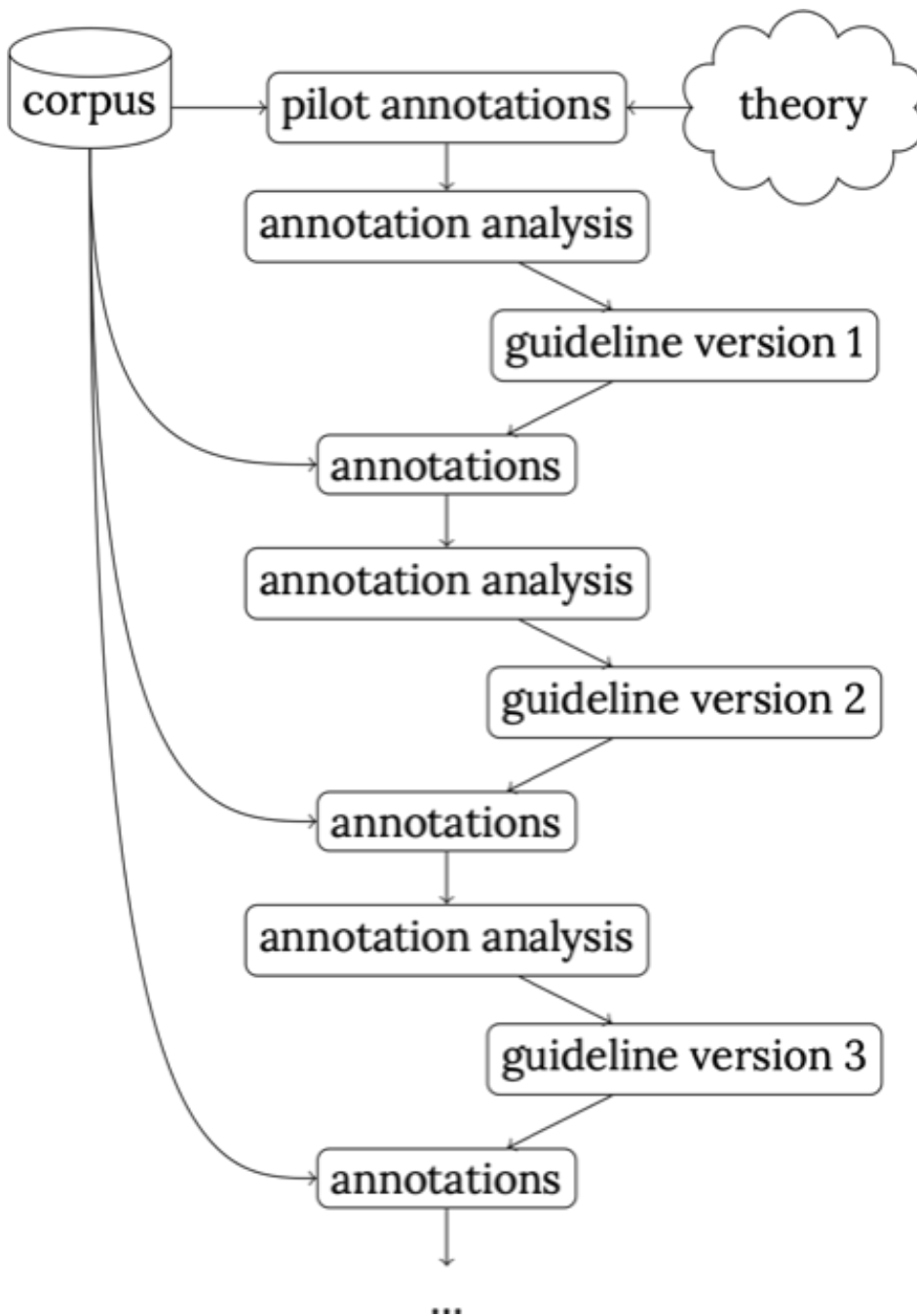
This notion of annotation is most similar to the linguistic notion of annotations of, for instance, coreference chains or semantic roles.<sup>4</sup> There are, however, a number of properties of narrative annotations that need to be taken into account for the annotation workflow: A narrative level may need to be annotated in a large portion of the entire text, while semantic role fillers are typically constrained to single noun phrases. In addition, the relevant context is typically much larger. For the linguistic annotation tasks that have been subject of shared tasks in the past, a context window of a single sentence is sufficient. Annotating coreference chains is the exception here, as it is typically considered a document level task and requires full document knowledge. Narrative annotations regularly consider the entire document as relevant co(n)text, thus requiring full text knowledge of the annotators. It is entirely conceivable (but not easy to implement in reality) to also consider text-external sources as relevant context (e.g., socio-historic conditions). This larger context has the potential to make narrative annotations more interpretative than linguistic ones.

---

<sup>3</sup><http://www.tei-c.org>

<sup>4</sup>Annotating coreference chains is the task of identifying which mentions of some entity refer to the same one (e.g., in “A house was bought by Mary. Peter loves her”, the pronoun “her” refers to Mary). Identifying semantic roles would tell us that Mary is the agent of the first sentence, and “a house” is the patient or theme (i.e., the thing that has been bought).

## Annotation Process



The annotation process that we have in mind is iterative and tightly connected to the development of a guideline. This iterative process is depicted in Figure 1 and is clearly related to the MATTER cycle.<sup>5</sup> In each step, we not only increase the amount of annotated texts, but the annotation guideline is improved as well. Of course, changes in the annotation guideline need to be reflected: They might change how previous portions of the texts should have been annotated, which

<sup>5</sup>J. Pustejovsky and A. Stubbs, *Natural Language Annotation for Machine Learning* (Sebastopol, CA: O'Reilly Media, Inc., 2013), 23ff.

should then be updated as well. The core idea in this annotation process—with the goal of producing coherent and inter-subjective annotations—is to have multiple annotators annotate the same texts in parallel, at least for some of the data. This allows the inspection and comparison of annotations and thus the identification of issues in the guideline. While asking the annotators for their impressions on the annotation process is valuable, not all issues are easily noticeable by annotators. Comparing annotations of the same texts with the same annotation guideline quickly reveals these possible issues. This annotation workflow has been employed by Evelyn Gius and Janina Jacke<sup>6</sup> on narrative time phenomena and it has obvious parallels to the “hermeneutic circle” that describes a general epistemological pattern in the humanities. If used in this way, the annotation workflow (and the iterative refinement of the annotation guidelines) has repercussions on the theoretical level and can be used productively for the development and refinements of theoretical concepts.<sup>7</sup>

### Annotation Guidelines

The goal of this annotation process is to produce coherent and systematic annotations. To this end, the annotations are done with the help of annotation guidelines. Annotation guidelines mediate between a specific theoretical understanding of concepts (like that of a narrative level) and the practical annotation of the concept in texts. They have multiple purposes, all of them directed towards the explication of theoretical concepts and/or the process of annotation:

1. Fill the gaps: Theories are often not specific enough to be used directly. In order to be as abstract as possible, they typically neglect many details and leave them underspecified (e.g. how to handle dashes marking insertions). These cannot be decided individually by annotators during the annotation process and thus need to be defined beforehand.
2. Provide examples: Ideally, an annotation guideline makes it possible for non-experts in narratology to also perform annotation. To this end, examples are provided, and/or replacement/insertion tests are formulated.
3. Make text-specific adaptations: Even for relatively simple phenomena in linguistics (e.g. parts of speech), existing annotation guidelines cannot be

---

<sup>6</sup>“The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis,” *International Journal of Humanities and Arts Computing* 11, no. 2 (October 2017): 233-54.

<sup>7</sup>Cf. Janis Pagel, Nils Reiter, Ina Rösiger, and Sarah Schulz, “A Unified Annotation Workflow for Diverse Goals,” in *Proceedings of the Workshop on Annotation in Digital Humanities, co-located with ESSLLI 2018*, ed. Sandra Kübler and Heike Zinsmeister (Sofia, Bulgaria, 2018) for a general-purpose workflow description.



expected to be all-encompassing, because the variability and creativity of human language production is enormous, and new text types are appearing constantly (consider part of speech tagging on twitter data). Annotation guidelines are a means to address phenomena that are text or genre specific.

4. Provide a log: Finally, most annotation processes accumulate a lot of procedural knowledge, as decisions on edge cases have to be made on a daily basis. An annotation guideline also serves the purpose of a log to document these decisions and make them traceable by other researchers.

### Annotation Analysis

Agreement between annotators is a major goal of this kind of annotation: Two annotators, who annotate the same text with the same annotation guideline are *generally* expected to produce the same annotations.<sup>8</sup> Inspecting annotations with respect to their achieved agreement is consequently a major component of the annotation analysis step in Figure 1.

The regular **discussion of annotation decisions** with the actual annotators is an effective way of learning about issues in the guideline. Asking annotators to explain their decisions (in particular if they have been diverging or difficult) not only keeps their attention up, it also reveals misunderstandings and/or highlights areas in which the guideline can be improved.

In addition, the amount of agreement between annotators can be quantified. This is known as inter-annotator agreement, and numerous metrics have been proposed for different kinds of annotation tasks.<sup>9</sup> All metrics aim at striking a balance between observed agreement and expected agreement. While the former

---

<sup>8</sup>There are exceptions, in particular regarding literary texts. In these cases, polyvalent text readings might lead to different annotations which constitute a case of justified disagreement. cf. Evelyn Gius and Janina Jacke, "The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis" *International Journal of Humanities and Arts Computing* 11, no. 2, (2017), 233-54.

<sup>9</sup>Joseph L. Fleiss, "Measuring Nominal Scale Agreement Among Many Raters," *Psychological Bulletin* 76, no. 5 (1971): 420-28; Jacob Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* 20, no. 1 (1960): 37-46; Chris Fournier, "Evaluating Text Segmentation Using Boundary Edit Distance," *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Sofia, Bulgaria: Association for Computational Linguistics, 2013), 1702-12, <http://aclweb.org/anthology/P13-1167>; Yann Mathet, Antoine Widlöcher, and Jean-Philippe Métivier, "The Unified and Holistic Method Gamma ( $\boxtimes$ ) for Inter-Annotator Agreement Measure and Alignment," *Computational Linguistics* 41, no. 3 (2015): 437-79; see Ron Artstein and Massimo Poesio, "Inter-Coder Agreement for Computational Linguistics," *Computational Linguistics* 34, no. 4 (2008) for an overview.

expresses how well real annotators agree, the latter expresses how much annotations would overlap if they were done at random. Thus, the actual, observed agreement is set in relation to the difficulty of the annotation task. The reasoning behind this is that, for instance, it is much easier to achieve agreement if there are only two categories than if there are 25 categories. Thus, the expected agreement (a.k.a. chance agreement) for two categories is higher than for 25 categories, which lowers the inter-annotator agreement if the observed agreement remains stable. Most inter-annotator agreement metrics are in the interval  $[-\infty:1]$ , in which values above zero express that the annotators agree more than chance agreement.

Measuring inter-annotator agreement for higher level tasks properly is not as easy as it sounds. This is due to the fact that many such tasks are actually composed of multiple subtasks and require the annotators to make multiple decisions in sequence. Annotating named entities, for instance, requires annotators to first find a segment that is a named entity, and secondly, to categorize this segment into a specific named entity category, such as person or location. The inter-annotator agreement metric needs to either take both decisions into account, which makes the exact calculation complex, or employ simplifying assumptions (e.g., to ignore overlapping spans).

In natural language processing, inter-annotator agreement is also often considered an upper boundary for machine performance. If humans only agree to a certain extent, we cannot expect machines to do better.

## The Subject of Analysis: Narrative Levels

The target concept in this shared task is the concept of narrative levels. Narrative levels are a ubiquitous phenomenon in narratology that is well-known to readers (and watchers) with and without an academic interest in narrative structure. They are a central element of every narrative. In some cases they are even a fundamental feature of a narration, as in the book *Arabian Nights* where Scheherazade tells a story every night in order not to be executed; or in Boccaccio's *Decameron* where a group of ten people having escaped from the great plague in Florence to the countryside help pass the time by telling stories. Even the TV show *How I Met Your Mother* consists of episodes narrated by the protagonist who tells stories from the past that eventually lead to his marriage. Very generally speaking, a narrative level is a separable part of a story within a story-narrative. The 'within' in the 'story within a story-narrative' is typically, but not necessarily, thought of as a subordination.

While the question of the status of these stories within stories as narrative levels depends on the actual definition of narrative levels and is thus disputable, all examples mentioned before show that narrative levels are a fundamental part of narratives, or, more precisely, that narratives can be seen as being constructed entirely of narrative levels. This also holds for narratives where the plot is not dependent on the integration of narrative levels as in these examples.

The introduction of additional narrators and their narrations is a typical aspect of the natural practise of storytelling and is therefore a frequent phenomenon in narratives, regardless of their function and mediality. Narrative levels are present in all narratives, in fictional texts as much as in self-narrations, journalistic writing, jokes, and many other text types. Additionally, narrative levels are not restricted to written text, but can also be observed in oral storytelling as well as in moving images, again both in fictional and non-fictional forms. Thus, narrative levels are highly relevant for all types of narrative analyses.

In our shared tasks, we focus on narrative levels in fictional texts, since the concept of narrative levels was originally developed for these, and they still constitute the major area of research. In addition, computational analysis in the context of narratives so far works best for written texts.

Please note: The remainder of this section provides an orientation for those who are interested in the role of narrative levels in manual and computational analysis and (literary) theory as well as our approach to it. The specific handling of narrative levels in the submitted guidelines and its evaluation will be discussed in Chapter 3 (“Annotation Guidelines Overview and Evaluation Results”).

## A Brief Narratological Background

Generally speaking, the major aspects of narrative levels are the narrator(s) and the horizontal and vertical embedding. The notion of narrative levels, as many other concepts in narratology, was introduced by Gérard Genette, one of the most famous narratologists. The phenomenon had already been described by others,<sup>10</sup> but it was Genette who coined the term *narrative level*.<sup>11</sup> In *Narrative Discourse*, he discusses a passage in Proust’s *A la recherche du temps perdu* (1913-1927), in which a character tells stories of their past loves to another character in an inn.

---

<sup>10</sup>For example by Bertil Romberg, *Studies in the Narrative Technique of the First-Person Novel* (Stockholm: Almqvist & Wiksell, 1962).

<sup>11</sup>Gérard Genette, “Discours Du Récit,” *Figures III*, 67-282. (Paris, 1972) (English translation: Gérard Genette. *Narrative Discourse: An Essay in Method* (Ithaca, N.Y: Cornell University Press, 1980).

Genette points out that it is not a distance in time or space that separates the narrated episodes from the inn, but rather "a sort of threshold represented by the narrating itself, a difference of *level*" and provides the following definition: "We will define this difference in level by saying that *any event a narrative recounts is at a diegetic level immediately higher than the level at which the narrating act producing this narrative is placed*".<sup>12</sup> According to this, narrative levels are produced by a narrating act, i.e., a new narrator is introduced in the narrative, recounting something as a new narrative. Or, as Pier puts it, "Narrative levels are most accurately thought of as diegetic levels, the levels at which the narrating act and the narratee are situated in relation to the narrated story."<sup>13</sup>

Many theorists have thought of narrative levels in terms of narrative framing or narrative embedding, some of them changing the terminology for the levels on which embedding occurs while doing so. Within narratology, there are several unresolved issues with the concept, ranging from terminological to categorical problems. For example, Wolf Schmid not only introduced a simpler nomenclature for the possible levels of embedding,<sup>14</sup> but also claimed that embedding can occur on all levels (Schmid, *Narratology*). In contrast, in Genette's view the introduction of a narrator is crucial for an embedded level and thus embedding can occur only within the so-called intradiegetic level. Another issue that is pointed out by Pier is that "intercalation" would be the more appropriate term for describing the relation between narrative levels. Framing and embedding are operations that involve inclusion, whereas levels are distributed vertically (Pier, *Narrative Levels*, 4). Some theorists include these seemingly contradictory concepts in their narrative level concept. Starting with Mieke Bal's approach,<sup>15</sup> a series of models<sup>16</sup> was developed that added horizontal embedding (where no change

<sup>12</sup>Genette, "Discours Du Récit," 228, emphasis in original.

<sup>13</sup>John Pier, "Narrative Levels," (revised version; uploaded 23 April 2014), Paragraph 1. *The Living Handbook of Narratology*. Hamburg: Hamburg University. <http://www.lhn.uni-hamburg.de/article/narrative-levels-revised-version-uploaded-23-april-2014> [last accessed 12 Feb 2019]. See also for a detailed discussion of narrative level concepts from a historical and a systematic perspective. The following overview in the text outlines the most important aspects Pier points out. Additional information can be found in Manfred Jahn, „N2.4. Narrative Levels," Manfred Jahn. *Narratology: A Guide to the Theory of Narrative*, (2017), and William Nelles „Embedding," David Herman, Manfred Jahn, and Marie-Laure Ryan (eds.). *Routledge Encyclopedia of Narrative Theory*. (London; New York: Routledge, 2005), 134-135.

<sup>14</sup>Schmid replaced Genette's terms extra-, intra- and metadiegetic, by primary, secondary and tertiary level of embedding (cf. Wolf Schmid, *Narratology. An Introduction*. (Berlin: de Gruyter, 2010), 67-70.

<sup>15</sup>Mieke Bal, *Narratology: Introduction to the Theory of Narrative*. (Toronto: U of Toronto Press, 1997), 43-66.

<sup>16</sup>Among other: William Nelles. *Frameworks: Narrative Levels and Embedded Narrative*. (New York: P. Lang, 1997), 121-158, and Marie-Laure Ryan. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. (Bloomington: Indiana University Press, 1991), 175-200.

of level takes place) to the Genettian vertical embedding or shift between levels. Therefore, horizontal embedding means that narratives are narrated by different narrators on the same level.<sup>17</sup> Based on its claimed background in artificial intelligence, Marie Laure Ryan's account would seem to be the most relevant for the context of this project (Ryan, *Possible Worlds*). Ryan approaches the question of narrative levels in terms of boundaries, frames and stacks. Still, her usage of terms is often not in their computational sense proper and thus does not provide a straightforward approach to operationalization and automation. It is rather her introduction of the now well-established concepts of ontological (semantic) boundaries and illocutionary (speech act) boundaries that seems promising for computational approaches and more general operationalization.

There are certainly other approaches to narrative levels in narratology that could be added to this overview. But it should have become clear that narrators and concepts related to embedding are the most important aspects to consider. Even though narrative levels have been debated for over 50 years, there are still open issues connected to the concept such as its relation to frame theory or its deployment in cognitive narratology (Pier, *Narrative Levels*, 31-32).

### **Relevance of Narrative Levels for Text Analysis**

We consider narrative levels as highly relevant for many text analysis projects because they are constitutive for narrative texts. This constitutivity makes them virtually ubiquitous in texts with narrative portions.

*Text analysis* refers to research steps that conduct a manual or automatic analysis of textual properties in relation to a specific research goal. Manual text analysis is usually a prerequisite for a hermeneutic interpretation of a literary text (and is typically not perceived as a distinct work step). Automatic text analysis employs methods from natural language processing (such as the detection of grammatical structure). On top of that, more "high-level" processing steps are typically added, one of them might be the detection of narrative structures such as levels. It is the ultimate research question that governs the kind and number of processing steps that need to be conducted for automatic text analysis. As we argue here, any text analysis with a focus on plot or character (i.e., analysis of the narrated content) requires the detection of narrative levels, as do some linguistically oriented pre-processing steps.

---

<sup>17</sup>Even though, in our view, this is already inherent in Genette's conception of voice that, among other, includes narrative levels and person. Therefore, even a shift of addressee may be interpreted as possible level change (cf. Evelyn Gius, *Erzählen über Konflikte. Ein Beitrag zur digitalen Narratologie* (Berlin: De Gruyter, 2015), 165-66).

The structure of narratives in terms of narrative levels can already be an interesting topic on its own. For example, one could be interested in the functions of frame stories in a certain period or the degree of nesting of narrative levels in fairytales in comparison to social novels. The way narrative levels are organized could in general may be a constituting element for a literary style in a broad sense,<sup>18</sup> and would be very hard to detect by existing stylometric approaches.

Still, the detection of narrative levels as a preparatory step for the analysis of a narrative is even more common and thus important. Gaining an understanding of the narrative levels present in a narration is a necessary prerequisite for its analysis. This applies to analyses concerned with the phenomena of the fictional world and to ones looking at the textual representation of the narrative, i.e., at phenomena related to what happens in the narrative (also known as the *what* of narration or *histoire*) or the very text (also known as the *how* of narration or *discours*).<sup>19</sup>

For the analysis of narratives one often needs to correctly conjoin narrative parts. Thus, it is necessary to have a prior understanding of the narrative levels present in a text. For example, when looking at character constellations, one should consider only characters in a more or less coherent space-time continuum, since interactions between characters are usually confined to coherent parts of the fictional world or story world and thus do not cross temporal or spatial borders. The identification of the narrative levels in a narrative and the analysis of their spatio-temporal features are therefore prerequisites to a proper character analysis.

Temporal or spatial coherence may also be relevant for the analysis of narrative representation. Thus, an analysis of the temporal relation between fictional world (*histoire*) and its representation (*discours*) is only sensible after having identified which narrative levels belong to which space-time continuum. The fictional world within a narrative is not necessarily coherent and can exhibit parts that are not connected to the main setting temporally or spatially as, for example, the world of a dream.<sup>20</sup> Therefore, a reconstruction of the order of events in the fictional world needs to first analyze which narrative levels belong to which parts of the fictional world and then analyze the temporal order only for the connected ones.

There are currently no published systems that detect narrative levels automatically. While there certainly is a need for text segmentation as a preparatory

---

<sup>18</sup>Berenike J. Herrmann, Karina van Dalen-Oskam, and Christof Schöch, "Revisiting style, a key concept in literary studies," *Journal of Literary Theory* 9, no. 1 (2015): 25-52.

<sup>19</sup>The terms *histoire* and *discours* have been coined by Genette (*Narrative Discourse*).

<sup>20</sup>Cf. Gius and Jacke "The Hermeneutic Profit of Annotation", 247-248.

step for subsequent processing of other phenomena,<sup>21</sup> segmentation is currently mainly accomplished by using textual surface phenomena. Features that can be derived directly from a text or its markup (e.g. in XML) are used as basis for segmentation (e.g., paragraphs, or, where available, chapters or other structural information encoded in the text). However, for more complex tasks the segmentation is more helpful the more meaningful it is. A division into chapters cannot be assumed to respect the structure of the events in the fictional world, as chapters are introduced for various reasons and some of them may have nothing to do with the plot of the narrative. Even worse, a division into, say, ten equal parts obviously is not related to the fictional world at all. As some segmentation is required for certain text analysis tasks, texts are often just segmented into parts of equal length, which is clearly not a usual procedure in literary studies. Hence, segments should be anchored in the narrated events rather than in chapter boundaries or, even worse, completely arbitrary segments of equal length in order to maximize their value for the analysis.

There are, however, some approaches to a content-related segmentation of texts in natural language processing.<sup>22</sup> Approaches in discourse analysis/processing<sup>23</sup> or topical segmentation<sup>24</sup> clearly feature related aspects to the ones needed to detect narrative levels, but a full-fledged and genuine automatic detection of narrative levels remains a desideratum. Moreover, existing approaches are typically tested and developed on texts such as news or wikipedia articles. As these texts differ in key areas from literary texts (fictionality, narrativity), the approaches cannot

---

<sup>21</sup> Cf. Nils Reiter, "Towards Annotating Narrative Segments," Kalliopi Zervanou, Marieke van Erp, and Beatrice Alex eds. *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, (Beijing, China: Association for Computational Linguistics, 2015), 34-38.

<sup>22</sup> Among others: Omri Koshorek, Adir Cohen, Noam Mor, Michael Rotman, and Jonathan Berant, "Text Segmentation as a Supervised Learning Task," *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, (Association for Computational Linguistics, 2018), 469-73, <https://doi.org/10.18653/v1/N18-2075>. Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto, "Unsupervised Text Segmentation Using Semantic Relatedness Graphs," *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, (Berlin, Germany: Association for Computational Linguistics, 2016), 125-30, <http://anthology.aclweb.org/S16-2016>.)

<sup>23</sup> For an introduction cf. Manfred Stede, *Discourse Processing* (San Rafael, California: Morgan & Claypool, 2012).

<sup>24</sup> Anna Kazantseva and Stan Szpakowicz, "Hierarchical Topical Segmentation with Affinity Propagation," *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, (Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 2014), 37-47. <http://www.aclweb.org/anthology/C14-1005>. Anna Kazantseva and Stan Szpakowicz. "Topical Segmentation: A Study of Human Performance and a New Measure of Quality," In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (Association for Computational Linguistics, 2012), 211-20. <http://aclweb.org/anthology/N12-1022>.

be directly applied to literary texts.

As we have already argued above, a segmentation based on the textual content is especially important when aiming for an analysis of more complex phenomena as addressed in literary studies, such as character constellation, point of view or the temporal structure of the plot. These are often only indirectly connected to the textual surface of narratives and it is a clear desideratum to base automatic segmentation more heavily on the content of narrations. For an application in literary studies or story-related analysis, narrative levels are the more adequate criterion for segmentation.

Generally speaking, the research goals related to text analysis often differ widely with literary scholars and computer scientists, but for research in both areas, narrative levels are an important asset. Most importantly, the analysis of narrative levels allows for a subsequent analysis of text qualities that may be structural, surface-related, or in the realm of narrative phenomena as well. Due to the integration of the latter, i.e., more content-related aspects of texts, a narrative level-based approach is much more adequate for text segmentation and subsequent analyses than a merely structural approach.

### **Narrative Levels in the Shared Task**

For this shared task, we did not specify any theoretical background for the narrative level concepts, thus not providing one of the specific narratological approaches discussed above, nor did we disclose our personal preference for a specific approach. Instead, the participants were encouraged to choose the approach they considered adequate. We provided a basic explanation as well as reading suggestions (categorized as introductory, basic, or advanced) on the homepage of the shared task,<sup>25</sup> but we did not intentionally prioritize any approach. Therefore, participants could use any (or even no) narratological theory as a basis for the operationalization in their guideline.

There are several reasons for this decision:

a) Even though there are only few well-established approaches to narrative levels in narratology and most of them overlap, there is no consensus about *the* narrative level concept. Narratologists tend to have strong and diverse opinions about the nature of narrative levels, and there are good arguments that can be made for

---

<sup>25</sup>Some of them have been discussed in section 3 above, for the comprehensive list see: <https://sharedtasksinthedh.github.io/levels/>



most available theories. Therefore, there was no way to select the most suitable concept for level annotation.

b) As in many humanities' disciplines, there is no established procedure of identifying the 'right' theory among coexisting approaches. The idea of something being right, true, objective etc. is hardly compatible with the humanities' disciplinary paradigm or matrix. Within the humanities paradigm, theories and interpretations typically exist alongside each other and may even contradict each other. This plurality is owed to the humanities and their often heavily interpretative analysis of ambiguous and multifaceted human artifacts. Since the overall process of understanding is rather complex and its parts are not completely intelligible, limiting the analysis of an artifact to the usage of specific theories can lead to a premature exclusion of approaches yielding relevant insights. Therefore, limiting the narrative level analysis in the shared task to one approach would have meant ignoring the process through which theoretical or methodological approaches were and are developed in literary theory.

c) Annotation guidelines barely play a role in contemporary narratology, and annotatability is—at this moment—not a criterion regularly considered. From a narratological point of view, the pure guideline creation is likely not that interesting, compared to a discussion/comparison of narratological theories. However, it was clear from the beginning that the participation of narratology experts would be of utmost importance to this shared task. Therefore, allowing different theoretical “flavors” to compete would also spark the interest of narratologists who may be new to the development of annotations.

Against this background, allowing for all possible theories was advantageous to the research process on multiple levels. Most importantly, it allowed us to stick to the humanities' paradigm and at the same time provide a framework for the exploration and testing of theories in this first shared task on guideline creation. This ensures a higher relevance of the automation's outcomes to their users.

## **The Shared Task**

### **Shared Tasks in Natural Language Processing**

Shared tasks are an established research format within the community of natural language processing (NLP) with the core idea being that multiple participants try to solve the same task given by the organizers ( e.g. automatic prediction of

part of speech tags). The solutions are then evaluated on the same data set with the same metric and thus directly comparable. Generally, a shared task works as follows: The organizers publish a call for participation in the task, describing the task as well as the associated data set in some detail. Shortly thereafter, the organizers publish a development and/or training data set. The dataset contains gold information, i.e., the categories to be identified are already annotated. This data set is then used by the participants to develop/train systems to automatically solve the defined task. After several months of development time, the organizers publish a second data set without the annotations: the test data. The participants apply their systems to the test data set (typically within a week) and send/upload the predictions made by their systems to the organizers. The organizers then evaluate all systems' predictions with the same evaluation script and against the same reference data. After this, a ranking of the systems can be generated, and a workshop is conducted to present the different systems and discuss the outcome.

## History

Within natural language processing, shared tasks have their roots in the Message Understanding Conference (MUC) community.<sup>26</sup> In this context, the goal has been to extract snippets of information from news reports (covering incidents of terrorist attacks in South America) or naval messages. The major contributions of the shared tasks in the context of MUC are categorized into three different categories by Beth M. Sundheim and Nancy A. Chinchor:<sup>27</sup> The first category, progress evaluation, refers to the progress in terms of raw system performance with a clearly defined evaluation metric, which can be used to express the current state of the art, for comparison to a previous performance or to measure progress towards human performance (given the same metrics can be applied to human and machine performance). The second category, adequacy evaluation, expresses the adequacy of the evaluation metrics: “it is not possible to translate the evaluation results directly into terms that reflect the specific requirements of any particular real-life applications.”<sup>28</sup> By applying evaluation metrics and scientific discourse about them, fostered by the MUC challenges, the community

---

<sup>26</sup>Cf. Beth M. Sundheim, “The Message Understanding Conferences,” *Proceedings of the Tipster Text Program: Phase I* (Fredericksburg, Virginia, USA: Association for Computational Linguistics, 1993), 5, <https://doi.org/10.3115/1119149.1119153>.

<sup>27</sup>“Survey of the Message Understanding Conferences,” in *HUMAN Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey* (Plainsboro, New Jersey, 1993), <http://aclweb.org/anthology/H93-1011>.

<sup>28</sup>Beth M. Sundheim and Nancy A. Chinchor, “Survey of the Message Understanding Conferences,” in *HUMAN Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey* (Plainsboro, New Jersey, 1993), 58.

gains insight into potential weaknesses of the evaluation metrics. Finally, the third category, diagnostic evaluation, refers to the fact that the MUC challenges also generate insights into reasons for over- and underperformance of certain systems. By participating in the challenges and inspecting the prediction errors, the system developers gain insight into possible bottlenecks and can find ways for improvements of the system. All three categories have been present in shared tasks in the years following Sundheim and Chinchor's publication.

Starting with the year 2000, the Conference on Natural Language Learning (CoNLL) has been the home for a series of shared tasks on various topics: Chunking,<sup>29</sup> clause identification,<sup>30</sup> language-independent named entity recognition,<sup>31</sup> various forms of syntactic parsing either multilingually or for specific languages<sup>32</sup> and semantic representation/role labeling.<sup>33</sup> Other conferences and venues have taken up the shared task concept as well, for instance, the PASCAL Recognizing Textual Entailment challenge,<sup>34</sup> which ran for eight years until

---

<sup>29</sup>Erik F. Tjong Kim Sang and Sabine Buchholz, "Introduction to the CoNLL-2000 Shared Task: Chunking," *Proceedings of Fourth Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop* (Lisbon, Portugal: acl, 2000).

<sup>30</sup>Erik F. Tjong Kim Sang and Hervé Déjean, "Introduction to the CoNLL-2001 Shared Task: Clause Identification," *Proceedings of the ACL 2001 Workshop on Computational Natural Language Learning*, (2001), <http://www.aclweb.org/anthology/W01-0708>.

<sup>31</sup>Erik F. Tjong Kim Sang and Fien de Meulder, "Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition," Walter Daelemans and Miles Osborne eds. *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, (2003), 142-47, <http://www.aclweb.org/anthology/W03-0419>.

<sup>32</sup>Sabine Buchholz and Erwin Marsi, "CoNLL-X Shared Task on Multilingual Dependency Parsing," *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNll-X)* (New York City: Association for Computational Linguistics, 2006), 149-64, <http://www.aclweb.org/anthology/W/W06/W06-2920>; Joakim Nivre et al., "The CoNLL 2007 Shared Task on Dependency Parsing," *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007* (Prague: Association for Computational Linguistics, 2007), 915-32, <http://www.aclweb.org/anthology/D/D07/D07-1096>; Sandra Kübler, "The PaGe 2008 Shared Task on Parsing German," *Proceedings of the Workshop on Parsing German* (Columbus, Ohio: Association for Computational Linguistics, 2008), 55-63, <http://www.aclweb.org/anthology/W/W08/W08-1008>; Jan Hajič et al., "The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages," *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task* (Boulder, Colorado: Association for Computational Linguistics, 2009), 1-18, <http://www.aclweb.org/anthology/W09-1201>.

<sup>33</sup>Xavier Carreras and Lluís Màrquez, "Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling," Hwee Tou Ng and Ellen Riloff eds. *HLT-NAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning*, (Boston, Massachusetts, USA: Association for Computational Linguistics, 2004), 89-97; Xavier Carreras and Lluís Màrquez, "Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling," *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)* (Ann Arbor, Michigan: Association for Computational Linguistics, 2005), 152-64, <http://www.aclweb.org/anthology/W/W05/W05-0620>; Johan Bos and Rodolfo Delmonte, eds., *Semantics in Text Processing: STEP 2008 Conference Proceedings 1, Research in Computational Semantics* (London, UK: College Publications, 2008).

<sup>34</sup>Ido Dagan, Oren Glickman, and Bernardo Magnini, "The PASCAL Recognising Textual Entail-

2013. Having started under the label SensEval in 2000,<sup>35</sup> the SemEval initiative now hosts many shared tasks every year concerning the evaluation of semantic analysis tools. For the first time, an open call was issued to propose shared tasks for the SemEval roof to be organized in 2018. No less than twelve shared tasks were offered by SemEval in 2018.<sup>36</sup>

Carla Parra Escartín et al.<sup>37</sup> discuss several reasons for the popularity and success of shared tasks in natural language processing: Apart from fostering development in a certain field, they also allow for direct comparison between systems. A number of de facto standards have evolved in shared tasks (e.g., the widely used CoNLL format for storing annotated data). Finally, curated data sets have been created along with the shared tasks and subsequently made available:<sup>38</sup> “shared tasks have proven themselves to be very effective in incentivising research in specialised areas.”<sup>39</sup>

Finally, ethical considerations about shared tasks have been pointed out by Escartín et al., mainly due to their competitive nature. Competition may lead to secretive behavior, hurt the relations of researchers with colleagues and lead to a general disregard for ethics. Escartín et al. identify a number of concrete scenarios which could directly impact the success story of shared tasks and might be a consequence of the competition. They propose that organizers follow a certain framework to minimize the negative impact of the competitive nature of shared tasks such as declaring early and explicitly if organizers or annotators are allowed

---

ment Challenge,” J. Quiñonero-Candela et al. eds. *Machine Learning Challenges. Lecture Notes in Computer Science*, (Springer, 2006).

<sup>35</sup>Adam Kilgarriff and Joseph Rosenzweig, “Framework and Results for English Senseval,” *Computers and the Humanities* 34, no. 1 (April 1, 2000): 15-48, <https://doi.org/10.1023/A:1002693207386>.

<sup>36</sup>Affect in Tweets, Multilingual Emoji Prediction, Irony Detection in English Tweets, Character Identification on Multiparty Dialogues, Counting Events and Participants within Highly Ambiguous Data covering a very long tail, Parsing Time Normalizations, Semantic Relation Extraction and Classification in Scientific Papers, Semantic Extraction from CybersecUrity REports using Natural Language Processing (SecureNLP), Hypernym Discovery, Capturing Discriminative Attributes, Machine Comprehension using Commonsense Knowledge, Argument Reasoning Comprehension Task. See <http://alt.qcri.org/semeval2018/index.php?id=tasks>.

<sup>37</sup>“Ethical Considerations in NLP Shared Tasks,” *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (Valencia, Spain: Association for Computational Linguistics, 2017), 66-73, <http://www.aclweb.org/anthology/W17-1608>.

<sup>38</sup>Carla Parra Escartín, Wessel Reijers, Teresa Lynn, Joss Moorkens, Andy Way, and Chao-Hong Liu, “Ethical Considerations in NLP Shared Tasks,” in *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (Valencia, Spain: Association for Computational Linguistics, 2017), 66-73.

<sup>39</sup>Carla Parra Escartín, Wessel Reijers, Teresa Lynn, Joss Moorkens, Andy Way, and Chao-Hong Liu, “Ethical Considerations in NLP Shared Tasks,” in *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (Valencia, Spain: Association for Computational Linguistics, 2017), 68.

to participate and if they will publish the system results under agreed license.

## Two Linked Shared Tasks

As the research practices, goals and not least communities in natural language processing and literary studies are clearly different, directly applying the shared task model known from natural language processing is not going to work. We therefore made several adjustments to the procedure. Our project consists of two shared tasks, and this volume appears after a milestone within the first one was reached: a guideline evaluation workshop with all participants. The two tasks have different goals, data sets, and target audiences, but both focus on the phenomenon of narrative levels. The goal of Task 1 is the generation of annotation guidelines which are then used to annotate a large corpus to be employed as training/testing data in Task 2. The second shared task is a ‘regular’ NLP shared task, i.e., its goal is to develop systems that automatically detect narrative levels.

### **Shared Task 1: Systematic Analysis of Narrative Texts through Annotation (SANTA)**

In the first shared task, the challenges of conceptualizing and defining narrative levels, as well as manually applying them to texts, are in focus. The main task of the participants was to develop annotation guidelines for narrative levels. As discussed above, we did not specify an exact theoretical background to be used for the guidelines, but we pointed the participants to a bibliography for further readings. We also provided a “how to”-article on our web page explaining how annotation guidelines can be developed, which contained the same information as the above section on annotations.

To foster the development of generic guidelines that do not make a lot of assumptions on the text type in question, we decided early on that the guidelines should be tested on an unspecified corpus, but it was stated that it would contain literary texts of certain genres. Each participant thus had to write the guideline without knowing the exact texts it would be applied to in the end. To ensure comparability of the guidelines, however, there needed to be some homogeneity in the corpus. We thus decided to provide the participants with a development set that they could use when writing the guidelines. The texts in the final test set were similar to the ones in the development set. This setup is inspired by the distinction

between development, train and test data used in machine learning.<sup>40</sup>

**Corpus considerations.** The corpus was compiled to cover as many of the suggested level phenomena as possible. It is heterogeneous with respect to genre, publication date, and text length.<sup>41</sup> However, representativity (whatever that means for literature) was not a guiding principle. All texts were made available in both English and German, some being translations from a third language.

The maximum length for a text to be included in this corpus was 2000 words. Since the constraint might limit the use of narrative levels, we also included longer texts to avoid this bias. We made these available in a shortened form, omitting passages that do not affect the overall narrative level structure in a substantial manner, according to the level definitions we suggested on our web page and our own judgement. A set of 17 texts had been made available as a development corpus, to be used by the participants during guideline development. Table 3 shows the texts with some metadata. The actual annotation experiment was conducted on a set of eight texts that were previously unknown to the participants. This list can be found in Table 4. All texts are freely available and can be accessed through our GitHub repository.<sup>42</sup>

**Creating Parallel Annotations.** Measuring inter-annotator agreement is an established way of gaining insight into the intersubjective applicability of an annotation guideline. In order to measure inter-annotator agreement, the same text(s) need to be annotated by multiple people, using the same guideline.

To implement this in the shared task, we asked each participating group to annotate that same test corpus using someone else’s guideline. In addition, a group of (paid) student assistants annotated with the same guideline. In this process, each guideline was used three times on the same set of texts (see Table 1 for an overview).

Label	Description
own	Annotations done by the guideline authors using their own guideline
foreign	Annotations done by the guideline authors using another guideline
student	Annotations done by a group of student assistants

*Table 1. Overview of the annotations*

**Timeline.** The full timeline of the various events of the shared task can be seen

<sup>40</sup>Cf. I. H. Witten and Eibe Frank, *Data Mining*, 2nd ed., *Practical Machine Learning Tools and Techniques* (Elsevier, 2005), 144ff.

