

# Detection of Scenes in Fiction

Narratology usually defines the event as the smallest unit of a narrative, but computational literary studies has problems to come up with a workable definition and operationalization of event for its purposes (Meister, 2003) . This has proven to be a serious obstacle for research on narrative texts, because similarity of plot is an important aspect for many operations, for example genre classification. Recent operationalizations like syuzhet (Jockers, 2015) capture the emotional arc but not the plot itself. But it is quite obvious even to non-narratologists that a main component especially of conventional fiction consists of parts which contain a lot of direct speech and detailed descriptions of actions. These parts are usually called 'scenes' and in our context they can be understood similarly to plays as "a division [...] during which the action takes place in a single place without a break in time" (Learner's Dictionary 2018) including the aspect that it is "a part of a play, movie, story, etc., in which a particular action or activity occurs" (Learner's Dictionary 2018).

A narrative text doesn't consist of scenes alone. It can be seen as a continuous alternation between summary and scene and usually also has some ellipses - that is events not told - and pauses , for example in descriptions. Genette, who proposed these four classes, famously based his definition of 'scene' on the aspect of time alone: "scene, most often in dialogue, which, as we have already observed, realizes conventionally the equality of time between narrative and story" (Genette, 1980) . It is worthwhile to point out that this equality is "only a kind of conventional equality between narrative time and story time" (Genette 1980: 87), with a lot of leeway. This reduction is useful in the context of his theory, but as the basic component of plot we need a definition which is more geared towards the representation of action and which includes more

information about action than just time.

So we start out from the fuller understanding of a scene outlined above. We are interested in a segment of the discours (presentation) of a narrative which presents a part of the histoire (chronologically ordered, causally connected events in the narrated world) in such a way that a) time is more or less equal in discours and histoire , b) place stays - more or less - the same c) it centers around a particular action, and d) the character configuration is - again: more or less - equal. We understand our panel as a first step towards modeling, annotating and automatically detecting scenes as such multi-dimensional phenomena. In this first step we will look into a) approaches for manually annotating scenes, b) the relation between just one aspect, character or time, with scenes and c) a machine learning approach for the detection of scenes. We start our research with a simpler form of literature: dime novels ('Heftromane'). Obviously the next step after detecting scenes would be to label them in an informative way, which would allow to see them as part of a plot, but that is beyond the scope of this panel.

## 1. Segment Annotation

The high complexity and context-dependence of the notion of 'scene' makes manual annotation and automatic detection difficult. Establishing manually annotated data, however, is a stepping stone and prerequisite for developing automatic detection systems (unless annotations can be inferred from other elements as shown below).

Intersubjective annotation of narrative segments is difficult to achieve. In a series of experiments, Reiter (2015) reports Fleiss'  $\kappa$  scores of below 0.3 for non-experts (crowd sourcing) and below 0.5 for students of literary studies (numbers are not directly comparable due to different task setups). The only setting that led to high agreement in Reiter (2015) is one in which the segmentation task is coupled with the task of summarizing the text.

In this section of the panel, we will report on annotation experiments that

implement the idea in Reiter (2015) more thoroughly and are currently conducted. The experiments are carried out on a corpus of dime novels. Dime novels have been selected for compatibility and comparability with the other sections of the panel. The aim in the experiments is to create segment annotations with high inter-annotator agreement, measured with boundary edit distance (Founier 2013).

All experiments are carried out with two annotators, recruited from undergraduate students of literary studies. After being trained, the general workflow is:

1. Read the entire text
2. Read a summary of the text
3. Align each sentence of the summary to a segment of the text

In a second stage, annotators first undergo an additional training round, in order to ensure that the summaries focus on the text (and leave out interpretative elements). For this training, we will use short stories that do not belong to either corpus under investigation. The annotators are also instructed that the (textual) quality of the summary does not matter.

After having written a summary for a narrative text, the annotators will – as before – link sentences/clauses of their summary with segments in the original text.

