Annotating Mystery Novels: Guidelines and Adaptations

Nuette Heyns
North-West University
Potchefstroom, South Africa
nuette.heyns@gmail.com

Menno van Zaanen
South African Centre for Digital Language Resources
Potchefstroom, South Africa
menno.vanzaanen@nwu.ac.za

Abstract

To understand how stories are structured, we would like to be able to analyze the architecture of narratives. This article reviews and compares existing annotation guidelines for scene and narrative level annotation. We propose new guidelines, based on existing ones, and show how these can be effectively extended from general-purpose to specialized contexts, such as mystery novels which feature unique narrative elements like red herrings and plot twists. This provides a controlled environment for examining genre-specific event structuring. Additionally, we present a newly annotated genre-specific dataset of mystery novels, offering valuable resources for training and evaluating models in narrative understanding. This study aims to enhance annotation practices and advance the development of computational models for narrative analysis.

1 Introduction

The process of narrative scene segmentation, which involves the identification of distinct scenes within a narrative, is a crucial task in the field of computational literary analysis. For instance, it allows researchers to better understand the structure and pacing of literary works, which can reveal insights about the author's stylistic choices and the overall narrative arc. Additionally, the ability to identify scenes can improve other tasks, such as summarisation (Droog Hayes et al., 2018), literary machine translation (Taivalkoski-Shilov, 2019), generation of narratives (Botelho, 2021; Lukin and Walker, 2019; Porteous et al., 2016), character interaction analysis (Agarwal, 2016; Chen and Bunescu, 2021; Fields et al., 2022; Lee, 2017; Macovei, 2017; Min and Park, 2016a,b,c; Porteous et al., 2016), and topic modelling (Schmidt, 2015).

Annotating literary texts presents challenges due to the often ambiguous and multifaceted nature of literary terms, which resist straightforward, computer-friendly definitions. To tackle this, standardized definitions and annotation schemes for key narratology elements such as narrative level, scene, focalization, and anachronisms (including flashbacks and flash-forwards) are essential.

Standardised guidelines ensure that these analyses are conducted in a consistent and comparable manner. They save time and effort in annotating texts, and enable the creation of reusable annotated datasets. Well-defined annotation guidelines are crucial to obtain high quality inter-annotator agreement (Alrashid and Gaizauskas, 2021). Despite previous attempts to standardize the guidelines for narrative scene segmentation (Alrashid and Gaizauskas, 2021; Gaizauskas and Alrashid, 2019; Kearns, 2020; Zehe et al., 2021a,b), there remains a need for a comprehensive and widely adopted set of best practices.

We begin by identifying essential concepts, such as narrative, narrative levels, anachronisms, focalization, scene, non-scene, and ellipsis. We then compare existing annotation guidelines, noting that similar concepts are often defined differently or annotated using various techniques. The aim is to consolidate concepts and streamline the annotation process. Next, we combine and integrate guidelines from previous work to create a cohesive annotation scheme. Finally, to investigate practical applicability, we apply the annotation guidelines to a new genre-specific dataset, focusing specifically on whodunits.

As mystery novels have specialized phases and the characters have specific roles, we propose to extend the novel annotation scheme (which is based on existing guidelines) in a modular fashion. Using this annotation scheme, we annotate a genre-specific dataset, and discuss how it complements the existing publicly available datasets. Note that other modular extensions may be proposed as well, which can be added and taken away as needed.

The contribution of this article is threefold.

Proceedings of the The 6th Workshop on Narrative Understanding, pages 55–66

First, we consolidate existing narrative annotation schemes. Second, we propose a modular extension of the annotation scheme. Modular extensions allow for annotation of specialized narrative genres, such as whodunits. Third, we apply the new annotation scheme with its modular extension to a set of narratives, showing its practical applicability. This results in an annotated data collection of whodunits.

2 Background

To form a solid basis for investigation into the annotation guidelines, we present the foundational concepts of narratology as defined in literary theory, particularly drawing from the works of Genette (e.g., Genette et al. (1980)). Additionally, we examine how previous work on computational narrative understanding has translated these literary concepts into computer usable definitions.

2.1 Narrative

According to Eisenberg and Finlayson (2021), a narrative is a linguistic representation that presents a coherent sequence of events involving specific characters and times, organized into a structured plot. It goes beyond commonsense coherence by employing elements such as climaxes and other plot structures. This aligns with Genette et al.'s concept of narrative discourse where the complexity of storytelling lies in the strategic choices of detail revelation, plot order, and narrative interruptions.

Stories are defined by the interaction of characters and events driving the plot forward. We can separate the chronological order of events (histoire, text) from the order in the story (récit, discourse) to understand both the "what" and "how" of a story. Additionally, narratives can appear contiguously as a single, solid text, or they can be embedded within another narrative, or even interrupt the preceding narrative (Eisenberg and Finlayson, 2021).

Segmenting narratives means identifying thresholds in the narrative that relate information about the structure/plot. The goal of these segmentation tasks is often to identify a scene or narrative level.