<sup>41</sup>*Genres*: anecdote, fable, folktale, literary fairy tale, novel, novella, narration, short story. *Publication date*: the majority of the texts were written in the 19th and 20th century. *Text length*: 2000 words maximum.

<sup>42</sup><https://github.com/SharedTasksInTheDH>

in Table 2.

Date	Event
October 6, 2017	Call for Participation
June 16, 2018	Submission of the guidelines
June 25, 2018	Submission of the annotations on test corpus, using own guideline
July 6, 2018	Submission of the annotations, using foreign guideline
September 17-19, 2018	Workshop

*Table 2. Timeline*

**Workshop.** As a milestone in the first shared task, all participants were invited to a workshop that took place in Hamburg, Germany. All but one team were physically present. The three-day event was structured as follows: The goal of the first day was for all participants to gain a better understanding of the other guidelines. This was realized in the form of brief presentations and a discussion to identify commonalities and differences. On the second day, the guidelines were evaluated in detail. To this end, a questionnaire was first presented and discussed. All questions could be answered in the form of a four point Likert scale. Each team was then asked to fill out the questionnaire (in digital form) for every guideline except their own. In addition, they were asked to keep notes on why they assigned which scores. We will cover the evaluation details in Chapter 3 of this volume. On the last day, the organizers presented the evaluation results as well as the inter-annotator agreement scores, and the entire group discussed the results and next steps.

### **Outlook: Shared Task 2—Automatic Detection of Narrative Levels**

The second shared task can be considered a “regular” NLP shared task, and is thus intended to primarily attract researchers in natural language processing. It is envisaged to take place in the summer of 2021. The annotated corpus will be split into development, training, and testing data sets, and will be made available to the participants at certain points in time. The final evaluation will then require participants to submit their automatic predictions to the organizers, who in turn will compare the predictions to the manual annotations of the test set. This shared task is planned to be organized with the SemEval community to attract a large enough number of participants. The participants are not required to be familiar with or experienced in literary studies, narratology, or digital humanities, as the task and its difficulties are encoded in the annotations. The result of the second shared task will be a comparison of automatic systems that detect narrative levels.

**Preparations.** After having completed the first, guideline-oriented shared task, the organizers will conduct an annotation phase. The goal of the annotation

phase is to provide an annotated corpus which is large enough to allow for methodological experiments, including machine learning.

This annotation phase will be executed using the best performing guideline of the first shared task as a starting point. It can be expected, however, that it will need updating during the annotation phase, as new phenomena are expected to arise. The final version of the guideline will be made available along with the annotated data for the second shared task.

Title (orig.)	Author	Title (en)	Genre	Year	Language(orig.)	Comment
Rosen-Alfen	Aesop	The Wolf and the Lamb	fable	600 BCE		
Kjærestefolkene [toppen og bolden]	Andersen, Hans-Christian	The Elf of the Rose	folktale	1839	dk	
Se una notte d'inverno un viaggiatore	Andersen, Hans Christian	The Top and Ball	folktale	1862	dk	
Мститель	Calvino, Italo	If on a Winter's Night a Traveller	novel	1979	it	Shortened
The Child's Story	Čechov, Anton Pavlovič	An Avenger	short story	1887	ru	
Die drei Federn	Dickens, Charles	The Child's Story	short story	1852	en	
Das wohlfeile Mittagessen	Grimm, Brüder	Feathers	folktale	1819	de	
Der geheilte Patient	Hebel, Johann Peter	The Cheap Meal	anecdote	1804	de	
Hills Like White Elephants	Hebel, Johann Peter	The Cured Patient	anecdote	1811	de	
How the Leopard got his Spots	Hemingway, Ernest	Hills Like White Elephants	short story	1920	en	
Beyond the Pale	Kipling, Rudyard	How the Leopard got his Spots	short story	1901	en	
Unwahrscheinliche Wahrhaftigkeiten	Kipling, Rudyard	Beyond the Pale	short story	1888	en	
The Cask of Amontillado	Kleist, Heinrich von	Improbable Veracities	anecdote	1810	de	
Frankenstein or The Modern Prometheus	Lagerlöf, Selma	Among the Climbing Roses	narration	1894	sv	
A Haunted House	Poe, Edgar Allen	The Cask of Amontillado	short story	1846	en	
	Shelley, Mary	Frankenstein or The Modern Prometheus	novel	1818	en	shortened
	Woolf, Virginia	A Haunted House	short story	1921	en	

Table 3. Overview of development corpus.

Title (orig.)	Author	Title (en)	Genre	Year	Language	Comment
Lenz	Büchner, Georg	Lenz	novella	1839	de	shortened
Выигрышный билет	Čechov, Anton Pavlovič	The Lottery Ticket	short story	1887	ru	
The Gift of the Magi	Henry, O.	The Gift of the Magi	short story	1905	en	
Kleine Fabel	Kafka, Franz	A Little Fable	fable	1831	de	
Der blonde Eckbert	Tieck, Ludwig	The White Egbert	literary fairy tale	1797	de	shortened
Der Schimmelreiter	Storm, Theodor	The Rider of the White Horse	novella	1888	de	shortened
Anekdote aus dem letzten preußischen Kriege	Kleist, Heinrich von	Anecdote from the Last Prussian War	anecdote	1810	de	
Herr Arnes penningar	Lagerlöf, Selma	The Treasure	narration	1904	sv	shortened

Table 4. Overview of the test corpus.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.



# A Shared Task for the Digital Humanities

## Chapter 2: Evaluating Annotation Guidelines

Evelyn Gius, Nils Reiter, Marcus Willand

11.05.19

*Article DOI: 10.22148/16.049*

*Journal ISSN: 2371-4549*

*Cite: Evelyn Gius, Nils Reiter, Marcus Willand, "A Shared Task for the Digital Humanities Chapter 2: Evaluating Annotation Guidelines," Journal of Cultural Analytics. November 4, 2019. doi: 10.22148/16.049*

In this section, we will discuss our idea of guideline evaluation and the underlying considerations. Evaluating annotation guidelines in this way is a fairly new endeavor, and we have developed the evaluation setup from the ground up. Although we do not claim our choices to be universally valid or applicable, we believe that this approach to guideline evaluation is relevant for similar settings and can be adapted to projects that might have other preferences and priorities.

### **Preliminaries and Challenges**

Our goal was to take into consideration requirements and principles from the humanities as well as from computational linguistics/natural language processing. Whichever evaluation method we would employ in the end, it needed to fulfill four basic requirements:

1. Establish a ranking: The method needs to be able to rank the guidelines. This ranking needs to be as clear as possible and avoid ties.
2. Be defined and explicit: The general design of shared tasks is a competition in which submissions are ranked according to an objective function. This

objective function needs to be defined in advance and as precisely as possible, in order for participants to know beforehand what they are getting into and so that it leaves little room for challenging this evaluation.

3. Be practical: The evaluation should be feasible to execute, within certain practical limitations. Concretely, we were aiming for an evaluation method that could be conducted within a two-day workshop.
4. Reflect our evaluation criteria: The evaluation method needs to reflect our evaluation standards, i.e., if a guideline contains aspects that are considered to be positive by the organizers, that guideline should be ranked higher than a guideline without these aspects. Defining positive/negative evaluation criteria was a decision that the organizers needed to make.

Those requirements are a consequence of aiming at creating annotation guidelines in a shared task. In shared tasks in natural language processing, the underlying intention is to reproduce the gold standard as closely as possible, which can then be measured in different ways, depending on the exact task (accuracy, f-score, MUC-score, ...). But there is no “ground truth” conceivable for annotation guidelines. Even measuring inter-annotator agreement would not necessarily be that straightforward, since there may be cases in the data in which different textual readings are possible, stemming from a legitimate ambiguity of the text. In such cases, disagreements between annotators would not indicate a flaw in the guideline.

In addition to these general requirements that any evaluation method for a shared task needs to fulfill, there are several challenges related to the specific nature of this one:

As this shared task is an **interdisciplinary** endeavor, a heterogeneous set of participants was to be expected. The notion of annotation plays a different role in different disciplines, and a diverse set of best practices, rules, and traditions has been established in each field. In literary studies, for instance, annotation is typically understood as note-taking while reading. In computational linguistics, annotation is typically done in parallel, digitally, and with a high intersubjective agreement as the most important goal. The latter does not matter at all for annotation in (traditional) literary studies, as the disciplinary approach to text analysis is rather focused on a not necessarily reproducible overall meaning of a text.

Thus, participants have different previous experiences and expectations with regards to the annotation process. Still, the evaluation we conduct in this shared task needs to be valid and functional across the different disciplines and be beneficial for each participant’s own discipline at the same time.

The **vagueness** of the source concepts provides another challenge. Narratology

represents a popular source for concepts in text-oriented DH, most likely because of its fundamental structuralist premises. Furthermore, narratologists and digital humanists agree on the idea that structural analyses expose interesting textual phenomena which remain hidden from purely content-related readings. However, as discussed above, the systematic application of narratological theory to texts also gives room for interpretation.

Given these considerations, we decided early on that the evaluation model needs to cover different perspectives. In particular, it should not ignore conceptual vagueness and complexity, but rather consider solution strategies for these problems.

## Evaluation Model

Generally, the evaluation was conducted in **three different dimensions**: conceptual coverage, applicability, and usefulness. Figure 1 schematically shows where the evaluation dimensions are situated with respect to research activities in the digital humanities. It projects them onto the course of the entire work process, from narratological theory to guideline creation to annotated texts, and finally to the insights that could be drawn from applying the annotated texts to understand single literary texts or whole corpora.

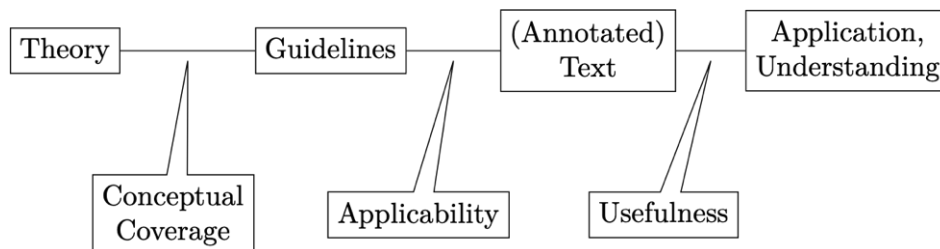


Figure 1: *The three evaluation dimensions connecting research areas in the digital humanities*

The dimension of **conceptual coverage** reflects how much of a theoretical basis is covered by an annotation guideline. If a guideline is explicitly based on a narratological theory, it might aim to fully implement every definition, rule, and exception of the theory. Another guideline based on the same theory might leave out some definitions or add others. This dimension is situated on the theoretical level, relating the guidelines to theory.

**Applicability** puts the guideline in relation to the text and reflects how well the guideline prepares annotators to do actual annotations, i. e., how well the guide-

line can be employed. A guideline's applicability may for instance be increased by thoughtful examples, a clear structure, and/or a careful use of terminology. The dimension of applicability also covers the achieved coherence and systematicity in the annotations.

Finally, the dimension of **usefulness** relates the annotated text to applications and understanding. "Applications", in this case, covers subsequent analysis steps as well as large scale analyses, while "understanding" refers to a hermeneutic interpretation of the text, that takes the annotations into account. Assuming corpora are annotated in accordance with the guideline (either manually or, in the case of large corpora, automatically), this dimension reflects how insightful they are, i.e., how "much" insight the annotations allow. Usefulness thus evaluates the insights gained by examining an annotated text or corpus.

The three dimensions allow a balanced evaluation of guidelines with diverse disciplinary and research backgrounds, aims, and understanding of narratological concepts. Focusing on only one of the dimensions will diminish the score in at least one other: A guideline addressing narratological theory exclusively might achieve a high score in the first dimension, but will be penalized in the second dimension, as mere theory is not very applicable. Optimizing for applicability could lead to guidelines that specify everything or nothing as narrative level, thus not being very useful. Finally, the blind optimization on usefulness will lead to guidelines that are unrelated to narratological theory. Thus, the challenge that this shared task poses to the guideline authors is to strike a balance between the three dimensions.

Arguably, an annotation guideline does not generally need to cover all three dimensions in order to be a useful guideline for a certain purpose. Guidelines that are detached or totally unrelated to a theoretical concept, for instance, could still address a relevant issue. Likewise, it is not always necessary to look at applications and aim, i.e., at the usefulness of a guideline. As guidelines and/or annotations are also an excellent tool for text analysis, their creation might be a sufficient research goal in its own right.

## Implementation

### A Multi-Purpose Questionnaire

In order to implement the three-dimensional evaluation model, we associated each dimension with a number of specific questions to be answered for each guideline. The questions represent different aspects of each dimension and

should be answerable directly for a guideline. Section 4 lists each question with a brief description.

The questions were made available to the participants before they submitted their guidelines. In the evaluation, they were used in two ways: Firstly, they provided a guide for qualitative evaluation. By following the online questionnaire we distributed during the workshop and discussing each question for each guideline, we assured that the same criteria were employed in the judgement of each guideline and that the same aspects were covered in the discussion. This is important to ensure both fairness and coherence in the evaluation. The discussion was quite extensive and thus difficult to document, but all teams described it as very helpful. The discussion gave rise to a number of guideline improvements, which will be documented in the second volume of this special issue.

Secondly, the questions were answered quantitatively. Each question was evaluated on a 4-point Likert scale, i.e., participants were asked to assign points for each guideline in each question with more points reflecting the more favorable choice. Thus, if guideline A has higher score than guideline B, it is considered the better guideline.

Our evaluation defined four questions for the dimensions of conceptual coverage and usefulness, and two questions for the dimension of applicability. In order to weigh the dimensions equally, two more scores regarding applicability were provided through the inter-annotator agreement score (see below), scaled to lie between one and four points. In the end, each guideline was given four scores in each dimension, which were added up, first by dimension and then to a total score. Each team evaluated all the other guidelines, leading to seven judgements and thus scores per question per guideline.

## Questionnaire

### Conceptual Coverage

1. Is the narrative level concept explicitly described?

Explanation: Narrative levels can be described or defined. This depends on the narratology used; some of them are structuralist, others are post-structuralist. Regardless of the mode, is the description/definition understandable and clear?

- 1: I did not understand what the guideline describes as “narrative level”.
- 4: I fully understood the concept described in the guideline.

2. Is the narrative level concept based on existing concepts?

Explanation: The level concepts can be self-designed, oriented on existing narratologies or copied from an existing level definition

- 1: The theory relation of the used level concept is not clear.
- 4: It is clearly mentioned whether the level concept is made up or (partially) based on a theory.

3. How comprehensive is the guideline with respect to aspects of the theory? Does it omit something?

Explanation: If the guideline is based on a theory or multiple theories, does it include the whole theory or only parts of it? Are there reasons mentioned why aspects are in-/excluded?

- 1: The guideline does not clearly state the extension of its dependence on theory/ies.
- 4: The guideline unambiguously states the scope of its theory-dependance.

4. How adequately is the narrative level concept implemented by this guideline in respect to narrative levels?

Explanation: Narratologies differ in their complexity. Firstly, you have to decide whether complexity or simplicity (in relation to x) is desirable, then you have to answer:

- 1: The guideline is too simple or too complex for narrative levels and thus not adequate.
- 4: The guideline's complexity is adequate.

### **Applicability**

1. How easy is it to apply the guideline for researchers with a narratological background?

Explanation: The question asks for an assessment of the ease of use of the guideline for an annotator with some narratological background. Indicators can be: Complexity of the concepts, length of the guideline, clarity of examples, clear structure, difficulty of finding special cases, etc.

- 1: Even as a narratology expert, I needed to read the guideline multiple times and/or read additional literature.
- 4: The guideline is very easy to apply, and I always knew what to do.

2. How easy is it to apply the guideline for researchers without a narratological background?

Explanation: The question asks for an assessment of the ease of use of the guideline if we assume an annotator who doesn't have a narratological background (e.g., an undergraduate student). Indicators can be: Complexity of the concepts, length of the guideline, use of terminology, clarity of examples, reference to examples only by citation, clear structure, difficulty of finding special cases, etc.

- 1: Non-experts have no chance to use this guideline.
- 4: The guideline is very easy to apply, and non-experts can use them straight away.

3./4. Inter-annotator agreement: gamma scores (see below)

### Usefulness

1. Thought experiment: Assuming that the narrative levels defined in the annotation guideline can be detected automatically on a huge corpus. How helpful are these narrative levels for an interesting corpus analysis?

Explanation: This question focuses on the relevance of the narrative level annotations for textual analysis of large amounts of texts, e.g., for the analysis of developments over time with regard to narrative levels or a classification of texts with regards to genre, based on narrative levels.

- 1: The narrative levels annotations are irrelevant for corpus analysis.
- 4: The annotations provide interesting data for corpus analysis.

2. How helpful are they as an input layer for subsequent corpus or single text analysis steps (that depend on narrative levels)?

Explanation: The analysis of some other textual phenomena depends on narrative levels, e.g., chronology should be analyzed within each narrative level before analyzing it for the whole text. This question asks whether the analysis of such phenomena is possible or even better when based on the narrative level annotations.

- 1: The usage of the narrative levels annotations makes no difference for subsequent analyses.
- 4: Subsequent analyses are possible only because of the narrative level annotations.

3. Do you gain new insights about narrative levels in texts by applying the foreign guideline, compared to the application of your own guideline?

Explanation: In most cases annotating a text in accordance to a guideline changes the evaluation of textual phenomena in the text, e.g., the quality (or quantity) of

narrative levels in the text.

- 1: It doesn't make a difference—I get no additional insights with the foreign guideline.
- 4: I gain a lot of new insights about narrative levels in texts based on this guideline.

4. Does the application of this guideline influence your interpretation of a text?

Explanation: Interpretations are normally based on the analysis of a text and thus on the observation of the presence (or absence) of certain textual phenomena. Therefore, the application of the guidelines may result in annotations that are relevant for your interpretation, e.g. the detection of a narrative level of a certain type may influence your interpretation of the reliability of a narrator.

- 1: My interpretation is independent from the annotations based on the guideline.
- 4: My interpretation is based primarily on the annotations based on the guideline.

### Measuring Inter-Annotator Agreement

In this shared task, we employed the metric  $\gamma$  (gamma) as developed by Yann Mathet, Antoine Widlöcher, and Jean-Philippe Métivier.<sup>1</sup> Its final score combines observed disagreement with chance disagreement ( $\gamma$  is thus calculated using *disagreements*, while most metrics are calculated using agreements). This is done in order to be able to compare evaluation schemes with different complexities and to avoid favouring more simple schemes (if the scheme is simpler, chance agreement is higher). Gamma is thus calculated as shown in equation 1, with  $\delta_o$  and  $\delta_e$  for the observed and expected disagreement respectively.

$$\gamma = 1 - \frac{\delta_o}{\delta_e} \quad (1)$$

**Chance Disagreement  $\delta_e$ .** For calculating the chance disagreement, Gamma takes the real annotations provided by an annotator, splits the text at a random point, and permutes the two parts. This is done repeatedly, until the disagreement in the permuted “text” approximates the real disagreement (in the entire population) with high confidence (above 95%). Based on these annotations, the

<sup>1</sup>“The Unified and Holistic Method Gamma ( $\gamma$ ) for Inter-Annotator Agreement Measure and Alignment,” *Computational Linguistics* 41, no. 3 (2015): 437-79.



chance disagreement can be calculated in the same way as observed disagreement.

**Observed Disagreement  $\delta_0$ .** Calculating the observed disagreement is based on an alignment and the pairwise comparison of the annotated segments. The alignment encodes which annotation of annotator 1 corresponds to which annotation of annotator 2 and is created in such a way that the overall inter-annotator agreement is maximal, i.e., all possible alignments are considered. For each possible alignment, the algorithm calculates an averaged observed disagreement by comparing the aligned segments.

For two aligned segments, Gamma considers both the positional and categorial disagreement. The *positional disagreement* expresses how different the position of two aligned segments is and is calculated as shown in equation 2. The functions  $end(x)$  and  $start(x)$  refer to the start and end position of the annotated segments, which is measured in token positions.

$$d_{pos}(u, v) = \left( \frac{|start(u) - start(v)| + |end(u) - end(v)|}{(end(u) - start(u)) + (end(v) - start(v))} \right)^2 \quad (2)$$

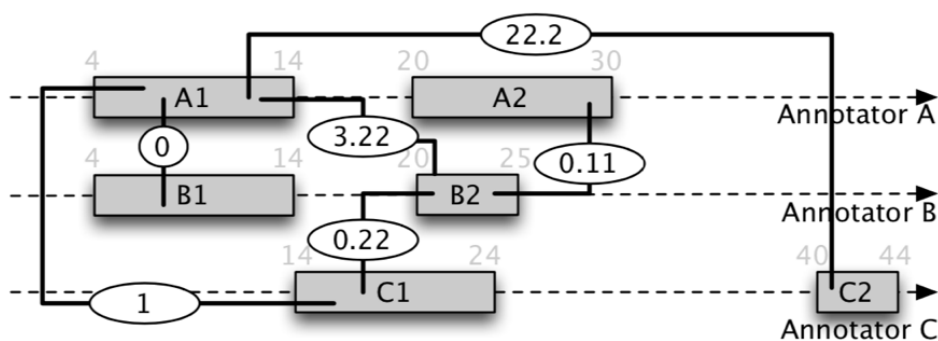


Figure 2: Example calculations of positional disagreement. Grey numbers show index positions, numbers in white oval shapes show the calculated disagreement (Mathet, Widlöcher, and Métivier, 451).

In equation 2, the numerator represents the difference between the starting and end positions of the two annotations, while the denominator incorporates the length of the respective annotations. Figure 2 shows several example situations and the resulting positional disagreement score. As can be seen, numbers between zero and one indicate some overlap; if  $d_{pos} > 1$ , the two annotations do not overlap. There is no upper limit on the positional disagreement. If the annotations differ widely in their position (e.g., are placed at the beginning and end of the text), they get a  $d_{pos}$ -value that is roughly as high as the text is long.

Incorporating *categorical disagreement* ( $d_{cat}$ ) allows to evaluate whether different annotation categories have been selected. If, for example, annotator 1 has assigned category A, while annotator 2 has assigned category B, a category disagreement is noted. Using a matrix, the difference between each pair of categories can be weighted by assigning a number between zero and one. Thus, a user of gamma can express that using category A instead of B is less severe than using A instead of C. There is, however, no way to automatically determine the severity of categorical disagreement. Instead, it is a preference that the user of Gamma has to provide.

In our evaluation, categories play a minor role and have thus been treated as equally distant: If a guideline specifies multiple categories, all pairs have been assigned a distance of one. Features' values attached to the annotation have been suffixed to the category name, so that differences between features are treated in the same way as category disagreement.

Finally, equation 3 shows how the two sub metrics are combined, using  $\alpha$  and  $\beta$  to express weighting (in our setting, both are set to one:  $\alpha = \beta = 1$ ).

$$d_{combi}(u, v) = \alpha d_{pos}(u, v) + \beta d_{cat}(u, v) \quad (3)$$

To measure Gamma, we employed an implementation provided by the developers on their web page. The way expected disagreement is calculated here leads to issues when annotations are sparse. If a single annotation covers the entire text, which is entirely plausible for narrative level annotation, there is no way to split the text and reshuffle the annotations. To circumvent this, we calculated Gamma individually for each text and on all eight texts concatenated together. The latter score was then used for the final ranking.

### Integration of the Evaluation Scores

The final score for each guideline was calculated as follows:

1. For each of the ten questions, the arithmetic mean over all answers is calculated. This results in ten values, distributed over three dimensions: four questions/values in the first dimension, two questions/values in the second and four questions/values in the third.
2. The Gamma scores are scaled to the interval of [1;4] and added twice as the scores of “virtual questions” in the second dimension. This results in four values per dimension, each in the interval [4;16].

3. In each dimension, all four (mean) values are added up.

This results in one score for each dimension, so guidelines can be ranked accordingly. As an overall score, we calculated the sum of the scores in all dimensions.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

# A Shared Task for the Digital Humanities

## Chapter 3: Description of Submitted Guidelines and Final Evaluation Results

Marcus Willand, Evelyn Gius, Nils Reiter

11.05.19

*Article DOI: 10.22148/16.050*

*Journal ISSN: 2371-4549*

*Cite: Marcus Willand, Evelyn Gius, Nils Reiter, "A Shared Task for the Digital Humanities Chapter 3: Description of Submitted Guidelines and Final Evaluation Results," Journal of Cultural Analytics. November 4, 2019. doi: 10.22148/16.050*

In this chapter, we give a descriptive overview of the annotation guidelines, their use of narrative level concepts, and the results of their quantitative evaluation. We will also connect some of the results to qualitative findings we uncovered during the workshop, although some references to the participants' objectives are conjecture. Finally, the chapter contains a reflection on the annotation and evaluation procedure.

### General Observations

Since this shared task was targeted at diverse audiences, the submissions and the disciplinary backgrounds of their authors are as diverse as expected. Table 1 shows key properties of the research teams.

Guideline	I	II	III	IV	V	VI	VII	VIII
Group size	2	1	2 (S)	4	1	2 (S)	3	1 (S)
Disciplinary background	Computational linguistics	English literature/ Data science	English literature	Literary studies/Digital humanities	Digital humanities /Literary studies	English literature	Computational linguistics	Literary studies
Country	U.S.A.	Ireland	Germany	Germany	Germany	Germany	Sweden	Canada

Table 1. Properties of the guideline authors. Group size indicates the number of authors of a guideline, (S) indicates that the guideline has been developed within a seminar or lecture. The disciplinary background is based on self-designation of the participants.

The participating teams also differ greatly in terms of age, gender, group size, academic level, research field, and disciplinary affiliation. This diversity is reflected in the submitted guidelines. They differ strongly in shapes and sizes: they range from 1 to 50 pages length as well as from theoretical essays to practical how-to’s (see Chapters 4-10 for the Guidelines).

While disciplinary differences come with diverging practical experiences and diverging genre knowledge about annotations and guidelines as such, there is no clear cut between computer scientists and humanists. Guidelines with authors from both areas aim at providing a mixture between conceptual definition and practical annotation instructions.

### Guidelines and Levels

The definition of this first task given by us organizers left it to the participants to select a useful/correct/reasonable theoretical basis for their guideline. To give an overview of the main theoretical foundations of the guidelines, Table 2 shows which publications and Table 3 shows which concepts are referred to in which guideline and/or rationale. Please note that different guidelines may employ different understandings of the same narratologist or concept (e.g., focalisation). Two guidelines referring to the same narratologist are not necessarily compatible, neither are two guidelines referring to a similar set of concepts. Please also note that this summary is a descriptive one and it is not our intention to suggest that guidelines *should* contain references to theoretical research.

### References to Research in Narratology <sup>1</sup>

---

1

Narratological publications	Guideline/Rationale							
	I	II	III	IV	V	VI	VII	VIII
Genette		x	x	x	x	x	x	
Jahn		x		x	x		x	x
Lahn/Meister				x	x			
Lämmert					x			
Mani								
Martínez/Scheffel				x				
Nelles			x (1997)		x (2005/1997)	x (1997)		
Neumann/Nünning					x			
Pier			x (2014)	x (2012)	x (2014)	x (2014)		
Rimmon-Kenan								
Romberg								
Ryan			x	x	x			
additional references		x	x	x	x	x	x	
own approach	x							x

*Table 2. Narratological publications operationalized by the guidelines. The assignment is based on the references in the guidelines and/or rationale.<sup>2</sup>*

As Table 2 shows, the most referenced publication was Genette (Narrative Discourse) with six guidelines referring to it. This is not surprising, since Genette introduced the concept of narrative levels and most other theorists relate to his work in some way. Other publications cited by three or more contributions that are thus comparably present in the guidelines are the introductory texts by Manfred Jahn (Narrative Levels) and John Pier (Narrative Levels), as well as Marie-Laure Ryan's account (Possible Worlds and others, see Endnote 2). While Jahn and Pier may have been chosen due to their introductory character—they don't develop anything new but summarize the most prominent existing approaches—Ryan is probably the most formalized approach among the suggested and may thus have been considered especially suited for the guideline development. The fact that no guideline refers to Mani (Computational Narratology) and Romberg (Narrative Technique of the First-Person Novel) is probably due to the rather

abstract description of computational narratology in the former and, on the contrary, the rather lengthy discussion and focus on first-person narrators in the latter. Seven out of eight contributions also cited additional research (cf. endnote 2). Guideline I developed their own approach. Even though the authors refer to certain concepts (cf. Table 3), there is no explicit reference to a theorist. Guideline I conceives of narrative as linguistic representation of a story and focuses on the identification of borders of narratives, introducing the notion of uninterrupted vs. embedded vs. interruptive narrative (referring also to analepsis and prolepsis). Guideline VIII gives its own definition of narratives and focuses on level changes that can be identified with a test (“Let me tell you a story”).

### References to Narratological Concepts

Narratological concept	Guideline/Rationale							
	I	II	III	IV	V	VI	VII	VIII
definition of “narrative”	x			x	x	x	x	x
narrative levels	x	x	x	x	x	x		x
discourse levels							x	
narrator (identity, relation to text)	x	x	x	x	x	x	x	x
narratee			x			x	x	
structure of the level change/embedding	x		x	x	x	x	x	
nature of boundary between levels (e. g. metalepsis)			x	x	x	x	x	
speaker (illocutionary boundary)			x	x	x		x	
change of world (ontological boundary)	x		x		x	x		
focalisation			x			x	x	
analepsis/prolepsis	x	x						
stream of consciousness/free indirect discourse		x						
extended/compressed time		x						

*Table 3. Narratological concepts operationalized by the guidelines. The assignment is based on the explicit reference to them in the guidelines.*

Table 3 gives a first impression of the concepts that were considered relevant by the guideline authors for the identification of narrative levels. The listed concepts can be divided into concepts connected directly to narrative levels (such as boundary and change related concepts) and concepts that typically co-occur with narrative levels (such as focalization or anachronies, i.e. analepses and prolepses). While directly connected concepts can be used for the operationalization of narrative levels, co-occurring concepts often appear within narrative levels and are interesting for further analysis. This intermingling of concepts is probably connected to theoretical openness as it is typical for the humanities and the vagueness of the narratological concepts.

Some discussed differences can be explained by the diverging research objectives of the participating teams; e.g., whether the narrative level annotation is supposedly used for narratological concept development (guideline IV), identifying narratological concepts other than levels in literary texts (as time, e.g. guideline II) or to recognize linguistic discourse levels (guideline VII).

## Results of the Evaluation

### Final Ranking

First of all, we would like to mention once again that both content (conceptual coverage, applicability, and usefulness) and method of the evaluation (questionnaire, IAA) arise from the specifics of a shared task in the humanities. The multi-dimensional approach allows for an evaluation of the guidelines irrespective of their disciplinary and research background, their aims, and their understanding of narratological concepts. As we have already pointed out, the knowledge before guideline writing and the aims of guideline application are crucial and lead to rather different guidelines. Evaluating such diverging guidelines in a fair manner requires the evaluation to be multi-dimensional and objective-agnostic. This is exactly what the three evaluation dimensions are supposed to capture without valuing one disciplinary paradigm over another.

Below, we present the final ranking for the overall evaluation and the three dimensions in tables 3 to 7. The tables show scores and standard deviation for each question in the questionnaire, grouped by dimension. The questions can be found in Chapter II: “Introduction to Annotation, Narrative Levels and Shared Tasks”.



Guideline	Dimension			Overall
	Conceptual Coverage	Applicability	Usefulness	
Guideline V	14.14	12.09	<b>12.88</b>	<b>39.10</b>
Guideline II	11.17	11.89	12.57	35.63
Guideline VI	12.33	11.01	11.37	34.71
Guideline IV	<b>14.43</b>	7.71	11.26	33.40
Guideline VIII	8.10	<b>14.14</b>	9.12	31.36
Guideline VII	11.60	9.82	9.77	31.18
Guideline I	7.83	10.39	10.00	28.22
Guideline III	10.29	6.48	10.95	27.72

Table 4. Final results of the evaluation (overall scores). The highest score in each dimension is shown in bold.

## Conceptual Coverage

Guideline	Rank	Question			
		Q1	Q2	Q3	Q4
Guideline I	8	3.00 ± 1.10	1.00 ± 0.00	1.00 ± 0.00	2.83 ± 0.98
Guideline II	5	2.67 ± 1.03	3.00 ± 1.10	2.83 ± 0.75	2.67 ± 1.21
Guideline III	6	2.43 ± 0.98	2.86 ± 1.07	2.71 ± 1.38	2.29 ± 0.76
Guideline IV	1	3.71 ± 0.49	4.00 ± 0.00	3.71 ± 0.49	3.00 ± 0.82
Guideline V	2	3.71 ± 0.49	4.00 ± 0.00	3.43 ± 0.53	3.00 ± 1.00
Guideline VI	3	3.33 ± 0.82	3.33 ± 0.82	2.67 ± 0.52	3.00 ± 0.89
Guideline VII	4	2.71 ± 1.11	3.33 ± 0.82	2.71 ± 0.95	2.83 ± 1.33
Guideline VIII	7	2.29 ± 1.11	1.00 ± 0.00	1.67 ± 1.21	3.14 ± 0.69

*Table 5. Evaluation results for conceptual coverage (dimension 1). The table shows mean and standard deviation for each question. The questions are shown in Chapter 2 (see above).*

The following reflections about the final results cover the three dimensions one after another. The results shown in Table 4 (conceptual coverage) are based on four questions. Guideline IV achieved the top position in this dimension. It is one of the guidelines that focuses on an in-depth description of the used narratological categories and definitions, which seems to be reflected by the positive evaluation of its conceptual coverage. On the opposite end of the spectrum is guideline I, which was ranked lowest. This coincides with the self-description and primary research interest of its authors, which is a “computational understanding of sto-

ries” (see guideline I, Chapter 4 of this issue). Their focus on future automation plans led to a guideline without ties to a theory and subsequently to a poor rating in this category. Guideline I also includes dialogical text genres (as scripts of TV shows and the transcripts of court cases) which caused some confusion for those who expected a narration to be necessarily narrated by a narrator. It is possible that the large number of narratologically versed participants penalized a deviation from established narratological consensus.

## Applicability

Guideline	Rank	Questions		Inter-Annotator Agreement	
		Q1	Q2	Evaluation score (counted twice)	Gamma
Guideline I	5	2.60 ± 0.55	2.67 ± 0.52	2.56	0.18
Guideline II	3	3.17 ± 0.41	2.17 ± 0.98	3.28	0.24
Guideline III	8	2.57 ± 1.13	1.43 ± 0.53	1.24	0.07
Guideline IV	7	3.57 ± 0.53	2.14 ± 0.38	1.00	0.05
Guideline V	2	3.00 ± 1.15	2.29 ± 0.95	3.40	0.25
Guideline VI	4	3.17 ± 1.17	2.00 ± 0.89	2.92	0.21
Guideline VII	6	2.33 ± 0.52	1.17 ± 0.41	3.16	0.23
Guideline VIII	1	2.71 ± 1.38	3.43 ± 0.79	4.00	0.30

Table 6. Evaluation results for applicability (dimension 2). For the two questions, the table shows mean and standard deviation. For the inter-annotator agreement,

*the table shows the scores used for the ranking (in the interval) as well as raw gamma scores. The questions are shown in Chapter 2 (see above).*

The applicability score is based on the inter-annotator agreement and two questions from the questionnaire (cf. Table 5). The first of the two questions queried how well an expert in narratology would be able to apply the guideline to a narrative text; the other question asked the same for laypersons. Guideline VIII is the overall winner in this dimension while guideline III scored the lowest points. For an interpretation of the results, it is worth looking at the scores for individual questions and the IAA separately. The fact that Guideline VIII was rated best raises the question of a relation between the guideline's relatively simple level concept and its applicability. The detailed results reveal that the first place is partly due to the guideline having the highest inter-annotator agreement score. It seems obvious to deduce that simplicity is related positively to applicability. However, this is not completely supported by the answers in the questionnaire for this dimension. Guideline VIII gained only a lower midfield position in the question about expert applicability, but it was considered to be the most applicable guideline for laypersons (note that this judgement was cast before the IAA scores were revealed). So, in practice, a basic level concept seems to be applicable with great congruence, but the results of the questionnaire suggest that simplicity is understood to come with restrictions for experts. We attribute this to the assumed incapacity of the guideline to do justice to the complexity of narrative levels.

Complexity and applicability thus seem to correlate negatively. However, the comparison to other guidelines raises doubts about the derivability of such a general rule. Guideline V, a guideline with a relatively complex level concept and the overall winner of the evaluation, achieved second rank not only in the first dimension of conceptual coverage, but also in the applicability dimension. In this dimension, the result is based on another second rank in the IAA and a third/fourth rank in the questions. Thus, the guidelines with the two best results (VIII and V) in the applicability dimension are very different in nature and there seems to be no direct correlation between guideline complexity and applicability.

But there are still interpretable results. As the first question refers to annotators *with* a narratological background, it is not surprising that the highest score was reached by guideline IV (the winning guideline of the conceptual coverage dimension) followed by Guideline V. The fact that both guidelines with best results in conceptual coverage scored only average points for layperson application (question two) might be explained by the consideration that laypersons can benefit from clear and explicit conceptual level description, but also run the risk of being overwhelmed by complexity.

The low rank of guideline III<sup>3</sup> might be explained by a combination of factors. The guideline covers a broad range of narratological concepts. At the same time, it neither defines these narratological concepts in depth nor does it give examples on how to apply its annotation categories. The mere description of annotation categories seems to lead to difficulties in their application (see the very low inter-annotator agreement). Conversely, it can be said that applicable guidelines should have to demonstrate their categories by way of example. This is what the top two guidelines do in great detail. Even though for the most part they only achieved average results in the questionnaire, they were elevated in the *applicability* dimension by relatively high inter-annotator agreement scores.

---

<sup>3</sup>Guideline III is not printed in this volume, as its authors withdrew their submission.

## Usefulness

Guideline	Rank	Question			
		Q1	Q2	Q3	Q4
Guideline I	6	3.17 ± 0.75	3.00 ± 0.71	2.17 ± 0.75	1.67 ± 0.52
Guideline II	2	3.50 ± 0.55	3.40 ± 0.89	3.00 ± 0.89	2.67 ± 0.52
Guideline III	5	3.33 ± 0.82	3.17 ± 0.98	2.29 ± 0.76	2.17 ± 0.75
Guideline IV	4	3.29 ± 0.76	2.83 ± 0.98	2.71 ± 0.76	2.43 ± 0.98
Guideline V	1	3.50 ± 0.55	3.40 ± 0.55	3.14 ± 0.69	2.83 ± 0.75
Guideline VI	3	3.40 ± 0.55	3.40 ± 0.89	2.17 ± 0.75	2.40 ± 0.55
Guideline VII	7	3.00 ± 0.63	2.60 ± 0.55	2.17 ± 0.98	2.00 ± 0.89
Guideline VIII	8	3.20 ± 0.84	2.80 ± 1.10	1.29 ± 0.76	1.83 ± 0.75

*Table 7. Evaluation results for usefulness (dimension 3). The table shows mean and standard deviation for each question. The questions are shown in Chapter 2 (see above).*

The top rank of guideline V in usefulness is most likely due to 1) the multifold examples of literary texts that illustrate the use of the elements to be annotated and 2) the very clear description of the research objectives of the guideline authors. This is also the case for guideline II, which ranked second in this dimension: Guideline II states that it is “designed for annotating analepsis, prolepsis, stream-of-consciousness, free indirect discourse, and narrative levels, with facility also for annotating instances of extended or compressed time, and for encoding the identity of the narrator” (see guideline II, Chapter 5). Since usefulness is the category that addresses a variety of possible cases in which the annotated texts might

allow further research, this result is quite interesting. It shows that expressing a specific application area in the guideline allows the evaluators to picture opportunities for use more clearly, but on the other hand might be more restricting compared to a guideline that does not express a specific application area.