Technically the experiments are done with the annotation tool CorefAnnotator. While it is not designed for this task, it allows annotating arbitrary text spans and since it is developed by one of the authors, problems can be addressed internally. We will also remove all typographical/visual segmentation indicators (e.g. chapters, horizontal lines) from the texts.

Next to an analysis of the annotation procedure, we will investigate (and discuss on the panel) the relation between the summary-based segments and scenes as discussed above. It can be expected that a number of segments created by the annotators are not scenes in the

above sense, but contain e.g., narrative summaries. An interesting question then is, whether the identification of scenes could be modeled as a two stage process – first segmenting the text, followed by a classification of the segments.

## **2. Character constellations as indicators for scenes - a narratological model**

In drama, the smallest unit is characterized by the change of character. It therefore seems obvious to transfer this directly to narrative texts. In the following, the problems that arise with a simple implementation of this concept to narrative texts are described. The starting point of our considerations is a model of reading in which the reader builds a mental model of the situation based on the text and world knowledge. Each new sentence contributes to enriching, supplementing and possibly correcting this model (Rautenberg and Schneider, 2015) . At the beginning of a chapter that is not explicitly marked as a continuation of the previous chapter, there is information that marks time, place and characters. This can be done explicitly in the mode of a description, i.e. without the beginning of the narrated time: "Where Soho is darkest, there lies the SHOCKING PALACE". (Dark, 1978) or as part of the narrative of the time by naming the characters and places: "'Hello, Joe!", I said to the old Henderson who manages the New York FBI's armory." (Cotton, 1963) . In the last example, there is the somewhat more complex situation where the 'I' refers to the fictional author of the novel series, FBI agent Jerry Cotton, and has to be resolved by the reader accordingly. The once established situation can then be supplemented by further information, at the same time knowledge about the situation can and must often be used to resolve the character references. It is not easy to model this process because it would require a distinction to be made between a mention of a character and a reference to a character present. Such a complete understanding of the text cannot be achieved at present, but it can be approximated by setting a counter to a value n in each figure reference - assuming a reasonably functioning coreference resolution - and then reducing the counter in each subsequent sentence until it

reaches 0, and removing the character from the situation model. A special feature at the beginning of a chapter must be taken into account: It may take a few sentences - k sentences - until the situation model is fully established. The values for n and k must be determined empirically, but preliminary investigations suggest that the values are text type-specific, i.e. are by no means the same even in the case of the dime novels. How do we determine the end of a scene? This is far more challenging based on character alone. We use two indicators: First we track words which point to someone leaving the group, like ‚weggehen‘, ‚sich verabschieden‘ usw. Secondly, we take an update of the situation which introduces new characters beyond the k sentences as an indicator of a new scene.

### **3. Temporal Changes as Scene Indicators**

This section of the panel explores the identification of scenes with regard to time. The broad conception of scene presented above defines scenes as segments of a narration displaying, among other, a more or less equal time. From an operationalizing point of view, this poses two problems: (1) What does equality mean? (2) What does time mean?

Equality of time is a graded feature of scenes as the ‚more or less‘ in the definition underlines. Additionally, the Genettian concept of equal time, i.e. scene as corresponding coverage of the time of the narrated and the time of the narrating, heavily depends on the not adequately operationalized concept of events (Bögel et al., 2015b). As a consequence, equality of time cannot be operationalized in a straightforward manner. For the detection of segments that are coherent in terms of time, detecting the borders of these segments is the better viable approach. Therefore, the focus on changes of time as indicators for segment borders seems more promising.