2.2 Narrative levels

Narrative levels refer to the hierarchical structure of a narrative, where the overall story can be composed of multiple nested levels of narrative. Genette et al. (1980) identify narrative levels in terms of the role the narrator plays in telling and ordering the story. According to Genette et al. (1980), there are three primary narrative levels: the *extradiegetic* level, which is the level of the narrator or implied author; the *(intra)diegetic* level, which is the level of the characters and the events they experience within the story; and the *metadiegetic* level, which is a secondary narrative embedded within the primary diegetic level.

On a practical level, annotating the narrative levels requires identification of clear thresholds between diegesis. From Genette et al.'s framework, we can infer that the threshold of a level is where the narrator changes. This leads to narrative levels in the form of embedded or interruption narratives (Eisenberg and Finlayson, 2021). *Embedded narrative* occurs when a plot event in the original narrative triggers the telling (i.e., embedding) of another narrative. This occurs, for example, when a character narrates a story in a dialogue of the main narrative. Embedded narratives typically occur on the metadiegetic level. *Interruptive narratives* interrupt the original narrator's narration. This is common in narratives where, for intance, each chapter has a different narrator.

2.3 Anachronisms

Anachronisms are deviations from the main temporal progression of the story (Kearns, 2021). We can identify a number of types of anachronisms (Eisenberg and Finlayson, 2021).

A *flashback* (analepsis) occurs when the time the events are told in, shifts from the present to the past, whereas a *flash-forward* (prolepsis) occurs in the form of visions, prophesies, or foreshadowing. Both can be either embedded or interrupted. *Embedded flashbacks* occur when the narrator is telling a story about the past, from the present time. In contrast, *interruptive flashbacks* replace the original narrative. Here, the original narrative ends and a new narrative starts (with events taking place before the original narrative). The narrator also moves in time, whereas the narrator is still in the present tense with embedded flashbacks.

Some research places anachronisms at a narrative level, but Ketschik et al. (2021) mention that anachronisms deal with the logical order of the discourse and do not leave the present narrative level. Because the narrator does not change in anachronistic narrations, narrative levels should be distinct from anachronisms.

2.4 Scene, non-scene, and ellipsis

Gius et al. (2019a) introduce the concept of a scene as a segment of narrative discourse that presents the histoire, so that time, place, and character constellation stay more or less the same. They do not explicitly relate scene to narrative levels, although a relationship must exist since both serve different (yet similar) aspects of narratives.

According to Gius et al. (2019a), there are four main aspects of a scene to define the boundaries: time, space, events, and characters. A scene often changes with a significant shift in *time*, such as when the narrative pace shifts from minutes to days. Similarly, a change in *space* or location triggers a scene change, though smaller locations may be grouped using the container principle (Gius et al., 2019a), which groups smaller rooms or locations together, allowing for small changes of location (within the same container) without scene changes. A scene also changes when a new action or *event* starts. Again, a container principle can be used here. Finally, a shift in *character* constellation (e.g., when characters join or leave) changes the scene. However, the scene does not change if the action remains the same. Also, a change of narrator does not necessarily cause a scene change if the main aspects remain the same (Gius et al., 2019a).

Gius et al. (2019a) also recognize *non-scenes*, which do not contain any acting characters. This mostly occurs as summaries, descriptions, or scenic passages. Non-scene information that briefly interrupts a scene or that occurs at the start or end of a scene, and which is too short to be considered a separate segment, is typically recognized as part of the scene.

Genette et al. (1980) define *ellipsis* as a form of narrative duration which omits certain events or periods of time within a narrative. This creates gaps that the reader must fill in, and is often used to accelerate the pacing or to focus on significant moments without detailing every occurrence.

Although Genette et al. (1980) does not specifically use the terms scene and non-scene, we can place the terms (along with ellipsis) as a form of narrative duration on the intradiegetic level of a text. Therefore, there can be several stories on an intradiegetic level consisting of scene, non-scene, and, indirectly, ellipsis (Ketschik et al., 2021).

2.5 Narrative perspective

Focalization is the perspective from which the narrative is seen (Wirén and Ek, 2021), or how much information the narrator has access to. We can identify different levels of focalization (Todorov, 1971). *Zero or unrestricted* focalization provides a fully omniscient perspective. The narrator knows more than any of the characters. *Internal* focalization is narrated from the perspective of a character in the story, where the narrator knows as much as the character. With *external* focalization, the perspective is outside the character in the story and the narrator knows less than any of the characters.

Narrative voice indicates the narrator's relationship with the text, and whether they are present in the text or not (Ketschik et al., 2021; Wirén and Ek, 2021). Narrative voice can be either *homodiegetic*, when the narrator appears in the story. They usually refer to themselves in the first person. Narrative voice can also be *heterodiegetic*, when the narrator does not appear in the story. The narration is mostly in the third person.