Finally, some remarks on guideline V, the **overall winner guideline** in the first shared task: Guideline V was ranked second best in conceptual coverage and applicability (as well as inter-annotator agreement); it also reached the top position in usefulness, which in combination makes it the overall winner with some distance to the second rank. Both the quantitative results from the evaluation as well as the qualitative results from the discussion suggest that guideline V defines narrative levels in quite some detail and with particular precision: The guideline distinguishes the narrative level concept from “narrative acts” and refers to other narratological concepts (such as narrator) in a way that is helpful for identifying levels. Abstract examples in the form of diagrams and tables are given, illustrating systematically how narrative levels are to be understood. Furthermore, concrete text examples are annotated with these concepts. Last but not least, a workflow for the practical annotation is given as well as a very clear description of the aims of the annotation of each concept.

## Observations on the Evaluation Process

As this is the first time such a shared task takes place, we think it is important to end this introductory chapter with an assessment of the evaluation process and the shared task as a whole. This includes our version of measuring inter-annotator agreement, the questionnaire and—the very heart of our evaluation—the three dimensional model.

### Inter-Annotator Agreement

Firstly, the role and calculation of the inter-annotator agreement should be reflected. The agreement scores are based on a comparatively low number of annotations done by annotators with no systematic training. Therefore, their level of expertise varied: The student annotators were basically untrained (although some had experience in other annotation tasks) and had virtually no knowledge of narratological theory. The foreign annotators naturally were trained on their own guidelines and may have had problems to disengage themselves from them. Both issues applied to all guidelines. Nevertheless, the participants stated differ-

ent degrees of satisfaction with the foreign annotations based on their guidelines in the discussion. Some participants did not feel adequately understood by the annotators, which was especially the case with narratologically complex guidelines. Therefore, this problem may be at least partially caused by the interdisciplinarity of the shared task.

In addition, it was observed that in the case of guideline IV the annotations made by student annotators had a higher agreement with the guideline authors than the annotations made by the other groups. Therefore, the question arose whether we should have taken into account the expected disciplinary competences when distributing the guidelines for foreign annotation among the participants. We do not believe that this is a viable way to go, but we believe that a profound revision of the mutual annotation model between participating teams (“foreign annotation”) is worth considering. Ultimately, this approach was born out of the need to obtain as many annotations as possible for each guideline to get data for the IAA. Given appropriate funding, it makes sense not to calculate the agreement on the basis of mutual guideline annotations, but to have it done exclusively by “external”, similarly trained annotators. In fact, this is what we will do in a second annotation round, the results of which will be published in the second volume of this special issue.

A further observation worth mentioning was found looking at the inter-annotator agreement scores for guideline VII, where the foreign and student annotations were more similar to each other than to the annotations by the guideline authors. This also points to the different disciplinary competences, in this case to the strong linguistic influence of this guideline. Since our annotators were aware of the narratological focus of most other guidelines, they might have translated the guideline into a narratological perspective in the same way the co-annotating participant(s) did.

### **Questionnaire**

We also want to highlight several issues that have been raised about the questionnaire itself during the workshop discussion: Filling in questions in the conceptual coverage dimension requires a broad narratological knowledge, which not all participants possessed. The two questions in the applicability dimension were intended to test the comprehensibility of the guideline for experts and non-experts, but as participant groups were homogeneous with respect to their expertise, one question could only be filled in with a grain of salt. As we have seen with guideline VIII, there is a clear mismatch between measuring applicability by us-



ing inter-annotator agreement, and by predicting applicability in a questionnaire. This is not surprising per se, but the magnitude of the mismatch is. Furthermore, it has been mentioned that clarity is such a relevant feature of guidelines that it should have been explicitly evaluated.

Some literary scholars voiced concerns about presenting the results of our complex evaluation in mere numbers. The dissatisfaction, though, did not refer to the results of the inter-annotator agreement of their own guideline, but was rather based on a general methodological skepticism, which was not shared to this extent by the participants with an affinity for automation. In the usefulness dimension, participants found it difficult to assign scores to guidelines they had not been working with intensively. Without this practical knowledge, the answers can only be conjecture. Nevertheless, despite the difficulties in filling in the questionnaire, the numeric scores are relatively homogeneous across the groups (low standard deviation, see above).

### The Three Dimensional Evaluation Model

Lastly, we would like to reflect on the nucleus of the whole process, the three dimensional evaluation model. Since each evaluation dimension favors characteristics that may be related to the disciplinary origin of the guideline authors and thus may lead to biases, the combination of the three dimensions was arranged in such a way that they cancel out those biases. For example, guideline IV, whose authors all have a background in literary studies, achieved the first position in the dimension of *conceptual coverage*, a midfield position in *usefulness* and the second to last position in *applicability*. Guideline I, written by researchers in natural language processing, was ranked last in *conceptual coverage*, but received average scores in the other two dimensions. This gives us reason to believe that disciplinary advantages and disadvantages are indeed offset by our evaluation approach. The fact that guideline IV and VIII reached inverted ranks in dimensions one and two also indicates that the evaluation dimensions neutralize disciplinary advantages. The guideline that was ranked highest overall received high scores in all dimensions, but was ranked first only in one dimension. This suggests that to succeed in general, one needs to strike a compromise between the dimensions. This is the effect we were aiming for when designing the evaluation scheme.

As an outlook, for our readers both the quantitative results and the distribution of the narratological concepts might serve another purpose: They provide a structure and categorization of the submitted guidelines. Researchers and scholars who are interested in narrative levels and/or their annotation for whatever pur-

pose can browse through the rationals, guidelines and short reviews, all published in the following. Those instructive documents as well as our introduction allow for an informed decision on any guideline or the combination of multiple guidelines, both to be used as a starting point for original, new work.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

## **Part II.**

# **Annotation Guidelines and Reviews**

# Annotation Guideline No. 1: Cover Sheet for Narrative Boundaries Annotation Guide

Joshua Eisenberg and Mark Finlayson

09.24.19

*Peer-Reviewed By: Meredith A. Martin*

*Article DOI: 10.22148/16.051*

*Journal ISSN: 2371-4549*

*Cite: Joshua Eisenberg and Mark Finlayson, "Annotation Guideline No. 1: Cover Sheet for Narrative Boundaries Annotation Guide," Journal of Cultural Analytics. November 20, 2019. doi: 10.22148/16.051*

## 1. Introduction

### 1.1 Purpose of the Project

Narratives and stories are found all over the world, in every culture, and they are used by every person every day. For computers to communicate with people in a natural and respectful manner, they need to understand stories. Unfortunately, computational understanding of stories is currently in its early stages: computers cannot yet identify even basic characteristics of a narrative, such as where it begins and ends. To train and test a computer's ability to identify the beginnings and endings of narratives (what we call here *narrative boundaries*) we are collecting human judgments.

### 1.2 What is Annotation?

Annotation is the process of explicitly encoding information about a text that would otherwise remain implicit. For this study, annotation is the record of hu-

man judgments identifying where a narrative begins and ends, which are called narrative boundaries. You will be highlighting spans of text in PDF documents to explicitly mark the boundaries of each narrative.

## 2. Narrative Boundaries

A narrative is a discourse presenting a coherent sequence of events which are causally related and purposely related, concern specific characters and times, and overall displays a level of organization beyond the commonsense coherence of the events themselves, such as that provided by a climax or other plot structure.

Narrative is a linguistic representation of a *story*. A *story* is a series of events effected by animate actors or characters. A story is an abstract construct, with two essential elements: plot (*fabula*) and characters (*dramatis personae*). The art of storytelling is much more complicated than merely listing events carried out by characters. There is great importance in the storyteller's choice of which details are revealed to the reader, the order in which plot events are told, whether to embed stories within each other, and whether to interrupt the telling of one story to make space for a new one. Even the choice of what details (character traits, setting, history) the author reveals to the reader is important.

Narrative is more concrete than story, in that narrative is made up of words, but a story is formed through the co-occurrence of characters who enact events which advance a plot forward. Narratives occupy spans of text, while stories are a more complex relationship involving characters and events. Throughout this document, we will say the *narrative* is the span of text that expresses the story, or simply *narrative text*.

These spans can appear in multiple forms: a narrative can appear contiguously as one solid span of text, or it might be embedded in another narrative, or it might even interrupt the preceding narrative. Many novels and short stories contain multiple instances of embedded and interruptive narratives, often with intricate combinations of the two phenomena. This is also true of scripts of TV shows and movies and the transcripts of court cases. The main goal of this study is to find and mark these narrative boundaries to enable us to develop, train, and test algorithms for this phenomenon.

Every narrative has at least two narrative boundaries: the *start point*—the position in text of the first character of the first word in the narration—and the *end point*—the position of the last character after the last word in the narration. The simplest kind of narrative is an uninterrupted one. The start point of such a nar-

ration is the first character of the text, and the end point the last character of the text. This text's narrative has only two boundaries.

1. I woke up early in the morning, checked the weather app on my phone and decided it would be a perfect day to go to the beach. I grabbed a book, a towel, and sunglasses, got in my car, and drove to the beach. I read my book, watched the waves, and took a quick swim. I dried off and drove home. It was a great day, even though I forgot to bring sun screen and got a sunburn.

(1) contains an uninterrupted narrative by a first-person narrator who tells the story of their trip to the beach. The narrator uses the first-person point of view to narrate. There are no shifts in time, and no interrupted narratives. The next three sections will define the different types of narrative boundaries. The texts that we will annotate in this study each have more than two narrative boundaries, and contain various arrangements of these boundaries. Both embedded and interruptive narrative can be found in any of the texts in this annotation study, including the TV and movie scripts, as well as the court transcripts.

## 2.1 Embedded Narratives

Narratives can be embedded in one another. An embedded narrative tells a story within a story. Before we discuss how embedded narratives occur in text, let's define how we refer to the relationship between the layers. The *original narrative* is the narrative in which the embedded narrative is told, and the original narrative contains an event (explicit or implied) that signals the telling of an embedded narrative. The *embedded narrative* is the narrative that is embedded within the original narrative.

Figure 1 contains a narrative boundary diagram for a text that contains an embedded narrative. The lower bar represents the span of text the original narrative appears in, while the upper bar represents the embedded narrative. The horizontal axis represents the text under consideration; the graph progressing from the left to the right represents the position in the text advancing from the first word to the last.

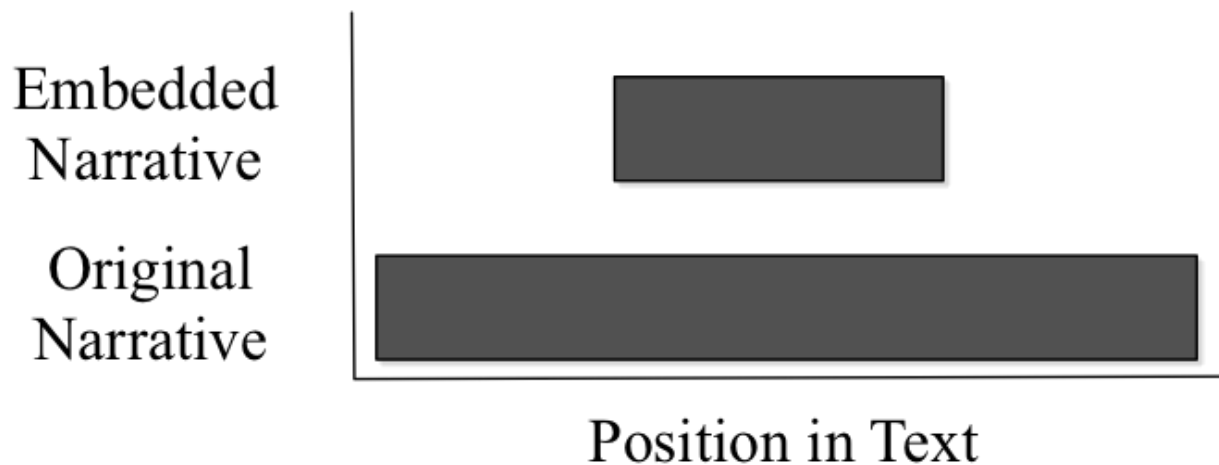


Figure 1. Narrative boundary diagram for an embedded narrative.

An embedded narrative occurs when there is a plot event in the original narrative that triggers the telling of another story in the story. The narrative that tells the second story is the embedded narrative. A typical example of this is when there is a conversation in the original narrative and in this dialogue one of the participants narrates a story. The plot event in the original narrative that signals the embedded narrative is the character telling a story.

Recall (1), a story about a day at the beach. Example 2, is an altered version of Example 1, one with an embedded narrative inserted. The span of text that contains the embedded narrative is surrounded by a brackets with a subscript of 1.<sup>1</sup> The narrative boundaries in Example 2 are graphically represented in the narrative boundaries diagram in Figure 1.

2. I woke up early in the morning, checked my weather app on my phone and decided it would be a perfect day to go to the beach. I grabbed a book, a towel, and sunglasses, got in my car, and drove to the beach. I read my book, watched the waves, and took a quick swim. As I emerged from the water a disheveled looking pirate washed ashore, [<sub>1</sub>“Aye Aye! I have just been washed ashore. I was the captain of the Shivering Sparrow, but there was a mutiny onboard. All of my crew including my parrot turned on me, and made me walk the plank. I clung onto a piece of driftwood for three days, and now I am here. Where am I?”]<sub>1</sub> I didn’t believe the pirate’s story, so I ignored him and walked away. I dried off and drove home. It was a great day, even though I forgot to bring sun screen and got a sunburn.

Example 2 contains a basic example of an embedded narrative. The story is al-

<sup>1</sup>[1...]<sub>1</sub>

most the same as Example 1's, except when the narrator gets out of the water, he encounters a pirate, who tells him a story about being abandoned at sea, clinging to a piece of wood, and washing ashore. The pirate's embedded narrative is surrounded by brackets with a subscript of 1, and is embedded in the narrative of the original narrator. The original narrative ends the same way in Example 1.

In Example 2's original narrative, the original narrator witnesses the pirate telling a story. The pirate's narration is a plot event in the lower level. This plot event in the original narrative triggers the start of the embedded narrative. The plot events of the pirate's story are a part of the embedded story, since they are told in the pirate's embedded narrative. The embedded narrative contains a story with events that are separate from the events in the story from the original narrative. It is also possible for the original narrator to tell an embedded narrative in the narrative text.<sup>2</sup> This type of narrative can occur via embedded flashbacks, which will be discussed in Section 2.3.

3. Her mojito glass was empty. She signaled the bartender and asked for a wine list, and, after some deliberation, she chose a glass of Napa Cabernet Sauvignon. Tsukuru had only drunk half his highball. The ice had melted, forming droplets on the outside of his glass. The paper coaster was wet and swollen.

[<sub>1</sub>“That was the first time in my life that anyone had rejected me so completely,”]<sub>1</sub> Tsukuru said. [<sub>1</sub>“And the ones who did it were the people I trusted the most, my four best friends in the world. I was so close to them that they had been like an extension of my own body. Searching for the reason, or correcting a misunderstanding, was beyond me. I was simply, and utterly, in shock. So much so that I thought I might never recover. It felt like something inside me had snapped.”]<sub>1</sub>

The bartender brought over the glass of wine and replenished the bowl of nuts. Once he'd left, Sara turned to Tsukuru.

“I've never experienced that myself, but I think I can imagine how stunned you must have been. I understand that you couldn't recover from it quickly. But still, after time had passed and the shock had worn off, wasn't there something you could have done? I mean, it was so unfair. Why didn't you challenge it? I don't see how you could stand it.”

---

<sup>2</sup>Here *narrative text* means the text the narrator uses to narrate to the reader. Narrative text does not include text in quotes or direct speech.



Example 3 is an excerpt from a novel (Murakami, 2014). This excerpt contains an embedded narrative, which is surrounded by brackets with the subscript of 1. In the original narrative, Tsukuru and Sara are at a bar, on a date, and Tsukuru is telling Sara a story about his past. The text surrounded in brackets with the subscript of 1 is part of the embedded narrative, since it contains Tsukuru telling a story about his previous rejection. This is a narration about his past; He is explaining how he felt, and why he acted a certain way. Note that only the bracketed text is part of the embedded narrative. The final paragraph is not part of the embedded narrative because it is not a telling of the embedded story. It is Sara trying to verbalize her empathy for Tsukuru and asking him a clarifying question.

It is important to note that the phrase “Tsukuru said” in the second paragraph is not part of the embedded narrative because it is an action that occurs in the original narrative. Tsukuru is having his conversation within the frame of the original narrative where he is on a date with Sara.

One final reminder, the entire span of text in Example 3 is part of the original narrative. Even the text of the embedded narrative, which is surrounded by brackets with subscript of 1, belongs to both the original and embedded narrative; it belongs to the original narrative because Tsukuru is saying these words to Sara in the bar while on his date. This is a part of the chain of events of the original narrative; it also belongs to the embedded narrative because the words he is saying tell a story that is separate and independent from the story in the original narrative.

Before we move to interruptive narratives, let’s talk about a canonical example of an embedded narrative: Joseph Conrad’s *Heart of Darkness*<sup>3</sup> where there is a homodiegetic narrator on a boat, listening to a story told by his shipmate Marlow. Marlow’s story, which is told in dialogue, is the main story of the novel. The original narrator’s story is quite simple, he is just a passenger on a boat listening to Marlow. The *real* story of *Heart of Darkness* is the story that Marlow is telling the original narrator, about Marlow’s experiences in Africa.

## 2.2 Interruptive narratives

Narratives that interrupt the original narrator’s narration are called **interruptive narratives**, which are different from embedded narratives. This is common in books where each chapter has a different narrator. For example, in the majority of the novel *1Q84*,<sup>4</sup> all the odd numbered chapters are narrated from the perspective of the heroine, Aomame, and the even numbered chapters are narrated from

<sup>3</sup>Joseph Conrad, *Heart of Darkness* (New York, NY: W. W. Norton & Company, 2016).

<sup>4</sup>Haruki Murakami, *1Q84* (New York, NY: Vintage International, 2011).

the perspective of the hero, Tengo. The boundaries at the end of each chapter in this novel mark interruptive narrative boundaries. For example, at the end of an odd numbered chapter, the narrator switches from the perspective of Aomame to Tengo, and at the end of each even numbered chapter, the narrator switches from the perspective of Aomame to Tengo.

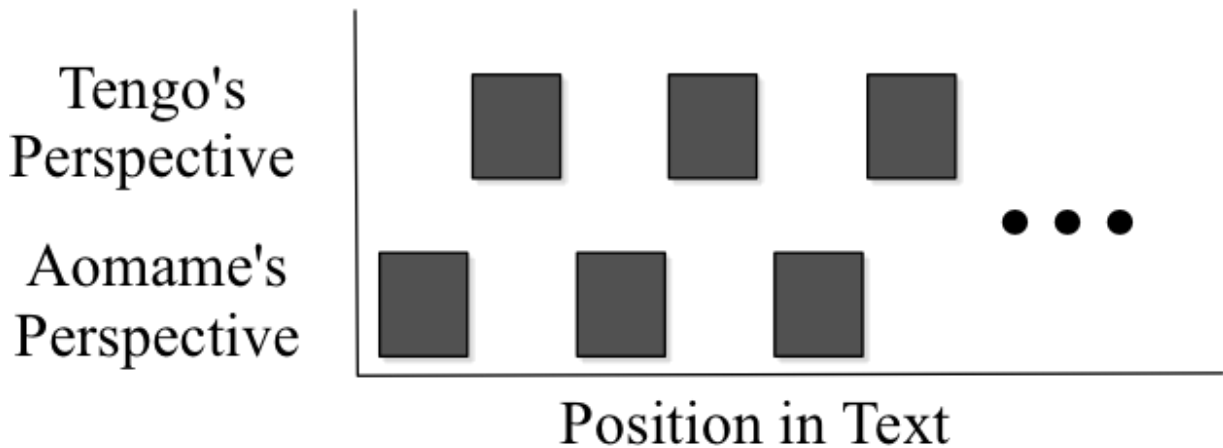


Figure 2. Narrative boundary diagram for the dual interruptive narrations from *1Q84*.

Interruptive narratives can occur within chapters, or, for our purposes, within short stories, chapters of novels, or in the dialogue of a script. Sometimes the person narrating will change; at other times, the original narrator is a first-person narrator, and then the narrator will suddenly shift to a third person impersonal narrator, or vice versa. If the narrator changes, there is usually an interruptive narrative boundary.

Sometimes there will be a section break which denote the change of narrator. Section breaks are visual markers that separate text. Sometimes a section break is signaled by a series of special characters, like an asterisk (\*) or a horizontal rule (a thin, horizontal line). Sometimes there will just be many blank lines in a section break. Note that the presence of a section break does not guarantee the presence of an interrupted narrative. For example, there can be a section break, and immediately after the break the narration is continued by the same narrator, from the same point in time that the narrative before the section break left off.

The difference between interruptive and embedded narratives may seem subtle, but there is a difference. In an embedded narrative, a plot event occurs in the story of the original narrative, which triggers the telling of an embedded narrative. An interruptive narrative is triggered by the original narrative stopping. The trigger of an interruptive narrative is not a plot event in the original narrative, instead it is more of a meta-event, where something more structural, about how the story

is being told, changes. Once the original narrative has stopped, the interrupting narrative begins to be told. The actual person narrating the story can change, or the narrator remains consistent but the time in which the story is told changes. If you are questioning whether a narrative is interruptive, you should ask yourself: Is the telling of the span in question a plot event in the original narrative? If it is, then the span in question is embedded. If not, then it is interruptive.

Let's consider an example of a story with an interruptive narrative. Below is Example 4. It is again an altered version of Examples 1 and 2. The story is like Example 2, in that the narrator goes to the beach, reads, goes for a swim, and encounters a pirate upon exiting the water. After the original narrator observes the pirate washing ashore, there is a section break signaled by three asterisks. Surrounded by brackets with subscript 1 is the interruptive narrative of the pirate, told in first person. The pirate telling this story is not an event in the original narrative, which is what happened in the embedded narrative of Example 2. There is no event, in the original narrative of Example 4, where the pirate tells a story. Instead, there is an interruption of the original narrative, the pirate tells his story, and then the original narrator begins telling his story. Figure 3 contains a narrative boundary diagram for this generic interruptive narrative.

4. I woke up early in the morning, checked my weather app on my phone and decided it would be a perfect day to go to the beach. I grabbed a book, a towel, and sunglasses, got in my car, and drove to the beach. I read my book, watched the waves, and took a quick swim. As I emerged from the water a disheveled looking pirate washed ashore.

\* \* \*

[<sub>1</sub> I have just been washed ashore. I was the captain of the Shivering Sparrow, but there was a mutiny onboard. All of my crew including my parrot turned on me, and made me walk the plank. I clung onto a piece of driftwood for three days, and now I am here.]<sub>1</sub>

\* \* \*

The pirate looked like he just went through a tragic ordeal, but he was a pirate, so I decided it was best to ignore him. I dried off and drove home. It was a great day, even though I forgot to bring sunscreen and got a sunburn.

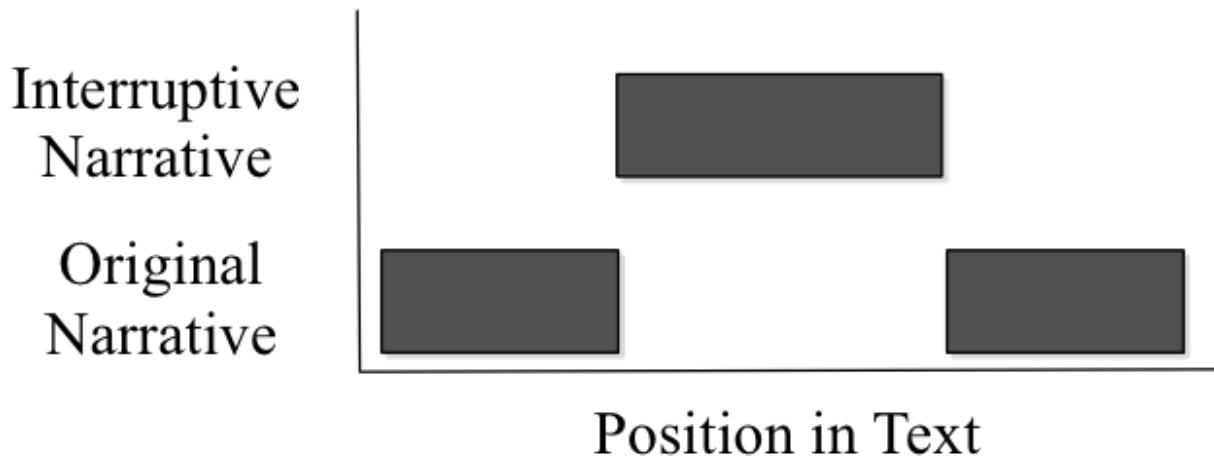


Figure 3. Narrative boundary diagram for an interruptive narrative

Next we will consider another excerpt from Murakami.<sup>5</sup> Example 5 contains two narrative levels, one surrounded by brackets with subscript 1 and one surrounded by brackets with subscript 2. The first narrative, brackets with subscript 1, is a continuation of the original narrative from Example 3, when Tsukuru is on a date with Sara. The first narrative is interrupted by a third person narrator, who tells a story about Tsukuru’s adolescence. There is a narrative break punctuating the two narratives. This is narrative is surrounded by brackets with subscript 2 and it is an instance of an interruptive flashback, which is discussed in the next section.

5. [<sub>1</sub>“You can hide memories, but you can’t erase the history that produced them.” Sara looked directly into his eyes. “If nothing else, you need to remember that. You can’t erase history, or change it. It would be like destroying yourself.”

“Why are we talking about this?” Tsukuru said, half to himself, trying to sound upbeat. “I’ve never talked to anybody about this before, and never planned to.”

Sara smiled faintly. “Maybe you needed to talk with somebody. More than you ever imagined.”]<sub>1</sub>

• • •

[<sub>2</sub> That summer, after he returned to Tokyo from Nagoya, Tsukuru was transfixed by the odd sensation that, physically, he was being completely transformed. Colors he’d once seen appeared completely different, as if they’d been covered by a special filter. He

<sup>5</sup>Haruki Murakami, *Colorless Tsukuru Tazaki and his years of pilgrimage: a novel* (New York, NY: Alfred A. Knopf, 2014).

heard sounds that he'd never heard before, and couldn't make out other noises that had always been familiar. When he moved, he felt clumsy and awkward, as if gravity were shifting around him.

For the five months after he returned to Tokyo, Tsukuru lived at death's door. He set up a tiny place to dwell, all by himself, on the rim of a dark abyss. A perilous spot, teetering on the edge, where, if he rolled over in his sleep, he might plunge into the depth of the void. Yet he wasn't afraid. All he thought about was how easy it would be to fall in.]<sub>2</sub>

### 2.3 Time shifts: Flashbacks and Flashforwards

There are two types of time shifts in story telling: *flashbacks*, (also known as *analepsis*), and *flashforwards* (also known as *prolepsis*). Both flashbacks and flashforwards are recurrent in storytelling. A flashback occurs when the time of the events told in the narration shift from the present to a time in the past. Flashbacks might occur when the narrator remembers something that happened in the past. A flashforward is similar, except the events are from the future. Flashforwards, can come in the form of visions or prophecies. Other times, flashforwards foreshadow or reveal key events that will occur in the future, even though the narrator might not know that these events will occur. Both flashbacks and flashforwards are popular storytelling devices in both literature and film. There are two ways flashbacks can be narrated:

*Embedded Flashbacks* are embedded in the original narrative. In the original narrative, the narrator is narrating a story about the present, and then the narrator will shift the subject of their narration to telling a story about events that happened in the past. Sometimes the retelling of past events will use verbs in the past tense. The narrator is telling a story about the past from the present time, in which the events of the original narrative are unfolding. This is similar to the case where an embedded narrative is told in dialogue (as in Example 2), except in flashbacks the embedded narrative is told in the narrative text; the audience of the flashback is the reader, not another character in the story. We will annotate this type of flashback in the same way as narratives embedded in dialogue.

*Interruptive Flashbacks* interrupt or replace the original narrative. The original narrative ends, and a new narrative of events occurring at a time before the original narrative begins. The key characteristic of the interruptive flashback, is that the narrator also moves in time. The narrator of the original narrative and the flashback do not have to be the same narrator. Sometimes the person who is nar-

rating the flashback will be a different character than the narrator of the original narrative. Sometimes the point of view of the flashback's narrator will be different than that of the original narrator. Other times, the narrator of the flashback is identical to the original narrator, the only difference being the events in the flashback happened in a time before the original narrative. Interruptive flashbacks break the telling of the original narrative: they are not embedded in any other narrative. We will annotate this type of flashback in the same way as an interrupted narrative. Remember that the excerpt in Example 5 contains an interruptive flashback. The original narrative is interrupted by a new narrative, which takes place at a time before the original narrative.

Flashforwards can also either be embedded or interruptive. Flashforwards tend to be interruptive though, since narrators typically do not know what will happen in the future, so the original narrative must be interrupted, to provide an account of events from a future time. Flashbacks can be embedded into speech, but this is usually either a telling of a vision, or it can be the telling of a hypothetical future.

## 2.4 Dreams and Visions

Many stories contain dreams. There are two types of dreams, and they are similar to the two types of flashbacks. Dreams are either embedded into the original narrative, or they interrupt it. Embedded dreams occur when the narrator is narrating about the memory of their experience of a past dream. Interruptive dreams occur when the narration is occurring from *within* the dream: the narrator is narrating as the dream unfolds.

Visions are similar to dreams. A vision could be a telling of the future, like a prophecy. The events of the prophetic vision may or may not come true, but the actual telling of the vision is distinct from the original narrative. Other types of visions can be sudden recollections of images or events from the past. Like dreams, visions can be either embedded in the original narrative, or interruptive of the original narrative.

6. [1 Haida got quite talkative when it came to music. He went on, delineating the special characteristics of Berman's performance of Liszt, but Tsukuru barely listened. Instead, a picture of Shiro performing the piece, a mental image, vivid and three-dimensional, welled up in his mind. As if those beautiful moments were steadily swimming back, through a waterway, against the legitimate pressure of time.]<sub>1</sub>

[<sub>2</sub> The Yamaha grand piano in the living room of her house. Reflecting Shiro's conscientiousness, it was always perfectly tuned. The lustrous exterior without a single smudge or fingerprint to mar its luster. The afternoon light filtering in through the window. Shadows cast in the garden by the cypress trees. The lace curtain wavering in the breeze. Teacups on the table. Her black hair, neatly tied back, her expression intent as she gazed at the score. Her ten long, lovely fingers on the keyboard. Her legs, as they precisely depressed the pedals, possessed a hidden strength that seemed unimaginable in other situations. Her calves were like glazed "porcelain, white and smooth. Whenever she was asked to play something, this piece was the one she most often chose." *Le mal du pays*." The groundless sadness called forth in a person's heart by a pastoral landscape. Homesickness. Melancholy.]<sub>2</sub>

[<sub>1</sub> As he lightly shut his eyes and gave himself up to the music, Tsukuru felt his chest tighten with a disconsolate, stifling feeling, as if, before he'd realized it, he'd swallowed a hard lump of cloud. The piece ended and went on to the next track, but he said nothing, simply allowing those scenes to wash over him. Haida shot him an occasional glance." ]<sub>2</sub>

The excerpt in Example 6 is also from Murakami (2014). The original narrative is surrounded by brackets with subscript of 1. This narrative is about Tsukuru talking to his friend Haida about classical music. Talking about classical music causes Tsukuru to have a vision, or a day dream, from his past. In Tsukuru's vision, which is surrounded by brackets with subscript of 2, he sees his old friend Shiro masterfully playing the piano in a very dreamy and vivid setting.

This vision interrupts the story told in the original narrative. The vision is not embedded because there is no action in the original narrative that triggers the telling of the vision. The last two sentences of the first paragraph inform the reader that Tsukuru is about to have a vision. These preparatory sentences are not part of the vision, since they describe events that are happening in the original narrative level, a "picture of Shiro...welled up in [Tsukuru's] mind." The actual vision is a departure from the original narrative. It describes what Tsukuru sees and feels when he is watching Shiro at the piano. This is not something that is *happening* at the time of the original narrative, it is something that Tsukuru is *experiencing*. The vision ends when the original third person narrator begins narrating about events that are actually happening in the present, "As he lightly shut his eyes and gave himself up to the music, Tsukuru felt his chest tighten with a disconsolate, stifling feeling..." The music then continues to play, and

Haida shoots Tsukuru “... an occasional glance.” These are events happening in the frame of the original narrative, and they signal the switch back to the original narrative from the interruptive vision.

### 3. Annotating Scripts

In addition to short stories and novels, we are interested in annotation narrative boundaries from scripts. Specifically, we will focus on the scripts of TV shows, and the transcripts of court proceedings. There are two types of text in a script: **dialogue** and **action**. Dialogue contains the words that actors (or people) speak, and the action gives direction for what the actors do, how they do it, and what happens in the world that the script describes. Scripts can either be used to tell actors what to say and how they should act, which dictates how they should behave during a performance, or scripts can be a recording of things that happened in real life, like a transcription of the dialogue in a court case.

#### 3.1 Dialogue

In the context of scripts, dialogue is a type of structured text. There are two components to dialogue: the character’s name and the character’s speech. In a script the character’s name will be stated. Typically, it will be bolded. Following the character’s name is the words that the character will speak. The character’s speech will not be in bold. Look at Example 7. This is an excerpt from the script of *Star Trek: Deep Space 9 - The Visitor*.<sup>6</sup> This excerpt portrays a conversation between two characters, Old Jake and Melanie. They are having a conversation about Old Jake’s writings and how Melanie enjoys his writing. In this excerpt, there are four utterances in the dialogue. Old Jake speaks first, Melanie speak speaks next, and then they each speak one more time.

7. [1 **OLD JAKE**

I didn’t realize people still read  
my books.

**MELANIE**

---

<sup>6</sup>Taylor, M. (Writer), & Livingston, D. (Director). (1995, July 31) The Visitor. [Television series episode] In Berman, R. (Executive Producer), *Star Trek: Deep Space 9*. New York, NY: CBS Television Distribution.



Of course they do. A friend  
recommended Anslem to me and I  
read it straight through, twice in  
one night.

**OLD JAKE**

Twice in one night... ?

**MELANIE**

It made me want to read everything  
you'd ever written, but when I  
looked, all I could find were your  
"Collected Stories." I couldn't  
believe it.

I'd finally found someone whose  
writing I really admired, and he'd  
only published two books.]<sub>1</sub>

Now let's think about the script of this conversation with respect to the narrative boundaries it contains. There are two narratives. The original narrative, where Old Jake and Melanie are having a conversation. This narrative makes up the entire span of text in Example 7. The span of the original narrative has been surrounded by brackets with subscript of 1.

It is important to note that the bolded character names have also been surrounded by brackets. The character names belong to the original narrative because this is a signal that a specific character will utter the proceeding text. The declaration of who is speaking in a script is like the phrase "He said..." or "Old Jake said..." in a novel or short story. The character names are included in the brackets of the original narratives since they mark the beginning of a character speaking, which is an action in the original narrative.

The next excerpt, Example 8, is also from the same episode of *Star Trek* (Taylor, 1995), and it contains an embedded narrative delivered by Melanie. The embedded narrative is surrounded by brackets with subscript of 1. Her narration is about her experience reading Old Jake's books, and how she reacted to his writing. In this embedded narrative, the bolded character names are not surrounded

by brackets. This is because the action of Melanie speaking belongs to the plot of the original narrative, and they do not belong to the plot of the narrative about Melanie's past. It is important to notice that Old Jake's speech is not part of the embedded narrative: he is not adding any information to the story of Melanie's past, he's just asking a clarifying question.

#### 8. OLD JAKE

I didn't realize people still read  
my books.

#### MELANIE

Of course they do. [1A friend  
recommended Anslem to me and I  
read it straight through, twice in  
one night.]1

#### OLD JAKE

Twice in one night... ?

#### MELANIE

[1 It made me want to read everything  
you'd ever written, but when I  
looked, all I could find were your  
"Collected Stories." I couldn't  
believe it.  
I'd finally found someone whose  
writing I really admired, and he'd  
only published two books.]1

### 3.2 Action

The action describes what is happening in the world that the script depicts. Typically, the action is written in present tense, since it describes what is happening in the present moment. Dialogue prescribes what each character says, and action

dictates what each character does, including the way they speak. Consider Example 9, where the action is surrounded by brackets with subscript of 1. Typically, the action in a script will be bolded, but it is not a requirement.

Now we will discuss the functions of each action sequence. The first sequence describes actions that Jake does before he speaks. The second action is during Jake's dialogue. It is a note for the actor playing Jake to take a moment to *consider* what he is saying. If the script is being read, then this stage direction allows the reader to imagine the character considering their actions. The third action sequence describes how Melanie reacts to what Jake says, and how she responds to him. The fourth action sequence instructs Melanie's next line to be said *softly*. The final action sequence describes an action Jake takes.

All of five of these actions sequences describe actions that occur in the original narrative of this script. When considering the narrative boundaries for this excerpt, each action sequence is a part of the original narrative. In fact, the entire span of text in Example 9 belongs to the original narrative. There are no embedded or interruptive narratives in this excerpt.

### 9. MELANIE

So that I could read them again...

like it was the first time.

[**<sub>1</sub> Jake smiles, nods that he understands. As he sits down with the tray...**]<sub>1</sub>

### OLD JAKE

There's only one "first time" for everything, isn't there?

[**<sub>1</sub>(considers)**]<sub>1</sub>

And only one last time, too. You think about that when you get to be my age. That today might be the last time you... sit in a favorite chair... watch the rain fall... enjoy a cup of tea.

[**<sub>1</sub> Melanie looks at him, then cautiously asks the question that brought her here.**]<sub>1</sub>

MELANIE

[<sub>1</sub>(softly)]<sub>1</sub>

Can I ask you something... ?

[<sub>1</sub> He nods that she go ahead... ]<sub>1</sub>

### 3.3 Structural Elements of Scripts

Structural elements are a final component of scripts that are separate from action and dialogue. They allow the readers or actors to distinguish between scenes and acts, and they give notes about the technical production for the performance, like a change of a camera angle. For our study, we will not include structural elements in our narratives. These elements are not part of the story being told, they just instruct the actors and crew when a scene begins or ends, and tell the camera operators logistics for how the scene is shot.

10. JAKE

Dad... ?

SISKO

What... what happened... ?

But before Jake can reply, Sisko's body starts to FLICKER and DISSOLVE like it did in the Defiant's Engineering room...

Jake watches as the terrible moment repeats itself... until Sisko completely DEMATERIALIZES once again...

Off Jake's confused, pained expression we...

[<sub>1</sub>FADE OUT.

END OF ACT ONE

DEEP SPACE NINE: "The Visitor" - REV. 08/04/95 - ACT TWO 20.

ACT TWO

FADE IN:]<sub>1</sub>

20 INT. JAKE'S HOUSE (DISTANT FUTURE)

Old Jake sits quietly, his thoughts far away in the past. Melanie watches him with great sympathy...

After a quiet beat...

OLD JAKE

I told Dax about what'd  
happened...

**Example 10** is another excerpt from *Star Trek: Deep Space 9* (Taylor, 1995), which has the structural elements surrounded by brackets with subscript of 1. In this example, the structural elements prescribe the camera fading out, the first act of the show ending, the second act beginning, and the camera fading back in. It is important to note that the action sequence "INT. JAKE'S HOUSE (DISTANT FUTURE)" is not a structural element, because it is telling the reader that the current scene is set at Jake's house. This is equivalent to author of a novel saying where the next scene occurs, which is an essential detail of the narrative, and not structural information. Following the location of the new scene, is a description of what is happening: Old Jake is sitting, and Melanie is watching him. Finally, the dialogue of the scene starts.

