

Narrative in Short German Prose: A Multi-Phenomenon Dataset for Computational Literary Analysis

Hans Ole Hatzel^{*1}, Haimo Stierner^{*2},
Evelyn Gius², Chris Biemann¹

University of Hamburg¹, TU Darmstadt²

{hans.ole.hatzel, chris.biemann}@uni-hamburg.de,

{haimo.stierner, evelyn.gius}@tu-darmstadt.de

Abstract

We present the novel dataset GermAnProse, an annotated corpus consisting of four German short prose texts accompanied by an extensive set of narrative-focused annotations. As part of this dataset, we contribute an annotation scheme for mentions, speech, and character agency: Characters in Action (ChiA). GermAnProse also contains information on narrative phenomena: narrativity, semantic verb classes, and plot keyness. Moreover, we include reader reception data in the form of timing information for audiobook performances, indicating pauses between sentences and the time taken to read a specific sentence in a performance. We release the dataset, which contains more than 18,000 manually created standoff annotations in JSON format, enabling researchers to utilize this resource for further exploratory applications.

1 Introduction

In recent years, the computational analysis of literature has gained increasing attention, as evidenced by the emerging field of Computational Literary Studies (CLS). Our novel corpus GermAnProse (short for **German annotated prose**), contains annotations of a series of narrative phenomena for four literary texts and is meant to serve the field. Figure 1 showcases our annotations on a toy example. Specifically, our dataset contains annotations for narrativity (Vauth et al., 2021), character mentions and agency, semantic verb classes (Hatzel et al., 2025), plot keyness (Hatzel et al., 2023), scenes (Zehe et al., 2021), and audiobook timing information. With this publication, we pursue the goal of providing a dataset with rich annotations for performing explorative studies of narrative phenomena with audiobook information, in the form of the time taken to read a sentence and the length of pauses between sentences, serving as a reader-reception-focused extension. In the established CLS process

^{*}Equal contribution.

Keyness: 0.5
Reading Speed: 1.2

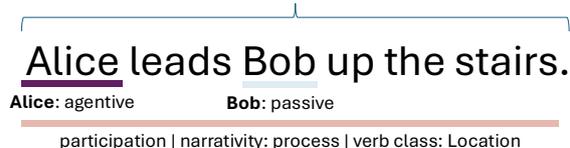


Figure 1: In our dataset, phenomena are annotated at a sentence level (plot keyness and reading speed) or at a phrase level (agency, narrativity, and verb classes). The roles of characters in a given agency annotation are identified on a span level; our scene annotations are not shown here. The example shows an interaction of *Alice* and *Bob* in which *Alice* is the active participant.

of operationalization, annotation, automation, and exploration (Pichler and Reiter, 2020), we envision our dataset to serve as an initial exploration resource. Researchers can use the annotated texts to identify interesting phenomena and explore their interaction, guiding their subsequent research on larger corpora. Moreover, the dataset can serve as ground truth for the development and evaluation of automated approaches. We see this work as part of a larger trend with drastic improvements in natural language processing, allowing the field to increasingly shift its focus from low-level local linguistic phenomena to higher-level narrative and semantic phenomena.

The dataset is provided in JSON format with text and corresponding character-level standoff annotations, allowing for a low barrier of entry and easy exploration in any scripting environment.¹

2 Related Work

Here, we will only discuss prior work that does not integrate into our dataset. Our dataset contains a range of previously released annotations that we merge into one unified dataset, see Section 3.1 for details on this data.

¹<https://github.com/forTEXT/GermAnProse>

CLS has a strong tradition of reusable datasets; largely, these datasets focus on providing collections of plain texts or texts in TEI-XML with formatting information in conjunction with limited metadata (Gius et al., 2021a; Schöch et al., 2021, e.g.). The DraCor initiative (Fischer et al., 2019), for example, focuses on drama and provides the texts in TEI-XML with metadata, such as publication year, and limited annotations, such as speaker identification. Often CLS researchers publish annotations for a single phenomenon of interest on a limited set of project-specific texts (Underwood, 2018; Piper and Bagga, 2022; Coll Ardanuy et al., 2020, e.g.). Hagedorn and Darányi (2022) set out to create a dataset of folk tales that is extensible by other researchers.

