ScANT: A Small Corpus of Scene-Annotated Narrative Texts

Tarfah Alrashid[1,2], Robert Gaizauskas[2]

[1]University of Sheffield, Sheffield, UK
[2]University of Jeddah, Jeddah, Saudi Arabia

Abstract

We present the first publicly available dataset of English narrative texts annotated in compliance with SceneML, a framework for annotating scenes in narrative text. The dataset is composed of selected chapters from six narrative texts – two children's stories and four novels from Project Gutenberg. We give a brief overview of SceneML, describe the corpus sources and the annotation process and provide details of the resulting annotations and inter-annotator agreement.

Keywords

SceneML, narrative text, scenes, text segmentation, corpus, dataset, annotation

1. Introduction and Related Work

Narrative, or storytelling, is a fundamental mode of human discourse, found across all cultures and all times, and in many different forms, including writing (both fiction and non-fiction), spoken storytelling, film, video games, and so on [1]. A basic structural unit of narrative is the *scene*, "a unit of a story in which the elements of time, location, and main characters are constant" [2]. Narratives tend to progress as a sequence of scenes, though of course the sequence of scenes in a narration need not be the same as the temporal sequence of the narrated events in the *storyworld* [3] the narrative is describing. Furthermore, one scene may be expressed in multiple non-contiguous text segments in the narration. Additionally, what are generally deemed narrative texts may include various non-narrative elements, e.g. authorial comment. Thus, the task of identifying those chunks in a narrative text which correspond to scenes in the storyworld and temporally ordering these chunks is a non-trivial challenge. It is an important challenge both for the insights it gives us into the structure of narratives and for possible applications, which include automatic story illustration, aligning books and movies, automatic generation of image descriptions and automatic generation of narratives.

In previous work we have introduced *SceneML* as a framework for annotating scenes in narrative text [4] and discussed issues arising in a pilot annotation exercise which focussed on the scene identification task [2]. In this paper we present ScANT, the first publicly available dataset of English narrative texts annotated in compliance with SceneML. While the corpus is small – just 14 chapters from 6 narrative sources – our hope is that the wider community

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): Proceedings of the Text2Story'23 Workshop, Dublin (Republic of Ireland), 2-April-2023
ttalrashid1@sheffield.ac.uk (T. Alrashid); r.gaizauskas@sheffield.ac.uk (R. Gaizauskas)

will find this useful both for converging on annotation standards for scene identification and for initial training and testing of automatic scene identification algorithms. The corpus and annotation guidelines are available at https://doi.org/10.15131/shef.data.21517908 and are made available under the CC By-NC 4.0 licence [1].

There has been a growing interest in computational analysis of narrative. Ranade *et al.* [5] provide a thorough overview of recent work on computational understanding of narrative and Santana *et al.* [6] provide an extended survey on narrative extraction from textual data. However, neither of these addresses the issue of identifying *scenes* in narrative text. The only other work on annotation of scenes in narrative texts of which we are aware is that of Zehe *et al.* [7]. Their work differs from ours in several respects. First, their definition of scene states that a scene is a segment in a narrative in which the time, place and characters remain constant and which centres around one action. This contrasts with our definition that does not take into account the actions in a scene and allows multiple actions to happen in one scene (see Gaizaukas and Alrashid [4] for discussion around our choice of definition). Secondly, their scheme is less comprehensive – it does not define narrative progression links between scenes or scene transition segments, and only distinguishes scene and non-scene segments. Thirdly, they follow a container principle (small places make up larger places) to detect a change in place, e.g. if the action of characters moves from a corridor to dining room that will not indicate a change in place as they both part of a hotel, where as our definition counts these as two different places. Finally, they work on German texts while we are working on English texts.

2. Methods and Resources

2.1. SceneML

SceneML is an evolving framework for annotating scenes in narrative text. The latest specification and annotation guidelines are available along with the corpus at the DOI referenced above. Here we summarise the core concepts in SceneML.

A *scene* is defined as a unit of narrative in which the time, location and principal characters are constant and in which specific events which constitute the narrative are recounted. Any change in time, location or characters indicates a change in the scene. A scene is realised in text (for written forms of narrative) through one or more *scene description segments (SDSs).* The SDS mechanism allows for the relation of one scene in a narrative to be embedded within another, as for example, in flashback or flashforward. The task of scene identification thus becomes the task of identifying the boundaries of SDSs, and linking SDSs for the same scene together [2]. SceneML also specifies a set of four *narrative progression relations* (sequence, analepsis, prolepsis and concurrence) that are used to capture the temporal relations between scenes.