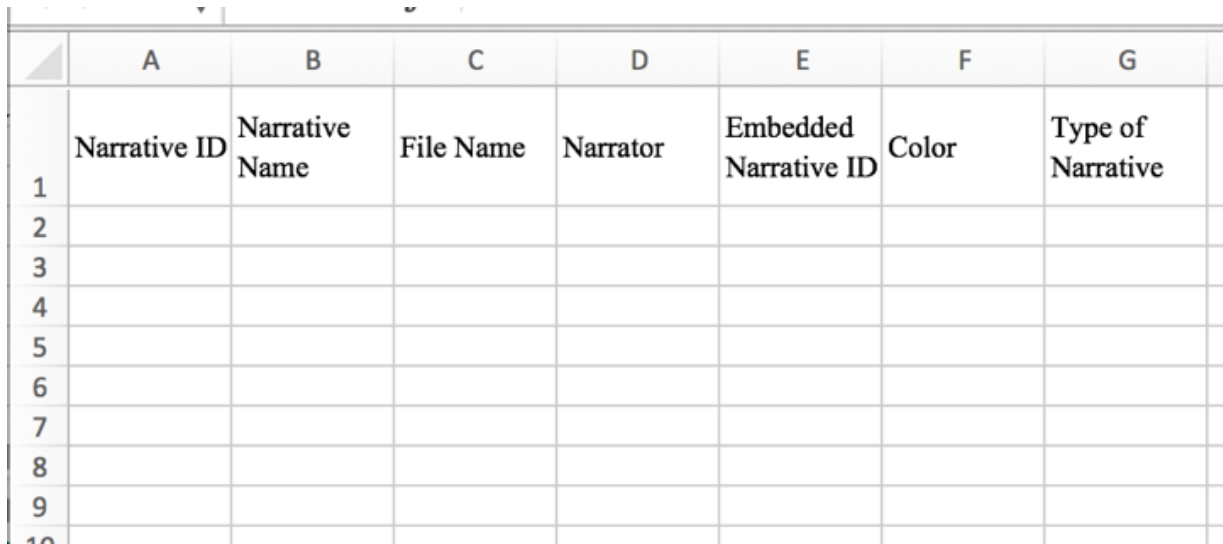
#### 4. Annotation Procedure

Each text will be provided to you as a PDF file. First read the text without making any annotations, reading just to understand; you can print the text out if you prefer reading from paper. Second, reread the story, and make a list of all the narratives. Third, go back to the beginning of the story. For each of the narratives you found, make a copy of the original PDF file, and in the corresponding file highlight the spans of text that the current narrative is told in. For clarity, if you found five narratives, you should make five copies of the PDF, one for each narrative. Then highlight the spans each narrative occupies in the corresponding file. To keep track of which narrative is annotated in which PDF file, please record the names of each file in the annotation metadata sheet, explained in the next subsection.

Note that in this annotation guide we talk about subscripted brackets surrounding narrative levels. When annotating texts, you should use the highlight function to distinguish which spans of text belong to a narrative level. Additionally, it is most important to annotate the spans of text that each narrative level belong to. The categories for *Narrator* and *Type of Narrative* are included just to help you think about narrative levels, and these characteristics will not be used to calculating agreement of narrative levels.

#### 4.1 The Annotation Metadata Sheet

For each short story, you will be provided with an Excel spreadsheet to fill out, in addition to the actual annotations that you will record by highlighting the PDFs.



	A	B	C	D	E	F	G
1	Narrative ID	Narrative Name	File Name	Narrator	Embedded Narrative ID	Color	Type of Narrative
2							
3							
4							
5							
6							
7							
8							
9							
10							

Figure 4. A screen shot from a blank narrative boundary annotation metadata spread sheet.

##### Narrative ID number

This is the ID of this row's narrative. Each narrative has an ID number. Assign the ID numbers in ascending order, start at 1. We will use these numbers at the end of the annotation PDFs to identify what narrative boundaries are highlighted in which document.

##### Narrative name

In this column, come up with a name for each narrative. The name can be a phrase or a sentence. This is mainly for helping refresh your memory while you are annotating, or when we meet for adjudications.

##### File name

For each narrative, you will create a copy of the original PDF file containing the narrative boundaries encoded as highlighted spans of text. In this column, you will write the file names of the PDF annotation files that correspond to each narrative

## Narrator

Each narrative is told by a narrator. In this column, please write who the narrator is. If the narrator has a name, write their name. If the narrator has no name, but annotates in first person point of view, write "1<sup>st</sup> person unnamed". If the narrator narrates from the third person point of view, write "3<sup>rd</sup> person". See "Narrative Characteristics Annotation Guide" for more specifics on determining the point of view of a narrator. Although this is not an annotation study on narrative point of view, it is sometimes useful to be aware of the changes in point of view throughout the text.

## Embedded narrative ID

If the narrative is embedded in an original narrative then put the ID number of the embedded narrative in this column. In Example 2, the pirate's narrative is embedded in the main narrative. In the row for the pirate's narrative, we write down the ID number "1" in the column for "embedded narrative ID", because this is the ID of the original narrative. If the narrative in question is not embedded in any other narratives, write "none" in the "embedded narrative ID" column. If a narrative is interruptive, then the embedded narrative ID is also "none".

## Color

In this column please put the color that you used to highlight the boundaries of the corresponding narrative. If possible, please use different colors for each narrative.

## Type of narrative

Put one of the following types of narrative that we have discussed in this guide:

• original	• Embedded flashback	• Interruptive flashback
• Embedded	• Embedded flashforward	• Interruptive flashforward
• Interruptive	• Embedded dream or vision	• Interruptive dream/vision

Note that these characteristics are not necessarily mutually exclusive, and they will not be used for calculating agreements. These categories are included so that you think about what type of narratives are being used.

## 4.2 Tips for Annotation

Don't do all the annotations in one sitting. Try to limit yourself to one to two hours at a time: more than that and you will become fatigued and the accuracy of your annotations will decrease. You should have this annotation guide handy while you are doing your annotations. This guide should serve as a reference and help you disambiguate tricky decisions.

## 4.3 Adjudication Procedure

Please do not speak to the other annotators about the specific annotations or the methods you use to make your decisions. This is because we are also trying to determine how clear the annotation guide and procedure are in and of themselves. We investigate this by measuring the agreement between different annotators, and if annotators talk with each other outside of the adjudication meetings about specific annotation decisions, then this results in artificially high agreement measures.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.



# Annotating Narrative Levels: Review of Guideline No. 1.

Meredith A. Martin

10.22.19

*Article DOI: 10.22148/16.053*

*Journal ISSN: 2371-4549*

*Cite: Meredith A. Martin, "Annotating Narrative Levels: Review of Guideline No. 1," Journal of Cultural Analytics. November 20, 2019. doi: 10.22148/16.053*

I am reviewing the Narrative Boundaries Annotation Guide, I had several concerns.

First, the authors are clear and concise when describing the project purpose: of training and testing a computer's ability to identify the beginnings and endings of narratives (which the authors call *narrative boundaries*). They are also clear and concise when describing what they mean by annotation in this project: "the record of human judgments identifying where a narrative begins and ends." When the authors arrive at the definition of narrative, however, their clarity and concision begin to unravel a bit. Since there are several theories of narrative, I suggest that the authors include some language like "for the purposes of this study" before launching into their definitions, which are necessarily oversimplified.

In the last paragraph of page two, the three forms could be elaborated more helpfully. Contiguous (or, later, un-interrupted), embedded, or interrupted narratives become important later on, so the guide could highlight that importance by bulleting the forms of narrative it deems important, and foreshadowing how deeply and how intricately the authors will attempt to distinguish between embedded or interrupted.

2.1 should be "Uninterrupted Narrative." I suggest they begin with "the simplest kind of narrative is an uninterrupted one. The section 1.2, titled "narrative boundaries," should end with "after the last word in the narration" and then continue

with “The next three sections” etc. (I have marked this change in the document using track change).

Embedded narratives were clearly explained and seems like a feasible for annotators, though arduous.

Marking boundaries for interrupted flashbacks and flashforwards seemed incredibly complicated. Interruptive narratives are the most complicated, and it seems that “Time Shifts” need their own heading (3) since they are not the three main kinds of narrative but may contain the three main kinds within them.

The example from 2.4 is too lengthy; I suggest adding a shorter example.

What happens if there are two types of narration happening (a flashback inside a dream)? Would the phrase “like I had done every morning since she left” be a flashback? The charts are useful but without multiple colors (and with only the example of Murakami to go on) I think that there is, at least as the guidelines are currently presented, quite a bit of room for interpretation and error. The human annotators would need to be a very large group, indeed, with a high level of fluency. For this to be a useful exercise, I can only imagine that the study would need hundreds of annotators.

Television scripts might be moved to an entirely different guideline - I feel that this guideline is already too complicated. I would eliminate scripts OR other kinds of narrative, but having both is too much.

Each section would benefit from additional examples drawn from authors other than Murakami. The beach / pirate example could be useful in each section, and since the levels of narratives accumulate, keeping the beach / pirate example active throughout and adding a literary text example alongside the beach / pirate would be a welcome addition and would more easily train the annotators. Or perhaps using a simpler narrative? Though the guidelines are helpful I would feel baffled by multiple levels of narration. I also can't imagine taking on this annotation task voluntarily.

The procedures seem complicated. How many short stories will the annotator receive? What does it mean to say “Each narrative has an ID number”? This makes no sense to me. I also don't understand “narrative name.” Examples here, again from the pirate / beach story or a simpler story than IQ84, would be helpful.

“If possible, please use different colors for each narrative”? I think the authors should assign colors. The authors could use a more robust bibliography for narrative theory.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

# Annotation Guideline No. 2: For Annotating Anachronies and Narrative Levels in Fiction

Edward Kearns

10.03.19

*Peer-Reviewed By: Tillman Köppe*

*Article DOI: 10.22148/16.052*

*Journal ISSN: 2371-4549*

*Cite: Edward Kearns, "Annotation Guideline No. 2: For Annotating Anachronies and Narrative Levels in Fiction," Journal of Cultural Analytics. November 20, 2019. doi: 10.22148/16.052*

## 0. Rationale

The following annotation guidelines were created with both the SANTA project and my own PhD research in mind. Their purpose is to provide tags for encoding narrative features related to time and perspective, as well as narrative levels. These things can interact with each other; for example, a story occurring within another story could also constitute a move backwards in time. This introduction will briefly discuss the rationale for the creation of the guidelines, the selection of the tags, and how the SANTA workshop has made me reflect on them and how they can be improved in the future.

My research project involves the use of a newly-created XML schema to encode analepsis, prolepsis (flashbacks and flashes forward in story time, respectively), instances of stream of consciousness and free indirect discourse narration, instances of extended or compressed time, changes in position of the narrator, and points in the text where the narrative level changes. I annotate these features in fiction texts, and will then use that encoded text to quantitatively compare those fictions to look for patterns, clusters, similarities, differences, and possible

lines of influence of one genre on another. The narrative concepts represented by most of the tags have been used for centuries, allowing for comparisons between fictions from many genres and time periods. In the case of my project, the two genres being compared are modernist novels, from the early twentieth century, and hypertext fictions from the late twentieth and early twenty-first centuries. These two genres have been compared for their shared use of narrative fragmentation, and some hypertext works allude to modernist texts; one of the aims of this project is to determine whether those comparisons are visible quantitatively. However the more important and primary aim was simply to translate these narratological terms into XML.

While the tag set was created in XML, it can be used in other formats as well. This has already been seen in the preparation for the first SANTA workshop, when the tag set was recreated in full in CATMA. The angle brackets from XML remain in these guidelines as a convenient way of designating a reference to a tag specifically rather than its corresponding narratological concept, but otherwise references to XML have been kept to a minimum in this version to maintain broader applicability. Excepting some small formatting changes and alterations in phrasing, these are the guidelines as they were circulated at the beginning of the SANTA workshop.

A particular feature of this annotation scheme is the inclusion of tags for stream of consciousness and free indirect discourse narration. The concepts and names for the tags in the set come largely from Gérard Genette but also other narratologists such as Shlomith Rimmon-Kenan, and indeed earlier literary criticism - as the term stream of consciousness comes from May Sinclair discussing the work of Dorothy Richardson. Stream of consciousness and free indirect discourse could be categorised, with extended and compressed time, as tags that describe how the story is being told stylistically, whereas the tags for narrator position, narrative levels, and anachronies describe more functional features of how the narrative is structured. The latter category maintains the focus on narrative features that have been used for centuries. Stream of consciousness and free indirect discourse have not always been used—they have been used from the nineteenth century onwards—but they are relevant stylistic features (not least in modernist and hypertext fiction) because they can correspond with the structural features of a text; an analepsis can occur within stream-of-consciousness narration; a deeper narrative level can occur within a character's mind.

Participating in the shared task has dovetailed with my work, adding a focus on narrative levels as well as narrative time. The workshop provided me with some excellent feedback on my annotation system from the other participants. That feedback, and the process of reviewing the other tag sets, has made me reflect on

how my own schema can be improved. As my system was focused on time and stylistic features as well as narrative levels, there is scope for more nuance and detail in providing the user with options for how a narrative level is encoded; for example the ability to annotate what the function of a narrative level is, relative to the other levels. This would allow the user to describe a narrative level using more than just the integer that is part of the tag already. Further narratological concepts such as narratee, focaliser, and narrative world would also allow for the user to more comprehensively describe the intricacies of the way a piece of fiction is told. Narratee and focaliser describe somewhat related features of narrative style to the stream of consciousness and free indirect discourse tags, while annotating narrative worlds can combine with the structural focus of annotating narrative levels. Incorporating these things will allow my schema to be more versatile in its application, more able to facilitate description of unusual cases. With an increased focus on narrative levels there will be a need for an application of an overall narrative theory to tie together all of the tags in the set, in the way that the time theories of Genette bring together some of the existing tags. This rationale for this version however was to focus on time as much as narrative levels, and it serves that purpose.

## 1. Introduction

The purpose of this document is to explain the elements of the SANTA 2 tag set and how they should be used. The tag set is based on concepts of narrative time suggested by Gérard Genette and other narratologists, and is designed for annotating analepsis, prolepsis, stream-of-consciousness, free indirect discourse, and narrative levels, with facility also for annotating instances of extended or compressed time, and for encoding the identity of the narrator and their position with respect to the levels of a narrative.

This annotation scheme is designed for XML, but can be used independently of that, for example if one is using CATMA to tag text. In the schema, the elements are nested inside one another in a way that allows the tags to be used in a manner that follows the way that the narrative constructs are used in fiction. For instance, analepsis and prolepsis often (but not always) occur within a character's mind, as part of the thought-process that is captured in prose by the stream of consciousness or free indirect discourse techniques. Accordingly, the analepsis and prolepsis tags here can be nested inside stream of consciousness and free indirect discourse tags, which in turn can be nested inside tags which annotate the appropriate narrative level, if this part of the narrative contains or is contained by another story. These tags do not have to be used all together, or in that or-

der; there are several possible combinations. For example an analepsis tag can be used inside a narrative level tag without the need for a stream of consciousness tag; the schema is flexible to meet the narrative structure of a piece of fiction.

All of this serves to allow the user to annotate changes in the temporal position of a narrative and changes in narrative levels, which can be related to one another; a character telling a story to the narrator about something that happened in the past would constitute both a move to a lower narrative level (because it is a story within a story) and an analepsis (because the telling of the story is a kind of flashback).

These narrative techniques have been used for centuries, allowing for comparisons of fiction texts across many genres and time periods. Once the text is encoded, it and the tags can be quantitatively analysed to aid those comparisons and look for patterns or discrepancies in narrative structures. The rest of this document will illustrate how the various tags should be applied to fiction.

## 2. <Text>

This tag is a by-product of the annotation scheme's initial focus on XML. It is a container for the rest of the encoded XML document. It is only opened at the beginning of the document, before any other elements, and then closed at the very end.

The purpose of this tag is to allow XML documents encoded with the schema to be valid, and to act as a frame within which all the other tags, and plain text, can be placed. When using this tag set in a format that is not XML, it is not necessary to use this tag, as its function no longer applies.

Once the text element is used, any other element from the schema can then be used, but there is no obligation to do this, as plain text will now be valid in XML, due to the fact that <text> is a complex type element with the mixed="true" attribute enabled. All other elements in which other elements can be nested also have this attribute.

## 3. <analepsis> and <prolepsis>

Analepsis and prolepsis are 'flashes' backwards and forwards in story time, respectively. Despite the use of the term flash, they are not necessarily brief. They

can last for many pages or even entire sections of novels. They are deviations from the main temporal progression of the story, disruptions between story time and narrative time. Story time is the actual series of events that occur in the novel, and like time in the real world it is linear. Narrative time can be nonlinear; it incorporates analepsis and prolepsis.

The terms were coined by Gérard Genette, who writes:

to avoid the psychological connotations of such terms as ‘anticipation’ or ‘retrospection,’ which automatically evoke subjective phenomena, we will eliminate these terms most of the time in favor of two others that are more neutral, designating as *prolepsis* any narrative maneuver that consists of narrating or evoking in advance an event that will take place later, designating as *analepsis* any evocation after the fact of an event that took place earlier than the point in the story where we are at any given moment, and reserving the general term *anachrony* to designate all forms of discordance between the two temporal orders of story and narrative.<sup>1</sup>

So, an instance of analepsis or prolepsis occurs when the narrative makes an instantaneous jump to another point in time, deviating from the current moment of the story in order to inform the reader about something that happened before or after that moment.

As such an analepsis or prolepsis tag should be opened at the point in the narrative where the narration jumps to another point in time, and closed when the story either returns to the moment left behind, or jumps again to another point in the story. The tags should not be used to merely annotate the regular, linear progression of time forwards in a narrative, because in that case there is a more proportional relation between narrative time and story time. The elements should only be used where there is a clear deviation of narrative time from a particular moment in the story.

### Example

An example of how the analepsis tag can be used is in the following XML encoded segment of *To the Lighthouse* by Virginia Woolf<sup>2</sup>:

She turned the page; <soc PERS="Mrs. Ramsay">there were only a few lines more, so that she would finish the story, though it was past bed-time. It was getting late. The light in the garden told her

<sup>1</sup>Gérard Genette, *Narrative Discourse: An Essay in Method* (Cornell University Press, 1980), 39-40.

<sup>2</sup>Virginia Woolf, "To the Lighthouse," *Project Gutenberg Australia*, 2008.



that; and the whitening of the flowers and something grey in the leaves conspired together, to rouse in her a feeling of anxiety. What it was about she could not think at first. Then she remembered; Paul and Minta and Andrew had not come back. She summoned before her again <analepsis>the little group on the terrace in front of the hall door, standing looking up into the sky. Andrew had his net and basket.</analepsis> That meant he was going to catch crabs and things. That meant he would climb out on to a rock; he would be cut off. Or coming back single file on one of those little paths above the cliff one of them might slip. He would roll and then crash. It was growing quite dark.</soc>

Here the narration moves from regular omniscient narration to stream of consciousness from the perspective of Mrs. Ramsay, with the instance of analepsis occurring within the stream of consciousness section. The analepsis tag is opened when Mrs. Ramsay remembers the image of Paul, Minta and Andrew standing in the doorway earlier in the evening, and is closed when she moves on from the memory to speculation about what they might be doing, which is disconnected from narrative time because it is not a thing that we know is definitely happening, whereas the event in the analepsis tags definitely did.

#### 4. <soc> (Stream of consciousness)

Stream of consciousness is a technique used to provide ostensibly subjective narration from the point of view of a character, rather than a detached omniscient observer. Its purpose is “to unfold the experience of a single mind . . . to emphasize, not the ego as such, but the moving, shifting, growing stream of consciousness confined within the walls of a single brain.”<sup>3</sup> The specific phrase is attributed to May Sinclair, who coined it almost exactly one hundred years ago while writing about Dorothy Richardson’s novels in *The Egoist*. Sinclair says about *Pilgrimage* that “in this series there is no drama, no situation, no set scene. Nothing happens. It is just life going on and on. It is Miriam Henderson’s stream of consciousness going on and on.”<sup>4</sup> Shirley Rose notes that William James had a hand in the phrase; in 1910 he does not quite put the words together in the order that Sinclair later does: “Consciousness . . . is nothing jointed; it flows. A ‘river’ or a ‘stream’ are the metaphors by which it is most naturally described. In talking of it here-

<sup>3</sup>Edith Rickert, “Some Straws in Contemporary Literature: Fiction in England and America,” *The English Journal*, vol. 12, no. 8 (1923), 509-10. JSTOR, doi:10.2307/801922.

<sup>4</sup>May Sinclair, “The Novels of Dorothy Richardson,” *The Egoist*, vol. 5, no. 4 (1918), 58.

after, let us call it the stream of thought, of consciousness, or of subjective life.”

<sup>5</sup> This helps to describe how the technique is used in fiction; it is the continuous reportage of thought, where one observation or memory flows into the next. It is meant to come directly from the character’s mind, without authorial intervention or translation into something more grammatical or contextually-informed; ironically it is one of the most contrived modes of narration used in fiction, such is its difference from the conventions of modern writing and grammar, and the effort the author must make to break from those conventions.

The <soc> tag should be used to fully surround each passage where this technique is used, beginning at the moment where the narration changes from a more detached, descriptive narration to the intimate reporting of a character’s thoughts. We see this kind of transition often in *Pointed Roofs* by Dorothy Richardson,<sup>6</sup> where the first sentence or two of a paragraph establishes a situation or setting in relatively objective terms, before the narration delves much deeper and provides Miriam Henderson’s subjective thoughts on the matter.

<analepsis> and <prolepsis> tags can be nested within <soc> tags.

### Example

In the following paragraph from *Pointed Roofs*, the opening sentence is an objective statement about Miriam’s reaction to something her sister Harriett has just told her. Similarly the second sentence, a description of Miriam’s physical movement in the room, is objective and would be observable by anyone else. By contrast, the impressions that come in the succeeding sentences are known only to Miriam, and, through her stream of consciousness, to the reader:

Miriam’s amazement silenced her. She stood back from the mirror.  
 <soc PERS=“Miriam”>She could not look into it until Harriett had gone. The phrases she had just heard rang in her head without meaning. But she knew she would remember all of them. She went on doing her hair with downcast eyes. She had seen Harriett vividly, and had longed to crush her in her arms and kiss her little round cheeks and the snub of her nose. Then she wanted her to be gone.</soc>

### PERS

This attribute is used to describe which character is currently providing the perspective for the stream-of-consciousness narration. The use of this attribute is technically optional, but it is recommended when the narration in one novel is

<sup>5</sup> Shirley Rose, “The Unmoving Center: Consciousness in Dorothy Richardson’s ‘Pilgrimage,’” *Contemporary Literature*, vol. 10, no. 3 (1969), 367. JSTOR, doi:10.2307/1207571.

<sup>6</sup> Dorothy Richardson, “Pointed Roofs,” in *Project Gutenberg*, 2018.

taken over by different characters' consciousnesses at different times, as in *To the Lighthouse*, and it can of course still be used when the narrative perspective only comes from one character, as in *Pointed Roofs*.

## 5. <fid> (Free Indirect Discourse)

Free indirect discourse is another technique which is used for “representing speech, thought, and perception.”<sup>7</sup> Brian McHale introduces indirect discourse (ID) as common narration where the narrator has a lot of control over what is being reported; they are, once again, detached and omniscient. Free indirect discourse is different however; it “handles person and tense as ID would . . . . On the other hand, it treats deixis as DD [direct discourse] would, reflecting the character's rather than the narrator's position . . . . Manifestly, it is contextual cues more than formal features that determine, in many cases, whether or not a sentence will be interpreted as a free indirect representation of speech, thought or perception.”<sup>8</sup> So, free indirect discourse shares features of more detached narration (including correct grammar), but it is coloured by the perceptions of a character, not to the radical extent of stream-of-consciousness, but in a more subtle way.

As with <soc>, <analepsis> and <prolepsis> tags can be nested inside <fid> tags.

### Example

The following paragraph from *Tender is the Night* by F. Scott Fitzgerald<sup>9</sup> features objective, omniscient third-person description in its first two sentences, until the em dash, when we switch to narration that, while still grammatically correct, not breaking the sentence, and still describing externally visible features of the physical world, is coloured by the assumptions and evaluations made by the character Rosemary Hoyt.

Rosemary swam back to the shore, where she threw her peignoir over her already sore shoulders and lay down again in the sun. The man with the jockey cap was now going from umbrella to umbrella carrying a bottle and little glasses in his hands; presently he and his friends grew livelier and closer together and now they were all under

<sup>7</sup>Brian McHale, “Speech Representation,” in *The Living Handbook of Narratology*, 10 June 2011, 4 <http://www.lhn.uni-hamburg.de/article/speech-representation>.

<sup>8</sup>McHale, “Speech Representation,” 4.

<sup>9</sup>F. Scott Fitzgerald, “Tender Is the Night,” in *Project Gutenberg Australia*, 2003.

a single assemblage of umbrellas–<fid PERS=“Rosemary”>she gathered that some one was leaving and that this was a last drink on the beach. Even the children knew that excitement was generating under that umbrella and turned toward it–and it seemed to Rosemary that it all came from the man in the jockey cap.</fid>

## PERS

Similarly to its use in <soc>, in <fid> the PERS attribute is used to annotate which character is influencing the free-indirect-discourse narration.

## 6. <level>

This tag is used to annotate a change in narrative level, if the piece of fiction has stories within stories. The use of the DEGREE attribute, while technically not necessary for the XML to be valid, must always be used in order for this element to have meaning. The DEGREE attribute allows for any integer to be assigned to it, because in theory there can be as many stories within stories as the writer cares to create. In narratology the term degree is used to assign numbers to these levels of story, and it is an attribute of the <level> element here. A first degree narrative level is “a narrative that is not embedded in any other narrative; a second-degree narrative is a narrative that is embedded in a first-degree narrative,”<sup>10</sup> and so on. A <level> tag with the attribute DEGREE=“1” should be opened when the top-level narrative first appears in a text, and then closed when the narration changes to a lower level. At this point a separate <level> tag should be opened with DEGREE=“2”. If the narrative returns to the higher, framing level, the <level> with DEGREE=“2” should be closed and a new <level> with DEGREE=“1” should be opened. If instead a further story is created within the second-degree narrative, this will be a third-degree narrative, and should be encoded with a <level> tag and DEGREE=“3” attribute.

An example of several degrees of narrative level can be seen in *Frankenstein* by Mary Shelley, where the novel begins with the narrative of the Arctic explorer Captain Walton (the novel’s first-degree narrative), who meets Dr Frankenstein, who tells Captain Walton his story (so Dr Frankenstein’s narrative is the second-degree narrative), which includes a story told to Dr Frankenstein by the monster (the third-degree narrative). In theory the levels could keep going further down, allowing for fourth-degree narratives and more.

<analepsis> and <prolepsis> can be nested inside <level> tags either on their own,

<sup>10</sup>Manfred Jahn, “Narrative Levels,” in *Narratology: A Guide to the Theory of Narrative*, May 2017.

or nested inside <soc> or <fid> tags which are in turn contained by the <level> element.

### Example

In this excerpt from *To the Lighthouse*, one can see how the level tags are used to encode a story within a story, as Mrs. Ramsay reads to her child. That is the second-degree narrative; we then return to the first degree narrative, the main story. When we do, we see a characteristic of Woolf's prose: the switching of stream-of-consciousness perspective from one character to another in the same paragraph - in this case from Mrs. Ramsay to her husband - revealing what each character is thinking about the other. There is also a brief instance of analepsis as Mrs. Ramsay remembers something that her husband had said earlier.

<level DEGREE="2">"The man's heart grew heavy,"</level> <level DEGREE="1"> she read aloud,</level> <level DEGREE="2">"and he would not go. He said to himself, 'It is not right,' and yet he went. And when he came to the sea the water was quite purple and dark blue, and grey and thick, and no longer so green and yellow, but it was still quiet. And he stood there and said—"</level>

<level DEGREE="1">Mrs. Ramsay could have wished that her husband had not chosen that moment to stop. <soc PERS="Mrs. Ramsay">Why had he not gone as <analepsis>he said</analepsis> to watch the children playing cricket?</soc> But he did not speak; he looked; he nodded; he approved; he went on. <soc PERS="Mr. Ramsay">He slipped, seeing before him that hedge which had over and over again rounded some pause, signified some conclusion, seeing his wife and child, seeing again the urns with the trailing of red geraniums which had so often decorated processes of thought, and bore, written up among their leaves, as if they were scraps of paper on which one scribbles notes in the rush of reading—he slipped, seeing all this, smoothly into speculation suggested by an article in THE TIMES about the number of Americans who visit Shakespeare's house every year . . . </soc> . . . </level>

## 7. <extendedTime>

The tag <extendedTime> is used when narrative time is extended relative to story time. For example this would occur if the reader is informed over several pages,

which take minutes to read, about a character's thoughts which fly through their mind in a matter of seconds. The tag should be closed when the narration returns to a more steady flow of time.

## 8. <compressedTime>

By contrast, <compressedTime> should be used when narrative time moves much faster than story time, for example if a story leaves its normal flow of narration, where narrative time and story time are in closer proportion, and details the events that occur over the span of a number of years, before returning to its initial, regular narrative flow.

## 9. <narrator TYPE= “extra”/“intra”/“meta”>

This tag and its attribute (TYPE) describe the narrator, the person or entity telling the story, although it doesn't describe them as an individual but instead their “relative situations and functions”<sup>11</sup> compared to other narrators on other levels. The terminology emanates from Genette, who states that the narrator of a first-degree narrative is an extradiegetic narrator, the narrator of a second-degree narrative is an intradiegetic narrator (inside the diegetic story of the first narrator), and the narrator of a third-degree narrative is a metadiegetic narrator.<sup>12</sup> The tag and attribute should be used only in a text where the narrator changes as the level changes, and the tag should be applied at the point in the text where this change occurs.

While any string could be entered as the value for the attribute TYPE, it is recommended that the only options used are either TYPE=“extra”, TYPE=“intra” or TYPE=“meta”, as appropriate for the narrator of a given section of fiction text.

The narrator tag should not be used when there is only one narrative perspective, even when there are multiple narrative levels. The narrative perspective must also change. For example, it would be appropriate to use the narrator tag to annotate *Heart of Darkness* by Joseph Conrad, for the first-level narrative is told by a crew member of the ship arriving in London, making him the extradiegetic

---

<sup>11</sup>Gérard Genette, *Narrative Discourse*, 229.

<sup>12</sup>Gérard Genette, *Narrative Discourse*, 228-9.

narrator, outside the main story but providing the frame narrative. The second-level narrative, and the main part of the novel, is told by Marlow, speaking to the first crew member, making Marlow the intradiegetic narrator. This dynamic is also present in *Wuthering Heights* by Emily Brontë, where Lockwood is the extradiegetic narrator, and Nelly Dean tells most of the novel to him, as the intradiegetic narrator, before we once again return to Lockwood's perspective for the final chapters. *Pale Fire* by Vladimir Nabokov may be quite a useful example, but only if one considers John Shade's poem as a kind of narrative level, making Shade the intradiegetic narrator and Charles Kinbote the extradiegetic narrator, with his introduction and endnotes forming a higher narrative level than Shade's poem. Indeed this annotation of the novel also only works if one considers Shade and Kinbote to be two different people, and that the whole story and Shade are not fabrications of Kinbote's mind. The monster in *Frankenstein* is an example of a metadiegetic narrator: a figure whose tale is relayed to the reader through not one but two other narrators.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

# Annotating Narrative Levels: Review of Guideline No. 2

Tilman Köppe

09.13.19

*Article DOI: 10.22148/16.054*

*Journal ISSN: 2371-4549*

*Cite: Tillman Köppe, "Annotating Narrative Levels: Review of Guideline No. 2," Journal of Cultural Analytics. November 4, 2019. doi: 10.22148/16.054*

In this brief review, I will assess the theoretical soundness of the categories as explained in the "SANTA 2 Annotation Guidelines" as well as the practical implementation of the theory through exemplary annotation choices.

All in all, the explanations of the categories reflect what can be found in many introductory textbooks to narratology. This is what is to be expected, of course, since the Annotation Guidelines do not provide the space for an extended, and critical, discussion of the more recent state of research. I shall therefore confine myself to commenting on where I think that the clarity and/or consistency of the explanations can be improved, and on whether the examples provided are controversial. My comments concern (1) *analepsis* and *prolepsis*, (2) *stream of consciousness*, (3) *free indirect discourse*, and (4) *level* respectively.

## **Analepsis and prolepsis**

In the example passage from *To the Lighthouse*, the following sentences are not treated as part of the analepsis: "That meant he was going to catch crabs and things. That meant he would climb out on to a rock; he would be cut off. Or coming back single file on one of those little paths above the cliff one of them



might slip. He would roll and then crash. It was growing quite dark.” The reason given for not treating these sentences as part of the analepsis is that Mrs. Ramsay “speculation [...] is not a thing that we know is definitely happening,” and therefore the sentences are taken to be “disconnected from narrative time.”

Now it is certainly true that we do not learn whether any of the things Mrs. Ramsay speculates *about* are actually happening in the story. But it seems to me that Mrs. Ramsay’s *speculation* about these things can be part of the analepsis. That is, she not merely ‘summons’ how the group looked when they left, but also what she thought at the moment when they left. In this case, the sentences quoted may be considered part of “narrative time” and hence part of the analepsis.—Note that we do not have reason to doubt *that* Mrs. Ramsay speculates about these things. So treating the passage as part of the analepsis is consistent with the criterion explicitly given: the speculation is something that “we know is definitely happening”. What we do not know, however, is the point in time when the speculation takes place; there are two possible interpretations both of which are consistent with the text. I’d therefore suggest using a clearer example, i.e. one that does not allow for two viable interpretations (only one of which is consistent with the proposed tagging of the example passage).

## Stream of consciousness

In the introductory paragraph the *stream of consciousness* is explained as follows: it “is meant to come directly from the character’s mind, without authorial intervention or translation into something more grammatical or contextually-informed”. The example however, clearly shows several marks of “authorial intervention”. Among these are the use of third person pronouns (“she”) and past tense. But Miriam neither thinks about herself using the third person nor using past tense, presumably. Moreover, in the passage that introduces the example, it is suggested that what is decisive about stream of consciousness is that what is said is neither “objective” nor “observable by anyone else”. These are criteria different from the ones introduced above (“without authorial intervention or translation into something more grammatical or contextually-informed”), and one wonders which of those is actually decisive or, in any case, how these different criteria relate to each other. It seems to me that the passage quoted from *Pointed Roofs* is indeed such that it is about something that is not observable by anyone but the character herself; but it does show several marks of authorial intervention. Actually, this passage is an instance of free indirect discourse rather than stream of consciousness.

## Free indirect discourse

Stream of consciousness is distinguished from free indirect discourse by claiming that the latter “is coloured by the perceptions of a character, not to the radical extent of stream-of-consciousness, but in a more subtle way”. It will be hard to judge whether the subjective “colouring” of a passage is “subtle” enough in order to distinguish it from stream of consciousness. The criterion could work only if one were able to *compare* two passages, such that one could be able to judge the *relative* ‘subjectivity’ of them. In standard cases, however, we do not have two passages for comparison. Moreover, the passage from *Tender is the Night* is introduced as containing “evaluations”, but I can’t find any in there.

## Level

The passage quoted from *To the Lighthouse* contains the phrase “that hedge which had over and over again rounded some pause, signified some conclusion, seeing his wife and child, seeing again the urns with the trailing of red geraniums which had so often decorated processes of thought, and bore, written up among their leaves, as if they were scraps of paper on which one scribbles notes in the rush of reading”. Given that “seeing” is a factive verb (in the sense that ‘seeing that *p*’ implies the truth of *p*), this strikes me as an analepsis too.

One final note: I think that it might be useful to distinguish clearly between the *definition* of a narratological term (such as *stream of consciousness*) on the one hand and the *criteria* which help us to decide whether the phenomenon in question is instantiated in a given narrative. The definition of a term captures the nature of the phenomenon or, to put it differently, its constitutive features. Criteria for the application of a term, in contrast, are mere guides to its application. Accordingly, if a constitutive feature of some narrative phenomenon is missing from some passage, we know that the passage does not feature that phenomenon. But if some criterion for the application of a term is not satisfied, that does not mean that the term *cannot* apply. (Criteria, in other words, need not have the status of necessary conditions.) To give but one example, the grammatical correctness of a passage may be a criterion for *stream of consciousness* such that the grammatical correctness of a passage counts against its being in this narrative mode. (In other words, it may be unlikely that a passage is both grammatically correct *and* in the stream of consciousness mode.) But the grammatical incorrectness is not a constitutive feature of stream of consciousness. It is certainly

possible that a grammatically correct passage features this very narrative mode.



# Annotation Guideline No. 4: Annotating Narrative Levels in Literature

Nora Ketschik, Benjamin Krautter, Sandra Murr, Yvonne Zimmermann

10.03.19

*Article DOI: 10.22148/16.055*

*Journal ISSN: 2371-4549*

*Cite: Nora Ketschik, Benjamin Krautter, Sandra Murr, Yvonne Zimmermann, "Annotation Guideline No. 4: Annotating Narrative Levels in Literature," Journal of Cultural Analytics. December 3, 2019. doi: 10.22148/16.055*

## Introduction

Our participation in the *Shared Task on the Analysis of Narrative Levels Through Annotation* was motivated by a theoretical and practical interest in narratological phenomena of literary texts. We are a group of four literary scholars, three of whom are also working in the field of Digital Humanities. Combining these two scientific perspectives seems to be a fruitful research approach to formalize concepts of narratology with a focus on intersubjectivity. Therefore, a shared task dealing with narrative levels was particularly appealing to us, since narrative levels are both a delimited aspect of narratological categories and a complex concept of literary theory that can be of great importance for a formal text analysis and the following interpretation.

When thinking about narrative levels in more detail, we noticed the necessity to first address the question of what a *narrative* is. Narratological concepts such as the distinction of exegesis and diegesis, the position of the narrator or the act of telling a story have been the starting point of our reflections. In a first step, our guidelines clarify the concepts that are fundamental to understand narrative levels. Since different narratological theories and traditions focus on different aspects and details of narrativity, it is important—from our point of view—to

explain the underlying theory and the ensuing concepts of the guidelines. Therefore, it was essential for us to define the characteristics and the exact scope of narrative levels according to our understanding. In addition, we sought to illustrate our definition with several examples in order to help annotate them adequately. Automation, up until now, has not been a part of our considerations.

After having analyzed and annotated several literary texts and having discussed the phenomenon at the *SANTA* workshop in Hamburg (2018), more specific and detailed questions regarding our guidelines arose. On the one hand, becoming more familiar with different narratological theories, scientific disciplines and approaches to annotation, we recognized that it would be beneficial for the applicability and the usefulness of our guidelines to rework some of their aspects, as they were written with a very specific theoretical background in mind. On the other hand, while annotating with our guidelines, we detected some elements that should be refined or clarified. However, most of these aspects remain open for discussion, and we will continue to think about them in more detail.

During the workshop, our guidelines were perceived as very dense and theory-laden. Thus, we seek to better connect our theoretical premises with the actual annotation guidelines in a revised version. Furthermore, we aim to provide more examples for standard cases, while also trying to explain the reasoning behind the annotation in those examples in a more detailed way. We noticed that some narratological terms have been utilized in different understandings by the research teams (e.g., the terms ‘narrator’ and ‘speaker’). Therefore, we will aim for a better defined terminology in a revised version of our guidelines.

Going into greater detail, there is at least one problematic aspect in our annotation tagset: the letter E (short for ‘exegesis’) that we used for annotating non-narrative passages should be replaced by the letter N, since non-narrative passages do not necessarily have to be linked to the exegesis.

Since the organizers’ explanation of the shared task focused on narrative levels, we have limited ourselves to defining and annotating these, although there are features such as the narrator, the position of the narrator, or speakers (who can become narrators in embedded narratives), which not only might help the annotators in using our guidelines, but also could be utilized as indicators in an automatic annotation process. In a revised version of our guidelines, we will think about adding other narratological categories to our tagset, as long as they help identify narrative levels. Those features might also be beneficial for subsequent corpus analyses of literary texts. But before we approach such questions, we still have to solve sophisticated and challenging theoretical issues that require a more detailed analysis and understanding of the phenomenon. For the future,

we will at least try to examine the following questions: How should we deal with different forms of imagination (as they appear, e.g., in Anton Chekhov's "The Lottery Ticket")? What about dreams or visions? Is it sufficient to characterize the narrator, or should we think about the narratee as well? Which specific criteria do we need in order to distinguish between analepses (flashbacks) and embedded narrative levels? Does it make a difference in this context, if an analepsis is completed or not, external or internal? If there is no change of narrator, how comprehensive must our criteria be for regarding presence of another set of characters, spatial distance and temporal distance to the subordinate narrative level? Is our assumption of at least two applying indicators too arbitrary? These and other questions will form the basis for further theoretical discussions and will be integrated in our revised guidelines.