Audiobooks have received limited attention as a resource in NLP research. Largely, approaches focused on audiobooks as a speech data source (Székely et al., 2012; Boeffard et al., 2012). Stiemer et al. (2025) introduced the idea of using audiobook information as a means of analysis for the original text to the CLS community. The public domain audiobooks on Librivox² have long attracted the interest of the research community in the humanities and natural language processing alike (Weber, 2021; Beilharz et al., 2020).

3 Corpus & Annotations

In our selection of texts, we largely follow the selection by Vauth et al. (2021), including representation of modern German literature while relying only on out-of-copyright texts (see Table 4 for details). This selection enables us to build upon their published annotations for our own work. Further, for resource reasons, we limit our selection to short narratives with the longest text consisting of just over 22,000 tokens. We annotate the texts in their entirety, rather than excerpts of long texts like some prior work (Bamman et al., 2020, e.g.); we want to, thereby, enable the analysis of entire self-contained narratives rather than focusing on local linguistic features in service of future automation. Here, we depart from the text selection by Vauth et al. (2021) in removing the annotated excerpt of the much longer text *Effi Briest*, including instead the full text *Der blonde Eckbert*; as a result, all our texts are fully annotated. The annotated texts originate either from the TextGrid corpus (TextGrid, 2021) or the d-Prose 1870-1920 corpus (Gius et al., 2021a). In Table 1

²<https://librivox.org>

we provide a full list of the works with relevant metadata.

Table 2 provides an overview of the source and quality of the annotations we discuss in detail below.

3.1 Pre-existing Data

In this section, we describe pre-existing published annotations that are included in our dataset and give a rationale for including each set of annotations.

3.1.1 Narrativity

Narrativity or eventfulness refers to the quantification of “how much” is happening in a narrative text passage. Our dataset contains narrativity annotations as operationalized in the EvENT concept by Gius and Vauth (2022). In the approach each finite verb in the texts is annotated as expressing one of four *event* types: change of state, process, state, or non-event. The event types, in the order listed above, are associated with decreasing narrativity (Vauth et al., 2021). Events are not derived from semantic units – as the concept is typically understood in natural language processing (see e.g. Doddington et al., 2004) – but are instead informed by their textual representation. As a result, the concept can be understood as the first step in a bottom-up approach to event semantics, ranging from simple and text surface-based phenomena to more complex semantic phenomena. Change of state events mark the most eventful moments in a text, while non-events indicate eventlessness (since they do not denote actual events in the fictional world but rather descriptive passages, questions, or counterfactual statements).

3.1.2 Semantic Verb Classes

The verb class data were generated to provide a coarse-grained semantic profile of individual literary texts based on the meanings of their verbs (Hatzel et al., 2025). Following the narrativity annotations (Vauth and Gius, 2021), verb phrases were used as the annotation unit, with each annotation semantically describing the finite verb in a phrase. Building on this event-centered segmentation, each main verb was classified according to the semantic field of its lemma in GermaNet, resulting in one of fifteen verb classes such as *communication*, *emotion*, *location*, *change*, or *perception*. Ambiguous or missing verb senses were resolved through contextual interpretation to maintain internal consistency across each annotated text. The classes represent broad categories of actions and processes

Text	Author	Published	Short Name
Der blonde Eckbert	Ludwig Tieck	1796	Eckbert
Das Erdbeben in Chili	Heinrich von Kleist	1807	Erdbeben
Krambambuli	Marie von Ebner-Eschenbach	1896	Krambambuli
Die Verwandlung	Franz Kafka	1915	Verwandlung

Table 1: We list the four texts in our dataset with their respective publication dates and original author names. For reference, we also supply short-form names that are used to identify the texts in some of our tables.

Annotations	Publication	Quality
Verbclasses	Hatzel et al. (2025)	
Keyness	Hatzel et al. (2023)	
Scenes	Ours	
Speech Time	Ours	
Agency ^{ChiA}	Ours	
Mentions ^{ChiA}	Ours	
Narrativity	Vauth et al. (2021)	

Table 2: We list the annotation source and quality across the different phenomena annotated in our dataset.  refers to gold data,  refers to automated annotations, and  refers to single annotator annotations. The ChiA superscript denotes data that is part of our novel annotation scheme.

and were selected to capture the central event types that shape narrative meaning.

For each literary work, Hatzel et al. (2025) aggregated the resulting annotations to produce text-specific verb class distributions. The resulting frequency profiles are used as interpretable summaries of a text’s semantic structure, highlighting tendencies such as an emphasis on emotional, social, or physical processes.

We expect that the verb classes can be used as a first order approximation of action semantics in exploratory text analysis.