Regarding the operationalization of time, the analysis of temporal expressions and tense would be an obvious, but somewhat reductionist approach. From a literary studies point of view, many phenomena in

narrations are related to time (e.g., cf. Meister and Schernus, 2011 for exemplary conceptions in narrative theory). These need to be considered in a time-based approach to scene detection. The presented approach tackles this by building on two prior narrative theory-based studies. In Gius (2015), time phenomena in narrations have been operationalized and analyzed manually for more than 40 categories capturing tense, temporal expressions, order, duration, frequency, the temporal relation of the narrator to the narrated and temporal perspective. These categories go well beyond linguistic temporal information and were subsequently automated by Bögel et al. (2015) for German literary texts. The detection of temporal signals was modelled by adapting and extending HeidelTime (Strötgen and Gertz, 2010) for the detection of expressions with temporal information considered relevant for narratological studies (Bögel et al., 2015a). Bögel et al. (2016) modelled order changes as approximation to anachronies (i.e., flashbacks —analepses— and flashforwards—prolepses). The detection of points in narrations where the order (i.e., chronology of the narrated) changes reached a comparably high F1 score of 81,4%. The 21 deployed features comprised temporal signals, aspects of tense, direct speech and structural features as well as features for the detection of narrative levels.

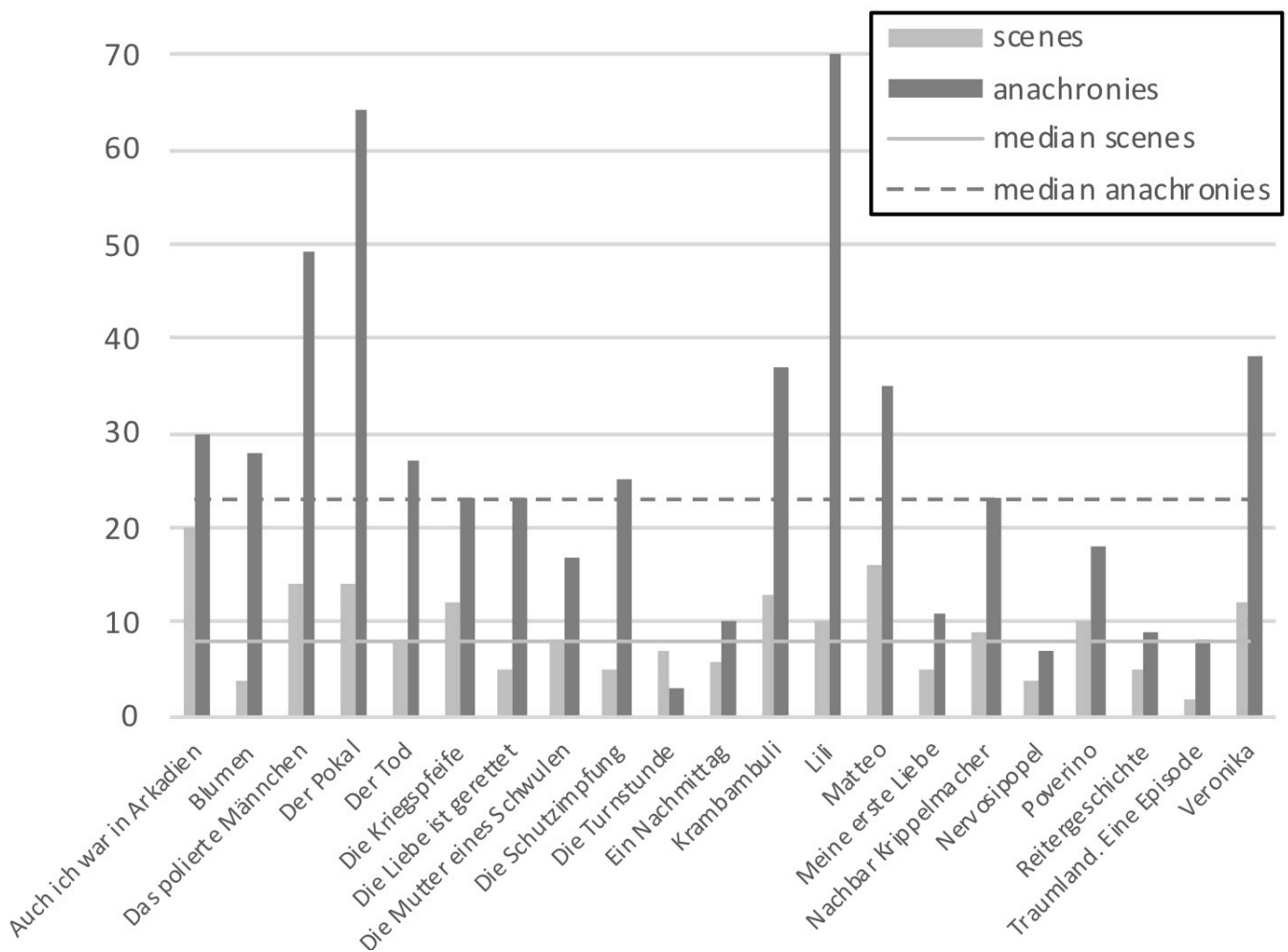


Figure 1. Occurrences of anachronies in Gius et al. (2017) and manually annotated scenes.

These so called 'order switches' can also be used as approximation to scene detection. This approach tackles both equality of time (as change of time-related phenomena) and the segmentation task. A closer examination of the annotated texts (Gius et al., 2017) suggests that the number of contained anachronies (cf. fig. 1) exceeds the number of scenes. Nevertheless, sometimes the borders of scenes seem to correspond to a border of an anachronistic text segment. Therefore, anachronies will be examined further as possible can be regarded as, though often more fine-grained segmentations of narrative texts connected to within scenes. Here, whereby anachronies encompassing more than one paragraph will be considered candidates for scenes (cf. fig. 2). Additionally, the relation between other temporal phenomena annotated in Gius et al. (2017) and scenes will be analyzed and the relevance of these temporal phenomena for the detection of scenes will be discussed.