## Submitted Guidelines for the Annotation of Narrative Levels

### I. Theoretical Introduction

Narrative theory—or Narratology—has been one of the central concerns of international literary studies since the early nineteen-sixties.<sup>1</sup> Narratology deals "with the general theory and practice of narrative,"<sup>2</sup> especially with different types of narrators and structural elements such as narrative levels. A fundamental interest of narratologists lies in the organization and structure of the literary plot. To describe both, the sequence of events in time and their implementation into an organized plot, Gérard Genette develops a systematic terminology that utilizes the terms *discours* and *histoire*<sup>3</sup> in order to differentiate between what is narrated and how it is narrated. While *histoire* subsumes the "totality of the narrated events,"<sup>4</sup> the "discours du récit" is the actual realization of the *histoire* in the narration, be it oral or written.

For the distinction of narrative levels, Genette proposes a classification of the narrator in *extradiegetic*, *intradiegetic* and *metadiegetic*. The *extradiegetic narrator* produces a "first narrative with its diegesis."<sup>5</sup> S/he is potentially followed by an *intradiegetic narrator*, a character that appears in the first narrative, who goes on to produce a second narrative, and so on.

In principle, we follow Genette's idea that a new narrative level needs a sufficiently

---

<sup>1</sup>Matías Martínez and Michael Scheffel, *Einführung in die Erzähltheorie* (München, 2007), 7.

<sup>2</sup>M. H. Abrams, "Narrative and Narratology," in *A Glossary of Literary Terms* (Orlando, 1999), 173.

<sup>3</sup>Gérard Genette, *Narrative Discourse Revisited* (Ithaca, 1988 [1983]), 13-15.

<sup>4</sup>Genette, *Narrative Discourse Revisited*, 13.

<sup>5</sup>Genette, *Narrative Discourse Revisited*, 84.

marked "threshold between one diegesis and another."<sup>6</sup> However, Genette ties a new narrative level to a new narrator. We would like to expand on this concept as, according to our understanding, literature has produced examples that show clear signs of being new narrative levels without exchanging the narrator.<sup>7</sup> Still, new narrative levels need to have clearly distinguishable diegeses. Thus, the crossing of illocutionary boundaries alone, i.e. speech acts that introduce a new speaker,<sup>8</sup> do not necessarily lead to the creation of a new narrative level. The extraction of direct speech is a separate annotating task that we will not apply within our *Annotation Guidelines* for narrative levels.

In order to understand and interpret a narrative text, we assume that it is necessary to analyze the structure and the form of the text to gain insight into the interrelation between form and content (e.g., Emil Staiger's "Gehalt-Gestalt-Gefüge"). Herein, narrative levels have a great relevance, as there is an important dependency regarding different narrators and different narrative levels within a given literary text. Possible research questions based on distinguishing narrative levels can focus on structural elements of a text, e.g., an overview of the different narrators and the stories they tell, the relationship between frame and embedded stories, or the importance of a narrative level according to its length. However, research questions can also address the content of narrative levels. Since narrative levels can be functionally related to each other, e.g., an embedded story that serves as an explanation for the frame story, it is important to interpret characters or the narrator's distribution of information with such interdependencies in mind. Furthermore, a more systematic analysis of crossovers reaching from one narrative level to another seems to be a productive goal that requires the recognition of narrative levels as fundamental.

## II. Terminology & Concepts

In order to get a basic grasp on the terminology that is used in our *Annotation Guidelines* (IV), we strive to explain some fundamental technical terms in a concise way. This should help to achieve a clearer understanding of our guidelines and the underlying literary concepts:

### 1. Narrative levels:

The terminology used to describe narrative levels is diverse and varies widely. Our basic approach is to define any new story that occurs within a given narrative text as a new narrative level (see III. 2 for a more detailed explanation).

---

<sup>6</sup>Genette, *Narrative Discourse Revisited*, 84.

<sup>7</sup>Silke Lahn and Jan Christoph Meister, *Einführung in die Erzähltextanalyse* (Stuttgart, 2013), 83.

<sup>8</sup>Marie-Laure Ryan, *Possible Worlds, Artificial Intelligence and Narrative Theory* (Bloomington, Indianapolis, 2001), 175-177.

Narrative levels can be interlaced. Within the **frame story** (superordinate level), several **embedded stories** with a different degree can occur. As an embedded story can become the frame story for another embedded story, we use the terms **first-, second-, third-, ... degree narrative** as an alternative terminology in order to avoid ambiguities.<sup>9</sup> Narrative levels can also be arranged sequentially. E.g., several embedded stories that belong to the same frame story are arranged next to each other.

## 2. Homodiegetic and heterodiegetic narrator:

With regard to the distinction of different narrators and, consequently, the change of a narrator in a single text, it is useful to determine her/his position in relation to the story s/he tells. In principle, it must be determined whether the narrator is part of the diegetic world or not. A **homodiegetic narrator** is a character in the story s/he tells. In contrast, a **heterodiegetic narrator** is not part of the story s/he tells.<sup>10</sup>

## 3. Exegesis and diegesis:

“[D]iegesis designates the level of the narrated world, and exegesis the level of the narrating.”<sup>11</sup> Consequently, a homodiegetic narrator (of a first-degree narrative) belongs to both levels: In her/his function as narrator, s/he belongs to the exegesis, but since s/he tells a story with herself/himself being a character in it, s/he is also part of the diegesis.<sup>12</sup> A heterodiegetic narrator, however, belongs only to the exegesis; the narrated world which s/he is not part of is the diegesis.

## 4. Narrating and experiencing “I”/self, experiencing space:

A homodiegetic narrator’s “I”/self is split into a narrating and an experiencing “I”/self. While the narrating “I”/self is located in the exegesis or on the superordinate level of the current narrative level, the experiencing “I”/self is located on the current narrative level as one character among others.<sup>13</sup> As a heterodiegetic narrator is not part of the story s/he tells, there is no experiencing “I”/self in the story. Therefore, we opt for the term “experiencing space” as an alternative. The experiencing space subsumes features of the narrative level regarding its time,

<sup>9</sup>Manfred Jahn, *Narratology: A Guide to the Theory of Narrative* (Universität Köln, 2017). <http://www.uni-koeln.de/%7Eame02/pppn.htm> (06/15/2018).

<sup>10</sup>Gérard Genette, *Narrative Discourse. An Essay in Method*, trans. Jane E. Lewin Oxford, 1980 [1972]), 244-245.

<sup>11</sup>Didier Coste and John Pier, “Narrative Levels,” in *The Living Handbook of Narratology*. (2016a [2014]). <http://www.lhn.uni-hamburg.de/article/narrative-levels-revised-version-uploaded-23-april-2014> (06/15/2018).

<sup>12</sup>Genette, *Narrative Discourse Revisited*, 84.

<sup>13</sup>Lahn and Meister, *Einführung in die Erzähltextanalyse*, 70; Monika Fludernik, *An Introduction to Narratology* (Abingdon, 2009), 90.



space and characters. The distinction between narrating “I”/self and experiencing “I”/self or experiencing space can help to identify narrative levels (see III. 4).

### 5. Projected teller role:

A projected teller role, i.e. “an agent whose sole involvement with the text is its material dissemination,”<sup>14</sup> always demands an additional narrative level (even if this level consists of only one sentence). The most prominent example for a projected teller role is the editor figure.

## III. Premises

1. We identify all narrative levels in a given narrative text. Our basic assumption is that each text has at least one narrative level.
2. A new story<sup>15</sup> within a text calls for a new narrative level.
3. A change of the narrator results in a change of the narrative level. However, a change of the narrative level does not necessarily have to be accompanied by a change of the narrator (cf. Max Frisch’s “I’m not Stiller”: homodiegetic narrator, who tells a fairy tale within her/his own narration). **Attention:** In our understanding, not every character speech is automatically a story. For this to be true, the criteria according to III. 2 have to be met.
4. What is needed for a change of narrative levels:
  - a. In a story that is narrated by a heterodiegetic narrator there is a clear distinction between the position of the narrator and the experiencing space of the characters.
  - b. In a story that is narrated by a homodiegetic narrator there is a clear distinction between the narrating “I”/self and the experiencing “I”/self.
  - c. In a new story that changes its narrator (e.g., a character telling an embedded story), there is a new narrating “I”/self. In a new story that does not change its narrator (e.g., a homodiegetic narrator telling an embedded story), the narrative “I”/self remains the same. Thus, for a change of narrative levels, there must be another experiencing “I”/self (experiencing space) perceptible.

---

<sup>14</sup>Marie-Laure Ryan, “The Narratorial Functions,” *Breaking Down a Theoretical Primitive*, *Contemporary Narratology* 9.2, (2001): 151.

<sup>15</sup>Genette’s term *histoire* is oftentimes translated as *story*. Our concept of *story*, however, does not coincide with Genette’s *histoire*.

d. Two of the following indicators, which point to a new experiencing space, must apply, if a new story—and thus a new narrative level—is created by the same narrator:<sup>16</sup>

- – the presence of another set of characters,
- a spatial distance to the first/current narrative level,
- a temporal distance to the first/current narrative level. However, it is also possible that a character narrates a storyline that takes place simultaneously. For this to be a new narrative level, the other two indicators (set of characters, spatial distance) have to apply **Attention:** In certain cases, the distance between the experiencing and the narrating “I”/self is seemingly removed.<sup>17</sup>

5. Embedded stories can be functionally related to their superordinate narrative level, their frame stories. Possible functions are:<sup>18</sup> a. explicative: The embedded story provides an explanation for elements of the frame story. b. actional: The embedded story is constitutive for the frame story. c. thematic: The embedded story is thematically related (analogies, correspondence, contrast, relationships) to the frame story.

6. Narrative levels can be interlaced (inclusion scheme) or arranged next to each other (sequential), see Fig. 1.

---

<sup>16</sup>Analepses (flashbacks), which create only temporal distance to the current story, and mental games (“what if”-scenarios) are not considered new narrative levels.

<sup>17</sup>Cf. stream of consciousness in Arthur Schnitzler: “Leutnant Gustl”, Lahn and Meister, *Einführung in die Erzähltextanalyse*, 72.

<sup>18</sup>Lahn and Meister, *Einführung in die Erzähltextanalyse*, 83-84.

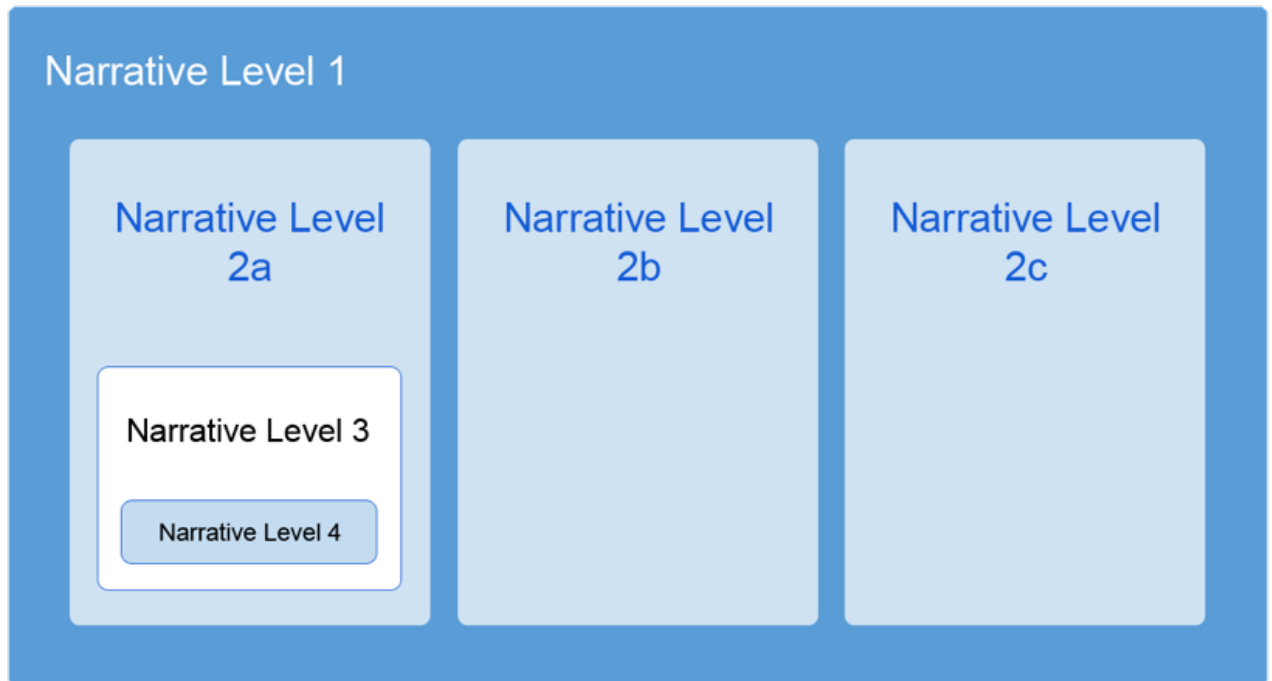


Figure 1. Interlaced and sequential arrangement of narrative levels within a literary text.

7. The location of the narrator (exegesis) in a first-degree narrative is not an independent narrative level. Aphorisms, mottos, comments, judgments, forms of address (fictitious recipient) and thoughts expressed by the narrator<sup>19</sup> do not form a new narrative level. They are part of the instantaneous narrative level. This is also true for expressions of narrators in a second-, third-, ... degree narrative, as long as they do not address an element of the superordinate level. Regardless, it is still possible to annotate such expressions (for further details see IV. 9).

#### IV. Annotation Guidelines

Before beginning the annotation process, the annotator has to read the entire text once. Following that, all narrative levels in the text are searched for as defined in the *Premises* (III. 1-7). They are annotated according to the following points:

1. All narrative levels are annotated with square brackets (opening bracket at the start and closing bracket at the end of a narrative level).<sup>20</sup>

<sup>19</sup>Wolf Schmid, *Elemente der Narratologie* (Berlin, Boston, 2014), 7.

<sup>20</sup>The annotation may be done in another way, too (e.g., with different colors marking the belonging to a certain narrative level), depending on the annotation tool that is used.

2. The narrative levels are annotated with numbers (1, 2, 3 etc.) as a first and letters (a, b, c etc.) as a second differentiator.
  - a. The numbers indicate the degree of the narrative level (inclusion scheme). E.g., **level 2** refers to a narrative level that is **embedded** into a superordinate level. Level 2 is a **second-degree narrative** or, in other words, an embedded story. Max Frisch's novel "Stiller" is one example: An embedded story (the fairytale of Rip van Winkle), is narrated by one of the novel's characters and enclosed into the **first-degree narrative** or frame story (level 1).<sup>21</sup>
  - b. Stories that are on the **same narrative level** (sequential arrangement) are identified by letters (a, b, c). Boccaccio's "Il Decamerone" is an example of sequentially arranged stories, which contain several separate novellas on the same narrative level (series of embedded stories); see Fig. 1 (the arrangement of the individual novellas that are embedded in the frame story would correspond to the numbering 2a, 2b, 2c etc.)
3. How to use the square brackets to separate the different narrative levels:
  - a. The brackets are marked with the number and, if applicable, a letter on both, the opening and closing brackets. E.g., [1 ... ]<sub>1</sub>
  - b. Inclusion scheme: The superordinate narrative level (e.g., first-degree narrative) starts before the embedded level (e.g., second-degree narrative). The brackets of the superordinate level close after the brackets of the embedded level: [1 ... [2 ... ]<sub>2</sub> ... ]<sub>1</sub>.
  - c. Sequential arrangement: The square bracket of the first sequentially ordered narrative level (e.g., 2a) closes before opening the square bracket of the second sequentially ordered narrative level (e.g., 2b): [1 ... [2a ... ]<sub>2a</sub> ... [2b ... ]<sub>2b</sub> ... ]<sub>1</sub>.
  - d. Punctuation is not separated from the preceding word.
    - (1) [1... [2 "*On a march in the Rhine campaign*", ]<sub>2</sub> *began the officer*, [2 "*I noticed, after a battle we had had with the enemy ...*" ]<sub>2</sub> ]<sub>1</sub> (Heinrich von Kleist: "Improbable Veracities")

**Note:** Each text usually has at least one narrative level and a corresponding number [1]. Letters are only used to denote a sequential arrangement and therefore not always utilized.

---

<sup>21</sup>Jahn 2017.

4. As a rule, a narrative text starts with the first narrative level (level 1) and may have other narrative levels embedded (level 2 and so on).

An exception to this rule are narrative texts with a projected teller role that requires its own narrative level. In this case, the projected teller role is annotated as a frame story (level 1), although this special case might only become visible at the end of the narrative text.

- (2) [<sub>1</sub> *The Editor believes the thing to be a just History of Fact; neither is there any Appearance of Fiction in it: And however thinks, because all such things are dispatch'd, that the Improvement of it, as well as the Diversion, as to the Instruction of the Reader, will be the same; and as such he thinks, without farther Compliment to the World, he does them a great Service in the Publication.*

*The Life and Adventures of Robinson Crusoe*

[<sub>2</sub> *I Was born in the Year 1632, in the City of York, of a good Family, tho' not of that Country, my Father being a Foreigner of Bremen, who settled first at Hull. ...* ]<sub>2</sub> ]<sub>1</sub> (Daniel Defoe: "Robinson Crusoe").

- (3) DECEMBER 6.

[<sub>1</sub> [<sub>2</sub> *How her image haunts me! Waking or asleep, she fills my entire soul! Soon as I close my eyes, here, in my brain, where all the nerves of vision are concentrated, her dark eyes are imprinted ...* ]<sub>2</sub>

THE EDITOR TO THE READER.

*It is a matter of extreme regret that we want original evidence of the last remarkable days of our friend; and we are, therefore, obliged to interrupt the progress of his correspondence, and to supply the deficiency by a connected narration ...* ]<sub>1</sub> (Johann Wolfgang von Goethe: "The sorrows of young Werther")

5. Paratexts<sup>22</sup> such as book titles, chapter headings and genre indications must not be annotated. If the narrative level remains the same, the square bracket of the narrative level is closed before a chapter heading and re-opened afterwards with the same label.

<sup>22</sup>G rard Genette, *Paratexts. Thresholds of Interpretation* (Cambridge, 1997).

- (4) [<sub>1</sub>*By reason of these things, then, the whaling voyage was welcome; the great flood-gates of the wonder-world swung open, and in the wild conceits that swayed me to my purpose, two and two there floated into my inmost soul, endless processions of the whale, and, mid most of them all, one grand hooded phantom, like a snow hill in the air.* ]<sub>1</sub>

### CHAPTER 2. *The Carpet-Bag.*

[<sub>1</sub> *I stuffed a shirt or two into my old carpet-bag, tucked it under my arm, and started for Cape Horn and the Pacific. Quitting the good city of old Manhatto, I duly arrived in New Bedford. It was a Saturday night in December. Much was I disappointed upon learning that the little packet for Nantucket had already sailed, and that no way of reaching that place would offer, till the following Monday.* ]<sub>1</sub> (Herman Melville: “Moby Dick”).

6. Headings that belong semantically and syntactically to a narrative level are exceptions to this rule. Those are assigned to the particular narrative level (normally to the superordinate level).

(5) [<sub>1</sub>... [<sub>E</sub> *After this, hear the true and graceful story of Lau, the beautiful water nymph.*

[<sub>2</sub> *In Swabia, on the Alb, near the little town of Blaubeuren, close behind the old monastery, you can see beside a sheer rock face the big round basin of a wondrous spring called the Blue Pool ...* ]<sub>2</sub> ]<sub>1</sub> (Eduard Mörike: “Das Stuttgarter Hutzelmännlein”, trans. by the authors).

7. Narrative levels can be interrupted and thwarted by other narrative levels. E.g., in a second-degree narrative, inserts from the first-degree narrative might occur. In this case, level 2 will be closed at the beginning of the insert and reopened after the insert with the same numbering.

(6)[<sub>1</sub> *The country gentleman was of the opinion that he knew how to choose well those stories that would verify his proposition.* [<sub>2c</sub> “The third story,” ]<sub>2c</sub> *the officer continued, [*<sub>2c</sub> *”took place in the war of independence of the Netherlands, at the siege of Antwerp by the Duke of Parma. The duke had blocked the Schelde river by means of a bridge of ships and the Antwerpens were working on their side, under the leadership of a talented Italian, to explode the bridge by means of fire boats that they launched against it. In that moment, ]*<sub>2c</sub> *gentlemen, [*<sub>2c</sub> *in which the vessels float down the Schelde*

*to the bridge, there stands, observe well, a cadet officer on the left bank of the Schelde right next to the Duke of Parma ...” ]<sub>2c</sub> Go to the Devil! shouted the country gentleman ... ]<sub>1</sub> (Heinrich von Kleist: “Improbable Veracities”)*

8. In rare cases, a text does not allow for the annotation of narrative levels. This will be the case, for example, if the narrative levels cannot be separated logically, a phenomenon that is called *metalepsis*.<sup>23</sup> In Italo Calvino’s “If on a winter’s night a traveler”, the world of the reader/narrator (exegesis) is so closely interwoven with the story (diegesis) that narrative levels can no longer be clearly distinguished from each other. In such cases, we do not annotate any narrative levels.

*(7) I am the man who comes and goes between the bar and the telephone booth. Or, rather: that man is called “I” and you know nothing else about him, just as this station is called only “station” and beyond it there exists nothing except the unanswered signal of a telephone ringing in a dark room of a distant city. I hang up the receiver, I await the rattling flush, down through the metallic throat, I push the glass door again, head toward the cups piled up to dry in a cloud of steam. The espresso machines in station cafés boast their kinship with the locomotives, the espresso machines of yesterday and today with the locomotives and steam engines of today and yesterday. It’s all very well for me to come and go, shift and turn: I am caught in a trap, in that nontemporal trap which all stations unfailingly set. A cloud of coal dust still hovers in the air of stations all these years after the lines have been totally electrified, and a novel that talks about trains and stations cannot help conveying this odor of smoke. For a couple of pages now you have been reading on, and this would be the time to tell you clearly whether this station where I have got off is a station of the past or a station of today; instead the sentences continue to move in vagueness, grayness, in a kind of no man’s land of experience reduced to the lowest common denominator. Watch out: it is surely a method of involving you gradually, capturing you in the story before you realize it’s a trap. (Italo Calvino: “If on a Winter’s Night a Traveler”).*

9. Sometimes, a narrator interrupts the story in order to comment on the story, such as to address the recipient (see III. 7). Aphorisms, mottos,

---

<sup>23</sup>John Pier, “Metalepsis,” in *The Living Handbook of Narratology* (2016b [2011]). <http://www.lhn.uni-hamburg.de/article/metalepsis-revised-version-uploaded-13-july-2016> (06/15/2018).

comments, judgments, forms of address and thoughts<sup>24</sup> expressed by the narrator are annotated as parts of the current narrative level and are not regarded as an independent narrative level. Since it may be beneficial for some cases (e.g., comparing judgments of the narrator to the plot), we nevertheless annotate those expressions as “non-narrative”:

To annotate non-narrative parts, we use square brackets followed by the letter E. This indicates that they do not form a narrative level. Opening brackets are used to signal the beginning and closing brackets to signal the end of the expression.

(8) [<sub>1</sub>*In the days when everybody started fair, [E Best Beloved ]E, the Leopard lived in a place called the High Veldt. [E ‘Member ]E it wasn’t the Low Veldt, or the Bush Veldt, or the Sour Veldt, but ...’ ]<sub>1</sub>*(Rudyard Kipling: “How the Leopard got his Spots”).

(9) [<sub>1</sub>... [<sub>2</sub> *The old woman often went out in the morning, and did not return till evening, when I used to go out with the little dog to meet her ... [E often and often as I must have repeated it, do what I will, I cannot call back again the singular name of the little dog. ]E ... ]<sub>2</sub> ]<sub>1</sub>*(Ludwig Tieck: “The white Egbert”).

(10)[<sub>1</sub>*That puzzled the Leopard and the Ethiopian, but they set off to look for the aboriginal Flora, and presently, after ever so many days, they saw a great, high, tall forest full of tree trunks all ’clusively speckled and sprottled and spottled, dotted and splashed and slashed and hatched and cross-hatched with shadows. [E (Say that quickly aloud, and you will see how very shadowy the forest must have been.) ]E ... ]<sub>1</sub>*(Rudyard Kipling: “How the Leopard got his Spots”).



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

<sup>24</sup>Schmid, *Elemente der Narratologie*, 7.



# Annotating Narrative Levels: Review of Guideline No. 4

J. Berenike Herrmann

09.17.19

*Article DOI: 10.22148/16.057*

*Journal ISSN: 2371-4549*

*Cite: J. Berenike Herrmann, "Annotating Narrative Levels: Review of Guideline No. 4," Journal of Cultural Analytics. December 3, 2019. doi: 10.22148/16.057*

## **Reviewer's note**

In the following I briefly detail my comments on the submitted guidelines for the annotation of narrative levels in the "SANTA 4 Annotation Guidelines".

The submission documents a comprehensive and thorough approach to annotating narrative levels, going for a theory-driven perspective. The guidelines themselves are well thought through. However, they should be more transparent with regard to theoretical premises and terminology, as well as more practically applicable through (1) more examples and (2) possibly an integration of the sections "premises" and "annotation guidelines." The link to computational application should receive more explicit attention.

Formal remarks: It is advisable, where possible, to use "international references" (published in/translated to English). Also, to use a gender-neutral language (e.g., establishing coreference to "narrator" not just by "he"). The text should be slightly revised for style and English idiomaticity. I would like to encourage the authors to be less tentative in their formulations. The shared task is not the place for discursively exploring complexity but for solving problems through straightforward guidelines - taking a positively reductive approach.

[I have provided more detailed remarks in the submitted guidelines for the authors' convenience].

## Theoretical Introduction

Within the context of the interdisciplinary scope of the shared task, the conceptualization should strike a better balance between brevity and a broader, but clearly delineated scope. Therefore, within the limits of an annotation manual, the authors should briefly situate and elucidate the particular concepts within the larger field (thus not limited to Genette only; and within an international frame). The terminology should be more precise, and more transparent through examples. Through this, the reader will get a first working knowledge and the particular approach taken will be motivated (“narrative”, “narrator”, and “narrative level”).

## Terminology & Concepts

The authors should flag out more precisely which theory they refer to (not tacitly assuming expert knowledge of reader). So far, formulations such as “we use the terms first-, second-, third-, ... degree narrative as an alternative terminology” leave open which specific theoretical frame is referred to.

### *Experiencing space*

The authors propose the useful term “experiencing space”. However, its definition “subsumes features of the narrative level with regard to its time, its space and its characters” (p. 6) is relatively vague. It should be further specified and accompanied by annotation criteria and one or more examples. Subsequently, in the part “premises” (p. 7: 4a-d) “experiencing space” appears as a good heuristic concept for annotation. It should be systematically applied in the procedure, but so far is not mentioned in 4b.

## Premises

The authors should define more precisely their “search” (“We search for all narrative levels in a given narrative text.” p. 6). By close reading on a word-by-word basis? By more loosely skimming the text? Are annotators allowed to use external references in this search, and if so, which (e.g., lexica—or Wikipedia etc.)?

The operational definition of story is “a self-contained action whose events and happenings are causally linked and cause a change of state.” (p. 7) Is a “story” really reducible to an “action” in your definition? What is the difference between “events” and “happenings”? As for “change of state” whose state does this refer to? The authors don’t mention actors, objects, etc.

The definition of story/level is given in 4d (p. 7). I suggest to provide it earlier, possibly together with that of the concept “experiencing space”. This would

solve the problem the reader confronts in 4c: without 4d, 4b leaves open how the authors distinguish narrating self and experiencing self practically. Also, 4c remains unclear as to whether “a new story” (new narrator/same narrator) is identical with “a new narrative level” - and thus whether a change of narrative levels needs a new narrator.)

Where addressing “embedding,” the authors may want to define the particular (spatial) model of levels—do they use Genette, “working up,” or others that work “down”? (the authors say “Embedded stories can be functionally related to their superordinate narrative level, their frame stories. Possible functions are <sup>1</sup>”) (p. 8). Generally, the model should be maximally precise - are “interlaced” ( p. 9) and “sequential” two types of “embedding”?

## Guidelines

For annotation of “non-narrative parts”, the tag E for “exegesis” may sometimes not be adequate (as the authors have pointed out themselves). Non-narrative passages do not necessarily have to be linked to the exegesis. The “teller level” may be named “level 0”, for conceptual, but also for practical reasons.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

---

<sup>1</sup>Cf. Silke Lahn and Jan Christoph Meister, *Einführung in die Erzähltextanalyse* (Stuttgart und Weimar: J. B. Metzler, 2008), 83-84.

# Annotation Guideline No. 5: Annotation Guidelines for Narrative Levels and Narrative Acts

Florian Barth

10.03.19

*Article DOI: 10.22148/16.056*

*Journal ISSN: 2371-4549*

*Cite: Florian Barth, "Annotation Guideline No. 5: Annotation Guidelines for Narrative Levels and Narrative Acts," Journal of Cultural Analytics. December 3, 2019. doi: 10.22148/16.056*

## 0 Rationale

The annotation guidelines for narrative levels and narrative acts have been developed in conjunction with my master thesis, the goal of which is to distinguish plot relevant settings from rather mentioned spaces in literary texts. For this, the determination of narrative levels is a requirement to precisely classify settings. Nevertheless, the guidelines itself were developed independently from the spatial classification task.

Since the notion of narrative levels encompasses both, level and narrative, the guidelines aim at a clear separation of these concepts. Therefore, they are designated as *narrative level* and *narrative act*, and both terms also serve as annotation tags. The narrative level gives the vertical dimension of the tagset and can hold a limitless amount of narrative acts on a horizontal axis, even on the first level. While narratological theory of levels broadly reflects on framing or embedding techniques and their specific function, these guidelines only focus on the determination of the vertical level structure of narrative acts or their horizontal succession. For this, the guidelines reflect on 1) nested narrators, as described by Genette, 2) possible types of level borders including Ryan's cross-classification of

illocutionary and ontological boundaries, or 3) techniques that cause a change of a narrative act without level switch, as Peer and Coste describe it for digression.

To specify textual characteristics in conjunction with the latter, each tag can be appended with property values that address, for example, the identity of the narrator, its presence in the diegesis or the relation and boundary to the upper level. Furthermore, the properties capture specific textual surfaces like letters or quotations of other literary works as well as metanarration and metafiction, which all indicate insertions of separate narrative acts on the same or a subordinate level. At last, properties give the annotator an opportunity to highlight metalepsis if there is a transgression between two or more narrative levels.

Overall, the guidelines are an attempt to deliver a simplistic and an easy to use set of tags with a clear hierarchical structure based on the distinction of narrative levels and narrative acts. In addition, property values include a more comprehensive perspective on the narratological background and also force the annotator to reflect on his annotation decision.

## 1. Tagset Conception

*Narrative levels*, as proposed by Genette, aim to describe the relations between an embedded narrative and the diegesis,<sup>1</sup> and indicate a clear hierarchical structure between these diegetic levels. Genette explicitly states his intention to systemize the existing notion of *embeddings*, which, according to him, lacks “the threshold between one diegesis and another” as well as the possibility to hierarchical structure a “second diegesis [...] within the first diegesis”.<sup>2</sup>

In these guidelines, the often co-occurring notion of embeddings and *framed narratives* is grouped under the term *narrative act*.<sup>3</sup> Since narrative acts can not always be considered in conjunction with vertical levels (e.g. William Nelles points out the possibility of horizontal embeddings),<sup>4</sup> we clearly separate narrative levels and narrative acts.

---

<sup>1</sup>G rard Genette, *Narrative discourse: An essay in method* (Cornell University Press, 1983), 227-231; John Pier and Didier Coste, “Narrative Levels (revised version)” *The living handbook of narratology*. Hamburg: Hamburg University Press. 2014.

<sup>2</sup>G rard Genette, *Narrative Discourse Revisited*, translated by Jane E. Lewin, 1988, 84.

<sup>3</sup>Conjunctions and delimitations between embeddings and frames are addressed in section 1.2.

<sup>4</sup>William Nelles, *Frameworks: Narrative levels and embedded narrative*, vol. 33 (Peter Lang Pub Incorporated, 1997), 132.

### 1.1 Tag: Narrative Levels

Typically, narrative levels arise “when a character in a story begins to tell a story of his or her own”, which creates a narrative act within a narrative act.<sup>5</sup> The change of a speaker is the most basic characteristic of levels and obligatory in Genette’s terminology, where for each narrative act on a certain level a different speaker occurs (figure 1).<sup>6</sup>

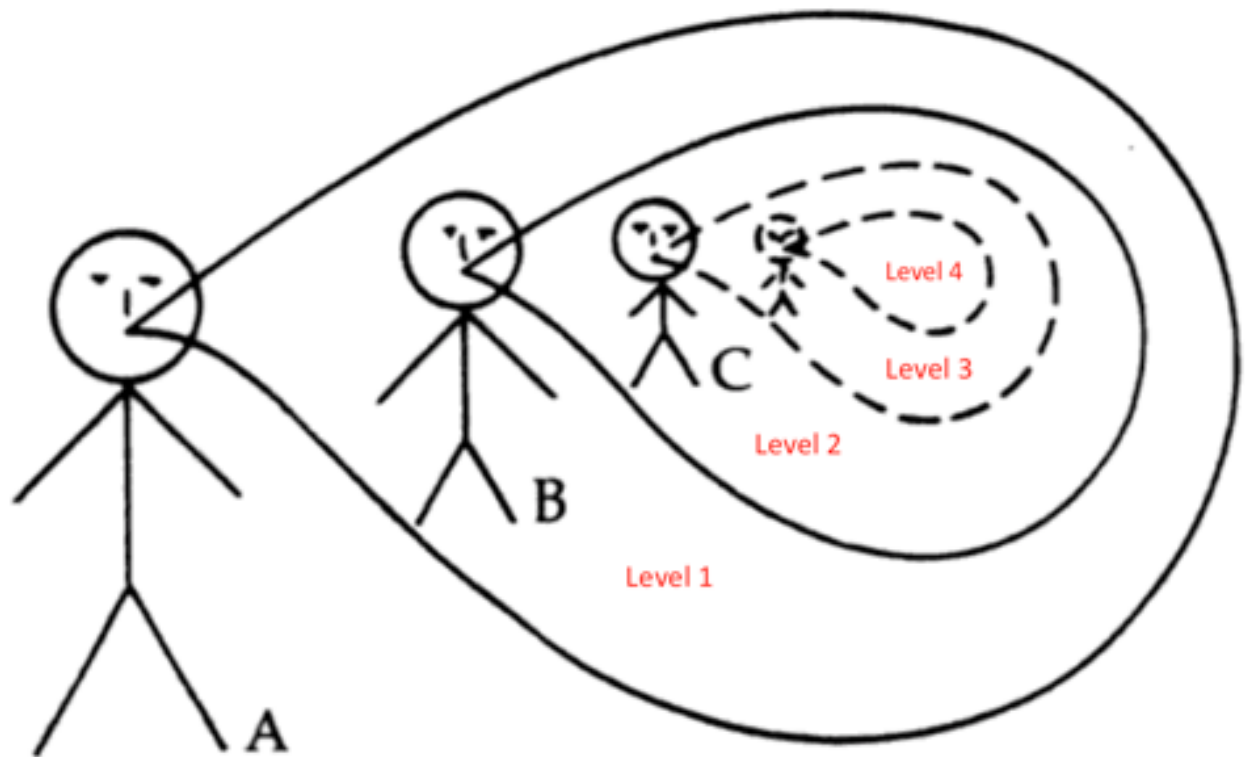


Figure 1. Narrative levels in conjunction with speech acts as proposed by Genette Marie-Laure Ryan describes the switch of speakers as an *illocutionary boundary*, which can be *crossed actually*, when a new voice like a character reports a story on the second level within a direct speech act.<sup>7</sup> Additionally, utterances of characters presented by the narrator as in indirect discourse (indirect speech, character

<sup>5</sup>Manfred Jahn, “Narratology: A guide to the theory of narrative,” *English Department, University of Cologne* 28 (2005).

<sup>6</sup>Jahn, “Narratology: A guide to the theory of narrative”; Silke Lahn and Jan Christoph Meister, *Einführung in die Erzähltextanalyse* (Stuttgart und Weimar: J. B. Metzler, 2008), 83; In Genette’s terminology, the narrating instance of a first level (speaker A in figure 1) is “extradiegetic by definition” (Genette, *Narrative discourse: An essay in method*, 229), therefore his story on level 1 is intradiegetic. An intradiegetic speaker (B) then tells a metadiegetic story (level 2), a metadiegetic speaker (C) a metametadiegetic narration (level 3) and so forth. Within the annotation, we only assign the level by a number, and for the speaker, we set a unique ID (cf. section 2.2 “Speaker: Identity”).

<sup>7</sup>Marie-Laure Ryan, *Possible worlds, artificial intelligence, and narrative theory* (Indiana University Press, 1991), 176.

thoughts) are considered as a *virtually crossed* illocutionary boundary.<sup>8</sup> Furthermore, Ryan highlights that levels not only arise through the switch of speakers but also if a “new system of reality is introduced” like in *Alice in Wonderland*, where “the primary reality of an everyday world” switches to “the dream world of Wonderland [...] in a continuous speech act.”<sup>9</sup> This is defined as the crossing of an *ontological boundary*. While *Alice in Wonderland* marks an actually crossed ontological boundary (the fictional characters indeed enter another form of reality), virtual crossing occurs in this case when the second reality “is anchored” in the primary one, e.g. if the plot of a movie is described from the perspective of the primary reality.<sup>10</sup> An ontological border is also crossed virtually, when the first level narrator cites an existing fictional narrative, like the quote of *Rip van Winkle* in Max Frisch’s *Stiller*. Both, illocutionary and ontological boundaries, can occur combined,<sup>11</sup> which leads to six possible boundaries (cf. figure 2) that are considered as a requirement for a new narrative level in these guidelines.

		<b>Illocutionary boundary</b>	
		—	+
<b>Ontological boundary</b>	—	<b>1</b>	<b>2a: Actually crossed 2b: Virtually crossed</b>
	+	<b>3a: Actually crossed 3b: Virtually crossed</b>	<b>4a: Actually crossed 4b: Virtually crossed</b>

Figure 2. Boundaries between narrative levels following Ryan

Ryan also indicates that each utterance of a new voice may create “its own seman-

<sup>8</sup>Ryan, *Possible worlds, artificial intelligence, and narrative theory*, 176.

<sup>9</sup>Ryan, *Possible worlds, artificial intelligence, and narrative theory*, 177.

<sup>10</sup>Ryan, *Possible worlds, artificial intelligence, and narrative theory*, 177.

<sup>11</sup>Ryan determines an actual crossed illocutionary and ontological boundary (4a in figure 2) as “a fiction within a fiction” told by different speakers, e.g. the stories of the intradiegetic narrator Scheherazade in *The Arabian Nights* (ibid.). Instead, virtual crossing for both boundaries (4b) would refer to a description of a metafictional story from the perspective of the first level speaker but including the mention of a second level speaker (Ryan, *Possible worlds, artificial intelligence, and narrative theory*, 177.). This rare constellation occurs in *Theme of the Traitor and the Hero* by Jorge Luis Borges, where the primary narrator tells his plan to write a story, whose narrator will be “Ryan”, but the first level narrator “never speaks as Ryan himself” (ibid.).

tic universe”, which potentially deviates from the primary reality of the narrative and therefore may establish a new narrative level.<sup>12</sup> Even though the theoretical assumption of a level switch through each crossing of an illocutionary boundary seems considerable, these guidelines only focus on levels, in which indeed a new narrative act is realized.<sup>13</sup>

## 1.2 Tag: Narrative Acts

As proposed above, narrative acts cover both, embedded and framed narratives. Framing is more a “presentational technique”, where the rather short frame narration encloses a more ample inner tale like a painting.<sup>14</sup> An example is Joseph Conrad’s novel *Heart of Darkness*, in which an extradiegetic narrator only introduces the character of Marlow that tells the story of his voyage up the Congo River on a second level (figure 3). In contrast, embeddings can be thought of as smaller insertings “within a larger unit,”<sup>15</sup> e.g. in Kleist’s short story *Improbable Veracities* an officer tells three stories that appear as independent narrative acts on the second level (figure 4). Practically, the border between the dominance of an inner tale and a frame narrative is fluent, and these guidelines do not aim to identify framing or embedding techniques, their specific function,<sup>16</sup> or a certain “main narrative” within several stacked narrative acts.<sup>17</sup>



Figure 3. Framing in Joseph Conrad’s *Heart of Darkness*

<sup>12</sup>Ryan, *Possible worlds, artificial intelligence, and narrative theory*, 175-176.