Typically, not all text in a narrative is part of a scene description. Some passages describe not one scene or another but rather the *transition* between scenes. For example, in Conan Doyle's *The Man With The Twisted Lip* the first scene takes place in Watson's house and the second

[1] https://creativecommons.org/licenses/by-nc/4.0/.

[2] Full scene annotation in SceneML also involves annotating the time, place and characters (named entities) in the scene, using existing annotation standards (ISO-TimeML,ISO-Space and the ACE NE guidelines). However, in ScANT we focus on scene segmentation only

in the East End of London, where Watson goes to seek a missing man. Between the two we have the short passage: "And so in ten minutes I had left my armchair and cheery sitting-room behind me, and was speeding eastward in a hansom on a strange errand...". Such elements SceneML refers to as *scene transition segments (STSs)*. Other sorts of non-scene elements are also present in narrative. These include general philosophising or opinion segments, background information segments, and narrative summary or narrative catchup. These passages serve a variety of functions but do not relate specific, situated events involving protagonists in the story. All such passages SceneML designates as *non-scene* elements.

2.2. Corpus Sources

The dataset is composed of selected chapters from children's stories and from out-of-copyright adult novels. The former were hypothesised as likely to have a simpler narrative structure and hence to be a good place to trial our approach; the latter as likely to possess more complex narrative structure and hence pose a more challenging test to our approach. The sources are: (1) *Bunnies from the Future*, a middle grade children's story by Joe Corcoran [3]. The author has personally granted permission for us to release annotated chapters of this work. (2) *The Wonderful Wizard of Oz*, originally released as part of the Brown Corpus [4] and free for non-commercial purposes. (3) *Pride and Prejudice*, *A Tale of Two Cities*, *The Adventures of Sherlock Holmes* and *The Great Gatsby* from Project Gutenberg [5]. These are out of copyright in the US and UK and freely re-distributable subject to Project Gutenberg's terms and conditions.

3. The Annotation Process

In an earlier pilot study [2] we investigated how well-defined the SceneML definitions and annotation framework were with respect to scene boundary identification. Analysis of the annotations in that study revealed several causes of observed disagreement: (1) lack of understanding of the guidelines and task, (2) lack of clarity or specificity in the guidelines, (3) failure of non-native English speakers to fully grasp the meaning of certain expressions (e.g. idioms). We have addressed these issues in the construction of ScANT through the following steps: (1) A more thorough training process that included both an initial training session with a presentation, demonstration and hands-on exercise for the trainees, plus a follow-on take-away exercise that was scored against gold-standard annotations produced by the authors and then discussed with the trainees, (2) Improvement of the initial guidelines to remove sources of confusion revealed in the earlier pilot, (3) recruitment of native English-speaking annotators with sensitivity to text analysis (two PhD students, one in English Literature and one in Computational Linguistics).

The annotation process was carried out through a web-based interface to a local instance of Brat Annotation Tool [6]. Annotators used swipe and click operations to annotate SDSs and STs. Multiple SDSs that are part of the same scene were linked using the Brat relation annotation

[3]https://freekidsbooks.org/author/joe-corcoran/
[4]https://www.nltk.org/nltk_data/
[5]https://www.gutenberg.org
[6]https://brat.nlplab.org

Table 1
Summary Statistics for the ScANT corpus, showing for each annotated text the count of sentences and words and of SDSs, STs, Scenes and Non-scene sentences (NSSs) for each annotator (A1 and A2).

Text	Sents	Words	SDSs		STs		Scenes		NSSs	
			A1	A2	A1	A2	A1	A2	A1	A2
Bunnies Ch3	124	2756	8	10	1	0	8	9	0	0
Bunnies Ch4	65	1775	10	8	0	0	9	7	0	0
Bunnies Ch5	173	3514	10	7	0	3	10	7	0	0
Bunnies Ch6	117	2911	10	10	0	6	10	10	0	0
WOZ CH2	132	2449	4	8	0	2	4	8	0	1
WOZ CH3	123	2361	9	8	1	7	9	8	1	1
Sherlock Holmes Ch1 P1	268	4200	11	10	0	4	11	10	17	0
Sherlock Holmes Ch1 P2	277	4784	23	11	1	6	20	11	0	0
Sherlock Holmes Ch1 P3	93	1333	8	6	0	5	8	6	2	0
Sherlock Holmes Ch6	561	10974	31	34	2	15	26	34	8	0
Pride and Prejudice Ch1	60	1018	1	3	0	2	1	3	6	0
Pride and Prejudice Ch3	86	1984	12	10	0	5	12	10	0	0
A Tale of Two Cities Ch1	19	1140	0	5	0	4	0	5	19	0
A Tale of Two Cities Ch3	73	1920	4	13	0	4	4	13	11	0
The Great Gatsby Ch1	337	7209	19	43	1	12	19	43	52	0
The Great Gatsby Ch3	288	5307	31	24	0	10	29	24	13	0
Total	2796	55635	191	210	6	85	180	208	129	2

tool to signal that a `same-scene-as` relation holds between them. The annotated data is stored and made available in BRAT standoff annotation format [7].