3 Existing datasets

To our knowledge, there are only three publicly available datasets annotated with narrative segmentation. Two of these were created within the SANTA (Systematic Analysis of Narrative Texts through Annotation) project (Gius et al., 2019b), which was a significant effort in developing annotation guidelines and annotating narrative structure. Several researchers (Barth, 2019; Bauer and Lahrsow, 2020; Eisenberg and Finlayson, 2021; Hammond, 2021; Kearns, 2019; Ketschik et al., 2019; Wirén and Ek, 2021) took part in this task by creating annotation guidelines. Barth (2021); Kearns (2021); Ketschik et al. (2021) later extended their guidelines.

Based on these results, the project established annotation guidelines for narrative levels, which were also applied to a corpus[1] in a shared task.

Note that the datasets of Chung et al. (2018); Kearns (2020); Newberry and Bailey (2019); Rogers et al. (2024) are not publicly available.

Gaizauskas and Alrashid (2019) proposed SceneML to annotate scenes, locations, characters, and time in narratives. Unfortunately, the annotation scheme is vague on how to treat narrative description and levels of narratives.

[1] https://github.com/SharedTasksInTheDH

4 Annotation guidelines

A number of standardised guidelines for annotating the key narratological elements have been proposed (many stemming from the SANTA shared task (Barth, 2019; Bauer and Lahrsow, 2020; Eisenberg and Finlayson, 2021; Hammond, 2021; Kearns, 2019; Ketschik et al., 2019; Wirén and Ek, 2021)). Note that some of these guidelines have been updated. We will only refer to the most recent version.

Gius et al. (2021) compare the SANTA annotation guidelines, highlighting the strengths of each set. Here, we analyze the guidelines in detail and select the best annotations to ensure a consistent and coherent annotation scheme. We focus on narrative levels, anachronisms, scenes, and focalizations as defined in Section 2. Additionally, we explore how guidelines propose handling metatext, paratext (i.e., the text that surrounds the narrative), and punctuation, while highlighting how guidelines use different terms to refer to similar concepts. We then provide an additional set of annotations, specifically for the annotation of mystery novels, that can be used in a modular fashion.

4.1 Narrative levels

As mentioned in Section 2.2, three levels of narrative are recognized: extradiegetic, intradiegetic, and metadiegetic.

Wirén and Ek (2021) introduce a guideline to annotate the extradiegetic level by using the tag `NARRATOR` combined with the numerical value 0. For the intradiegetic level, they use the same tag, but combine it with the numerical value 1.

Metatextuality occurs on the extradiegetic level as moments where the text comments on itself or the act of storytelling (Genette et al., 1980). Barth (2021) classifies these sections as "metanarrative" or "metafiction". However, Ketschik et al. (2021) argue that the exegesis and diegesis become intertwined and suggest not annotating any levels here, although they add a "non-narrative" tag. Wirén and Ek (2021) annotate a form of metatextuality simply as narrator's discourse on an extradiegetic level regardless of the degree of overtness (and hence use the value 0).

The *metadiegetic level* refers to embedded stories, often told by characters within the intradiegetic level. Wirén and Ek (2021) use the term "narrator discourse", which can be embedded. This embedded narrator discourse is the equivalent of what we refer to as narrative level (as their criteria for a threshold is a switch in the narrator).

As a critique to using the narrator as a threshold for level changes, Barth (2021); Ketschik et al. (2021) argue that not all narrator changes cause a change in level. New narrative levels can be introduced without a prototypical change of narrator. Also, the introduction of a new speaker does not necessarily signal a level change as the speaker would have to narrate a separate story (Ketschik et al., 2021). Similarly, Ketschik et al. (2021) argue that homodiegetic narrators can tell embedded stories, they are not part of. In this case, narrators remain the same, but the level changes, as there is a change in the narrator's position in relation to the story they tell.

Barth (2021) collectively refer to embeddings and framed narratives as acts (which are placed on a horizontal level) and separate them from narrative levels (which they place on a vertical level). They use Genette's requirement of a narrator change to induce an act change. The main distinction is that the different narrators of acts are on the same narrative level. Barth (2021) state that "a new narrative act at least diverges in time, setting or the corresponding characters from the previous one", which is similar to scenes as proposed by Gius et al. (2019b). Ketschik et al. (2021) make a similar distinction between vertical and horizontal thresholds. However, they use the terms story (horizontal) and level (vertical). Here, story is defined as a self-contained action whose events and happenings are casually linked and cause a change of state. Hammond (2021) also makes the distinction between vertical and horizontal levels but refers to "frames".

Eisenberg and Finlayson (2021); Hammond (2021); Ketschik et al. (2021) suggest using numbers to indicate the vertical degree of the narrative level and letters to indicate the horizontal, sequential arrangement of acts. Similarly, Kearns (2019) uses a level tag to indicate an embedded narrative, but they also use a numerical value to indicate the sequential acts.

Barth (2021); Ketschik et al. (2021) also distinguish between illocutionary (e.g., speaker change) and ontological (narrator change) boundaries as introduced by Ryan (1992). They add that boundaries can also be crossed actually or virtually. However, they only mention this and do not include these concepts in their annotation guidelines.