<sup>13</sup>If only the boundaries for potential narrative levels are of interest, this may lead to tasks like the detection of direct and indirect speech acts that has been done separately, cf. Annelen Brunner, “Automatic recognition of speech, thought, and writing representation in German narrative texts,” *Literary and linguistic computing* 28, no. 4 (2013): 563-575.

<sup>14</sup>Inner tale refers to the term *Binnenerzählung* in German literary discourse (Lahn and Meister, *Einführung in die Erzähltextanalyse*, 79); Pier and Coste, “Narrative Levels (revised version)”.

<sup>15</sup>Lahn and Meister, *Einführung in die Erzähltextanalyse*, 79.

<sup>16</sup>Cf. Shlomith Rimmon-Kenan, *Narrative fiction: Contemporary poetics* (Taylor & Francis e-Library, 2005), 95; Lahn and Meister, *Einführung in die Erzähltextanalyse*, 87-90.

<sup>17</sup>Evelyn Gius, *Erzählen über Konflikte: Ein Beitrag zur digitalen Narratologie*, vol. 46 (Walter de Gruyter GmbH & Co KG, 2015), 164.



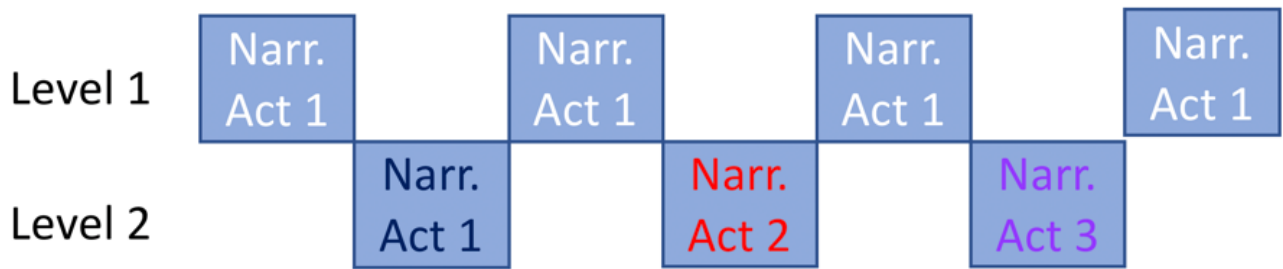


Figure 4. Multiple embeddings of independent narrative acts in Kleist’s *Improbable Veracities*<sup>18</sup>

As opposed to the “vertical” arrangement of narrative acts within levels, Nelles describes “horizontal” embedded narrative acts, which appear at the same level.<sup>19</sup> This happens, when texts by different narrators are presented next to each other without an upper frame narrator. For example, in J. M. R. Lenz epistolary novel *Der Waldbruder* several letters by alternating characters are presented on the same diegetic level (figure 5).

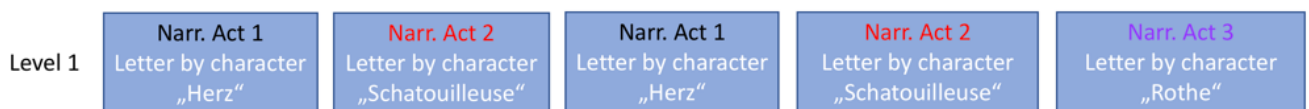


Figure 5. First five letters of the epistolary novel *Der Waldbruder*

Moreover, Pier and Coste describe *digression* as a form of embedding without the switch of levels.<sup>20</sup> This includes *excursus*, e.g. if the narrator directly addresses the reader,<sup>21</sup> which occurs, for example, in Houellebecq’s novel *Extension du domaine de la lutte*, where the narrator states:

1. The pages that follow constitute a novel; I mean, a succession of anecdotes in which I am the hero. [...] There are some authors who employ their talent in the delicate description of varying states of soul, character traits, etc. I shall not be counted among these.

Additionally, Bernard Duyfhuizen describes *intercalation* as a form of digres-

<sup>18</sup>In our terminology, we count narrative acts separately on each level. Narrative act 1 to 3 on the second level represent the embedded stories, while narrative act 1 on the first level marks the gathering, in which the officer tells these stories.

<sup>19</sup>Nelles, *Frameworks: Narrative levels and embedded narrative*, 132; Nelles also defines the term *modal embedding* for dream worlds (William Nelles, “Embedding,” *Routledge Encyclopedia of Narrative Theory*, 2010). In contrast to Ryan, he doesn’t see a level switch here, even though he states a shift in the ‘reality’ of the fictional world. Still, for our guidelines the assumption of a subordinate level for crossing ontological boundaries seems more accurate (cf. Ryan, *Possible worlds, artificial intelligence, and narrative theory*).

<sup>20</sup>Pier and Coste, “Narrative Levels (revised version)”.

<sup>21</sup>Excursus also corresponds with metanarration, which is captured as property in caption 2.2.

sion.<sup>22</sup> This includes intercalated apologues like in Aesop’s fable *The Wolf and the Lamb* that closes with a moral statement:

2. The tyrant can always find an excuse for his tyranny. The unjust will not listen to the reasoning of the innocent.

In summary, a new narrative act is indicated by a level switch (illocutionary or ontological boundary) or by horizontal insertings (letters without framing instance; apologues). Besides such formal criteria for narrative acts, Eberhard Lämmert indicates that a new narrative act at least diverges in time, setting or the corresponding characters from the previous one.<sup>23</sup>

## 2. Annotation Scheme

### 2.1 Inclusion and Stacking of Narrative Acts

The main focus of the annotation is to determine the relationship between vertical stacked or horizontal structured narrative acts, which happens by associating the narrative level. Therefore, no limits of inclusion exist, narrative acts can have multiple embeddings and on each level several independent narrative acts can occur. Therefore, embedded narrative acts can frame stories and vice versa. Ryan illustrates this by means of *The Arabian Nights*, where the framing narrative act of Scheherazade and the Sultan directly includes the stories of *Ali Baba and The Three Ladies of Baghdad* told by Scheherazade on level 2.<sup>24</sup> Moreover, the latter story includes several independent narrative acts on level 3 like *Amina’s tale* (stories 4, 5, 7, 8 in figure 7), which also contains *The young Man’s Tale* on level 4 (figure 8).

To represent the vertical structure of narrative levels, each of which can include a limitless amount of narrative acts, we use the following nested structure of tags:

- Level 1
  - Narrative act 1
  - Narrative act 2
  - ...
  - Narrative act n
- Level 2
  - Narrative act 1

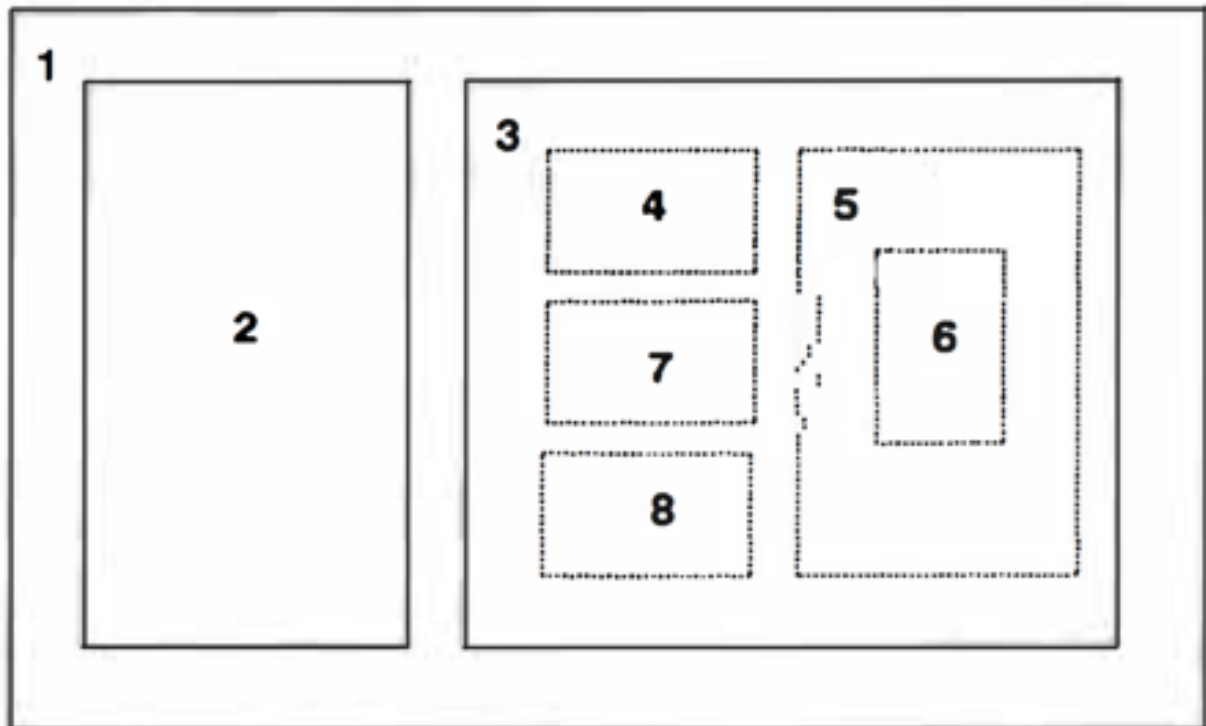
<sup>22</sup>Bernard Duyfhuizen, “Framed narrative,” *Routledge encyclopedia of narrative theory*, 2005, 187.

<sup>23</sup>Eberhard Lämmert, *Bauformen des Erzählens* (Stuttgart: Metzler, 1955).

<sup>24</sup>Ryan, *Possible worlds, artificial intelligence, and narrative theory*; Marie-Laure Ryan, “Stacks, Frames, and Boundaries,” in *Narrative Dynamics: Essays on Time, Plot, Closure, and Frames*, ed. Brian Richardson (Ohio State University Press, 2002), 366.

- Narrative act 2
- ...
- Narrative act n
- ...
- Level n

The span of the annotation can cover whole chapters but also single paragraphs, complete sentences or clauses.



- 1: *The Arabian Nights***
- 2: *"Ali Baba"***
- 3: *"The Three Ladies of Baghdad"***
- 4: *The Porter's tale***

- 5: *Amina's tale***
- 6: *The Young Man's tale***
- 7: *Safia's tale***
- 8: *Zubaida's tale***

Figure 6. Inclusion scheme for the *Arabian Nights* by Ryan

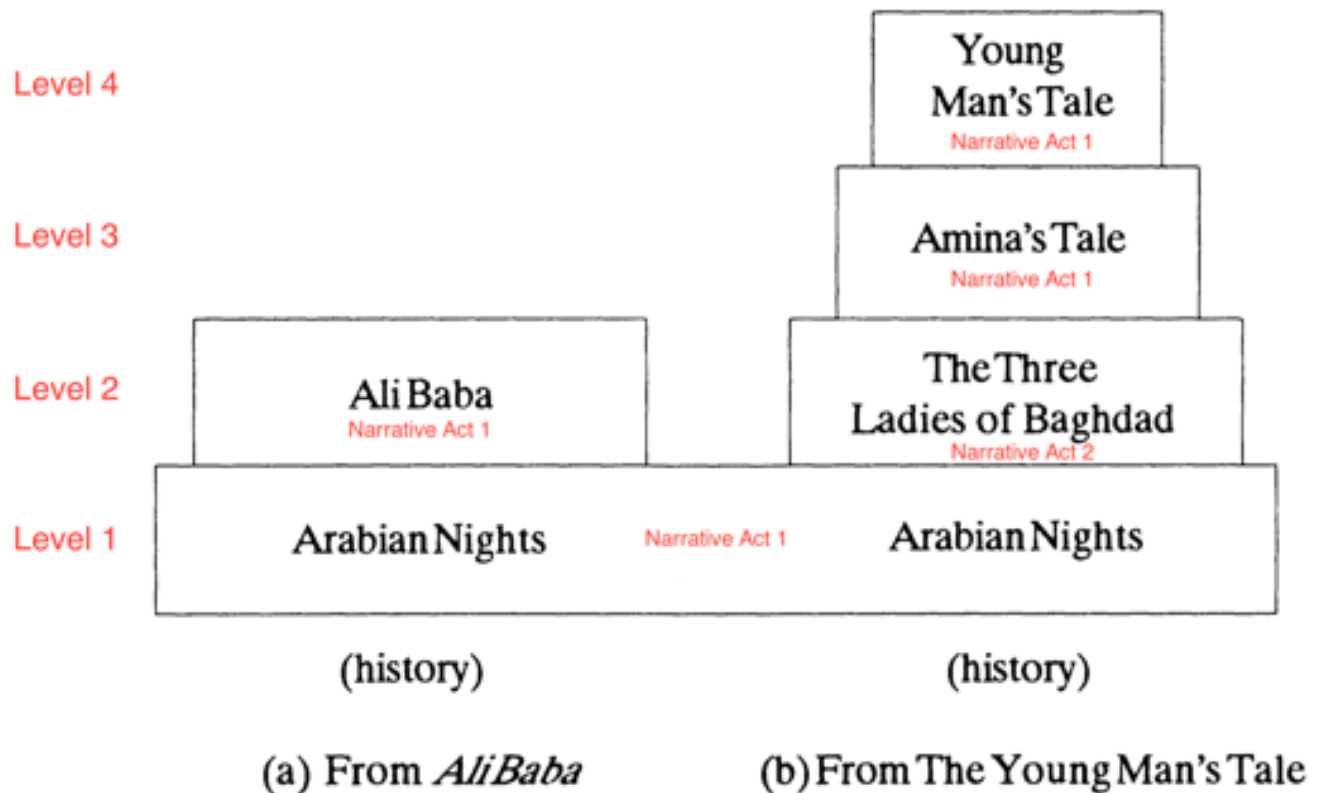


Figure 7. Stacking of narrative levels in the *Arabian Nights* (adapted from Ryan)

## 2.2 Properties

Properties aim to reflect on the annotation decision and give further information about the relation of narrative acts and levels.

### Upper Level: Boundary

This property indicates the boundary between narrative levels following Ryan (cf. figure 2). As mentioned above, illocutionary and ontological boundaries can be combined.

- Illocutionary boundary (actual)
- Illocutionary boundary (virtual)
- Ontological (actual crossed)
- Ontological (virtual crossed)

### Upper Level: Head of former level

The annotator should indicate the narrative act of the former level, in which the current narrative act is embedded. For example, the head of *Amina's Tale* is *The Three Ladies of Baghdad*, which is narrative act 2 on level 2 (cf. figure 8).

### Speaker: Identity

Since stacked narrative levels can have multiple narrators, we capture the identity of each speaker. This is done by alphabetic ID's for each speaker identity:<sup>25</sup>

- speaker entity a
- speaker entity b
- ...
- speaker entity n

For example, in Mary Shelley's *Frankenstein; or, The Modern Prometheus* a different narrator occurs on each level: Robert Walton recounts in his journal the meeting with Victor Frankenstein and quotes the oral narration of Frankenstein, who cites the metadiegetic narration of his creature.<sup>26</sup>

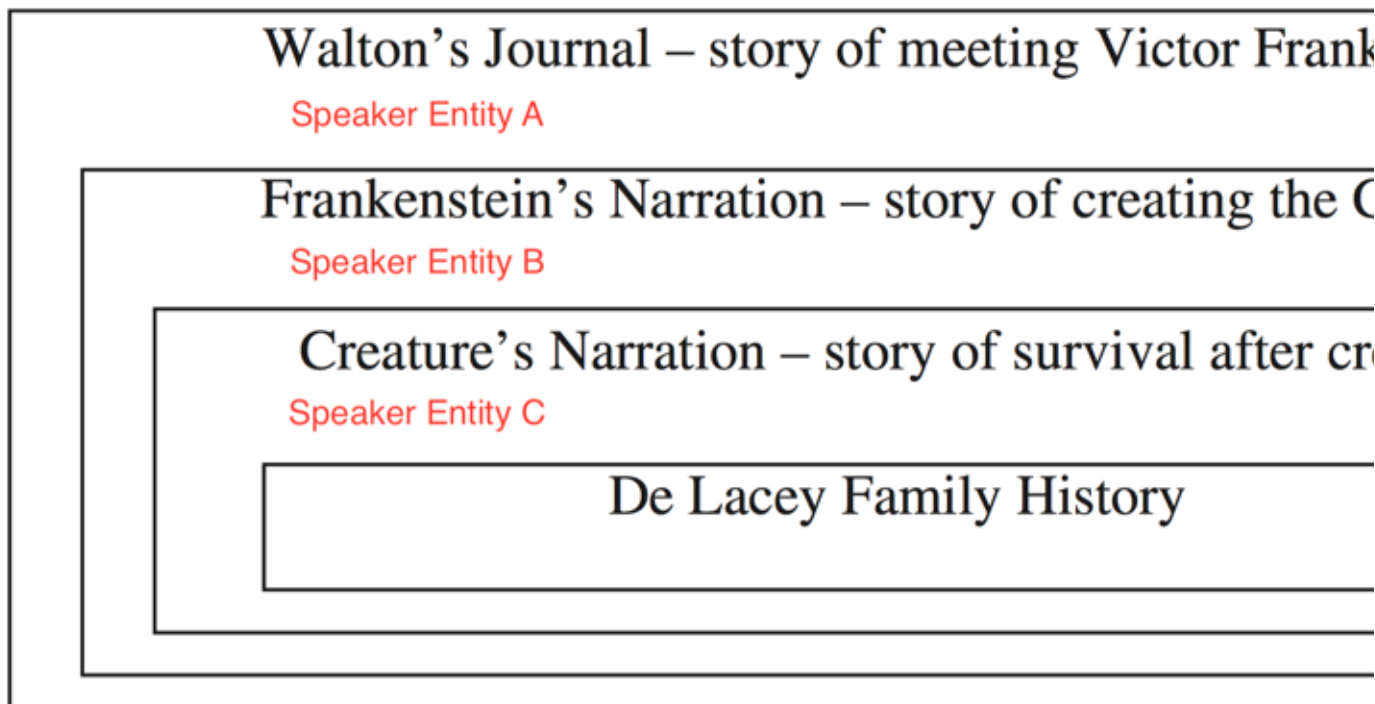


Figure 8. Different narrators for each level in *Frankenstein; or, The Modern Prometheus*

In contrast, when the reporting voice remains constant between level 1 and 2 (e.g. if the same narrator reports a dream, which corresponds with the crossing of an ontological border), it should be annotated as the same speaker entity.

<sup>25</sup>We do not use Genette's terminology for speakers (extradiegetic, intradiegetic, metadiegetic) since they only capture the level of a speaker, not his identity.

<sup>26</sup>Cf. Duyfhuizen, "Framed narrative," 187; Another example would be Theodor Storm's *Der Schimmelreiter*, cf. Lahn and Meister, *Einführung in die Erzähltextanalyse*, 85-87.

### Speaker: Story Presence

This property captures if a speaker is present in the story or not. We use the terms defined by Genette:

- homodiegetic (speaker is part of the diegesis)
- heterodiegetic (speaker isn't part of the diegesis)

### Narrative: Type

To record the type of narrative or speech act of an intradiegetic character, we annotate the textual type of a narrative act. Predefined are:

- undefined (This applies to the most extradiegetic narrators on level 1.)
- direct speech act (cf. *Heart of Darkness* in figure 3)
- indirect speech act (cf. the example of Chekhov's *An Avenger* below)
- quotation of a literary work (e.g. the quote of *Rip van Winkle* in Max Frisch's *Stiller*)
- letter (for example, the letters in *Waldbruder* [figure 5] or *Frankenstein; or, The Modern Prometheus* [figure 9]).
- transcribed speech (This also occurs in *Frankenstein; or, The Modern Prometheus*, since Walton transcribes Frankenstein's narration in his letters.)

Example: Speaker switch within one narrative act in Chekhov's *An Avenger*

Following Ryan, we consider indirect utterances or thoughts of characters presented by the narrator as an implication for a switch of levels (cf. above: virtually crossed illocutionary boundary in section 1.2). Therefore, it happens that two speakers occur within a single narrative act like in *An Avenger*. First, the thoughts of Fyodor Fyodorovitch Sigaev are uttered within direct speech and secondly expressed by the frame narrator:

3. ["Shouldn't I challenge him to a duel?"]<sub>1</sub> [flashed through Sigaev's mind.]<sub>2</sub>  
 ["It's doing him too much honour, though. . . . Beasts like that are killed like dogs. . . ."]<sub>1</sub>

[His imagination pictured how he would blow out their brains, how blood would flow in streams over the rug and the parquet, how the traitress's legs would twitch in her last agony. . . . But that was not enough for his indignant soul. The picture of blood, wailing, and horror did not satisfy him. He must think of something more terrible.]<sub>2</sub>

1: level 2; narrative act 1; speaker entity 2

2: level 2; narrative act 1; speaker entity 1

This passage is embedded within the narration of the extradiegetic narrator of level 1. Therefore, the direct speech is assigned the property “speaker entity 2”, while “speaker entity 1” in the second paragraph refers back to the narrator of the first level.

### Metanarration & Metafiction

Both, metanarration and metafiction, address self-reflexive utterances. While metanarration covers “the narrator’s reflections on the act or process of narration” (like in the example of Houellebecq’s novel in section 1.2), metafiction rather concerns “comments on the fictionality and/or constructedness of the narrative.”<sup>27</sup> Metafiction occurs in Italo Calvino’s *If on a winter’s night a traveler*, where narrator describes the reading process in second person. Each chapter contains another version of how the novel could be written (each is a separate narrative act), but none of these stories gets finished.

As mentioned in section 1.2, metanarration and metafiction are supposed to be annotated on the same level, in which they occur, but they create a new narrative act. These narrative acts can be marked by the property values “metanarration” or “metafiction”.

### Metalepsis

We capture metaleptic intrusions of the upper or the lower level.<sup>28</sup> For example, if a metafictional character from level 2 appears in a narrative act on level 1 (by violating ontological boundaries), we add the property value “intrusion by level 2” to the annotation of the narrative act on level 1.<sup>29</sup>



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

<sup>27</sup>Birgit Neumann and Ansgar Nünning, “Metanarration and metafiction”, *Handbook of Narratology*, 2015, 204-211.

<sup>28</sup>Cf. Lahn and Meister, *Einführung in die Erzähltextanalyse*, 90.

<sup>29</sup>Originally, Genette’s concept of metalepsis includes “any intrusion by the extradiegetic narrator into the diegetic world” (Genette, *Narrative discourse: An essay in method*, 234-235). For example, if two intradiegetic characters on level 1 speak about the narrator, who writes the story (like in Flann O’Brian’s *At Swim-Two-Birds*), this refers to the extradiegetic point of view of the narrator and is captured within our guidelines by setting the property value “metanarration” (see above).

# Annotating Narrative Levels: Review of Guideline No. 5

Jan Horstman

08.29.19

*Article DOI: 10.22148/16.058*

*Journal ISSN: 2371-4549*

*Cite: Jan Horstman, "Annotating Narrative Levels: Review of Guideline No. 5," Journal of Cultural Analytics. December 3, 2019. doi: 10.22148/16.058*

The guideline is based on a clear and plausible distinction between narrative levels and narrative acts. Narrative levels make up the vertical axis of the developed tagset, on which there can be  $n$  narrative acts on each of its levels. The narrative acts in turn form the horizontal axes of the tagset. The amount of narrative acts ( $n$ ) on each level (which can be embedded, framed, juxtaposed) is principally unlimited. This general distinction takes into account the fact that a change of speaker/narrator can take place without the changing of narrative level, i.e., can happen in the same diegesis.

The main theoretical foundations of the proposed tagset are Gérard Genette's works on narrative levels (1983: *Narrative discourse: An essay in method* and 1988: *Narrative Discourse Revisited*) and Marie-Laure Ryan's framework for the actual and/or virtual forms of crossing narrative boundaries (i.e. illocutionary or ontological) as proposed in *Possible worlds, artificial intelligence, and narrative theory* (1991). Both contributions stem from classical narratology (Ryan's with a more transmedial angle to it than Genette's) and are well-established in the field. Despite their respective complexity, the guideline aptly explains the theories, and - more importantly - takes them as they are without criticizing them for aspects that could be seen as inconsistent or unintuitive (for example the fact that a first level narrator for Genette is "extradiegetic by definition"); a pragmatic decision which clearly puts the focus of the guideline on the operationalization of narrative levels as discussed in theory rather than letting it become a contribution to



these theoretical discussions.

A point that could be stated more clearly is that the guideline decidedly tries to operationalize only selected parts of the discussed theories and e.g. does not consider every crossing of an illocutionary boundary as outlined by Ryan as a case in which “indeed a new narrative act is realized” (p. 3). At the bottom of this specific choice there seems to be a differentiation between speakers and narrators. The underlying understanding of narrativity, however, unfortunately remains rather unspecified.

A very useful differentiation is established with regards to narrative acts: the guideline considers embedded and framed narratives to both be cases of several narrative acts alternating in different ways. Whereas embedded narratives are considered to be shorter narratives within a larger story, framed narratives are described as a longer story that is framed by a short narrative (which e.g. narrates the situation of the telling). Since there does not always have to be a definite border between inner and frame narratives, the comprehensible aim of the proposed tagset is not to identify techniques or functions of embedding or framing.

A further positive aspect of this guideline are the cross-categorical properties that allow the annotators to justify their decisions during the annotation process and to exemplify possible relations between narrative levels and narrative acts (i.e. the specific way of boundary crossing, the respective narrative act of the former level, the speakers’ - or narrators’? - identities, types of narrative-like direct speech acts, quotations of a literary work, letters etc., metanarration and metafiction as well as metalepsis). Genette’s differentiation between homo- and heterodiegetic narrators occurs as properties as well; the difficulties that these supposedly binary categories can bring (or even the different understandings that exist within academic discussions), however, are not reflected upon and thus these two properties could lead to irregularities in the annotation process and lower inter-annotator agreement.

The intelligible, theory-based guideline frequently operates with examples from literary texts, which makes it much easier to capture the explained categories and in return enhances the usability of the proposed tagset. The properties in practice should help to tell whether the annotation of narrative acts helps to identify narrative levels or not, and how the manifold relations between acts and levels can be used to operationalize the detection of narrative levels.



Unless otherwise specified, all work in this journal is licensed under a Creative

Commons Attribution 4.0 International License.

# Annotation Guideline No. 6: SANTA 6 Collaborative Annotation as a Teaching Tool Between Theory and Practice

Matthias Bauer and Miriam Lahrswow

01.15.20

Article DOI: [10.22148/001c.11747](https://doi.org/10.22148/001c.11747)

Journal ISSN: 2371-4549

Cite: Matthias Bauer and Miriam Lahrswow, "Annotation Guideline No. 6: SANTA 6 Collaborative Annotation as a Teaching Tool Between Theory and Practice," *Journal of Cultural Analytics*. January 15, 2020. doi: [10.22148/001c.11747](https://doi.org/10.22148/001c.11747)

## Preliminary Remarks

These guidelines were developed in our seminar "Digital Methods in Literary Studies", which was aimed at M.A. students and advanced B.A. students.<sup>1</sup> At the beginning of the seminar, students were introduced to the aims and challenges of digital annotating in general as well as to different narratological theories (including Genette, Ryan, Nelles, and Füredy). Due to its narratologically challenging nature, Mary Shelley's *Frankenstein* was chosen as a text against which we could test our guidelines and which triggered their modification. In *Frankenstein* many changes (e.g. of narrator and narratee) occur at the beginning of chapters. Even though such changes can, of course, also be found in the middle of chapters, annotators should pay special attention to the beginning of chapters, because they often coincide with a change in narrator, narratee, or narrated world.

---

<sup>1</sup>We would like to thank the organisers of SANTA and our anonymous reviewers for their detailed and valuable feedback. We would also like to thank our students Elisabeth Bleaß, Berit Boehling, Kristina Burghardt, Aylin El Damhougi, Leonie Greß, Yani Hu, Alia Luley, Günay Mammadova, Sandrina Kimberly Müller, Jona Odza, Yasemin Özalp, Laila Prota, Jonathan Schneider, Sarah Schneidewind, Andra Cristina Sterian, Lilien Sztudinka, Amina Tschubajew, Panagiotis Tzatsos, Ella Ujhe-lyi, Ningxi Xie, Yicong Xu, and Karmen Zeiler for their great participation and input.

One problem we debated in class was *how* to annotate in the first place: Should we only annotate the place in which the change occurs, e.g. the point between two different narrative levels, or should we annotate the whole passage belonging on one level? In the end, we decided to use a combined model, i.e. to allow both the use of paired brackets and the annotation of the point in-between two contrasting passages.

The in-class discussions soon drew our attention to fundamental problems that arise when trying to transform vague or even contradictory narratological theories into unambiguous, widely applicable annotation categories. The first issue was the definition of narrative itself. In particular, when does a dialogue, which is part of a narrative, become a narrative of its own? For example, is the statement “I went to the supermarket and bought some fruit” already a narrative? As a simple working definition we decided to chose “a report of connected events.”<sup>2</sup>This is important because, for example, Ryan has an even wider definition,<sup>3</sup> which leads, as we think, to obscuring matters by a proliferation of narratives. The example, however, indicates a wider problem: there needs to be a clearly defined research question *before* starting to define and annotate narrative levels. For example, when one wants to find out whether novels from the eighteenth century tend to have more embedded narratives than twentieth-century novels, using annotation guidelines that are primarily based on Ryan’s theory (see 1 and 4 below) might distort one’s results because the crossing of an illocutionary or an ontological boundary does not necessarily establish an embedded narrative. Hence, even within the field of embedded narrative, there is no such thing as a ‘universally marked-up text’ that has to be annotated once and then can be re-used for many different research purposes.

The discussion of *Frankenstein* alerted us to another problem, namely the question of who, actually, is the narrator in a given passage: In the novel, Walton does not hear the Creature relate its own story; instead, it is filtered through Frankenstein. Who, then, is the narrator of the passages concerning the early life of the Creature? The Creature who related them to Victor, Victor who tells them to Walton (and maybe slightly manipulates them), or Walton who writes them down (and maybe does not transcribe Victor’s tale verbatim)? For the sake of simplicity, we decided to go for the original source and assumed that the Creature is the most relevant narrator of its own tale. We also suggest to consider mediated documents (e.g. letters that are transcribed or read aloud by characters) as embedded narratives. The aspect of time (e.g. whether a certain part of

---

<sup>2</sup>“Narrative,” Wikipedia, last modified September 22, 2018, <https://en.wikipedia.org/wiki/Narrative>.

<sup>3</sup>Marie-Laure Ryan, “Embedded Narratives and Tellability,” *Style* 20 (1986): 319-40.

the narrative occurs in a prolepsis) also had to be discarded since otherwise our guidelines would have become too complex. Furthermore, when contemplating how to annotate two narrative levels that describe different worlds, we decided not to use separate tags for dreams, beliefs, delusions, and the like. This would have led to a proliferation of tags and would have made annotation too dependent on the interpretation of the text (e.g. we sometimes cannot be sure whether a character is hallucinating/dreaming or not). Instead, according to our guidelines, annotators need only indicate whether the world depicted in the narration of a lower level is factually dependent on the world depicted in the higher level or not (see 4 below). We also agreed that it would be helpful to annotate whether a narrative on a lower level is embedded in, or framed by, the narrative of the higher level. (For the theoretical background see 5.1 below.) The problem was, again, one of drawing a clear line between framing and embedding. For example, when the narrated passage on the lower level is just as long as the narrated passage on the higher level, is the former embedded in, or framed by, the latter?

Hence, in our systematization of narrative levels we focused on the features that define narratives within narratives: the narrator (position) (see 2), the narratee (see 3) and the (in)dependence of the narrated world (see 4). We furthermore determined whether the narrative within a narrative is (quantitatively) the main narrative of the whole text or not and if it is fully enclosed (see 5). Last but not least, we took into account if the boundary between narrative levels is strictly observed or if there are cases in which, although we may notice a separate level of narration in some respects, the boundary is transcended in others (see 6).

A question that came up time and again during our discussions was which aspects our guidelines should cover in the first place. We might try to only annotate features that can be identified without much prior interpretation but this would mean to exclude exactly those issues that make literary analysis so intriguing. The students also wondered whether it is possible to develop guidelines that can be used for *all* literary texts. When we annotated the short texts provided by the organisers of SANTA, we soon realised that some of the phenomena that we included in our guidelines were not to be found in these texts, whereas some features that we identified in the texts were not covered by our guidelines. Hence, developing guidelines that are too specifically tailored to one text or genre will make the guidelines useless for analysing other texts, but when the guidelines are too general, they tend not to yield interesting results.

During our in-class discussions, it became clear to what extent annotation depends on definitions and interpretations. Students pointed out that, in the future, they would never rely to studies based on corpora without first considering the guidelines that were used to annotate them. Even though many of them were

critical as to the applicability of annotation for their purposes as literary scholars, they appreciated the development and use of annotation guidelines as a tool for close reading: Rather than let an ambiguous text stay ambiguous, they simply *had* to decide for one option in order to be able to annotate a passage and had to justify their choice with reference to the whole text or to adapt the guidelines in order to address and document the ambiguity. Likewise, they had to precisely identify the location of changes (e.g. of level or narratee) in the text. Students also liked the idea of creating guidelines that were to be used by others as it provided a welcome contrast to writing term papers that no one but their lecturer would read. However, they would have appreciated to get the guidelines and annotated texts of all other participants and to receive feedback on their own guidelines (either by the organisers or by the participants who used them to annotate).

The biggest problem was that it was not really clear which research question the guidelines were designed to tackle. Depending on this, we could have shifted the focus of our guidelines by adding or omitting certain categories. Overall, our students enjoyed the SANTA competition because it enabled them to practice their close reading skills as well as to learn and critically evaluate a new method of conducting literary studies.

## 1. Change of narrative levels (Genette)

### Theoretical Explanation

Change of narrative levels,<sup>4</sup> a threshold between the one and the other: according to Genette, strictly speaking only a second narrative (metadiegetic level) within the first one (the intradiegetic one).

#### 1.1 an actual change of narrator

an actual change of narrator (one of the narrated characters tells a story etc.); cf. Ryan's illocutionary boundary: a different speaker<sup>5</sup>

#### 1.2 no change of narrator

no change of narrator

Definition of the three possible narrative levels:

1. Level within the global text at which the telling of the narrator-characters story occurs
2. The Level at which the primary narrators discourse occurs

<sup>4</sup>G rard Genette, *Narrative Discourse*, trans. Jane E. Lewin (Oxford: Blackwell, 1980).

<sup>5</sup>Ryan, "Embedded Narratives and Tellability."

3. The Level outside of the narrative act situated outside the primary narrator's discourse

### Categories, Attributes, Values

Category	Definition of Category	Attribute	Possible Values
narrative_level	to indicate which narrative of the three described above is presented	number	1 2 3 etc.
level_change	to define if there is a change of narrative level	value	Yes No
narrator_change	to define whether a change in narrator is happening as well	value	Yes No

### Examples <sup>6</sup>

#### 1. LEVEL CHANGE:

[In the first sentence, the narrator is Walton, who is writing a letter to his sister. In the second sentence below, the narrator is the Creature, who is telling his story to Frankenstein, who, in turn, is telling it to Walton.]

</level\_change value="no"> So strange an accident has happened to us, that I cannot forbear recording it, although it is very probable that you will see me before these papers can come to your possession.

[...]

</level\_change value = "yes"> It is with considerable difficulty that I remember the original æra of my being: all the events of that period appear confused and indistinct.

#### 2. NARRATIVE LEVELS:

[In this example, we have Walton's narrative on Level 1, Frankenstein's embedded narrative on Level 2, and the Creatures' narrative, which is embedded in Frankenstein's, on Level 3.]

This manuscript will doubtless afford you the greatest pleasure: but to me, who know him, and who hear it from his own lips, with what interest and sympathy shall I read it in some future day! [...] </narrative\_level> <narrative\_level number = "2"> I am by birth a Genevese; and my family is one of the most distinguished of that republic. My ancestors had been for many years' counsellors and syndics; and my father had filled several public situations with honour and reputation. [...] </narrative\_level> <narrative\_level number = "3"> I lay on my straw, but I

<sup>6</sup>Unless otherwise indicated, all examples are drawn from *Frankenstein*. Invented examples are marked with *inv.* after the number of the example, e.g.: (4, *inv.*) <Invented example>.

could not sleep. I thought of the occurrences of the day. What chiefly struck me was the gentle manners of these people; and I longed to join them, but dared not.

### 3. NARRATOR CHANGE:

[In this example, the sentence marks a change of narrator from Frankenstein to the Creature.]

</narrator\_change value= “yes”> It is with considerable difficulty that I remember the original æra of my being: all the events of that period appear confused and indistinct.

[In this example, we have marked that the narrator stays the same in a new chapter.]

#### Chapter VIII

Thus spoke my prophetic soul, as, torn by remorse, horror, and despair, I beheld those I loved spend vain sorrows upon the graves of William and Justine, the first hapless victims of my unhallowed arts.

#### Chapter IX

</narrator\_change value=“no”> Nothing is more painful to the human mind, than, after the feelings have been worked up by a quick succession of events, the dead calmness of inaction and certainty which follows, and deprives the soul both of hope and fear.

## 2. Narrator’s Position and Part in the Narrative (Genette)

### Theoretical Explanation<sup>7</sup>

2.1 The narrator is either part of the narration or not, i.e. s/he is:

2.1.1 Heterodiegetic narrator

2.1.2 Homodiegetic narrator

2.1.2.1 Autodiegetic narrator (special case of 2.1.2)

2.2 Narrator can also be identified according to their position with respect to the narrative levels:

2.2.1 Extradiegetic narrator

2.2.2 Intradiegetic narrator

---

<sup>7</sup>Gérard Genette, Narrative Discourse.



## 2.3 Narrator Participation

2.3.1 Homodiegetic Narrator : The Narrator is part of the actual narration

2.3.2 Heterodiegetic Narrator: The narrator is not part of the actual narration

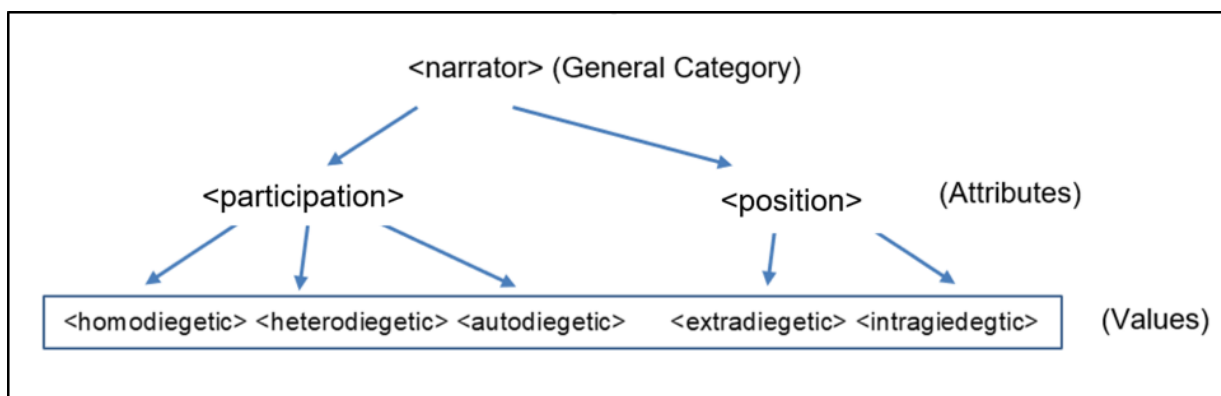
2.3.3 Autodiegetic Narrator: The narrator is part of the narration and is also the protagonist of the story

## 2.4 Narrator Position

2.4.1 Extradiegetic Narrator: Extradiegetic narrative level = level at which intradiegetic events are described; literary act. An extradiegetic narrator does not appear as narrator within a diegesis.

2.4.2 Intradiegetic Narrator: Intradiegetic events are described within the first level of the narrative. There is also an intradiegetic narrator: s/he is already a character in a narrative that is not his/her own.