3.1.3 Plot Keyness

With “plot keyness” we refer to a given sentence’s relevance to the plot of the overall story. For example, a passage describing the natural beauty of the scenery will, in most cases, not contain the most important actions of the story. The data captures plot keyness by using a corpus of summaries for measuring how many of them reference a given passage of text (Hatzel et al., 2023). Conceptually, this data builds on the ideas of the Pyramid method (Nenkova and Passonneau, 2004), asserting that the most relevant events are mentioned in the greatest number of summaries. For the plot keyness data, we use text passages instead of atomic event units. In

the downstream analysis of our dataset, we expect this to be used to identify low-level textual features for relevance to the overall story.

Specifically, our plot keyness measure is the fraction of summaries in our dataset that reference a given sentence. A value of 1.0 means all summaries refer to a specific sentence, whereas a value of 0.5 means only half of all summaries reference the events in the sentence.

The plot keyness and its relation to concepts like narrativity is a phenomenon of interest for CLS. We showcase the exploration of this relationship on our data (see Section 4).

3.2 Novel Annotations

In this section, we describe new annotations we contribute in this publication that are based on existing annotation schemes. The newly created annotations are marked as *ours* in Table 2.

3.3 Scenes

To enable the aggregation and contextual analysis of the data compiled in this dataset across selected text segments, we divided the corpus texts into scenes. The guidelines applied for this purpose (Gius et al., 2021b) are based on the narratological understanding of a scene as a narrative segment that constitutes a coherent unit of time, space, action, and character constellation—roughly corresponding to a film shot in terms of its continuity.

The guidelines specify how scene boundaries can be identified by examining changes in these four aspects. They also provide clear criteria for distinguishing between narrative scenes and non-scenic passages such as summaries, reflections, or general descriptions. To ensure reliable and consistent annotations, the guidelines assign different weights to each aspect, prioritizing the coherence of events and character constellations over the temporal or spatial continuity.

In addition, the guidelines define lexical and structural indicators of scene changes (e.g., “suddenly”, “the next morning”), recommend the use

of the film-shot analogy as an intuitive test, and explain how to handle transitional or ambiguous passages. The resulting annotations enable the segmentation of literary texts into meaningful narrative units, thereby allowing for a systematic analysis of plot structure and narrative rhythm across different genres and styles. Scenes are one of the aspects of our work that can potentially be extended to large corpora.

3.4 Audiobook Timing

Audiobook information serves as additional reception-based data in our dataset, allowing us to connect purely narrative phenomena to a simple form of reception representation. Potentially, readers’ interpretations of a text can hint at deeper phenomena, and analyzing the relationship between narrative phenomena and reader presentation is a worthwhile endeavor. We obtain our data from amateur readers (from the platform Librivox or read by student assistants) and professional readers in commercial audiobooks. Each text and its audio performances are put through a forced alignment pipeline to obtain token-level timing information.³ We break this information down to the sentence level and provide a reading speed comparison as measured against a text-to-speech (TTS) system that is based on public training data (Müller and Kreutz, 2021).⁴ While the speech synthesis has dealt with explicitly modeling speech speed given phonemes (Ni et al., 2019), we take a comparatively simple approach in relying on the duration of a single speech generation. Our approach is potentially sensitive to individual terms that the TTS system fails to reproduce accurately on a story-analysis level. If such a term is central to a given text and, for example, co-occurs with major events at an increased rate, it is conceivable that the approach could adversely affect downstream analysis.

3.5 Mentions, Direct Speech and Character Agency

We developed annotation guidelines for systematically capturing how characters act and interact as agents within narrative texts; the guidelines are referred to as characters in actions (ChiA). To maintain compatibility with the narrativity and semantic verb class data, we follow the two annotation schemes in regard to the annotation unit. That is to

say, for each verbal phrase, the character actions associated with the verb are modeled. Our guidelines specify that these actions encompass all forms of activity within the fictional world, including mental and emotional processes, speech, and physically manifested actions.

Our annotations chiefly focus on different degrees of agency, describing the extent to which a character actively shapes an event: ranging from intentional and controlled actions (agentive) to involuntary or imagined actions (low-agency), and further to situations in which the character is merely affected by or influenced by events (passive). By distinguishing between these levels, the guidelines enable the tracing of narrative dynamics and the influence of characters throughout the text. We refer to these annotations that capture co-presence and the agentivity of characters on a verbal phrase level as the **agency** annotations.