The corpus consists of fourteen chapters from six different narrative sources [8]. In each chapter SDSs, STs and `same-scene-as` relations were annotated by two annotators and saved in a separate text file. Both annotators' annotations are supplied with the corpus. Further annotations together with a consensus annotation may be made available in the future.

4. The ScANT Corpus

4.1. Corpus Statistics

Table 1 shows summary statistics for the ScANT corpus and associated annotations. Note that the relation between scenes and SDSs is largely one-to-one. With one exception this is always true for the children's stories, while there is somewhat more variation, suggesting more complex narrative form, in the adult novels.

The variation between annotators A1 and A2 is relatively small in terms of SDSs and Scenes. However, they are far apart regarding both scene transition segments and non-scene segments. We discuss this further below in Section 4.3. First, however, we examine inter-annotator agreement regarding SDSs in more detail.

[7] It can be converted to JSONL format using the tool at: https://github.com/astutic/brat-standoff-to-json/.
[8] As one of the chapters is quite long it has been divided into three parts for analysis.

Table 2
IAA Results, showing Cohen's kappa under varying degrees of leniency, where N indicates the number of sentences apart SDS boundaries may be to count as a match or, where $N = 30\%$, the number of sentences expressed as a percentage of the median SDS sentence length for that text.

Chapter	SDS Median	N = 30%	N = 0	N = 1	N = 3	N = 5
Bunnies Chapter 3	7	0.79	0.74	0.74	0.79	0.79
Bunnies Chapter 4	5.5	0.60	0.60	0.60	0.60	0.60
Bunnies Chapter 5	14.5	0.45	0.29	0.29	0.45	0.45
Bunnies Chapter 6	5.75	0.68	0.47	0.47	0.72	0.76
WOZ Chapter 2	19.75	0.47	0.11	0.11	0.27	0.41
WOZ Chapter 3	11	0.77	0.57	0.57	0.72	0.77
Sherlock Holmes Chapter1 P1	10.5	0.16	0.07	0.07	0.16	0.16
Sherlock Holmes Chapter1 P2	8	0.53	0.42	0.42	0.53	0.53
Sherlock Holmes Chapter1 P3	8.75	0.75	0.75	0.75	0.75	0.75
Sherlock Holmes Chapter 6	8.5	0.37	0.25	0.25	0.37	0.40
Pride and Prejudice Chapter 1	16	0.65	0.65	0.65	0.65	0.65
Pride and Prejudice Chapter 3	6.25	0.65	0.43	0.43	0.65	0.65
Tale of Two Cities Chapter 3	6.5	0.21	0.01	0.01	0.30	0.39
The Great Gatsby Chapter 1	7.5	0.34	0.22	0.22	0.34	0.39
The Great Gatsby Chapter 3	6	0.49	0.38	0.38	0.58	0.66
Average	8.94	0.53	0.40	0.40	0.53	0.56

4.2. Inter-annotator Agreement

Table 2 shows inter-annotator agreement results for SDSs using Cohen's Kappa [8]. To calculate Kappa, each sentence is given a tag, 1 for sentences on the boundary of an SDS (either beginning or end) and 0 otherwise. Boundaries of STs are ignored as it is clear the two annotators' understanding of the task is so different that precise quantitative analysis is not merited.

Aside from calculating only exact matches as agreement ($N = 0$ in Table 2) we also investigated a more lenient approach to calculate the agreement in which annotators are deemed to agree if they place a sentence boundary within N sentences. This was prompted by the observation that in many cases annotators seemed to be placing SDS boundaries relatively close to each other, but not exactly in the same place. Kappa scores have been calculated for various N sizes: $N = 30\%$ of the median SDS sentence length in each chapter, *N=1*, *N=3* and *N=5*. We have omitted one chapter from Table 2 – *A Tale of Two Cities*, chapter 1, because one annotator believed it contained nothing but non-scene segments, while the other thought it contained 5 scenes. This gave a kappa score of 0, which skewed the rest of the results.