## Categories, Attributes, Values



## Example

(4) (Beginning of Chapter 7 of *Frankenstein*)

< narrator participation “homodiegetic narrator” > On my return, I found the following letter from my father: </narrator>

< narrator position “intradiegetic narrator” > ”My dear Victor,

”You have probably waited impatiently for a letter to fix the date of your return to us; and I was at first tempted to write only a few lines, merely mentioning the day on which I should expect you. But that would be a cruel kindness, and I dare not do it. What would be your surprise, my son, when you expected a happy and glad welcome, to behold, on the contrary, tears and wretchedness? </narrator>

### 3. Narratee (Nelles)

#### Theoretical Explanation<sup>8</sup>

We have included this category since sometimes narrative levels are only to be distinguished by a change of narratee. In other words, the narrator may remain the same, and the narrated world (see 4 below) may remain the same but the person to whom the story is told may become a different one. (E.g. when the autodiegetic narrator of the first-level narrative tells a story to a specific person within that narrative.)

#### 3.1 Change of narratee

#### 3.2 No change of narratee

#### Categories, Attributes, Values

Category	Attribute	Possible Values
change_narratee	Value	Yes/No

### 4. Change of narrated worlds

#### Theoretical Explanation

We have included this category since it is a key to providing significant information about the relation of the different narratives to each other: do they depend on each other or are they fictions within fictions? Just as fictional texts are counterfactually independent of the actual world,<sup>9</sup> second-level narratives may be counterfactually independent of the world of the first-level narrative. Examples are inserted narratives (as in the *Decamerone* or the *Canterbury Tales*).

Ryan describes in her theory the crossing of boundaries, either illocutionary or ontological. An ontological crossing of boundaries refers to a change of reality. These kinds of reality shifts affect the narratological structure and are therefore important for our guidelines. A shift of reality occurs when narratives refer to two different worlds that are not dependent on each other.

Our category of narrated worlds is similar but not identical with Ryan's "ontological boundary,"<sup>10</sup> which is, however, not strictly logical and therefore impractica-

<sup>8</sup>William Nelles, *Frameworks: Narrative Levels and Embedded Narrative* (New York: Lang, 1997).

<sup>9</sup>Matthias Bauer and Sigrid Beck, "On the Meaning of Fictional Texts," in *Approaches to Meaning: Composition, Values, and Interpretation*, ed. Daniel Gutzmann, Jan Köpping and Cécile Meier (Leiden: Brill, 2014), 250-75.

<sup>10</sup>Ryan, "Embedded Narratives and Tellability."

ble. In the case of narrated dreams it may sometimes be difficult to decide if there is a change of worlds, but even though in dream worlds different physical laws might apply, the dream world is dependent on the narrative world, either due to influence of the experiences of the dreamer or due to their prophetic character. This is why we recommend tagging dreams, as a rule, as “same world”.

## Categories, Attributes, Values

Category	Attribute	Possible Values
change_reality	value	Yes No

### Example

[Even though both the narrator and the narratee change here, the narrated world does not change.]

(5) <narrative\_level number=“1”> This manuscript will doubtless afford you the greatest pleasure: but to me, who know him, and who hear it from his own lips, with what interest and sympathy shall I read it in some future day! </narrative\_level> [...]

</change\_narrator value=“yes”></change\_narratee value=“yes”></change\_reality value=“no”>

<narrative\_level number “2”> I AM by birth a Genevese; and my family is one of the most distinguished of that republic. My ancestors had been for many years counsellors and syndics; and my father had filled several public situations with honour and reputation.

## 5. The nature of the level-change structure

### 5.1 Embedding vs framing narrative

#### Theoretical Explanation

1. Description of the theory: the initial idea of using this tag to mark a level-change is framing of embedding. Embedding can be thought of as inserting or placing something within a larger unit, thus the main story is the embedding one. Framing is generally regarded as a presentational technique: the frame tale is of limited length and varying significance, serving to render the ampler inset or inner tale (Binnenerzählung) accessible

and/or to authenticate it, imbuing it with a “narratorial illusionism,”<sup>11</sup> particularly in simulations of oral storytelling,<sup>12</sup> in which case the main story is the embedded one. However, there is no strict definition distinguishing how large a lower should be when it is called the embedding story, and similarly, how long a higher level should be when the lower level is called a framing story. Besides, if one identifies framing or embedding by finding which level the main story belongs to, the result could depend largely on interpretation.

Here, we provide an alternative by giving the number of words in each level which can be used to compare the length of levels without using the ambiguous term “framing” and “embedding”.

2. Whenever there is a level change in the text, which should be tagged following the instruction in “(2) Change of narrative levels (Genette)”, read the following guidelines to add the information of level length.

When counting the words of “level n”, first count the number of words “Ln” between the tag `<narrative_level number=“n”>` and the first end of tag `</narrative_level number=“n”>` after it (so that you do not count any other parallel level n that does not belong to the same narrative).

If there is no “level n+1” within “level n”,  $L=Ln$ .

If there is “level n+1” within “level n”, count the number of words “Lma”, “Lmb”, “Lmc”, etc. between each pair of beginning tag `<narrative_level number=“n+1”>` and the its corresponding end tag `</narrative_level number=“n+1”>` respectively.  $Lm=Lma+Lmb+Lmc\dots$   $L=Ln-Lm$

Put the tag `</narrative_level words=“L”>` after the corresponding level tag `<narrative_level number=“n”>`

## Categories, Attributes, Values

Category	Attribute	Possible values
Narrative_level	Words	[counted number of the words in Arabic letters]

<sup>11</sup> Ansgar Nünning, “On Metanarrative: Towards a Definition, a Typology and an Outline of the Functions of Metanarrative Commentary,” in *The Dynamics of Narrative Form: Studies in Anglo-American Narratology*, ed. John Pier (Berlin: de Gruyter, 2004), 11-57. 17.

<sup>12</sup> John Pier, “Narrative Levels,” in *The Living Handbook of Narratology*, last revised October 10, 2016. <http://www.lhn.uni-hamburg.de/article/narrative-levels-revised-version-uploaded-23-april-2014>.

## Example

(6, *inv.*) `<narrative_level number="1"></narrative_level words="13">`  
 Dear Mary, I had a conversation with a strange boy about frogs yesterday.`<narrative_level number="2"></narrative_level words="6">` I have much interest in frogs. `</narrative_level number="2"></narrative_level number="1">`

## 5.2 Opened vs closed narratives

### Theoretical Explanation

Both framing and embedding mentioned in 5.1 can have three kinds of structures concerning if they are complete: opened and closed, opened but never closed, and closed but never opened.

### Categories, Attributes, Values

Category	Attribute	Possible Values
narrative_levelchange	completion	Complete Never closed Never opened

### When to use which value

#### 5.2.1 opened and closed.

When there is text between `<narrative_level number="n">` and `<narrative_level number="n+1">`, and there is text between `</narrative_level number="n+1">` and `</narrative_level number="n">`, the structure is opened and closed.

Put the tag `</narrative_levelchange completion="complete">` before `<narrative_level number="n+1">`

#### 5.2.2 opened but never closed

When there is text between `<narrative_level number="n">` and `<narrative_level number="n+1">`, but there is no text between `</narrative_level number="n+1">` and `</narrative_level number="n">`, the structure is opened and closed.

Put the tag `</narrative_levelchange completion="never closed">` before `<narrative_level number="n+1">`

#### 5.2.3 closed but never opened

When there is no text between `<narrative_level number="n">` and `<narrative_level number="n+1">` , but there is text between `</narrative_level number="n+1">` and `</narrative_level number="n">`, the structure is closed but never opened.

Put the tag `</narrative_levelchange completion=" never opened">` after `</narrative_level number="n+1">`

## Examples

[In example 7, both the beginning and the end of the embedded narrative are marked and present in the text.]

(7) `<narrative_level number="1">` Yesterday the stranger said to me, 'You may easily perceive, Captain Walton, that I have suffered great and unparalleled misfortunes.' [...] `</narrative_levelchange completion="complete">``<narrative_level number="2">` I am by birth a Genevese; my family is one of the most distinguished of that public. [...] `</narrative_level number="2">` You have heard this strange and terrific story, Margaret; and do you not feel your blood congealed with horror, like that which even now curdles mine? `</narrative_level number "1">`

[In example 8, the switch from the frame narrative to the embedded narrative is included, but we never switch back to the frame narrative.]

(8, *inv.*) `<narrative_level number"1">` Dear sister, I confronted a strange person yesterday and heard a thrill story from him. `</narrative_levelchange complete="never closed">``<narrative_level number"2">` I created a monster who has already killed several people. `</narrative_level number"2">``</narrative_level number"1">`

[In example 9, the switch back from the embedded narrative to the frame narrative is included, but the text began with the embedded rather than the frame narrative.]

(9, *inv.*) `<narrative_level number="1">``<narrative_level number="2">` A flying elephant is playing with a pink monkey. `</narrative_level number"2">``</narrative_levelchangecomplete="never opened">` Mom, I had an interesting dream last night. `</narrative_level number="1">`

## 6. The nature of the boundary between the levels (Füredy)

### Theoretical Explanation<sup>13</sup>

This category is optional and should only be applied if there is at least one metalepsis (6.2) that can be clearly identified in a text.

#### 6.1 Strictly observed

Strict boundary between narrative levels. (<boundary transgression="no">)

EXPLANATION: This category is applied when the boundary between narrative levels are respected. Strictly put: It is applied when a metalepsis (6.2) does *not* occur and therefore *can not* be applied.

#### 6.2 Metalepsis<sup>14</sup>

EXPLANATION: A Metalepsis is identified according to Genette's terminology. Therefore, this category is only applied in instances where a transition between narrative levels can be identified and only if the following condition is fulfilled: Any intrusion by a narrator or narratee from outside of the particular narrative level that transgresses its internal logic. This can occur when an author (or his reader) introduces himself into the fictive action of the narrative, or when a character in a narrative intrudes into the narrative level of the author (or reader). Such intrusions disturb the distinction between levels.

#### 6.3 Pseudo-diegetic narration (cf. Genette: second-level narrative told as first-level narrative)<sup>15</sup>

"a narrative second in origin but which, lacking a diegetic relay, is narrated as though it were diegetic"<sup>16</sup>

## Categories, Attributes, Values

Category	Attribute	Possible Values
Boundary	Transgression	No Metalepsis Pseudo

<sup>13</sup>Viveca Füredy, "A Structural Model of Phenomena with Embedding in Literature and Other Arts," *Poetics Today* 10 (1989): 745-69.

<sup>14</sup>Genette, *Narrative Discourse*.

<sup>15</sup>Genette, *Narrative Discourse*.

<sup>16</sup>Pier, "Narrative Levels."

## Examples

(10) <boundary transgression="no">this is a passage with a strict boundary.  
</boundary>

(11) <boundary transgression="metalepsis">this is the passage with the metalepsis.  
</boundary>



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.



# Annotating Narrative Levels: Review of Guideline No. 6

Natalie M. Houston

01.15.20

*Article DOI: 10.22148/001c.11774*

*Journal ISSN: 2371-4549*

*Cite: Natalie M. Houston, "Annotating Narrative Levels: Review of Guideline No. 6," Journal of Cultural Analytics. January 15, 2020. doi: 10.22148/001c.11774*

The framing of Guideline VI within the pedagogical situation of a class on "Digital Methods in Literary Studies" is helpful in pointing out some of the ways in which the theory and practice of annotation can serve students of literature, as well as eventually contributing to computational analysis. Above all, annotation necessitates firm decisions, as the authors describe: "Rather than let an ambiguous text stay ambiguous, they simply *had* to decide for one option in order to be able to annotate a passage and had to justify their choice with reference to the whole text or to adapt the guidelines in order to address and document the ambiguity" (3). This remark highlights the challenge in developing annotation guidelines so that they can be used consistently by different communities of users without modifications. The authors note several points of debate within the class that are relevant to the overall shared task and its evaluation: the feasibility of developing annotation guidelines that could be applied to a wide range of literary texts; the involved levels of textual interpretation that some kinds of annotation require; and the effect of prior study or knowledge on an annotator's ability to discern or interpret narrative levels. As the shared task proceeds, it may be necessary to specify the applicability of the annotation guidelines to works from particular genres, time periods, or languages.

This guideline usefully deploys key concepts from narratological theory, including Genette's outline of narrative levels and types of narrators, Ryan's focus on the illocutionary and ontological boundaries, and Nelles's notion of the narratee.

Translating these theories into an outline of annotation choices is useful. But the textual examples that are provided are not sufficiently explained, and so do not adequately serve to guide a user's potential application of these categories to another text. Although Shelley's *Frankenstein* is a well-known novel, it presents several levels of narratological complexity that ought to be better explained if it is to be used as an exemplar. To readers who have not spent weeks immersed in the task of annotating *Frankenstein*, many of these examples of level change or narrator change will not be clear.

In particular, the following aspects of this Guideline could be revised for greater clarity and applicability:

(1) The authors state that they decided to annotate “the point between two different narrative levels,” yet the examples show the use of paired tags around sections of the narrative, which seems to contradict this statement.

(2) Some of the examples are drawn from *Frankenstein*, and others appear to be invented examples. The sources and selection of example texts should be clearly identified and explained.

(3) The examples are presented together as a block paragraph that often suggests they are taken sequentially from the novel, when in fact they are not. (i.e., #3 Narrator Change presents extracts from chapter 11 and then chapter 9). What are the closing tags closing? This formatting error can lead a reader unfamiliar with Shelley's novel to think all this text comes from one sequence in the novel.

(4) Most of the examples taken from *Frankenstein* are drawn from the opening lines of chapters or letters presented within chapters. No discussion is offered in the Guideline about whether or how particular attention ought to be paid to chapter or section boundaries in the literary text.

This Guideline also raises a theoretical point that ought to be considered in the shared task. The authors raise the question about the mediation of embedded narratives when they discuss whether the Creature should be considered the narrator of its own story, which is doubly mediated by other narrators in the novel *Frankenstein*. Many nineteenth-century novels, such as *The Moonstone* and *Dracula*, deliberately exploit textual mediation by presenting documents or letters that are ostensibly transcribed or read aloud by other characters. Whether such mediations constitute a form of narrative level or not should be considered in the annotation guidelines.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

# Annotation Guideline No. 7: Guidelines for annotation of narrative structure

Mats Wirén, Adam Ek and Anna Kasaty

01.15.20

Article DOI: [10.22148/001c.11772](https://doi.org/10.22148/001c.11772)

Journal ISSN: 2371-4549

Cite: Mats Wirén, Adam Ek and Anna Kasaty, “Annotation Guideline No. 7: Guidelines for annotation of narrative structure,” *Journal of Cultural Analytics*. January 15, 2020. doi: [10.22148/001c.11772](https://doi.org/10.22148/001c.11772)

## Rationale

### Background

Analysis of narrative structure can be said to answer the question “Who tells what, and how?”<sup>1</sup> The first part of the question thus concerns aspects such as who is narrating, whether it is a character in the story or not, and if it is a first-person or third-person narrator. The second part is related to the story and its basic elements: characters and events, and how the sequence of events forms a plot. The third part concerns how the narrative text is constructed: ordering of the events, the perspective from which the story is seen, how much information the narrator has access to, etc.

The key part of our annotation scheme is related to the “who?”, in other words, keeping track of who is doing the telling (showing, speaking), and to whom. To this end, our annotation scheme is grounded in the basic levels of narrative transmission: author-reader (highest level), narrator-narratee (intermediate level) and

---

<sup>1</sup>Manfred Jahn, “N2. The narratological framework”, in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018). Compare also the classification of “narration” (which is about the “who?”), “story” (“what?”) and “text” (“how?”) in Shlomith Rimmon-Kenan, *Narrative fiction: Contemporary poetics* (New York: Routledge, 2002).

character-character (lowest level).<sup>2</sup> Typically, works of literary fiction consist of alternations between the second and third of these levels: narration and fictional dialogue. As for narration, we annotate *voice*, which corresponds to the narrator's relationship to the story, and specifically whether the narrator is ever present in the story or not. The most detailed part of our annotation concerns the dialogue, however, for which we keep track of turns, lines, speakers and addressees.

Our annotation of the “what?” is rather rudimentary. We have a notion of *scene* which is meant to capture a coherent set of events and characters at a particular time and place, but no means of ordering sequences of scenes timewise or causally, and hence no notion of plot. Also, only characters that take part in fictional dialogue become part of the annotation.

Our annotation of the “how?” is also relatively rudimentary. We annotate *focalisation*, that is, the perspective from which the narrative is seen and specifically how much information the narrator has access to. Furthermore, we annotate *narrative levels* in the sense of stories within stories by using embeddings within the intermediate level of narrative transmission.

Our emphasis of the “who?” no doubt reflects the fact that we have approached the problem of narrative structure from linguistics. In particular, given the apparent resemblances between fictional dialogue and everyday conversation, it makes sense to apply linguistic models in the analysis of the former.<sup>3</sup> Suitable models that have been mentioned in the literature include speech-act theory and Conversation Analysis, but computational methods for the study of dialogue may also be useful. In contrast, the language of higher narrative-transmission levels can only “create an illusion, an effect, a semblance of mimesis,”<sup>4</sup> which means that the role of linguistic analysis is less straightforward.

In developing our annotation scheme, we have used the following criteria as guiding principles:

- *Simplicity and readability.* We have opted for a simple annotation scheme whose result should be easy to read together with the original texts. To this end, we use embedded (in-line) annotation in the original text.
- *Hierarchical tagset.* To increase inter-annotator agreement, we use a hierarchical tagset with mutually exclusive tag categories in the same layers.

---

<sup>2</sup>Manfred Jahn, “N2.4. Narrative Levels”, in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>3</sup>As pointed out by Aino Koivisto and Elise Nykänen, “Introduction: Approaches to fictional dialogue”, in *International Journal of Literary Linguistics* 5, no. 2 (2016), <http://www.ijll.uni-mainz.de/index.php/ijll/article/view/56> (accessed September 30, 2018).

<sup>4</sup>Shlomith Rimmon-Kenan, *Narrative fiction: Contemporary poetics* (New York: Routledge, 2002), 109.

- *Minimal interpretation.* Ideally, we would like an annotation which represents as objectively as possible the basic events, the characters involved and the discourse levels through which the narrative is transmitted, without undue subjective interpretation. This is arguably the most difficult principle to attain, however.
- *Relation to linguistic annotation.* We assume that a machine-learning model for predicting narrative structure will be based on annotation of both the narrative and linguistic structures of the text. The latter might build on what is produced by a standard automatic linguistic-analysis pipeline, involving, for example, sentence segmentation, tokenisation, part-of-speech tagging, named-entity recognition, syntactic parsing and co-reference resolution.<sup>5</sup> In contrast to narrative annotation, we consider linguistic annotation to be a means and not a goal in itself, and therefore do not include discussion of this here.

## Overview of annotation layers

Our tagset is hierarchically structured in four annotation layers, ordered by an inclusion relation. The top layer encodes *voice*, using the tags <VOICE\_1> and <VOICE\_3> that correspond to whether the narrator is present in the story or not, respectively. These are the opening tags; the corresponding closing tags are written with a slash, </VOICE\_1> and </VOICE\_3> (and similarly for other tags throughout). The second layer encodes focalisation, that is, the perspective from which the narrative is seen and how much information the narrator has access to. It includes the tags <FOC\_UNR>, <FOC\_INT> and <FOC\_EXT>, corresponding to unrestricted, internal and external focalisation, respectively. The fact that the annotation of focalisation is included in that of voice means that a change of voice requires resetting focalisation, even though the value of the focalisation may not change.

A focalisation includes one or more scenes in the third layer, <SCENE>, each of which is a coherent set of events at a particular interval in time and place, with a more or less constant set of characters. A scene typically consists of alternations between narration and dialogue, annotated using <NARRATOR> and <CHARACTERS>, but may also contain <AUTHOR>. In addition, these types of discourses may be embedded in each other (to be detailed below).

---

<sup>5</sup>For example, Stanford CoreNLP (for English), <https://stanfordnlp.github.io/CoreNLP/> (accessed September 30, 2018); efselab (for Swedish), <https://github.com/robertostling/efselab> (accessed September 30, 2018).

A characters' discourse, <CHARACTERS>, consists of one or more turns, each of which is associated with one speaker (or a chorus of several speakers) and one or more addressees (who may vary). Thus, whereas the speaker (or set of speakers) is immutable throughout a turn, this does not necessarily hold for the addressee(s). Finally, each turn consists of one or more lines, each of which is associated with one speaker (or a chorus of several speakers) and one addressee (or a set of addressees). In addition, a line consists of one or more utterances, that is, sentences or fragments typically distinguished by full stops in the text. Utterances are not part of the annotation, however. Note that only lines are tagged with speaker(s) and addressee(s), and that there is no notion of opening or closing tags in this case. Note also that we make a difference between addressees and listeners: addressees are the recipients of lines, whereas listeners overhear lines without being recipients. We only annotate addressees. In sum, a characters' discourse is a sequence of turns uninterrupted by narrator's discourses (unless they are embedded). A description of the layers and discourse levels is shown in Table 1.

Layer	Tag	Description
1	<VOICE_1>, <VOICE_3>	Narrator's presence in the story (yes, no)
2	<FOC_UNR>, <FOC_INT>, <FOC_EXT>	Perspective of the narrator (unrestricted, internal, external)
3	<SCENE>	Coherent set of events and characters at particular interval of time and place
4.1	<AUTHOR>	Narrative transmission: Highest level
4.2	<NARRATOR>	Narrative transmission: Intermediate level
4.3	<CHARACTERS>	Narrative transmission: Lowest level
4.3.1	<TURN>	Turn: One or several lines
4.3.2	<Speaker-Addressee>	Line: One or several utterances. Tagged only with speaker(s) and addressee(s)

Table 1. Hierarchical structure of the annotation scheme.

We use a deliberately simple criterion for delimiting the scope of a narrator's discourse, namely, letting each paragraph that begins with narration correspond to one narrator's discourse by enclosing it with opening and closing tags (<NARRATOR> and </NARRATOR>). As for dialogue, each turn is often put in a paragraph of its own in fictional works. But when we have a sequence of turns, whether each line is in a paragraph of its own or not, we let the entire sequence be enclosed by the corresponding opening and closing tags, that is, <CHARACTERS> and </CHARACTERS>.

## Discourse levels

### Introduction

We assume that a text in its entirety can be divided into occurrences of the following three discourse levels:<sup>6</sup>

- Highest level: Transmission from the author to a (typically) imagined, but explicitly referred reader of the work (for an example, see Section Transmission across levels). We refer to this as *author's discourse*.
- Intermediate level: Transmission from a narrator to a narratee. The latter can be visible or invisible, but is distinct from (the imagined) reader of the work. We refer to this as *narrator's discourse*.
- Lowest level: Transmission between characters in the story in the form of dialogue, whether direct or indirect discourse.<sup>7</sup> We refer to this as *characters' discourse*. The dialogue is typically spoken, but we also take soliloquy and interior monologue to be possible types of characters' discourse.<sup>8</sup>

Assuming that transmission between author and reader is the exception in fiction, a narrative typically consists of passages alternating between the two lower levels.<sup>9</sup>

To annotate discourses in the three levels, we use opening and closing tags indicating both the type of speaker and addressee at the respective level, namely, <AUTHOR-READER>, <NARRATOR-NARRATEE> and <CHARACTERS-CHARACTERS>. For convenience, and when there is no ambiguity, we may abbreviate <NARRATOR-NARRATEE> as <NARRATOR> and <CHARACTERS-CHARACTERS> as <CHARACTERS>. An example text with alternating narrator's and characters' discourses is shown in Example (1).<sup>10</sup> (Note that this author uses neither dashes nor quotation marks, but only paragraph breaks, to indicate turns.)

- (1) Melissa had her camera on the table and occasionally lifted it to take a photograph, laughing self-deprecatingly about being a 'work addict'. She

<sup>6</sup>Manfred Jahn, "N2.3. Narrative communication", in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>7</sup>Manfred Jahn, "N4", in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>8</sup>For a listing of different representations of speech, see Shlomith Rimmon-Kenan, *Narrative fiction: Contemporary poetics* (New York: Routledge, 2002).

<sup>9</sup>As stated in Lubomír Doležel, *Narrative Modes in Czech Literature* (Toronto: University of Toronto Press, 1973), quoted in Manfred Jahn, "N8.1", in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018): "Every narrative text T is a concatenation and alternation of [narrator's discourse] and [characters' discourse]".

<sup>10</sup>From Sally Rooney, *Conversations with Friends* (London: Faber & Faber, 2017), 4-5.



lit a cigarette and tipped the ash into a kitschy-looking glass ashtray. The house didn't smell of smoke at all and I wondered if she usually smoked in there or not.

I made some new friends, she said.

Her husband was in the kitchen doorway. He held up his hand to acknowledge us and the dog started yelping and whining and running around in circles.

This is Frances, said Melissa. And this is Bobbi. They're poets.

He took a bottle of beer out of the fridge and opened it on the countertop.

Come and sit with us, Melissa said.

Yeah, I'd love to, he said, but I should try and get some sleep before this flight.

Example (2) shows how (1) would be annotated for narrator's and characters' discourses. The latter includes turns, whereas lines are introduced further below. Vertical space corresponds to alternations of discourse level for readability.

(2) <NARRATOR>

Melissa had her camera on the table and occasionally lifted it to take a photograph, laughing self-deprecatingly about being a 'work addict'. She lit a cigarette and tipped the ash into a kitschy-looking glass ashtray. The house didn't smell of smoke at all and I wondered if she usually smoked in there or not.

</NARRATOR>

<CHARACTERS>

<TURN>

I made some new friends, she said.

</TURN>

</CHARACTERS>

<NARRATOR>

Her husband was in the kitchen doorway. He held up his hand to acknowledge us and the dog started yelping and whining and running around in circles.

</NARRATOR>

<CHARACTERS>

<TURN>

This is Frances, said Melissa. And this is Bobbi. They're poets.

</TURN>

</CHARACTERS>

<NARRATOR>

He took a bottle of beer out of the fridge and opened it on the countertop.

</NARRATOR>

<CHARACTERS>

<TURN>

Come and sit with us, Melissa said.

Yeah, I'd love to, he said, but I should try and get some sleep before this flight.

</TURN>

</CHARACTERS>

## Characters' discourses

Whereas narrator's discourses capture narrations in the text, characters' discourses corresponds to fictional dialogues between characters. A characters' discourse consists of one or more turns, each of which consists of one or more lines as defined above. A characters' discourse is thus a sequence of turns uninterrupted by narrator's discourses.

To represent transmission between characters, we annotate each line with its speaker and addressee, as in example (3).<sup>11</sup> Specifically, we assume that a turn has a single speaker, but that different lines within a turn may have different addressees. Also, a speaker may address more than one character simultaneously, which means that one line can have several addressees. We only annotate the intended recipient(s) of the utterance as addressees, whereas listeners who only overhear a line are not annotated.

(3) <CHARACTERS>

<TURN>

We're all on the same side here, Derek said. <Derek-ALL>

<sup>11</sup>From Sally Rooney, *Conversations with Friends* (London: Faber & Faber, 2017), 112.

Nick, you're an oppressive white male, you back me up. <Derek–Nick>

</TURN>

<TURN>

I actually quite agree with Bobbi, said Nick. Oppressive though I certainly am.

<Nick –Derek>

</TURN>

</CHARACTERS>

The first turn is divided into two lines since there is a change in addressees. The second turn consists of one line which consists of two utterances.

Note that we do not have opening and closing tags that surround lines, but just a speaker—addressee tag at the end of each line. The scope of the speaker—addressee tag is the current line.

What we refer to as a line consists of a line proper, which is what is actually spoken by the character, and optionally a speech-verb construction, which indicates who the speaker (and possibly the addressee) is. Example (4) shows a line proper with the direct speech of the officer, whereas Example (5) shows the speech-verb construction that identifies the speaker.

(4) “That was the first story,”

(5) “said the officer,”

We have chosen to include the speech-verb construction in the line tag to avoid cluttering the annotation, and because these constructions follow a predictable pattern.

In the case where a line has multiple addressees and (some of) these are not identifiable, this is annotated using the keyword SEVERAL. This is shown in Example (6):<sup>12</sup>

(6) <CHARACTERS>

<TURN>

“That was the first story,” said the officer,

<Officer–SEVERAL>

</TURN>

<sup>12</sup>From Heinrich von Kleist, *Improbable Veracities*, <https://www.gutenberg.org/> (accessed September 30, 2018).

<NARRATOR>

as he took a pinch of snuff and became silent.

</NARRATOR>

</CHARACTERS>

Example (6) also shows how a narrative construction within a line is represented as an embedded narrator's discourse (further detailed below).

In contrast, if there are multiple identifiable addressees, these are listed within parentheses as exemplified below:<sup>13</sup>

(7) <CHARACTERS>

<TURN>

'Four reales.' 'We want two Anis del Toro.' <Man–Woman>

</TURN>

<TURN>

'With water?' <Woman–(Man,Girl)>

</TURN>

[...]

</CHARACTERS>

A line may also have multiple speakers talking in chorus. This is annotated with a parenthesised expression as in Example (8), analogously with multiple addressees.<sup>14</sup>

(8) <CHARACTERS>

<TURN>

'Now watch,' said the Zebra and the Giraffe. 'This is the way it's done. One–two–three! And where's your breakfast?' <(Zebra, Giraffe)–Leopard>

</TURN>

<sup>13</sup>From Ernest Hemingway, *Hills Like White Elephants*, <https://www.gutenberg.org/> (accessed September 30, 2018).

<sup>14</sup>From Rudyard Kipling, *How the Leopard got his Spots*, <https://www.gutenberg.org/> (accessed September 30, 2018).

</CHARACTERS>

## Embeddings

### Characters' discourse embedded in characters' discourse

Discourse levels can be embedded into each other. In particular, when a character is quoting or recounting a dialogue with someone else, this is represented by embedding that characters' discourse into the present one. This is annotated as an additional opening of CHARACTERS inside the present characters' discourse, as in Example (9).<sup>15</sup>

(9) <CHARACTERS>

<TURN>

Did you ever get that thing with the car sorted? Nick said to Evelyn. <Nick–Evelyn>

</TURN>

<TURN>

No, Derek won't let me talk to the dealership, she said. He's <Evelyn–Nick>

<CHARACTERS>

<TURN>

'taking care of it'. <Derek–Evelyn>

</TURN>

</CHARACTERS>

</TURN>

</CHARACTERS>

Similarly:<sup>16</sup>

(10) <CHARACTERS>

<TURN>

I think your wife is a little on edge today, said Bobbi. <Bobbi–Nick>

<sup>15</sup>From Sally Rooney, *Conversations with Friends* (London: Faber & Faber, 2017), 138.

<sup>16</sup>From Rooney, Sally, *Conversations with Friends* (London: Faber & Faber, 2017), 145.

She was not impressed with my linen-folding technique earlier. Also, <Bobbi-Nick>

<CHARACTERS>

<TURN>

she told me she didn't want me 'making any snide remarks about rich people' when Valerie gets here. Quote. <Melissa-Bobbi>

</TURN>

</CHARACTERS>

</TURN>

</CHARACTERS>

### **Narrator's discourse embedded in characters' discourse**

Elements of narrator's discourse may be interspersed, typically in a fragmentary way, inside lines in a characters' discourse without breaking the flow of the dialogue. We represent this by an embedding of the narrator's discourse in the present characters' discourse.

Example (11) illustrates this, where the line " 'It tastes like liquorice,' the girl said" is followed by the narrator's description, "and put the glass down", which is thus embedded within the characters' discourse.<sup>17</sup>

(11) <CHARACTERS>

<TURN>

'Four reales.' <Man-Girl>

'We want two Anis del Toro.' <Man-Woman>

</TURN>

<TURN>

'With water?' <Man-Woman>

</TURN>

<TURN>

---

<sup>17</sup>From Ernest Hemingway, *Hills Like White Elephants*, <https://www.gutenberg.org/> (accessed September 30, 2018).

'Do you want it with water?' <Man-Girl>  
</TURN>  
<TURN>  
'I don't know,' the girl said. 'Is it good with water?' <Girl-Man>  
</TURN>  
<TURN>  
'It's all right.' <Man-Girl>  
</TURN>  
<TURN>  
'You want them with water?' asked the woman. <Woman-(Man,Girl?)>"  
</TURN>  
<TURN>  
'Yes, with water.' <Man-Woman>  
</TURN>  
<TURN>  
'It tastes like liquorice,' the girl said <Girl-Man>  
<NARRATOR>  
and put the glass down.  
</NARRATOR>  
</TURN>  
<TURN>  
'That's the way with everything.' <Man-Girl>  
</TURN>  
</CHARACTERS>

Note that, in accordance with what was said at the end of the introduction section, an embedding is a change of discourse level within another level. This means that if a paragraph begins with dialogues and ends with a narration, the narration is embedded in the character discourse.

### Characters' discourse embedded in narrator's discourse

We represent *indirect discourse* - broadly speaking, dialogue rendered narratively - by embedding a characters' discourse into the present narrator's discourse.<sup>18</sup> Example (12) shows two subsequent characters' discourses being embedded in one narrator's discourse, which is in turn followed by another characters' discourse.<sup>19</sup>

(12) <NARRATOR>

The car had been sitting in the sun all morning and we had to roll the windows down before we could even get in. Inside it smelled like dust and heated plastic. I sat in the back and Bobbi leaned her little face out the passenger window like a terrier. Nick switched on the radio and Bobbi withdrew her face from the window

<CHARACTERS>

<TURN>

and said, do you not have a CD player? <Bobbi-Nick>

Can we listen to music? <Bobbi-Nick>

</TURN>

<TURN>

Nick said: sure, okay. <Nick-Bobbi>

</TURN>

</CHARACTERS>

Bobbi started looking through the CDs then

<CHARACTERS>

<TURN>

and saying whether she thought they were his or Melissa's. <Bobbi-Nick>

</TURN>

</CHARACTERS>

</NARRATOR>

<CHARACTERS>

<sup>18</sup>Manfred Jahn, "N8.4", in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>19</sup>From Sally Rooney, *Conversations with Friends* (London: Faber & Faber, 2017), 106.



<TURN>

Who likes Animal Collective, you or Melissa? she said. <Bobbi–Nick>

</TURN>

<TURN>

I think we both like them. <Nick–Bobbi>

</TURN>

<TURN>

But who bought the CD? <Bobbi–Nick>

</TURN>

<TURN>

I don't remember, he said. You know, we share those things, I don't remember whose is whose. <Nick–Bobbi>

<TURN>

</CHARACTERS>

In sum, by embedding the characters' discourse in a narrator's discourse, we represent the fact that it's being rendered indirectly through the narration, and not directly as in a (non-embedded) characters' discourse appearing at the (top) level of alternating narrator's and characters' discourses. The reason that we still represent this using an (embedded) characters' discourse is that we want to capture all transmission between the characters, whether it occurs directly at the lowest discourse level or is rendered indirectly at the intermediate level.

### **Narrator's discourse embedded in narrator's discourse**

A narrator's discourse embedded in a narrator's discourse corresponds to what has been called *narrative level*, in other words, a story within a story.<sup>20</sup> Example (13) shows a story imagined by the character within the main story.<sup>21</sup>

(13) <NARRATOR>

<sup>20</sup>Genette, Gérard, *Narrative Discourse: An Essay in Method* (Ithaca, N.Y.: Cornell University Press, 1983), 189; Jahn, Manfred, *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>21</sup>From Anton Chekov, *The Avenger*, <https://www.gutenberg.org/> (accessed September 30, 2018).

The shopman, swaying gracefully and tripping to and fro on his little feet, still smiling and chattering, displayed before him a heap of revolvers.

The most inviting and impressive of all was the Smith and Wesson's.

Sigaev picked up a pistol of that pattern, gazed blankly at it, and sank into brooding.

<NARRATOR>

His imagination pictured how he would blow out their brains, how blood would flow in streams over the rug and the parquet, how the traitress's legs would twitch in her last agony....

</NARRATOR>

But that was not enough for his indignant soul.

The picture of blood, wailing, and horror did not satisfy him.

He must think of something more terrible.

</NARRATOR>

In the Example (14), the narrator addresses the reader and tells something about the main story.<sup>22</sup>

(14) <NARRATOR>

I am the man who comes and goes between the bar and the telephone booth.

<NARRATOR-READER>

Or, rather: that man is called "I" and you know nothing else about him, just as this station is called only "station" and beyond it there exists nothing except the unanswered signal of a telephone ringing in a dark room of a distant city.

</NARRATOR-READER>

I hang up the receiver, I await the rattling flush, down through the metallic throat, I push the glass door again, head toward the cups piled up to dry in a cloud of steam.

</NARRATOR>

---

<sup>22</sup>From Italo Calvino, *If on a Winter's Night a Traveler*, <https://www.gutenberg.org/> (accessed September 30, 2018).

### An example of multiple embeddings

There may be multiple layers of discourse level embeddings. The following example illustrates this.<sup>23</sup> In Example (15), the narrator's discourse has an embedded characters' discourse, which in turn has an embedded narrator's discourse.

(15) <NARRATOR>

But it wasn't that you woke us.

Oh, no.

<CHARACTERS>

<TURN>

'They're looking for it; they're drawing the curtain,' one might say,

</TURN>

<NARRATOR>

and so read on a page or two.

</NARRATOR>

</CHARACTERS>

[...]

</NARRATOR>

### Transmission across levels

Transmission typically occurs within a single discourse level, but it may also cut across levels,<sup>24</sup> a phenomenon referred to as *metalepsis*.<sup>25</sup> For example, the narrator may explicitly address the (supposed) reader. We annotate this by including the addressee explicitly (here READER) in the discourse level tag (<NARRATOR>) as in Example (16).

(16) <NARRATOR-READER>

<sup>23</sup>From Virginia Woolf, *A Haunted House*, <https://www.gutenberg.org/> (accessed September 30, 2018).

<sup>24</sup>Manfred Jahn, *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>25</sup>Gérard Genette, *Narrative Discourse: An Essay in Method* (Ithaca, N.Y.: Cornell University Press, 1983).

You are about to begin reading Italo Calvino's new novel, *If on a winter's night a traveler*. Relax. Concentrate. Dispel every other thought. Let the world around you fade. Best to close the door; the TV is always on in the next room. [...]

</NARRATOR-READER>

We do not worry about the ontological status of an explicitly addressed reader, for example, whether it is an implied reader or an identifiable physical person.

We distinguish this from the case when a narratee is addressed, typically using second-person pronouns, but no explicit reference to a reader or an act of reading is being made, as in Example (17).<sup>26</sup>

(17) <NARRATOR>

If you really want to hear about it, the first thing you'll probably want to know is where I was born, and what my lousy childhood was like, and how my parents were occupied and all before they had me, and all that David Copperfield kind of crap, but I don't feel like going into it, if you want to know the truth. [...]

</NARRATOR>

Here, it might be tempting to think of "you" as an imagined reader, but it might just as well be an (imagined, third-person) listener to whom Holden Caulfield is telling his story. In the absence of information linking "you" with a reader, we assume the latter possibility. <NARRATOR> is thus the short form for <NARRATOR-NARRATEE>.

## Scenes

To represent the basic progression of events in a narrative, we use a notion of *scenes*, inspired from film. We take a scene to be a coherent set of events at a particular interval in time and place, with a more or less constant set of characters. Furthermore, we take a scene to be the minimal unit in this respect, anticipating that we will later be able to use scenes as primitives in higher-level structures, such as plot. Consider Example (18) about the protagonist of a novel (Frances) who is seeing her father.<sup>27</sup>

(18)           After dinner I told my mother I would visit him. She kneaded my shoulder and told me she thought it was a good idea. It's a great idea, she said. Good woman.

<sup>26</sup>From J. D. Salinger, *The Catcher In The Rye* (Boston: Little, Brown and Company, 1951), where this is a frequent phenomenon.

<sup>27</sup>From Sally Rooney, *Conversations with Friends* (London: Faber & Faber, 2017), 51-52.

I walked through town with my hands in my jacket pockets. The sun was setting and I wondered what would be on television. I could feel a headache developing, like it was coming down from the sky directly into my brain. I tried stamping my feet as loudly as I could to distract myself from bad thoughts, but people gave me curious looks and I felt cowed. I knew that was weak of me. Bobbi was never cowed by strangers.

My father lived in a little terraced house near the petrol station. I rang the doorbell and put my hands back in my pockets. Nothing happened. I rang again and then I tried the handle, which felt greasy. The door opened up and I stepped in.

Dad? I said. Hello?

[...]

I'm off, I said.

You're away, are you?

That bin needs taking out.

See you again, my father said.

Our basic criterion for dividing a narrative into scenes is to think about it in terms of a faithful rendering as a film or a play. In the example above, we would obtain three scenes: The first one includes Frances and her mother at her mother's place talking to each other, the second is Frances's walking on her own to her father's place, and the third one includes Frances and her father's meeting at her father's place (further developed in the story). We would thus annotate the passage above as in Example (19).