We suggest keeping with the style of Wirén and Ek (2021) to annotate all three levels with the tag `NARRATOR`. Similar to Hammond (2021); Ketschik

et al. (2021), numbers are used to indicate the degree, with 0 for the extradiegetic level, 1 for the intradiegetic level, and 2 for the metadiegetic level. Letters can be used to indicate the sequential arrangement on the INTRADIEGETIC property and METADIEGETIC property. Furthermore, the value meta can be used (in addition to the 0 value) to indicate metatextuality in the EXTRADIEGETIC property.

4.2 Anachronisms

As we discussed in Section 2, flashbacks and flashforwards can be embedded or interruptive. Eisenberg and Finlayson (2021) differentiate between these and include tags accordingly. Similarly, Kearns (2021) proposes using the tags ANALEPSIS and PROLEPSIS. These should be used when a new narrative or point in time starts.

Ketschik et al. (2019) emphasise that the narrative level does not change with prolepsis or analepsis and similarly with character thoughts, dreams, and visions. However, they do not seem to include tags for any of these cases in their guidelines.

Following Eisenberg and Finlayson (2021); Kearns (2021), we propose using the tags ANALEPSIS and PROLEPSIS and assigning the properties EMBEDDED or INTERRUPTIVE where needed.

4.3 Scene, non-scene, and ellipsis

Existing guidelines do not explicitly position scenes, non-scenes, and ellipses in relation to the narrative levels. Our understanding is that scenes, non-scenes, and ellipses can occur within either the intradiegetic or metadiegetic levels, but this does not imply a strict hierarchy between scenes and diegetic levels. In fact, there is not necessarily a strict hierarchical relationship between diegetic levels themselves. The intradiegetic level and the metadiegetic level might have a hierarchical connection since the metadiegetic level can only occur within the intradiegetic level. However, the extradiegetic level operates independently of this hierarchy, as it can exist outside or within the intradiegetic or metadiegetic levels. As a result, scenes can appear on both the intradiegetic and metadiegetic levels, with multiple scenes potentially existing on the same level. Additionally, the extradiegetic narrator may comment within a scene that is otherwise situated on an intradiegetic level.

The most widely accepted definition of a scene is where time, location, and main characters are constant and focus on one action. Alrashid and Gaizauskas (2021) suggest a scene can contain multiple actions by grouping "scene description segments" (SDS), or continuous spans of text. This resembles the idea of multiple events in a scene, as long as place, time, and characters remain the same. A scene can reference past or future events, similar to embedded narratives, where a character tells another story. Alrashid and Gaizauskas (2023) also propose scene transition segments (STS) to refer to text segments where the action shifts between locations as the narrative transitions from one scene to another.

Alrashid and Gaizauskas (2023); Gius et al. (2019a) distinguish between scene and non-scene. Alrashid and Gaizauskas (2023), however, only include SCENE and NON-SCENE tags and do not differentiate between different types of non-scene. The only other work that annotates a form of non-scene is Kearns (2019). They annotate extended (when time is extended relative to story time) or compressed time (when narrative time moves faster than story time) using the tags by the same names. Similar to how Gius et al. (2019a) use the weight that an aspect carries to determine a threshold, time should in this context also be evaluated with respect to the overall text.

We use the definition of scene as provided by Gius et al. (2019a) and annotate this using the SCENE tag. The SCENE tag also allows for the properties TIME, PLACE, and CHARACTER_CONSTELLATION. We do not annotate events or SDS, but if needed, these annotations can easily be added as values to properties of the SCENE tag.

We propose adding a TRANSITION tag to mark STSs. Additionally, we use the container principle, where several (smaller) locations can be contained in a larger one. Furthermore, we extend this container principle to STSs: if the transition text carries significant weight in the overall narrative, it is marked as a transition. However, if the transition occurs between places that are contained within the same scene, then the transition is not annotated.

For simplicity, instead of using a separate tag for compressed time and extended time, to indicate the accelerated speed of narration (or summaries as defined in Section 2), or descriptive passages, we use the tag NON-SCENE and assigning a property of SUMMARY, DESCRIPTIVE_PASSAGE, or SCENIC_PASSAGE.

To our knowledge, no work includes ellipsis in their annotation guidelines. We advocate the anno-

tation of omitted time in narratives as it plays an important role in the pacing of a narrative. We can annotate this with the tag ELLIPSIS.

4.4 Narrative perspective

Focalization provides important information of the narrator's perspective and the extent of their knowledge within a narrative.

In addition to information on the perspective of the narrator, dreams, visions, fantasies, and thoughts are forms of focalization. Even though they do not represent a change in narrative level (Ketschik et al., 2021), they are essential for understanding the narrator's role and the narrative's structure. Eisenberg and Finlayson (2021) similarly categorise dreams and visions in the same way as flashbacks and flash-forwards, labelling them as either as embedded or interruptive based on their function within the narrative.

Wirén and Ek (2021) offer a detailed framework for annotating focalization. In this framework, character discourse is broken down into "turns", which include a single speaker addressing multiple addressees, and "lines", which correspond to a single addressee or set of addressees. To enhance the precision of dialogue annotation, characters are assigned numeric values, and narrative construction (NC) tags are used to mark speech-framing constructions within lines.