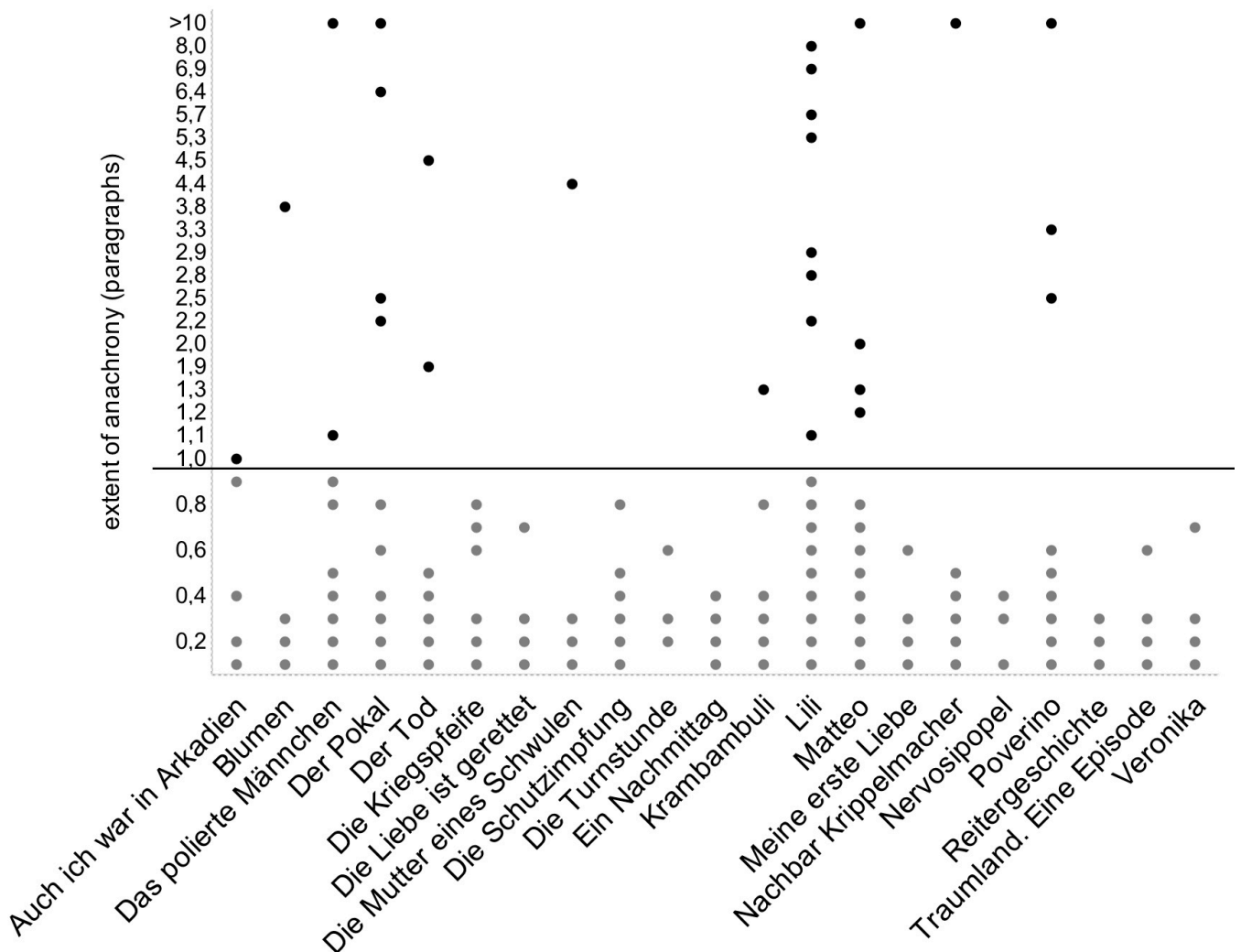


Figure 2. Anachronies with  $\geq 1$  paragraph extent

## 4. Scene Segmentation as a Classification Problem

In this section we will talk about our strategies to detect scenes using machine learning for this new task. Previous talks in this panel underline the importance of character constellation, narrative time and the fictional location for the definition of a literary scene. These form three scene detection markers ( sdm ) assumed to heavily correlate with scenes. We therefore propose to use a two-stage approach to detect scene borders: First detect the sdm s and then use these to detect scene borders.

Lacking manually labelled datasets, we used 70 Jerry Cotton novels, where segment borders are marked with "\*\*\*\*". Manual inspection has shown that these segments are more coarse-grained than traditional definitions of scenes: they do not mark all changes between scenes, but at every segment marker a new scene begins. A second dataset was



manually annotated: we randomly sampled 1000 paragraphs from 100 novels from TextGrid and labelled them as to whether or not they contain one of the sdm s, that is change in character constellation, time or place.

By using FastText (Joulin et al., 2017) for document classification, we reached an F1-score of about 60% only on the class sdms (baseline: 36%), showing that this classification can be done using standard methods.

To show the influence of sdm s on scene detection, we first attempted to detect scene borders in Shakespeare's dramas, where the sdm s are explicitly given. In dramas, scene markers are also directly found in the text, enabling us to extract scene annotations automatically. In order to train our network, we removed the scene markers as well as the words "exeunt" and "exit".

For the prediction of scenes, we used a neural network-architecture consisting of two LSTM encoders and a prediction layer, depicted in figure 3.

More specifically, we encode four sentences on each side of a position in the text as context using separate BiLSTMs and ask the model to predict whether or not a new segment begins at this position. We encode the LSTMs' input using word2vec-embeddings pre-trained on a large corpus of German novels. The output of the LSTMs is then fed into a fully connected layer to perform the classification.

We tested this model on both Jerry Cotton and Shakespeare's dramas. As anticipated, the model performed well on dramas, where sdm s are explicitly present: we reached an F1-score of 62%, with high recall of 93% and moderate precision of 46%. This shows that our system captures almost all scene changes, but tends to predict too many scenes, on average splitting every scene into two.