In addition to character agency, we annotate character mentions in texts. For the mention extents, we adopt an existing annotation scheme for character mention span extents (Krug, 2020). Beyond that, we differ from prior work in that we only annotate individual characters (potentially multiple for one span) rather than annotating collective mentions on a group level (e.g. “the three friends” would refer to three separate characters in our annotation scheme). Further, we introduce two special characters *background character* and *crowd* that are used for non-specific or non-reoccurring characters and unspecified groups of people, respectively. Figure 1 illustrates a toy example of an active and a passive character. For real world example from our dataset see Section 3.7 As illustrated, character mentions are also annotated with a canonical name for the corresponding character, as the mention’s surface form does not always unambiguously identify a character.

In addition to character mentions and agency, direct speech is also annotated, and text inside speech acts is not considered for agency annotations. All three tags are linked to their corresponding characters; in the case of agency, the particular roles (agentive, low-agency, passive) of each character are also annotated. Taken together, these annotations provide a structured representation of how characters appear, act, and speak, thereby enabling a systematic analysis of their narrative functions and their contribution to plot development.

³<https://github.com/feldberlin/timething>

⁴<https://github.com/OHF-Voice/piper1-gpl>

3.5.1 Agreement

For the purposes of calculating inter-annotator agreement, we treat the identification of agentive, low-agency, and passive participants as a binary classification problem for each character in each annotation unit. In terms of Krippendorff’s Alpha, in a weighted average setup, we achieve agreements of up to 0.6 in identifying characters in the agentive class. For the low-agency and passive classes, by contrast, we observe much lower agreement, ranging from 0.35 to 0.51 (see Table 3 for a breakdown by class and text). Qualitatively speaking, we can attribute many of the errors to the special *background character* and *crowd* characters; we observe that it can be difficult to discern, especially for larger groups of people, who, besides the main characters, are included in an action.

For the mutual accuracy metric, we calculate the accuracy across all pairs of span-wise matching annotations, macro-averaging across annotator pairs. This results in a perhaps more interpretable measure: showing that annotators agree on the exact set of characters in approximately 75% of cases.

3.6 Annotation Workflow

Both the scene and ChiA annotations were manually created using the web-based tool CATMA (Gius et al., 2025) in conjunction with the GitMA package (Vauth et al., 2022) for programmatic processing. All annotators were students of literary studies and employed as student assistants or were full-time researchers and co-authors of this paper.

For the ChiA guidelines, extensive pilot annotations accompanied by annotator training were performed, iteratively improving the guidelines. At least two annotators annotated each text, and for all but one text, at least two annotators annotated the text in its entirety, in addition to another annotator creating the gold standard. Regular coordination meetings were held to discuss and resolve ambiguous annotation cases, which were additionally documented using a dedicated uncertainty tag. Based on the resulting annotations, gold-standard annotations were produced by an additional annotator who reconciled inconsistencies. To facilitate this process, we developed a semi-automatic adjudication pipeline that inserted annotations on which both annotators agreed (on both the spans and properties of an annotation) into a new collection in CATMA. All annotations automatically created in this manner were subsequently carefully reviewed. We found

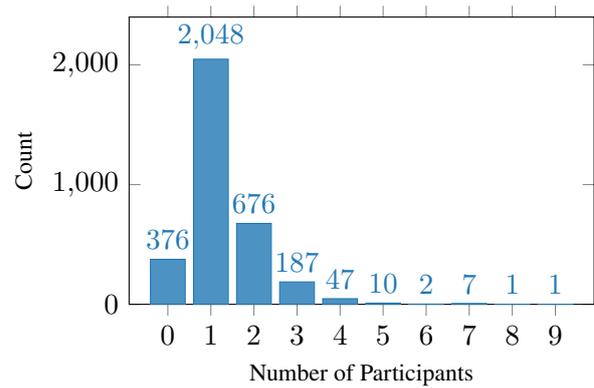


Figure 2: The distribution of the number of participants in each agency annotation in our data is shown. Each annotation includes an average of 1.27 participants. Very few annotations have more than three participants.

the approach to vastly speed up the creation of the gold labels.

We have around 3,400 gold agency annotations across the four texts in our dataset. For each agency annotation in our gold data, 0.40 agentive characters, 0.57 low-agency characters, and 0.30 passive characters are annotated on average; this results in a total of 1.27 characters per annotation, see Figure 2 for the distribution. The maximum number of characters participating in one agency annotated phrase in our dataset is nine, occurring in a sentence “When **they** arrived in the Dominicans’ church” from *Das Erdbeben in Chili*, in which a large portion of the story’s cast is involved.