Although these detailed annotations help capture the nuances of character interactions and narrator shifts, we have chosen not to use them, as they seem to be overly detailed for the scope of most studies. Moreover, as Ketschik et al. (2021) advises, it is important not to overemphasise focalization, so a more balanced approach will be adopted.

We propose the FOCALIZATION tag with the possible properties EMBEDDED and INTERRUPTIVE. As mentioned in Section 2, narrative voice distinguishes whether a speaker is present in the narrative or not. Following the suggestions of Barth (2021), we introduce the tag VOICE with the properties HOMODIEGETIC and HETERODIEGETIC.

4.5 Punctuation and paratext

When considering how to treat punctuation marks and paratext in the annotation process, the aim is to maintain a clear distinction between the narrative elements central to the story and the textual features that serve a more structural or contextual role.

Gius et al. (2019b) suggest adding punctuation marks inside the annotated segment, but punctuation marks that structure the text, such as asterisks, should be placed outside of the annotated segment. Furthermore, Ketschik et al. (2021) suggests not annotating paratexts, such as titles, forewords, chapter headings, and genre indications. While important for understanding the broader context of the work, these elements are typically considered external to the narrative itself and thus are excluded from the core annotation process.

We propose not to annotate punctuation and paratext. The annotation focuses specifically on the narrative itself.

4.6 Genre specific annotations

The predictable structure of classic whodunit mysteries makes them ideal for analyzing how narrative elements unfold and for employing digital tools to annotate texts. In this section, we introduce tags to capture the essential elements that are specific to whodunit mystery novels.

According to Cawelti (2014), whodunits typically include six phases: introduction of the detective, crime and clues, investigation, announcement of the solution, explanation of the solution, and denouement.

We propose to use the following structural tags: INTRODUCTION for the detective's arrival at the crime scene. INVESTIGATION for scenes where the detective gathers clues or interrogates suspects. This combines the crime and clues, and investigation phrases. CONFRONTATION for the solution announcement. CONFESSION deals with the explanation of the solution and the confession. REVEAL for explaining the solution, while AFTERMATH annotates the denouement.

Additionally, whodunits make use of specific concepts that can be identified. The crime scenes are annotated by specifying the value of the PLACE property of the SCENE tag, and clues are tagged with CLUE, with optional IDENTIFIED and REFERRED properties to distinguish when a clue is first found and when it is referenced later respectively.

Given the genre's reliance on the detective's thought process, we add a DETECTIVE_THOUGHT property to the focalization tag. This includes not just direct thoughts, but also gestures reflecting the detective's thinking, especially when the narrator is another character.

5 Annotation scheme

The teams that took part in the SANTA project used different annotation schemes ranging from XML to Excel documents. Eventually, all the annotations were translated to CATMA annotations[2].

CATMA offers several advantages that make it a suitable tool for annotating texts with narrative elements. Its flexibility and customisation allow researchers to create and adjust annotation categories and schemes to fit the specific needs of their analysis, making it particularly useful when dealing with genre-specific texts. The platform supports multiple annotation levels, enabling the tagging of narrative elements without a set hierarchical structure. This is especially useful in this context where it is oftentimes unclear what the order of tags should be. For example, is the diegetic level the outer layer that can include scenes or are scenes the outer layer and can contain a diegetic level? Although it is possible to add attributes to an XML file to deal with these situations, CATMA provides a seamless approach to the structure. Additionally, CATMA provide textual analysis integration to examine the data. The tool also supports XML export, ensuring that the annotated data can be easily shared, reused, and integrated with other tools or systems, which is crucial for collaborative research and future studies.

Appendix A provides the complete list of tags, properties, and values used during the annotation process following the CATMA annotation scheme.

6 Dataset

For this study, we will use short mystery stories, allowing us to observe the annotation process across complete texts. We have selected *The jewel robbery at the grand metropolitan* (C1) and, *The adventures of the Italian nobleman* (C2) both Agatha Christie stories and *A case of identity* (D1) and *The red-headed league* (D2) which are Sir Arthur Conan Doyle stories. To ensure that the annotated dataset can be made publicly available, we use texts in the public domain from Project Gutenberg[3]. Information on the texts can be found in Table 1.

An overview of the number of tags annotated for each of the selected stories will be provided in Appendix B (due to space restrictions).

Currently, we have only annotated four texts, but plan to expand the dataset in the future. Table 2 shows a size comparison between our dataset and the SANTA[4] dataset. While our dataset is smaller at 26825 words across four texts, its strength lies in containing full narratives, allowing for a thorough analysis of story flow and development, which shorter text extracts may miss. This provides a better overview of narrative techniques from start to finish.

Text	# words	AWS	# sentences	# tags
C1	5029	9.5	526	41
C2	3783	10.3	364	59
D1	6990	17.5	399	97
D2	9115	15.9	574	82

Table 1: Properties of the annotated texts. AWS represents the average number of words per sentence. C1 is *The adventures of the Italian nobleman*, C2 is *The jewel robbery at the grand metropolitan*, *A case of identity* and *The red-headed league* are D1 and D2 respectively.