(19) <SCENE>

<NARRATOR>

After dinner I told my mother I would visit him. She kneaded my shoulder

<CHARACTERS>

<TURN>

and told me she thought it was a good idea. <Mother–Frances>

It's a great idea, she said. <Mother–Frances>

Good woman. <Mother–Frances>

</TURN>

</CHARACTERS>

</NARRATOR>

</SCENE>

<SCENE>

<NARRATOR>

I walked through town with my hands in my jacket pockets. The sun was setting and I wondered what would be on television. I could feel a headache developing, like it was coming down from the sky directly into my brain. I tried stamping my feet as loudly as I could to distract myself from bad thoughts, but people gave me curious looks and I felt cowed. I knew that was weak of me. Bobbi was never cowed by strangers.

</NARRATOR>

</SCENE>

<SCENE>

<NARRATOR>

My father lived in a little terraced house near the petrol station. I rang the doorbell and put my hands back in my pockets. Nothing happened. I rang again and then I tried the handle, which felt greasy. The door opened up and I stepped in.

</NARRATOR>

<CHARACTERS>

<TURN>

Dad? I said. Hello? <Frances–Dad> [...]

</TURN>

</CHARACTERS>

[...]

<CHARACTERS>

<TURN>

I'm off, I said. <Frances–Dad>

You're away, are you? <Dad–Frances>

That bin needs taking out. <Frances–Dad>

See you again, my father said. <Dad–Frances>

</TURN>

</CHARACTERS>

</SCENE>

In this example, there are three scenes. The first scene covers the first <NARRATOR> and <CHARACTERS> elements. The second scene covers the second <NARRATOR> element, and the third scene covers the third <NARRATOR> element and the second and third <CHARACTERS> elements.

## Narrative situation

This section describes two notions related to the position and perspective of the narrator. We refer collectively to these as *narrative situation*.<sup>28</sup>

### Voice

The notion of *voice* concerns the narrator's relationship to the story, and more specifically whether the narrator is ever present in the story or not.<sup>29</sup> If the narrator appears in the story at some point, we say that we have a *homodiegetic* narrative. Such narrators usually refer to themselves in the first person, but there are exceptions to this, such as Caesar's *De Bello Gallico* in which the narrator refers to himself in the third person. In contrast, if the narrator is never present in the story, we say that we have a *heterodiegetic* narrative. Such narrators usually refer to themselves in the third person, but again there are exceptions.

We annotate this binary distinction using <VOICE\_1> for a homodiegetic narrator and <VOICE\_3> for a heterodiegetic narrator, with corresponding closing tags </VOICE\_1> and </VOICE\_3>, respectively.

### Focalisation

We take *focalisation* to correspond to the perspective from which the narrative is seen, and specifically how much information the narrator has access to; alterna-

<sup>28</sup>Inspired by Gérard Genette, *Narrative Discourse: An Essay in Method* (Ithaca, N.Y.: Cornell University Press, 1983), 188, but we do not use the term in the full meaning developed there.

<sup>29</sup>Gérard Genette, *Narrative Discourse: An Essay in Method* (Ithaca, N.Y.: Cornell University Press, 1983), Chapter 5.

tively, in what ways this information is restricted.<sup>30</sup> We distinguish the following types:

- *Zero* or *unrestricted* The story is narrated from a fully unrestricted or omniscient perspective. This often involves helicopter views of the story that no single character would be capable of, but it could also involve taking the perspectives or looking into the souls of the individual characters. The narrator knows more than any of the characters, symbolised by *Narrator* > *Character*.<sup>31</sup> We annotate this as <FOC\_UNR> with a corresponding closing tag </FOC\_UNR>.
- *Internal* The story is narrated from the inside perspective of a character in the story, limited by the perception and feelings of that character. The narrator knows only as much as this character, symbolised by *Narrator* = *Character*.<sup>32</sup> We annotate this as <FOC\_INT> with a corresponding closing tag </FOC\_INT>.
- *External* The story is narrated from a perspective outside of the characters in the story, like using a camera, but without an omniscient perspective, as in <FOC\_UNR>. Typically, the main components of such narratives are dialogues and narrations in the form of neutral descriptions of events. The narrator knows less than any of the characters, symbolised by *Narrator* < *Character*.<sup>33</sup> We annotate this as <FOC\_EXT> with a corresponding closing tag </FOC\_EXT>.

## Definitions

We end with a list of definitions of some of our central terminology.

**Author's discourse:** The highest level of transmission in a narrative, from the author to a (typically) imagined, but explicitly referred reader of a story. Consists of text.

---

<sup>30</sup>G rard Genette, *Narrative Discourse: An Essay in Method* (Ithaca, N.Y.: Cornell University Press, 1983), 189.

<sup>31</sup>As put by Tzvetan Todorov, "Les cat gories du r cit litt raire", in *Communications*, 8 (1966): 125-151, cited in G rard Genette, *Narrative Discourse: An Essay in Method* (Ithaca, N.Y.: Cornell University Press, 1983), 188.

<sup>32</sup>As put by Tzvetan Todorov, "Les cat gories du r cit litt raire", in *Communications*, 8 (1966): 125-151.

<sup>33</sup>As put by Tzvetan Todorov, "Les cat gories du r cit litt raire", in *Communications*, 8 (1966): 125-151.



**Characters' discourse:** The lowest level of transmission in a narrative, from character to character. Consists of one or more turns.

**Dialogue:** The text corresponding to a characters' discourse. This is not limited to spoken dialogue, but could also be soliloquy, interior monologue, thoughts, etc.

**Layer:** This refers to the overall annotation, which is ordered by an inclusion relation in four hierarchical layers.

**Level:** This refers to the type of narrative transmission in the fourth annotation layer. We distinguish between the highest level (author's discourse), the intermediate level (narrator's discourse) and the lowest level (characters' discourse).

**Narration:** The text corresponding to a narrator's discourse.

**Narrator:** The teller of the narrative; the person who articulates ("speaks") the narrative.<sup>34</sup>

**Narrator's discourse:** The intermediate level of transmission in a narrative, from narrator to narratee. Consists of text.

**Narrative:** Anything that tells or presents a story.<sup>35</sup>

**Narrative situation:** A collective name for voice and focalisation.<sup>36</sup>

**Story:** A sequence of events involving characters.<sup>37</sup>



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

---

<sup>34</sup>Quoted from Manfred Jahn, "N1.2", in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>35</sup>Quoted from Manfred Jahn, "N1.2", in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

<sup>36</sup>Note that our use of the term does not capture the full meaning of that used by Gérard Genette, *Narrative Discourse: An Essay in Method* (Ithaca, N.Y.: Cornell University Press, 1983), 188.

<sup>37</sup>Quoted from Manfred Jahn, "N1.2", in *Narratology: A guide to the theory of narrative*, <http://www.uni-koeln.de/~ame02/pppn> (accessed September 30, 2018).

# Annotating Narrative Levels: Review of Guideline No. 7

Gunther Martens

01.15.20

*Article DOI: 10.22148/001c.11775*

*Journal ISSN: 2371-4549*

*Cite: Gunther Martens, "Annotating Narrative Levels: Review of Guideline No. 7," Journal of Cultural Analytics. January 15, 2020. doi: 10.22148/001c.11775*

The guideline under review builds on the acquired knowledge of the field of narrative theory. Its main references are to classical structuralist narratology, both in terms of definitions (Todorov, Genette, Dolezel) and by way of its guiding principles, which strive for simplicity, hierarchy, minimal interpretation and a strict focus on the annotation of text-intrinsic, linguistic aspects of narrative. Most recent attempts to do "computational narratology" have been similarly "structuralist" in outlook, albeit with a stronger focus on aspects of *story grammar*: the basis constituents of the story are *to some extent hard-coded into the language of any story*, and are thus more easily formalized. The present guideline goes well beyond this restriction to story grammar. In fact, the guideline promises to tackle aspects of narrative transmission from the highest level (author) to the lowest (character), but also demarcation of scenes at the level of plot, as well as focalisation. Thus, the guideline can be said to be very wide in scope.

The shared task to which this guideline responds focuses on identifying and reaching consensus on the demarcation of narrative *levels*. In standard narratological parlance, shifts in level correlate to shifts in the information distribution from one narrative agent to another. In keeping with film terminology, these acts, including of the act of taking charge of the narration itself, are taken to be acts of framing constitutive of distinctive levels: "We will define this difference in level by saying that any event a narrative recounts is at a diegetic level immediately higher than the level at which the narrating act producing

this narrative is placed.”<sup>1</sup> In Genette’s view, narrative levels lead to an intricate nesting or embedding effect of speakers and viewers.

While the more comprehensive approach of the guideline will be more palatable to scholars trained in literary theory, it is to some extent undecided as to what it takes “levels” to mean. Though the guideline addresses a broad set of narrative features, it is ultimately geared towards annotating the most conspicuous shifts in narrative levels: the turn-taking in dialogues between characters and switches in voice from narrator to character and vice versa. This is certainly the part of the guideline most easily to be operationalized. It should be pointed out that the guideline chose to restrict its interaction with the shared task corpus to a minimum: only three of the texts are briefly cited, and the bulk of the examples stems from Sally Rooney’s novel *Conversations with friends*. It is stated that: “The main components of such narratives are dialogues”, which may help to explain why the annotation schema is more focused on reported speech than on reported thought.

While the current guideline takes its cue mainly from the tried-and-trusted toolkit of (textual) narrative theory, it is also informed by Digital Humanities. This can be seen when aspects of the paratext (Genette’s short-hand notation for any extra-textual element that frames texts and guides their reception) are taken into account, for instance when the typographic make-up of chapters, paragraphs and quotation is considered as a machine-readable index of narrative levels. Likewise, aspects of the guideline go beyond structuralism when it undertakes to consider narratees and addressees. This extension of the narratological toolbox is in keeping with recent redefinitions of style in the area of Digital Humanities, as epitomized by the following definition:

In Digital Humanities, ›style‹ is seen as anything that can be measured in the linguistic form of a text, such as vocabulary, punctuation marks, sentence length, word length, the use of character strings.<sup>2</sup>

The adoption of this line of reasoning becomes evident when the guideline draws on the layout of the texts: “alternations between discourse levels are usually signalled by paragraph breaks.” It is certainly necessary and helpful to consider such material underpinnings of narrative structure. Yet, there is a wide variety in national and historical print cultures to be considered in this regard, so these apparently stable markers of narrative level should be handled with care and flexibility.

---

<sup>1</sup>G rard Genette, *Narrative Discourse: An Essay in Method* (Ithaca: Cornell University Press, 1983), 228.

<sup>2</sup>J. Berenike Herrmann, Karina van Dalen-Oskam, and Christof Sch och, “Revisiting Style, a Key Concept in Literary Studies,” *Journal of Literary Theory* 9, no. 1 (2015): 25-52.

The guideline claims that it seeks to make the annotation amenable to machine learning so as to “predict narrative structure”. While this is certainly a laudable ambition, it remains to be seen whether the guideline’s heuristic focus actually allows for this. The current guideline is rather hybrid in nature. On the one hand, it caters to the hermeneutic strengths of human annotators. Especially the attempt to annotate the addressee(s) of specific utterances presupposes a lot of interpretation, as it hinges on implication and logical deduction rather than on actual mentions. Likewise, the guidelines for annotating focalisation strike me as undecided. The main reference here is Todorov, which is somewhat dated in view of the lengthy debates on various conceptualisations of focalisation and the question of its transferability to specific media. Focalisation is restricted to “perspective of the narrator”. It would seem that even more semantics would be required to demarcate other types of focalisation. The ambition to cover these areas may run counter to the manual’s declared adherence to structuralist tenets, as both rely on interpretation and semantics. Co-reference resolution of unstructured textual data (like fictional narratives) is notoriously difficult.<sup>3</sup> Currently, automatic event detection on the basis of machine learning has proven most successful with regard to text genres that involve a lot of referential anchoring (e.g. news articles).<sup>4</sup> The current state-of-the-art allows machine learning to predict structure “in the wild” only over a limited span of semi-structured text.<sup>5</sup> Annotating the intricacies of implied audiences presupposes an even more extensive degree of co-reference resolution.

I would like to take issue with another specific decision: The guideline argues in favour of handling tags as cleanly as possible, in order to provide a visual analogy to levels that it demarcates. For instance, it encloses the markers that attribute discourse to specific characters within the tags that demarcate that very content. These attributive markers typically involve *verba dicendi* in the so-called inquit-formulae. The main rationale for “includ[ing] the speech-verb construction in the line tag” is “to avoid cluttering the annotation”. I am not convinced that this is a workable decision. This might seem to be an issue of lesser importance with regard to texts that keep this attributive marking to an absolute minimum, as is the case in the samples from the contemporary novel. Yet, if the focus of the shared task is indeed on identifying levels in a wide range of narrative texts, this decision is counterproductive. It undermines the attempt to identify levels and,

---

<sup>3</sup>“R/ProgrammerHumor - When Do We Want What?,” reddit, accessed June 3, 2019.

<sup>4</sup>Tommaso Caselli and Oana Inel, “Crowdsourcing StoryLines: Harnessing the Crowd for Causal Relation Annotation,” in *Proceedings of the Workshop Events and Stories in the News 2018*, 2018, 44-54.

<sup>5</sup>Markus Krug et al., “Rule-Based Coreference Resolution in German Historic Novels,” in *CLfL@NAACL-HLT*, 2015, 98-104; S. Malec et al., “Landing Propp in Interaction Space: First Steps Toward Scalable Open Domain Narrative Analysis With Predication-Based Semantic Indexing,” in *DIVA*, 2015.

especially to extricate from sentences chunks that allow machines to identify patterns indicative of shifts in level. While the concatenation of discourse with discourse markers is in line with a fairly recent trend in postclassical narratology, as I discussed elsewhere,<sup>6</sup> it would seem that these tags are kept to a minimum for the sake of human readability. Chunking at higher-order levels such as scenes is not necessarily the way to go when aiming for machine readability. In order to annotate narrative levels, it is mandatory to provide tagging at the micro-level of words rather than of sentences, paragraphs or even scenes. This will inevitably lead to a cluttered view to the human eye, but such a nesting of annotations is much more likely to lead to transfer learning. Much more meta-information is needed with regard to the framing verbs. These tags could then be linked with existing tag-sets that deliberately aim to target and/or attenuate contextual ambiguity, such as PropBank and FrameNet. Similar efforts are under way. A brief look at [www.redewiedergabe.net](http://www.redewiedergabe.net) might suffice to illustrate what such micro-coding may afford in terms of the detection of narrative levels.<sup>7</sup>

It is certainly laudable that the guidelines undertakes to emulate the structuralist annotation of complex aspects of narrative levels. It remains to be seen whether the textualist and bottom-up focus of this guideline warrants for a basis representative enough to provide a gold standard in order to extrapolate from. Granted, this is a dilemma that currently most attempts at doing computational narratology with roots in literary narrative theory are facing. While the adherence of the guideline to structuralist tenets can be lauded for its principled nature, there is much to be learned from the extension of the narratological toolkit in the direction of multimodality and paralinguistics. While references to time and co-reference can be resolved with a high degree of confidence in formulaic genres like news articles or scientific articles, especially co-reference resolution in elliptic fictional texts like Virginia Woolf's can probably only be solved by looking at interactions of readers and other users with the text (e.g. through eye tracking<sup>8</sup> or the study of adaption in other media<sup>9</sup>). Notwithstanding the many conceptual challenges of doing transmedia comparisons, one may profit from compar-

---

<sup>6</sup>Gunther Martens, "Narrative and Stylistic Agency: The Case of Overt Narration," in *Point of View, Perspective, and Focalization. Modeling Mediation in Narrative*, ed. Peter Hühn, Wolf Schmid, and Jörg Schönert, Narratologia (Berlin: De Gruyter, 2009), 99-118.

<sup>7</sup>Annelen Brunner et al., "Das Redewiedergabe-Korpus. Eine neue Ressource," in *Digital Humanities: multimedial & multimodal. 6. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e.V. (DHd 2019)*, ed. Patrick Sahle (Frankfurt am Main, 2019), 103 - 106.

<sup>8</sup>Geert Brône and Bert Oben, *Eye-Tracking in Interaction: Studies on the Role of Eye Gaze in Dialogue* (Amsterdam: John Benjamins, 2018).

<sup>9</sup>Alexander Dunst, Jochen Laubrock, and Janina Wildfeuer, *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods* (London: Routledge, 2018).

ing with retellings<sup>10</sup> and film adaptations<sup>11</sup> to gauge more safely which words are imagined as spoken by what character (and to what music).<sup>12</sup> The powers of machine learning can be harnessed more productively through learning from transfer and actual reception.

Hence, I am under the impression that a purely text-based, bottom-up approach will not suffice to reach the declared goal of prediction. Narratology has already taken advantage of ongoing research in the fields of multimodality and paralinguistics. Also annotation schemata should go beyond purely text-intrinsic formalism and accommodate for drawing on the ways in which users process and interact with complex narratives.<sup>13</sup> This may involve annotating for semantic properties in tandem with strictly formal properties. This is a dilemma faced by all of those seeking to reconcile with cultural analytics. High-profile advances in the study of large amounts of narrative text, however, have been achieved without any reference to narratology or to (at least a customary understanding of) narrative aspects of the texts at hand ( e.g. authorship attribution in the cases of J.K. Rowling and Elena Ferrante). These experiments do away with the nitty-gritty of conventional narratological analysis at the advantage of ruthless, yet highly principled reductions of complexity in order to make hidden patterns visible. At the same time, it should be clear that narratology's toolkit has a lot in store to bring to the table of cultural analytics. Annotating for narrative structures of reported speech and variations in ontological modalities may help to reveal that apparently unstructured text is far more structured and/or narrative than has often been taken for granted. Narratologists should also be aware that a mere transposition of these tried-and-trusted methods onto large amounts of unlabelled data necessitates compromise and conceptual tweaking. Hence, this annotation guideline is a productive invitation to a much-needed continuation of the dialogue between narratology and Digital Humanities.



<sup>10</sup>Fritz Breithaupt et al., "Fact vs. Affect in the Telephone Game: All Levels of Surprise Are Retold With High Accuracy, Even Independently of Facts," *Frontiers in Psychology* 9 (November 20, 2018).

<sup>11</sup>Katalin Bálint and András Bálint Kovács, "Focalization, Attachment, and Film Viewers' Responses to Film Characters: Experimental Design with Qualitative Data Collection," in *Making Sense of Cinema: Empirical Studies into Film Spectators and Spectatorship*, ed. CarrieLynn D. Reinhard and Christopher J. Olson (Bloomsbury Publishing USA, 2016), 187-210.

<sup>12</sup>Joakim Tillman, "Solo Instruments and Internal Focalization in Dario Marianelli's *Pride & Prejudice* and *Atonement*," in *Contemporary Film Music: Investigating Cinema Narratives and Composition*, ed. Lindsay Coleman and Joakim Tillman (London: Palgrave Macmillan UK, 2017), 155-86.

<sup>13</sup>Susanna Salem, Thomas Weskott, and Anke Holler, "On the Processing of Free Indirect Discourse," *Linguistic Foundations of Narration in Spoken and Sign Languages* 247 (2018): 143.

Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

# Annotation Guideline No. 8: Annotation Guidelines for Narrative Levels

Adam Hammond

01.15.20

Article DOI: [10.22148/001c.11773](https://doi.org/10.22148/001c.11773)

Journal ISSN: 2371-4549

Cite: Adam Hammond, "Annotation Guideline No. 8: Annotation Guidelines for Narrative Levels," *Journal of Cultural Analytics*. January 15, 2020. doi: [10.22148/001c.11773](https://doi.org/10.22148/001c.11773)

## 1. Rationale

I first became aware of the SANTA project at the Digital Humanities conference in Montreal in the summer of 2017. I had just been assigned a 90-student second-year undergraduate Digital Humanities undergraduate English Literature class, set to begin in January 2018,<sup>1</sup> and I was looking for a group annotation project for my students. In previous iterations of the course, I had carried out several annotation projects focused on the narrative phenomenon of free indirect discourse (FID) in texts by Virginia Woolf and James Joyce.<sup>2</sup> What made these projects successful, from my perspective, was that FID is a complex phenomenon (by definition, a passage in which it is difficult or impossible to say for certain whether a character or narrator is speaking certain words) which is however relatively easy to represent in machine language (for instance, with the TEI <said> element and a few value-attribute pairs). The challenge in the assignment, in other words, was literary rather than technical: while it was easy to learn the TEI tagging, it

---

<sup>1</sup>The syllabus for this class, ENG 287, "The Digital Text," is available at <http://www.adamhammond.com/eng287s18/>

<sup>2</sup>See Adam Hammond, Julian Brooke, Graeme Hirst, "Modeling Modernist Dialogism: Close Reading with Big Data," *Reading Modernism with Machines: Digital Humanities and Modernist Literature*, eds. Shawna Ross and James O'Sullivan (Palgrave Macmillan, 2016): 49-78 and Julian Brooke, Adam Hammond, Graeme Hirst, "Using Models of Lexical Style to Quantify Free Indirect Discourse in Modernist Fiction," *Digital Scholarship in the Humanities* 32.2 (June 2017): 234-250.



was hard to say for certain whether a passage from *To the Lighthouse* was in direct discourse or FID, or to identify who exactly was speaking. To my mind, this made the assignment a meaningful one for my students, teaching them a technical skill while also bringing them into closer contact with the sometimes-irresolvable complexities of literary language.

Listening to the SANTA presentation at DH2017, it struck me that the phenomenon of narrative levels would make for a similarly meaningful annotation project. On the one hand, narrative levels could be represented fairly easily with a single XML element and through XML's nesting structure (its "ordered hierarchy of content objects"). On the other hand, definitions of what a narrative is, and what a "narrative level" might be, were sufficiently complex that the annotation even of a relatively simple text would present an interpretive challenge to my students. By the time that I had begun planning my course, the SANTA group had published a more detailed set of instructions on their website, including suggestions for theoretical readings on narrative levels.<sup>3</sup> They organized these readings in three levels: Introductory, Basics, and Advanced. Reading through these texts, I was struck by three things. First, that the concept of narrative levels, relatively intuitive at first glance, becomes more complex the more one looks at it. Second, that there was significant disagreement among narratologists concerning even basic categories (such as the distinction between a "narrative level" and a "narrative frame"). Third, that many of my second-year undergraduate students would be deeply confused even by the recommended texts at the simplest, "Introductory," level.

In light of this, I decided to keep my definitions as simple as possible — as close as possible to the level of the "intuitive," and free from explicit discussion of the theoretical disagreements that preoccupy narratologists who study the phenomenon. Since my motivation in preparing annotation projects is to find tasks that are simple technically but make my students reflect deeply about literary phenomena, I would keep my tagging scheme as simple as possible and restrict my definitions to the points on which all narratologists basically agree. This led to the very short guidelines that you see here — the shortest, by some margin, in this group. Although it could be argued that their brevity might lead to unnecessary disagreement among annotators — that by offloading so much of the literary work to my students, I was deliberately reducing the likelihood that the guidelines would produce annotations with useable levels of inter-annotator agreement — my suspicion from the beginning was that any greater detail would in fact simply confuse my student annotators and reduce inter-annotator agreement. (Analysis of the first round of SANTA annotation schemes confirms this suspicion to some

---

<sup>3</sup>See <https://sharedtasksinthedh.github.io/levels/>

extent.)

My guidelines depend on annotators' mastering three relatively simple concepts. The first is the concept of a narrative, which I define, drawing on Porter Abbott's *Cambridge Introduction to Narrative*, as "a representation of a story (an event or series of events) by a narrator."<sup>4</sup> The next is the notion that a given text can contain more than one narrative, and that narratives can be embedded within one another. I provide a rule of my own devising for helping students to decide whether they have reached a moment at which one narrative is embedded within another: if they could plausibly insert the phrase "Let me tell you a story" (a phrase which captures both sides of my simple definition of a narrative, the narrator ["me"] and the story itself) at the beginning of the proposed embedded narrative, then they should mark the beginning of a new narrative. The third concept is that of *degrees* of embeddedness, borrowing terminology from Shlomith Rimmon-Kenan via Manfred Jahn.<sup>5</sup> In my original guidelines, the annotation is described in terms of XML tags, which makes the discussion of embedding somewhat simpler, in that I can simply import XML's model of embedding and make the assumption that narrative levels also form an "ordered hierarchy of content objects."<sup>6</sup>

A further benefit of this simple annotation scheme is that it serves to focus the eventual computational task. Though annotations produced according my guidelines could not be used to train machine learning models in all narrative phenomena related to narrative levels, they could help to keep attention focused on three crucial and related tasks: identifying moments where one narrative yields to another; identifying the speaker of each; and placing these narratives in hierarchical relation to one another.

I carried out my annotation project as follows. First, I assigned Henry James's *The Turn of the Screw*, and presented a two-hour lecture focused in large part on how James's complicated framing structure serves to complicate (rather than resolve) the text's many "narrative gaps." The next week, in another two-hour

---

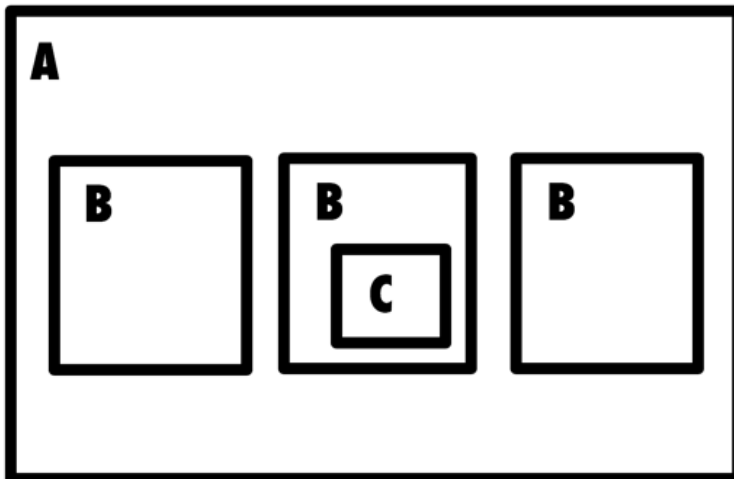
<sup>4</sup>H. Porter Abbott, *The Cambridge Introduction to Narrative*, 2<sup>nd</sup> edn. (Cambridge: Cambridge UP, 2014). Abbott defines a narrative as "the representation of a story (an event or series of events)" (237). He excludes the necessity of a narrator from his definition of a narrative on the basis that this would exclude most drama and film. Since I was working exclusively with prose fiction, this exclusion was not necessary for my own schema.

<sup>5</sup>Manfred Jahn, "N2.4. Narrative Levels," *Narratology: A Guide to the Theory of Narrative* (English Department, University of Cologne, 2017), <http://www.uni-koeln.de/~ame02/pppn.htm#N2.4>

<sup>6</sup>The actual guidelines distributed to students — which describe the annotation project in terms of XML — are available at <http://www.adamhammond.com/wp-content/uploads/2018/10/narrative-frames-annotation-guidelines.pdf> The main difference between these guidelines and the tool-agnostic version included here are the necessary addition of the "open" attribute, which is a less elegant method than the method described in note 8 below.

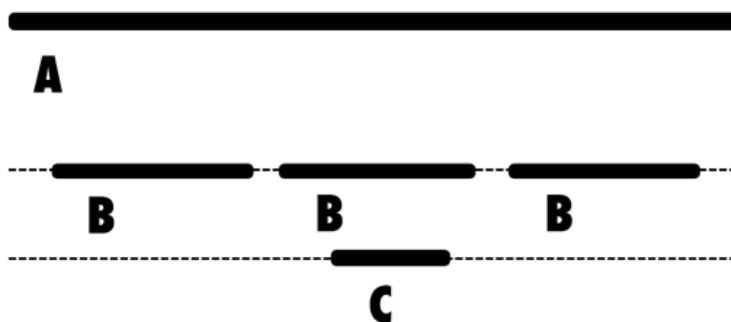
lecture, I introduced the students to XML and to the project itself.<sup>7</sup> In this lecture, I provided slightly more detail than I provide in the guidelines themselves. For example, I showed students Genette's speech bubble doodle and discussed its implications. I introduced box diagrams for representing narratives within narratives, for instance in the *Thousand and One Nights*, as follows:

## **THOUSAND AND ONE NIGHTS**



I also provided corresponding diagrams emphasizing the stratified levels of narrative — the “degrees” of narrative — in such box diagrams:

## **THOUSAND AND ONE NIGHTS**



## **FIVE NARRATIVES THREE NARRATIVE LEVELS (A+B+C)**

I provided additional diagrams for *Hamlet*, *The Taming of the Shrew*, and *The Turn of the Screw*. For the latter text, I emphasized that there were multiple valid ways of interpreting the text's structure: for instance, the Governess's tale could be

<sup>7</sup>The slides for this lecture are available at [http://www.adamhammond.com/wp-content/uploads/2018/10/eng287\\_narrative\\_levels\\_lecture.pdf](http://www.adamhammond.com/wp-content/uploads/2018/10/eng287_narrative_levels_lecture.pdf)

envisioned as a third-degree narrative embedded within Douglas's and the outer narrator's, or could be seen as embedded only within that of the outer narrator; further, certain stories that the Governess tells Mrs. Grose could be marked as separate narratives, though one could argue that they are simply part of the Governess's narrative, not independent of it. I also used *Turn of the Screw* to introduce the notion of "open frames."<sup>8</sup> I next explained "Mise-en-Abyme" or recursive narratives. I concluded the lecture by explaining the process students would use to annotate their assigned stories. I next posted an instructional video explaining the annotation procedure, which students accomplished with the Sublime Text editor.<sup>9</sup> In practice, the project seems to have been a success in the context of undergraduate pedagogy. In the 74 annotations received in the project, there were only three coding errors — evidence that, as desired, the technical challenge was minimal. Although we have yet to perform detailed investigation of inter-annotator agreement among my students, informal evaluations performed in the context of grading students' work revealed that disagreement occurred primarily in instances where literary interpretations might reasonably differ — evidence that the literary questions asked of students were meaningful ones. Going forward and revising these guidelines for use beyond my classroom, I would add more explicit and theoretically-grounded definitions and include diagrams like those depicted above.

## 1. OVERVIEW

A set of narrative texts are to be annotated for narrative levels. Any span of text containing a narrative is to be marked with the **nframe** category marker. For the purpose of our task, a narrative is defined as a representation of a story (an event or series of events) by a narrator. The texts in our annotation set may contain a single narrative (and thus a single nframe category) or may contain multiple narratives embedded within one another (nframe categories within nframe categories). If you come to a point in a text where you are uncertain whether to indicate a shift in narrative levels, imagine inserting the phrase "Let me tell you a story" right after the proposed division point. If the phrase fits, you should

---

<sup>8</sup>In my original guidelines, these open frames indicated by a deliberate XML error — withholding an end-tag — which is not practical but which I believe perfectly captures a reader's feeling at the end of a story like *The Turn of the Screw*, where it is as if the author had made a coding error, omitting crucial information that allows us to properly process the conclusion of the story.

<sup>9</sup>The instructional video is available at <https://www.youtube.com/watch?v=DsEEUlcYfSU> The texts I assigned to students were mostly those proposed by the project, though I made several substitutions based on various factors, including stereotyped representations of racialized characters in certain supplied texts. For instance, I replaced Rudyard Kipling's "Beyond the Pale" and "How the Leopard Got Its Spots" with Wallace Thurman's "Cordelia the Crude" and an abridged version of Zora Neale Hurston's *Their Eyes Were Watching God*.

likely mark a new narrative level. The `nframe` category has two necessary and one optional attribute.

- **level attribute**

The level attribute is used to express the degree of embedding of a narrative. If the narrative is not embedded within any others, it is a top-level or *first-degree* narrative and should be given the attribute value of “A”. A narrative embedded within an “A”-level narrative — a narrative within a narrative, or *second-degree* narrative — is given the attribute value of “B”. A narrative embedded within a “B”-level narrative — a narrative within a narrative within a narrative, or *third-degree* narrative — is given the attribute value “C”, and so on. Note that a text may contain multiple narratives at each level. For instance, the *Thousand and One Nights* contains hundreds (in some tellings, exactly 1,001) of “B”-level narratives — some of which contain “C”-level narratives of their own.

- **narr attribute**
- **open attribute**

The `narr` attribute keeps track of the narrator who conveys the narrative. We will represent these with numbers. The first narrator you encounter should be numbered “1”, the second “2”, the third “3”, and so on. If the narrator of a “B”-level narrative is the same as the narrator of the “A” level, both are numbered “1”. If the narrator of a “B”-level narrative is different from the narrator at the “A” level, the first is numbered “1” and the second “2.” And so on. Some writers choose deliberately to leave frames “open.” For example, in Henry James’s *The Turn of the Screw*, the governess’s “C”-level tale is framed within a Christmas fireside storytelling session by two narrators, the “A”-level “I” and the “B”-level Douglas. Yet after the governess finishes her tale, James does not return to the “A” or “B” levels to explicitly close them. Instead, they are left hanging. Indicate “open” by setting the “open” attribute to “true” (if not indicated, it will be assumed that the frame is “closed”).

## 2. SAMPLE ANNOTATIONS

A simple text containing only one narrative might be annotated as followed, using XML markup as an example:

```
<nframe level="A" narr="1">It was a dark and stormy night.
The wind blew and the wolf howled. The wind blew open
my window and the wolf entered. The wolf bit me and I
died.</nframe>
```

A text containing a single “B”-level narrative might be annotated as follows. (Since the narrator of the “B”-level narrative is different

from that of the “A”-level narrative, it is given the narrator attribute of “2”.)

```
<nframe level="A" narr="1">It was a dark and stormy night.
The wind blew and the wolf howled. The wind blew open my
window and the wolf entered. The wolf opened his mouth and
spoke. <nframe level="B" narr="2">"Once upon a time, when
I was but a young pup, a wizard appeared before me and
predicted my fate. He told me that one day, I would leap
through a window and eat a man whole. After enduring many
hardships, I have come to enact my fate."</nframe> He bit
me and I died.</nframe>
```

A text containing two “B”-level narratives and a single “C”-level narrative might be tagged as follows. (Since the narrator of the second “B”-level narrative is the same as the “A”-level narrative, they share the narrator attribute of “1”.)

```
<nframe level="A" narr="1">It was a dark and stormy night.
The wind blew and the wolf howled. The wind blew open my
window and the wolf entered. The wolf opened his mouth and
spoke. <nframe level="B" narr="2">"Once upon a time, when
I was but a young pup, a wizard appeared before me and
predicted my fate. The wizard told me, <nframe level="C"
narr="3">'I was born in the east. My father was a plumber
and my mother an auto mechanic. From a young age, it was
clear that I had little talent for either profession, so
I set off for the wizard academy. My expert wizardry has
brought me here to you. You, dear wolf, will some day leap
through a window and eat a man whole.'

```

- Special Case: “Open Frames”

In the following example, the “A”-level narrative is not explicitly closed by narrator 1 (presumably because he has been eaten and is unable to write) and thus the attribute “open” attribute has been set to “true”

```
<nframe level="A" narr="1" open="true">It was a dark and
```

stormy night. The wind blew and the wolf howled. The wind blew open my window and the wolf entered. The wolf opened his mouth and spoke. `<nframe level="B" narr="2">"Once upon a time, when I was but a young pup, a wizard appeared before me and predicted my fate. He told me that one day, I would leap through a window and eat a man whole. After enduring many hardships, I have come to enact my fate."``</nframe></nframe>`

- **Special Case: “Mise-en-Abyme” Narratives**

Some narratives, especially popular with postmodern writers, paradoxically embed a story within itself. This paradoxical situation can be represented by showing a series of “A”-level narratives embedded within one another:

```
<nframe level="A" narr="1">It was a dark and stormy night. The band of robbers huddled together around the fire. When he had finished eating, the first bandit said, <nframe level="A" narr="1">"Let me tell you a story. It was a dark and stormy night and a band of robbers huddled together around the fire. When he had finished eating, the first bandit said: <nframe level="A" narr="1">'Let me tell you a story. It was a dark and stormy night and...' "</nframe></nframe></nfr
```

### 3. OTHER NOTES

If a shift in narrative level occurs around a chapter break and you’re unsure whether to put your nframe category marker before or after the chapter header, put it after.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

# Annotating Narrative Levels: Review of Guideline No. 8

Tom McEnaney

01.15.20

*Article DOI: 10.22148/001c.11776*

*Journal ISSN: 2371-4549*

*Cite: Tom McEnaney, "Annotating Narrative Levels: Review of Guideline No. 8," Journal of Cultural Analytics. January 15, 2020. doi: 10.22148/001c.11776*

"Let me tell you a story." The proposed guidelines suggest that this phrase serve as the heuristic that readers supply at the beginning of any possible embedded narrative to identify a shift in narrative frames or levels. (The difference between "frame" and "level," although perhaps confusing in the history of narratology, does not seem like an important distinction at this stage of the project.) This simple phrase, the author suggests, can replace a field of narrative theory they feel would "simply confuse my student annotators." However simple the phrase might seem, however, it, in fact, conceals a number of key narratological issues: focalization, temporal indices, diction / register, person, fictional paratexts, duration, and, no doubt, others. The question for the guidelines is whether one can leapfrog the particularity of these issues if students use the above phrase to annotate texts with XML tags and produce operational scripts that identify the nested narratives. As it currently stands, students seem capable of learning the basic idea of nested narratives and tagging changes in narrative frames, but there are no real results to confirm the project's success, as the author reports they are not yet able to confirm any inter-annotation agreement.

How does one "identif[y] moments where one narrative yields to another"? We might have an intuitive sense of this change, or we might see obvious di-acritical markers (new sets of quotation marks, for instance), but teaching the machine would seem to require more specific categories. Certain classic narratological keywords ("story" vs "discourse") and debates ("Narrate or Describe?")



might play a part in teaching what is at stake in these narrative shifts, but they do not seem necessary to identify changes in narrative footing. On the other hand, the key categories mentioned in the first paragraph seem useful for writing a program that would include a robust sense of narrative theory. For instance, a reader might notice a change in person (first, second, third, singular or plural), or a character name. Or a reader might realize that while the previous narrative was told in the present tense, the next strip of discourse appears in the imperfect. Additionally, there might not be any change in person or tense, but a new register might take over the text. All of these concepts might fall under the umbrella term of “focalization,” and, if the author of the proposed guidelines does indeed have a sturdy model for sorting moments of free indirect discourse, as they suggest in the proposal, perhaps that script could address these issues. But what if a change in person and / or tense and / or register occurs for only a sentence or two in the course of a dialogue between two characters. Is this a new narrative? Duration, the number of words that pertain to a shift in person and / or tense, seems like a relevant concept to help in classification, as well. Lastly, how would such a classifier independently account for fictional paratexts such as epistolary introductions, fictional prefaces, or other frames that might not differ in person, diction, or tense? Can we limit the search to just one or a few of these key categories and still write a script that successfully identifies the transition from one narrative frame to another?

Operationalizing these issues would seem necessary in order to fulfill the author’s proposal to find a “use beyond the classroom” for these guidelines. On the other hand, the guidelines do seem useful as a pedagogical assignment to draw attention to issues of focalization and other key features of narrative for students confused by the jargon of narrative theory or unconvinced or unexcited by non-operational hand annotation (i.e. circling moments where they’ve identified a change in narrative frame). Moreover, the proposed assignment seems like an excellent introduction to the very idea and process of operationalizing literary concepts.

In a classic of popular narratology, Umberto Eco’s *Six Walks in the Narrative Woods* (delivered as the 1992-1993 Norton Lectures at Harvard), Eco repeatedly turns to the 19<sup>th</sup> century French writer Gérard de Nerval’s *Sylvie* to explain his theory of nested narratives, the differences between an “author,” “model author” and “narrator,” and the temporal problems that arise alongside or at an intersection with these categories. Citing Bal, Barthes, Booth, Chatman, Cohn, Genette, Greimas, Ricoeur, Todorov, and White, Eco concludes that “a text is a lazy machine that demands the reader do part of its work.” Can the proposed guidelines make a more active and flexible machine? A machine that will easily identify

shifts in narrative levels or frames? And will the results help readers to attend to the meaningfulness of these changes in perspective, acting alongside one machine reading another? The proposal's success will depend on whether it can answer these questions positively.



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.