In terms of co-occurrence, there is a sizable number (376) of agency annotations with no characters involved, as each verbal phrase is an annotation unit regardless of its semantics. For example, the phrase “The weather is nice.” has no participants.

3.7 Example

Figure 3 shows an example of real annotations from our dataset. Note that the single sentences consist of multiple verbal phrases, each with its own narrativity and participation annotations. Further, the example illustrates that not every participant must have a surface representation in the phrase in which they appear. The phrase “ließ aber eine Flasche [...] bringen”, for example has the character *Hopp* perform an action. In practice, unlike in the visualization, our annotators label character identities rather than their spans. However, since all character mentions are separately annotated, the corresponding text mention spans—as depicted in our graphics—can still be retrieved.

	Mutual Accuracy			Krippendorff’s Alpha		
	Erdbeben	Verwandlung	Krambambuli	Erdbeben	Verwandlung	Krambambuli
passive	0.722	0.762	0.731	0.381	0.444	0.509
low-agency	0.684	0.653	0.674	0.354	0.443	0.409
agentive	0.730	0.785	0.732	0.510	0.603	0.590

Table 3: The agreements for annotating character agency as measured by mutual accuracy and Krippendorff’s Alpha, show that generally speaking, the agentive class is annotated most consistently.

Text	Number of						
	Tokens	Sentences	Scenes	Agency	Mentions	Speech Acts	Narrativity
Der blonde Eckbert	7508	314	14	259	1205	20	930
Das Erdbeben in Chili	6499	224	14	627	1064	39	712
Krambambuli	4573	226	16	477	628	39	571
Die Verwandlung	22299	850	26	1992	2884	106	2363
Total	40879	1614	70	3355	5781	204	4576

Table 4: Annotation Overview. All annotation schemes cover each of the four texts; we omit some of the annotation schemes for brevity.

4 Data Exploration

After laying out the components of our dataset, we perform a preliminary analysis of the data. Table 4 shows a range of dataset statistics.

Narrativity and Plot Keyness For narrativity, as measured by event scores (Vauth et al., 2021), we find a very weak correlation with plot keyness ($r=.067$, $p<0.05$). The results indicate some connection between the two phenomena, but, as captured by our data, there is no clear discriminative value for either variable in predicting the other. Narratologically speaking, we expect to find a limited correlation as keyness captures plot relevance, where narrativity only captures narration at a local level; yet, one might expect non-narrative passages to only rarely constitute key passages. Hatzel et al. (2023) came to a similar conclusion, based on comparable data.

Semantic Verb Classes and Narrativity In Figure 4, we illustrate the distribution of event types in ascending order of narrativity with each semantic verb class. We find that some classes, particularly *Stative Verbs*, *Cognition*, and *Emotion*, are associated with lower narrativity. Perhaps most notably, the classes *Location*, *Communication*, and *Contact* are associated with higher narrativity.

Pauses, Reading Speed, and Plot Keyness We examine the phenomenon of pauses and reading speed in the audiobook timing information contained in ChiA-Seeds. The idea was initially ex-

plored by Stiemer et al. (2025), and we expand on their analysis using additional texts and introduce the concept of reading speed to the analysis.

In our data, professionals make longer pauses than amateur readers, with an average length of 0.87 and 0.64 seconds, respectively. The two groups read at an equal pace when considering only the sentences themselves, without taking into account the pauses (1.25 times faster than the TTS system for the professional readers and 1.26 times faster for the amateurs). For all readers in our data, we observe a very small but significant correlation between the pause after a sentence and its plot keyness ($r=.05$, $p<0.05$). The same effect can be observed for the correlation of reading speed and plot keyness, but it does not rise to statistical significance. Sentences that contain direct speech are also read slightly faster in our dataset ($r=.04$, $p<0.05$). Further work is required to identify potential differences in presentations by amateur readers compared to professional readers.