Dataset	# texts	Max words	# words
SANTA	25	2000	50 000
Our dataset	4	9115	26 825

Table 2: Comparison with the SANTA dataset ranked by total word length.

7 Annotation process

All annotations were manually done by one of the authors using CATMA. As such, no compensation was received for the annotations. The manual annotations provide a foundation for the development of a method to automatically annotate similar texts, which we plan to explore in a follow-up publication.

8 Discussion

Throughout the annotation process, several challenges arose that required careful consideration. These challenges were often linked to the application of the container principle, the tagging of transition segments, and the differentiation between dialogue and embedded sections.

The container principle proved valuable, but it introduced some inconsistencies in the annotation process. Deciding when to apply this principle was challenging, particularly when characters left and then returned (e.g., *The man returned shortly;*

[2] https://catma.de/

[3] https://www.gutenberg.org/

[4] https://github.com/SharedTasksInTheDH

with him came the manager.) or when the narrative moved between locations within the same building. The significance of a place or character constellation is not always straightforward to determine. For instance, in text C2, the story shifts between rooms within a hotel, such as the hallway, lobby, and lift. We concluded that spaces like the hotel room and kitchen were significant enough to justify scene changes, while the hallway, lobby, and lift were often transitions between these key spaces.

However, this raised the question of how to group these transitional spaces. When a character moves from the lobby to the elevator, then into the hallway, and finally enters a room, we want to split the scenes between the hallway and the room. This requires determining the precise moment the character crosses the threshold from one space to another. For example, consider the sentence: *The manager produced the key without more ado, and we all entered the flat.* Here, the first half of the sentence places the characters in the hallway, but by the end, they are inside the room. In such cases, we might annotate the sentence as a transition segment, although it could be seen as either part of the previous scene or the beginning of a new one. It can also be seen as a separate scene with the place value set to "transition". We took the latter approach where we annotate a scene where characters move between two places, such as travelling in an elevator or a car between locations as a separate scene with the place value "transition".

Distinguishing between embedded focalization and simple dialogue posed significant challenges. We define embedded focalization as a shift in perspective to a different character, or a different temporal or spatial point within the story, while still remaining within the broader narrative framework. For example, the main narrator might describe a scene, and within that scene, a character recalls a past event from their own perspective. However, the focalization can sometimes shift between the character's recollection and the narrator's interruptions, making it difficult to determine where the embedded focalization begins and ends.

Similarly, it can be challenging to differentiate between an embedded flashback and dialogue that briefly references a past event. For instance, if a character mentions something that happened the previous day in just one sentence, this might not seem significant enough to be tagged as an embedded flashback. An example from text C2 illustrates this: *"... You—in company with a friend—visited the late Count Foscatini on the morning of Tuesday the 9th—" The Italian made an angry gesture.* In this case, the brief mention of a past event feels more like a part of the dialogue in the present moment rather than a true flashback. However, if that past event is described in more detail, expanding into a paragraph or more, it begins to take on the characteristics of an embedded flashback.

Another challenge was understanding the interplay between scenes and summaries. Often, a subscene (an event or moment that is part of a larger scene) can be narrated as a summary. In such cases, the boundary between scene and summary is not clear-cut. For instance, a narrative may describe the actions within a scene in detail, then briefly summarize the events that followed within the same scene. This overlap suggests that scenes and summaries are not mutually exclusive; rather, they can coexist, with a sub-scene being narrated through summary within the broader scene.

Our annotation process involved multiple passes through the text to ensure accuracy and consistency. In the initial round, we concentrated on identifying and tagging scenes, non-scenes, and ellipses. This foundational layer allowed us to establish the basic structure of the narrative. In the second pass, we focused on annotating diegetic levels, focalization, narrative voice, and anachronisms. In a final pass, we tag clues, detective thought and the different acts within the narrative.

9 Conclusion and future work

This work addresses the challenges of narrative scene segmentation by consolidating existing narratological annotation schemes and proposing a modular extension for a genre-specific dataset. In Section 8 we outlined the challenges encountered during the annotation process and offered potential solutions.

For future work, we aim to expand the dataset and encourage the development of additional genre-specific annotations and datasets. We aim to contribute to a more comprehensive and widely adopted set of best practices for narrative annotation. Additionally, we plan to conduct experiments with this dataset to identify types of features that help with automatic annotation. It is currently unclear what kind of features will be useful, e.g., lexical, syntactic, semantic, pragmatic features or properties from, for instance, character and location networks.

References

Apoorv Agarwal. 2016. *Social Network Extraction from Text*. Ph.D. thesis, Columbia University.

T. Alrashid and R. Gaizauskas. 2023. ScANT: A Small Corpus of Scene-Annotated Narrative Texts. In *Text2Story@ECIR*.

Tarfah Alrashid and Robert J. Gaizauskas. 2021. A Pilot Study on Annotating Scenes in Narrative Text using SceneML. In *Text2Story@ECIR*.

Florian Barth. 2019. Annotation Guideline No. 5: Annotation Guidelines for Narrative Levels and Narrative Acts. *Journal of Cultural Analytics*, 4(3).