4.3. Discussion

Regarding differences between our annotators, it is clear that the two annotators have a different conception of what STs and NSSs are. This is probably due to the fact that the children's stories which we used as training materials contain very few of either of these, particulary of NSSs (in fact this appears to be an interesting difference between children's and adult narrative).

Any future annotation effort should ensure that these concepts are more clearly understood by annotators. On examination our view is that A1 has followed the guidelines much more closely regarding STs and NSSs and therefore, if one is to train or test a classifier on these materials, our recommendation would be to use the A1 annotations only. However recent work such as that reported in Uma *et al.* [9] highlights the potential value of learning from disagreement, so we have included both sets of annotations in the corpus.

Concerning the kappa scores for SDS agreement, they fall in the range that has been interpreted as ”fair” or ”fair to good”. However, kappa scores are known to be lower when there are fewer labels (just two in our case) and where the labels are not equiprobably occurring (also true in our case, since 0 labels are much more frequent than 1’s) so results should be viewed in this light [10]. Percentage agreement scores for SDS annotations are around 90%. Note that kappa scores rise significantly if we are prepared to allow some leniency in terms of non-exact matching. How legitimate this is needs further examination to determine whether the improvement is a reflection of genuine uncertainty about the precise boundary between what the annotators clearly agree are distinct scenes or whether it is the result of conflating separate scenes, according to the different annotators’ perceptions.

While the corpus is too small to start making generalisations about stylistic differences between different authors, it is worth noting that the amount of non-scene content in the adult novels (6.21% of the total sentences if we accept A1’s NSS annotations, which we believe are more accurate) is vastly greater than that in the children’s stories (0.14%), suggesting that much beyond simple event narration goes on in adult fiction.

5. Conclusion and Future Work

We have presented the first dataset of English narrative texts annotated in compliance with SceneML. The dataset consists of fourteen chapters of novels and children’s stories annotated for scene description segments and scene transitions segments as defined in SceneML. A total of almost 200 scenes have been annotated.

Future work plans include various activities. These include:

1. gathering further annotations for the ScANT source texts to increase robustness of the annotations and guidelines;
2. extending the annotation to include SceneML narrative progression links;
3. training a model on the corpus to investigate automating the task of scene boundary detection and to ascertain the sufficiency of ScANT for this task;
4. adding other text types to the annotated dataset, such as biography, plays and film scripts;
5. expanding the dataset to include texts in other languages.
6. exploring whether there is interest in a shared task challenge on scene boundary detection.

We hope the community finds ScANT of use and welcome comment on our work.

Acknowledgments

The authors thank the Text2Story reviewers for their helpful comments. The first author acknowledges support from the University of Jeddah in the form of a PhD studentship.

References

[1] Wikipedia, Narrative, 2023. URL: https://en.wikipedia.org/wiki/Narrative, last accessed 25 March, 2023.

[2] T. Alrashid, R. Gaizauskas, A pilot study on annotating scenes in narrative text using SceneML, in: Proceedings of the 4th international workshop on narrative extraction from texts (Text2Story 2021), 2021, pp. 7–14.

[3] W. Schmid, Narratology: An introduction, Walter de Gruyter, Berlin, 2010.

[4] R. Gaizauskas, T. Alrashid, SceneML: A proposal for annotating scenes in narrative text, in: Proceedings of the 15th Workshop on Interoperable Semantic Annotation (ISA-15), Gothenburg, Sweden, 2019.

[5] P. Ranade, S. Dey, A. Joshi, T. Finin, Computational understanding of narratives: A survey, IEEE Access 10 (2022) 101575–101594.

[6] B. Santana, R. Campos, E. Amorim, A. Jorge, P. Silvano, S. Nunes, A survey on narrative extraction from textual data, Artificial Intelligence Review (2023) 1–43.

[7] A. Zehe, L. Konle, L. K. Dümpelmann, E. Gius, A. Hotho, F. Jannidis, L. Kaufmann, M. Krug, F. Puppe, N. Reiter, et al., 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, 2021, pp. 3167–3177.

[8] J. Cohen, A coefficient of agreement for nominal scales, Educational and psychological measurement 20 (1960) 37–46.

[9] A. N. Uma, T. Fornaciari, D. Hovy, S. Paun, B. Plank, M. Poesio, Learning from disagreement: A survey, J. Artif. Int. Res. 72 (2022) 1385–1470. URL: https://doi.org/10.1613/jair.1.12752. doi:10.1613/jair.1.12752.

[10] Wikipedia, Cohen's kappa, 2023. URL: https://en.wikipedia.org/wiki/Cohen%27s_kappa, last accessed 26 March, 2022.