Character Interaction As our character agency annotations capture the co-presence of characters in verbal phrases, we can use the annotations to build a novel kind of character network. In conjunction with the scene annotations, the data can also be used to create dynamic character networks, changing with each scene. We show an instance of a character network for a subset of the scenes in *Das Erdbeben in Chili* in Figure 5. So far, high-quality data for performing character network visualization on prose texts has not been available, with many

Herr Hopp sagte vorerst kein Wort von dem Wohlgefallen, das er an dem Hunde gefunden hatte,

participation | Communication | non-event

Hobb: passive

participation | Cognition | stative event

Hobb: agentive

Krambambuli: passive

ließ aber eine Flasche von dem guten Danziger Kirschbranntwein bringen,

participation | Social | process | Hopp: agentive

den der Löwenwirt damals führte, und schenkte dem Vazierenden fleißig ein.

participation | Social | process

Wirt: low agentive

participation | Change | process | Hopp: agentive

Forstgehilfe: passive

Figure 3: Example annotations for a single sentence from *Krambambuli*. The sentence has a keyness of 0.25 and a reading speed of 1.36 in our data. Translation: “Mr. Hopp said for the time being not a word of the pleasure that he had found in the dog, but had a bottle of the good Danzig cherry brandy brought, which the Lion-innkeeper carried at that time, and poured for the wanderer diligently.”

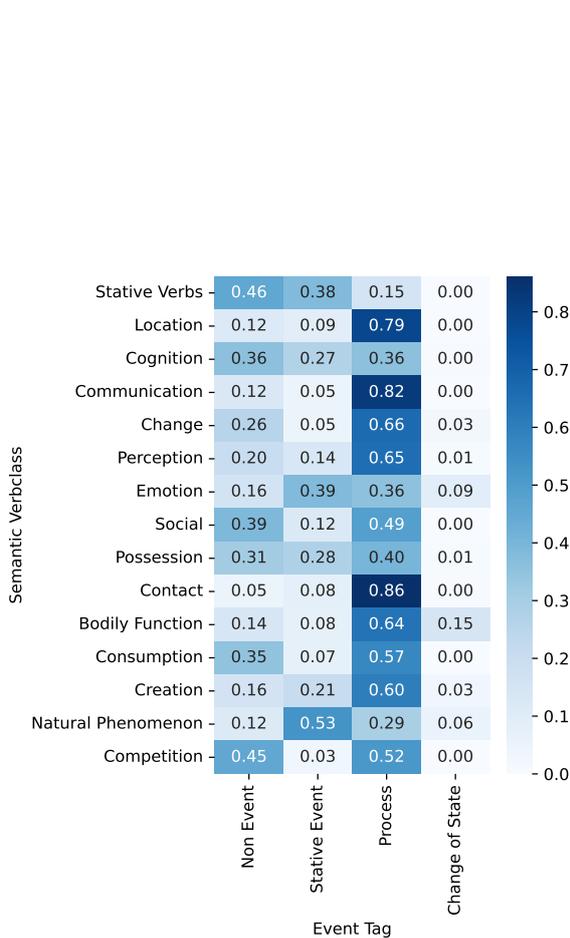


Figure 4: The event tags are listed in ascending order of narrativity. For each verb class, we show the relative distribution of event tags within its occurrences.

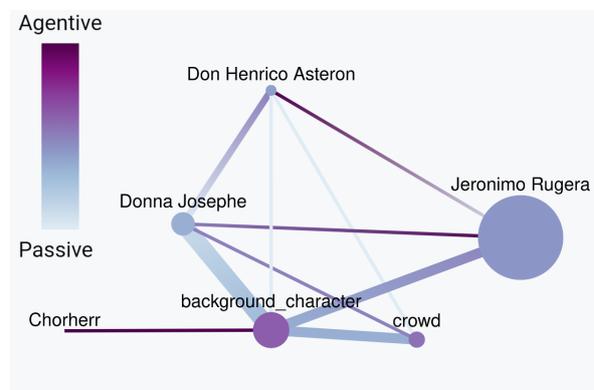


Figure 5: In our character networks with agency data, the color of nodes describes character agency, while the colors at the ends of each describe each character’s agency in only that specific interaction. Here, we show the character network for the first five scenes in *Das Erdbeben in Chili*.

prior approaches falling back on using only specific mentions (e.g., explicit names) and their co-occurrence (Ketschik, 2025) or using hybrid models that include coreference resolution with name clustering or additional rules (Perri et al., 2022, e.g.). Ketschik (2025) also points out that many prior approaches fail to account for differences in character mentions, a problem that our data, at least partially, resolves, as direct speech is not included, and our agency data captures character interactions (or, in the case of low-agency, co-presence).