Florian Barth. 2021. Annotation Guidelines for Narrative Levels and Narrative Acts V2. *Journal of Cultural Analytics*, 6.

Matthias Bauer and Miriam Lahrsow. 2020. Annotation Guideline No. 6: SANTA 6 Collaborative Annotation as a Teaching Tool Between Theory and Practice. *Journal of Cultural Analytics*, 4(3).

Luís Miguel Botelho. 2021. A guided journey through non-interactive automatic story generation. *ArXiv*, abs/2110.11167.

J. G. Cawelti. 2014. *Adventure, Mystery, and Romance Formula Stories as Art and Popular Culture*. The University of Chicago Press.

Mike Chen and Razvan C. Bunescu. 2021. Changing the Narrative Perspective: From Deictic to Anaphoric Point of View. *Inf. Process. Manag.*, 58:102559.

John Joon Young Chung, Joseph Jay Williams, and Juho Kim. 2018. Collaborative crowdsourcing between experts and crowds for chronological ordering of narrative events. *Korean HCI Society Conference*, pages 621–626.

Maximilian Droog Hayes, Geraint Wiggins, and Matthew Purver. 2018. Automatic Detection of Narrative Structure for High-Level Story Representation. In *5th AISB Symposium on Computational Creativity*.

Joshua D. Eisenberg and Mark Finlayson. 2021. Narrative Boundaries Annotation Guide. *Journal of Cultural Analytics*, 6.

Sam Fields, Camille Cole, Catherine Oei, and Annie Chen. 2022. Using named entity recognition and network analysis to distinguish personal networks from the social milieu in nineteenth-century Ottoman–Iraqi personal diaries. *Digital Scholarship in the Humanities*, 38.

R. Gaizauskas and T. Alrashid. 2019. SceneML: A Proposal for Annotating Scenes in Narrative Text. In *Workshop on Interoperable Semantic Annotation*, page 13.

G. Genette, J. E. Lewin, and J. D. Culler. 1980. Narrative discourse : an essay in method. *Comparative Literature*, 32:413.

E. Gius, F. Jannidis, M. Krug, A. Zehe, A. Hotho, F. Puppe, J. Krebs, N. Reiter, N. Wiedmer, and L. Konle. 2019a. Detection of Scenes in Fiction. In *Book of Abstracts of the Digital Humanities conference*, Utrecht, Netherlands.

E. Gius, M. Willand, and N. Reiter. 2021. On organizing a shared task for the digital humanities – conclusions and future paths. *Journal of Cultural Analytics*, 6(4):1–28.

Evelyn Gius, Nils Reiter, and Willand Marcus. 2019b. Cultural Analytics: A Shared Task for the Digital Humanities: Annotating Narrative Levels. *Journal of Cultural Analytics*, 4(3).

Adam Hammond. 2021. Annotation Guidelines for Narrative Levels. *Journal of Cultural Analytics*, 6.

Edward Kearns. 2019. For Annotating Anachronies and Narrative Levels in Fiction. *Journal of Cultural Analytics*, 4.

Edward Kearns. 2020. Annotating and quantifying narrative time disruptions in modernist and hypertext fiction. In *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*, pages 72–77, Online. Association for Computational Linguistics.

Edward Kearns. 2021. Annotation Guidelines For Narrative Levels, Time Features, and Subjective Narration Styles in Fiction (SANTA 2). *Journal of Cultural Analytics*, 6.

N. Ketschik, B. Krautter, S. Murr, and Y. Zimmermann. 2021. On the theory of narrative levels and their annotation in the digital context. *Journal of Cultural Analytics*, 6:69–97.

Nora Ketschik, Benjamin Krautter, Sandra Murr, and Yvonne Zimmermann. 2019. Annotating Narrative Levels in Literature. *Journal of Cultural Analytics*, 4.

James Lee. 2017. Shakespeare's Tragic Social Network; or Why All the World's a Stage. *DHQ: Digital Humanities Quarterly*, 11(2).

Stephanie Lukin and Marilyn Walker. 2019. A Narrative Sentence Planner and Structurer for Domain Independent, Parameterizable Storytelling. *Dialogue and Discourse*, 10:34–86.

Andreea Macovei. 2017. Capturing the Evolution of Characters in Novels. In *2017 21st International Conference on Control Systems and Computer Science (CSCS)*, pages 545–552.

Semi Min and Juyong Park. 2016a. *Complex Networks VII*, volume 644 of *Studies in Computational Intelligence*, chapter Network Science and Narratives: Basic Model and Application to Victor Hugo's Les Misérables. Springer.

Semi Min and Juyong Park. 2016b. Mapping Out Narrative Structures and Dynamics Using Networks and Textual Information. *CoRR*, abs/1604.03029.

Semi Min and Juyong Park. 2016c. Narrative as a Complex Network : A Study of Victor Hugo's Les Misérables. In *PROCEEDINGS OF HCI KOREA 2016*, pages 100–107.

K. M. Newberry and H. R. Bailey. 2019. Does semantic knowledge influence event segmentation and recall of text? *Memory and Cognition*, 47:1173–1187.