For Jerry Cotton, the system performed poorly, reaching an F1-Score of 17%. This confirms our expectation that sdms are immensely useful for

the detection of scenes and thus combining our two stages, the detection of sdms and the classification of scenes using a neural network, is a promising approach to this task.

Furthermore, while manual labeling of scenes is a task with notoriously low inter-annotator agreement, we noticed that labeling paragraphs for sdm s could produce good inter-annotator agreement and enable training classifiers for a very helpful first step in a scene detection pipeline.

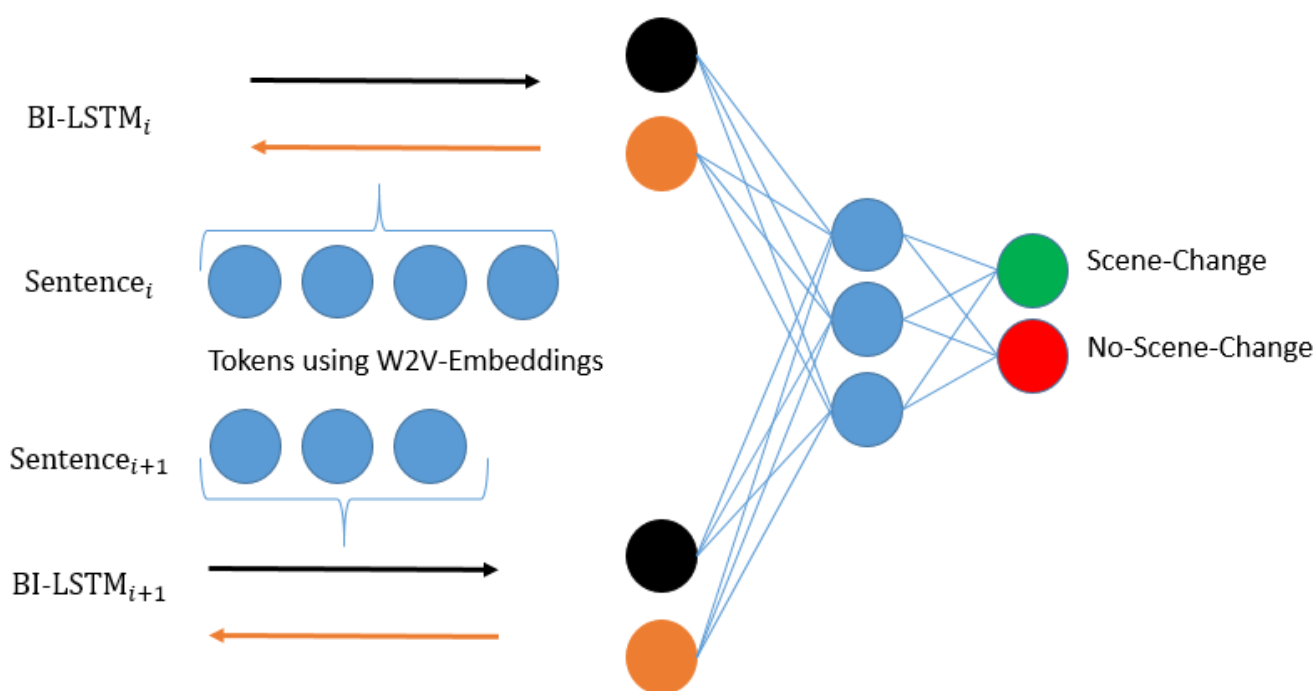


Figure 3. Neural Network architecture to detect scene changes. An instance is created between two sentences and encoded using a context of four sentences before and after the sentence. Both contexts are embedded using a separate BiLSTM, which are concatenated and after an additional fully connected layer a prediction layer decides whether a scene change occurs.