5 Conclusion

In this work, we introduce GermAnProse, a novel dataset for exploring narrative phenomena in short narratives. Our extensive annotations cover a total of over 40,000 tokens, with more than 18,000 manually created span annotations in various categories,

out of which more than 9,000 are newly created for this work. In addition, we provide automated sentence timing information for audiobooks as aligned with the original texts, allowing a form of reader studies to be performed on our data.

In our data exploration, we show initial examples of identifying phenomena that warrant further investigation. We provide a novel dataset allowing for the exploration of narrative texts.

Despite its focus on narrative exploration, we believe that our dataset will also prove useful for established computational approaches such as character coreference resolution (Roesiger et al., 2018) or the attribution of direct speech (Petersen-Frey and Biemann, 2024). Likewise, we expect that our dataset can aid in the creation of automation for the character agency annotations. We recommend that all such approaches adopt text-level leave-one-out cross-validation schemes to verify their generalizability.

6 Limitations

We see the main limitation of our work in the limited size of the dataset. Resource-wise, creating a much larger dataset would not be feasible. In our view, the dataset is sufficiently large to enable the identification of potentially interesting effects, but might not support robust quantitative conclusions in most cases.

As mentioned, our audio speech assessments are potentially sensitive to systematic errors in the speech synthesis system we use. Further analysis, comparing across readers and potentially with other speech synthesis approaches, would be beneficial.

7 Ethical Considerations

In terms of legal considerations, all primary annotated works are out of copyright. The audiobooks we incorporated into the analysis are, in part, commercial products. We only report derived data (i.e., timing information) intended for analysis that does not allow a reconstruction of the original performance.

In terms of gender representation in our dataset, one of four texts is written by a female author (*Krambambuli* by Marie Ebner von Eschenbach).

References

David Bamman, Olivia Lewke, and Anya Mansoor. 2020. An Annotated Dataset of Coreference in English Lit-

erature. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 44–54, Marseille, France. European Language Resources Association.

Benjamin Beilharz, Xin Sun, Sariya Karimova, and Stefan Riezler. 2020. LibriVoxDeEn: A Corpus for German-to-English Speech Translation and German Speech Recognition. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3590–3594, Marseille, France. European Language Resources Association.

Olivier Boeffard, Laure Charonnat, Sébastien Le Maguer, and Damien Lolive. 2012. Towards Fully Automatic Annotation of Audio Books for TTS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 975–980, Istanbul, Turkey. European Language Resources Association (ELRA).

Mariona Coll Ardanuy, Federico Nanni, Kaspar Beelen, Kasra Hosseini, Ruth Ahnert, Jon Lawrence, Katherine McDonough, Giorgia Tolfo, Daniel CS Wilson, and Barbara McGillivray. 2020. *Living Machines: A study of atypical animacy*. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4534–4545, Barcelona, Spain (Online). International Committee on Computational Linguistics.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The Automatic Content Extraction (ACE) Program – Tasks, Data, and Evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Frank Fischer, Ingo Börner, Mathias Göbel, Angelika Hechtel, Christopher Kittel, Carsten Milling, and Peer Trilcke. 2019. *Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama*. In *Digital Humanities 2019: "Complexities" (DH2019)*, Utrecht. Zenodo.

Evelyn Gius, Svenja Guhr, and Benedikt Adelman. 2021a. *d-Prose 1870-1920*.

Evelyn Gius, Jan Christoph Meister, Malte Meister, Marco Petris, Dominik Gerstorfer, Mari Akazawa, and Stefanie Messner. 2025. *CATMA*.

Evelyn Gius, Carla Sökefeld, Lea Dümpelmann, Lucas Kaufmann, Annekea Schreiber, Svenja Guhr, Nathalie Wiedmer, and Fotis Jannidis. 2021b. *Guidelines for detection of scenes*.

Evelyn Gius and Michael Vauth. 2022. *Towards an event based plot model. a computational narratology approach*. *Journal of Computational Literary Studies*, 1(1).