J. Porteous, F. Charles, and M. Cavazza. 2016. Plan-based narrative generation with coordinated subplots. *Frontiers in Artificial Intelligence and Applications*, 285:846–854.

Anna Rogers, Marzena Karpinska, Ankita Gupta, Vladislav Lialin, Gregory Smelkov, and Anna Rumshisky. 2024. NarrativeTime: Dense Temporal Annotation on a Timeline. In *LREC-COLING 2024*, pages 12053–12073. ELRA Language Resource Association.

M. Ryan. 1992. Possible Worlds, Artificial Intelligence, and Narrative Theory. *Special Issue: Between Science & Literature*, 23(2):135–139.

Benjamin M. Schmidt. 2015. Plot arceology: A vector-space model of narrative structure. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 1667–1672.

Kristiina Taivalkoski-Shilov. 2019. Free indirect discourse: an insurmountable challenge for literary MT systems? In *Proceedings of the Qualities of Literary Machine Translation*, pages 35–39, Dublin, Ireland. European Association for Machine Translation.

Tzvetan Todorov. 1971. The 2 Principles of Narrative. *Diacritics*, 1(1):37.

Mats Wirén and Adam Ek. 2021. Annotation Guideline No. 7 (Revised): Guidelines for Annotation of Narrative Structure. *Journal of Cultural Analytics*, 6.

Albin Zehe, Leonard Konle, Lea Katharina Dümpelmann, Evelyn Gius, Andreas Hotho, Fotis Jannidis, Lucas Kaufmann, Markus Krug, Frank Puppe, Nils Reiter, Annekea Schreiber, and Nathalie Wiedmer. 2021a. Detecting Scenes in Fiction: A new Segmentation Task. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3167–3177. Association for Computational Linguistics.

Albin Zehe, Leonard Konle, Svenja Guhr, Lea Katharina Dümpelmann, Evelyn Gius, Andreas Hotho, Fotis Jannidis, Lucas Kaufmann, Markus Krug, Frank Puppe, Nils Reiter, and Annekea Schreiber. 2021b. Shared Task on Scene Segmentation @ KONVENS 2021. In *STSS@KONVENS*.

A Final tagset

Tagset	Tags	Properties	Values
Diegetic level*	NARRATOR	EXTRADIEGETIC*	0*, meta
		INTRADIEGETIC	1a*, 1b, 1c, ...
		METADIEGETIC	2a*, 2b, ..., 3a, 3b, ...
Segment	SCENE	TIME*	e.g., evening
		PLACE*	e.g., crime scene
		CHARACTER_CONSTELLATION*	characters in scene
	NON-SCENE	SUMMARY	
		SCENIC_PASSAGE	
		DESCRIPTIVE_PASSAGE	
	ELLIPSIS		
Anachronisms	ANALEPSIS	EMBEDDED	
		INTERRUPTIVE	
	PROLEPSIS	EMBEDDED	
		INTERRUPTIVE	
Perspective*	FOCALIZATION	EMBEDDED	
		INTERRUPTIVE	
		DETECTIVE_THOUGHT†	
	VOICE*	HOMODIEGETIC	
		HETRODIEGETIC	
Misc†	CLUE	IDENTIFIED	e.g., murder weapon
		REFERRED	e.g., murder weapon
Acts†	INTRODUCTION		
	INVESTIGATION		
	CONFRONTATION		
	CONFESSION		
	REVEAL		
	AFTERMATH		

Tagsets and the DETECTIVE_THOUGHT tag marked with a dagger (†) are modular and specific to whodunit texts. Tagsets marked with an asterisk (∗) are compulsory for each text. Properties marked with an asterisk (∗) are compulsory if the related tag was chosen and values marked with an asterisk (∗) are compulsory if the related property was chosen.

B Distribution of tags in annotated texts

Property	C1	C2	D1	D2
Diagetic level				
Extradiegetic level – 0	1‡	1‡	1‡	1‡
Extradiegetic level – Meta	1	0	2	1
Intradiagetic level	2	3	19	10
Metadiagetic level	1	2	18	9
Segment				
Scene	11	12	5	10
Non-Scene – Summary	3	5	5	1
Non-Scene – Description	0	1	1	0
Non-Scene – Scenic passage	0	0	0	0
Ellipsis	3	2	1	3
Anachronism				
Analepsis – Embedded	1	4	3	5
Perspective				
Voice – Homodiegetic	1‡	1‡	1‡	1‡
Embedded focalization	1	4	18	9
Detective thoughts	4	10	4	9
Acts				
Introduction	1	1	1	1
Investigation	1	1	1	1
Confrontation	0	1	1	1
Confession	0	0	0	1
Reveal	1	1	1	1
Aftermath	1	1	1	0
Misc				
Clues	8	9	14	18

Entries marked with a double dagger (‡) are assigned to the entire text. C1 is *The adventures of the Italian nobleman*, C2 is *The jewel robbery at the grand metropolitan*, D1 is *A case of identity*, and D2 is *The red-headed league*.