## Appendix A

### Bibliography

1. Scene. 2018. Learner's Dictionary. Retrieved from <http://www.learnersdictionary.com/definition/scene> (Nov 2018)
2. Thomas Bögel, Michael Gertz, Evelyn Gius, Janina Jacke, Meister

- Jan Christoph, Marco Petris, and Jannik Strötgen. 2015a. Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative. *DHCommons Journal* , 1.
3. Thomas Bögel, Evelyn Gius, Janina Jacke, and Jannik Strötgen. 2016. From Order to Order Switch. Mediating between Complexity and Reproducibility in the Context of Automated Literary Annotation. In *Digital Humanities 2016 Conference Abstracts* , pages 379–382.
  4. Thomas Bögel, Jannik Strötgen, and Michael Gertz. 2015b. A Hybrid Approach to Extract Temporal Signals from Narratives. In *Proceedings of the Int. Conference of the German Society for Computational Linguistics and Language Technology* .
  5. Jerry Cotton. 1963. *Jerry Cotton: Ein teuflischer Plan* . Bastei Entertainment.
  6. Jason Dark. 1978. *John Sinclair: Im Nachtclub der Vampire* . Bastei.
  7. Chris Fournier. 2013. Evaluating Text Segmentation using Boundary Edit Distance. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* , pages 1702–1712. Association for Computational Linguistics.
  8. Gerard Genette. 1980. *Narrative Discours. An Essay in Method* . Cornell University Press, Ithaca N.Y.
  9. Evelyn Gius. 2015. Erzählen über Konflikte. Ein Beitrag zur digitalen Narratologie. In *Narratologia* , volume 46. De Gruyter.
  10. Evelyn Gius, Janina Jacke, Jan Christoph Meister, and Marco Petris. 2017. *Heurecléa Time Annotations Compared Public: V1.1* . Zenodo, February.
  11. Matthew Jockers. 2015. Revealing sentiment and plot arcs with the *syuzhet* package.
  12. Armand Joulin, Edouard Grave, Piotr Bojakowski, and Tomas Mikolov.

2017. Bag of tricks for efficient text classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics , volume 2, pages 427–431, Valencia. ACL.

13. Jan Christoph Meister. 2003. Computing Action: A Narratological Approach . de Gruyter, Berlin.
14. Jan Christoph Meister and Wilhelm Schernus. 2011. Time. From Concept to Narrative construct: A Reader . de Gruyter.
15. Ursula Rautenberg and Ute Schneider. 2015. Lesen, Ein interdisziplinäres Handbuch . De Gruyter, Berlin, Boston.
16. Nils Reiter. 2015. Towards Annotating Narrative Segments. In Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities , pages 24–28, Beijing.
17. Jannik Strötgen and Michael Gertz. 2010. HeidelTime: High Quality Rule-Based Extraction and Normalization of Temporal Expressions. In Proceedings of the 5th International Workshop on Semantic Evaluation , pages 321–324, Uppsala. ACL.

*Evelyn Gius (evelyn.gius@uni-hamburg.de), Hamburg University, Germany und Fotis Jannidis (fotis.jannidis@uni-wuerzburg.de), Würzburg University, Germany und Markus Krug (markus.krug@uni-wuerzburg.de), Würzburg University, Germany und Albin Zehe (zehe@informatik.uni-wuerzburg.de), Würzburg University, Germany und Andreas Hotho (hotho@informatik.uni-wuerzburg.de), Würzburg University, Germany und Frank Puppe (frank.puppe@uni-wuerzburg.de), Würzburg University, Germany und Jonathan Krebs (jonathan.krebs@uni-wuerzburg.de), Würzburg University, Germany und Nils Reiter (nils.reiter@ims.uni-stuttgart.de), Stuttgart University, Germany und Nathalie Wiedmer (nathaliewiedmer@yahoo.de), Stuttgart University, Germany und Leonard Konle (leonard.konle@uni-wuerzburg.de), Würzburg University, Germany*