- Joshua Hagedorn and Sándor Darányi. 2022. [Bearing a Bag-of-Tales: An Open Corpus of Annotated Folktales for Reproducible Research](#). *Journal of Open Humanities Data*.
- Hans Ole Hatzel, Evelyn Gius, Haimo Stierner, and Chris Biemann. 2023. [Narrativität und Handlung: Zum Verhältnis von Handlungszusammenfassungen und relevanten Ereignissen](#). In *DHd2023: Open Humanities, Open Culture*, Trier, Germany/Belval, Luxembourg. Zenodo.
- Hans Ole Hatzel, Haimo Stierner, Evelyn Gius, and Chris Biemann. 2025. Scalable Verb-Based Literary Semantics. In *Proceedings of the Workshop on Computational Humanities Research (CHR 2025)*.
- Nora Ketschik. 2025. [Netzwerkanalysen narrativer Texte - ein Vorgehensmodell](#). In *DHd 2025: Under Construction*, Bielefeld, Germany. Zenodo.
- Markus Krug. 2020. [Techniques for the automatic extraction of character networks in german historic novels](#). Released Online via OPUS Würzburg, accessed 2025-10-11.
- Thorsten Müller and Dominik Kreutz. 2021. [Thorsten - open german voice \(neutral\) dataset](#).
- Ani Nenkova and Rebecca Passonneau. 2004. Evaluating Content Selection in Summarization: The Pyramid Method. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, pages 145–152, Boston, Massachusetts, USA. Association for Computational Linguistics.
- Jinfu Ni, Yoshinori Shiga, and Hisashi Kawai. 2019. [Duration Modeling with Global Phoneme-Duration Vectors](#). In *Proc. Interspeech 2019*, pages 4465–4469, Graz, Switzerland.
- Vincenzo Perri, Lisi Qarkaxhija, Albin Zehe, Andreas Hotho, and Ingo Scholtes. 2022. One Graph to Rule them All: Using NLP and Graph Neural Networks to analyse Tolkien’s Legendarium. In *Proceedings of the Computational Humanities Research Conference 2022*, volume 3290 of *CEUR Workshop Proceedings*, pages 291–317, Antwerp, Belgium. CEUR.
- Fynn Petersen-Frey and Chris Biemann. 2024. Fine-grained quotation detection and attribution in German news articles. In *Proceedings of the 20th Conference on Natural Language Processing (KONVENS 2024)*, pages 196–208, Vienna, Austria. Association for Computational Linguistics.
- Axel Pichler and Nils Reiter. 2020. [Reflektierte Textanalyse](#). In *Reflektierte algorithmische Textanalyse: Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, pages 43–60. De Gruyter.
- Andrew Piper and Sunyam Bagga. 2022. Toward a Data-Driven Theory of Narrativity. *New Literary History*, 54(1):879–901.
- Ina Roesiger, Sarah Schulz, and Nils Reiter. 2018. Towards Coreference for Literary Text: Analyzing Domain-Specific Phenomena. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 129–138, Santa Fe, New Mexico. Association for Computational Linguistics.
- Christof Schöch, Roxana Patras, Tomaž Erjavec, and Diana Santos. 2021. [Creating the European Literary Text Collection \(ELTeC\): Challenges and Perspectives](#). *Modern Languages Open*, 0(1):25–44.
- Haimo Stierner, Hans Ole Hatzel, Chris Biemann, and Evelyn Gius. 2025. [Pause im Text. Zur Exploration semantisch konditionierter Sprechpausen in Hörbüchern](#). In *Book of Abstracts - DHd 2025*, Bielefeld, Germany.
- Éva Székely, Joao Paulo Cabral, Mohamed Abou-Zleikha, Peter Cahill, and Julie Carson-Berndsen. 2012. Evaluating expressive speech synthesis from audiobook corpora for conversational phrases. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3335–3339, Istanbul, Turkey. European Language Resources Association (ELRA).
- TextGrid. 2021. Die digitale bibliothek bei textgrid. <https://textgrid.de/de/digitale-bibliothek>.
- Ted Underwood. 2018. Why Literary Time is Measured in Minutes. *ELH*, 85(2):341–365.
- Michael Vauth and Evelyn Gius. 2021. [Richtlinien für die Annotation narratologischer Ereigniskonzepte](#). Zenodo.
- Michael Vauth, Hans Ole Hatzel, Evelyn Gius, and Chris Biemann. 2021. Automated Event Annotation in Literary Texts. In *CHR 2021: Computational Humanities Research Conference*, volume 2989 of *CEUR Workshop Proceedings*, pages 333–345, Amsterdam, The Netherlands.
- Michael Vauth, Malte Meister, Hans Ole Hatzel, Dominik Gerstorfer, and Evelyn Gius. 2022. [GitMA](#).
- Millicent Weber. 2021. “Reading” the Public Domain: Narrating and Listening to Librivox Audiobooks. *Book History*, 24(1):209–243.
- 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. 2021. 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, Online. Association for Computational Linguistics.