A Pilot Study on Annotating Scenes in Narrative Text using SceneML

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

[1] Department of Computer Science, University of Sheffield, Sheffield, UK
{ttalrashid1,r.gaizauskas}@sheffield.ac.uk
[2] University of Jeddah, Jeddah, Saudi Arabia

Abstract

SceneML is a framework for annotating scenes in narratives, along with their attributes and relations [GA19]. It adopts the widely held view of scenes as narrative elements that exhibit continuity of time, location and character. Broadly, SceneML posits *scenes* as abstract discourse elements that comprise one or more *scene description segments* – contiguous sequences of sentences that are the textual realisation of the scene – and have associated with them a *location*, a *time* and a set of *characters*. A change in any of these three elements signals a change of scene. Additionally, scenes stand in narrative progression relations with other scenes, relations that indicate the temporal relations between scenes. In this paper we describe a first small-scale, multi-annotator pilot study on annotating selected SceneML elements in real narrative texts. Results show reasonable agreement on some but not all aspects of the annotation. Quantitative and qualitative analysis of the results suggest how the task definition and guidelines should be improved.

1 Introduction

We all have an informal idea of what constitutes a *scene* in a narrative, such as in literature or in film: the story moves to a different location; or one set of characters exits the story and another set enters; or we are taken forwards or backwards in time. Scenes are the fundamental building blocks of extended narrative, the chunks into which narrative naturally divides.

Despite the ubiquity of the notion in literary studies, in particular in narrative theory and in drama studies, there has been little work on formalising a notion of scene or on developing an annotation framework for scenes such that automated approaches to scene segmentation might be developed. Why might one want to do this? One reason is that as in any area of literary or linguistic studies, operationalising a theoretical model in a computer program and applying to data allows one to verify or, if necessary, to refine the theory to give it empirical support. Another reason is that there are several potential applications for automated scene segmentation capability. These include: (1) *automatic text illustration* [JWL04, AGKK11, FL10], since scenes are the likely discourse units with which to associate illustrations; (2) *aligning books with movies* [ZKZ+15], since scenes are the high level units to be aligned; (3) *automatic generation of image descriptions* [KPD+11, DFUL17, YTDIA11],

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Finlayson (eds.): Proceedings of the Text2Story'21 Workshop, Online, 1-April-2021, published at http://ceur-ws.org

where scene-segmented narratives could provide background knowledge about what sorts of descriptive elements should be mentioned in descriptions of particular scene-types; and, (4) *automatic narrative generation* [CL02], which could benefit from a corpus of scene-segmented narratives on which to train models.

To address the lack of a formal model of scene annotation in narrative text Gaizauskas and Alrashid [GA19] proposed an annotation framework called *SceneML*. That paper is an initial, theoretical proposal for an annotation framework comprising annotations for entities, including scenes, scene description segments, times, locations and characters, and for relations between scenes, such as narrative progression relations. While that paper laid the foundation for SceneML as a framework for scene annotation, it did not report any empirical work on annotating a collection of texts, nor did it discuss levels of agreement obtainable by annotators. This paper addresses these considerations, reporting the first pilot study on applying SceneML to real narrative texts in which multiple annotators applied SceneML to annotate selected SceneML elements in several chapters from a children's story.

The rest of the paper is organised as follows. Section 2 briefly reviews the SceneML annotation framework to enable this paper to be read stand-alone. Section 3 describes the pilot study carried out and the results, both quantitative and qualitative, of an analysis of inter-annotator agreement. It also discusses the implications of the pilot study for SceneML and what steps need to be taken to improve inter-annotator agreement. Section 4 summarises our findings and discusses future work.

2 SceneML

This section provides a compressed overview of the SceneML framework, as introduced in [GA19]. All references to SceneML in the following are to that paper.

2.1 The Annotation Framework

Adopting the most widely accepted definition of *scene* in the literature, we treat a scene as a unit of a story in which the elements of time, location, and main characters are constant. Any change in these elements indicates a change of scene. A scene is an abstract discourse element, not a specific span of text. It consists of a location or setting, a time and characters who are involved in the events that take place in the scene. These elements exist in the real or fictive world, i.e. the *storyworld* as per narrative theory [Sch10], within which the narrative unfolds. "The scene itself is an abstraction away from the potentially infinitely complex detail of that real or fictive world, a cognitive construct that makes finite, focussed narrative possible" (SceneML, §3.1).

A scene in a textual narrative is realised through one or more *scene description segments* (SDS). An SDS is "a contiguous span of text that, possibly together with other SDSs, expresses a scene" (SceneML, §3.1). Typically a scene consists of a single SDS. However, a scene may reference other scenes, e.g. scenes involving past or future events. The SDSs describing these events may be embedded within the description of the embedding scene, causing its textual realisation to be split into multiple, non-textually adjacent SDSs. A scene may also be realised through multiple SDSs if "the author is employing the narrative device of rotating between multiple concurrent scenes each of which is advancing a distinct storyline (a common technique in action movies)" (SceneML, §3.1).

It is important to define what each element of a scene (*characters*, *time*, *location*) is to facilitate the annotation process and make it easier to detect scene changes if any of the elements change. We propose to adopt the definitions and annotation standards for the elements *time*, *location*, and *spatial entities* from Iso-TimeML and Iso-Space [1]. As for characters, we will adopt the definition and annotation standards for named entities of type person from the ACE program, recently used in the TAC 2018 entity discovery and linking task [2].

These previous standards facilitate the annotation process for all mentions of times, locations/spatial entities and persons represented in the text. However, we are interested in just the specific *characters*, *time*, and *location* that define a specific scene.

2.2 SceneML Elements

SceneML elements are categorised into two main categories:

1. **Entities:** entities consists of scenes, SDSs, characters, times, and locations;

[1] www.iso.org/standard/37331.html and www.iso.org/obp/ui/#iso:std:60779.

[2] See http://nlp.cs.rpi.edu/kbp/2018/ and https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-entities-guidelines-v5.6.6.pdf.

2. **Relations:** scene-scene narrative progression links, relational links connecting times, characters, locations to the scenes in which they are participant elements and connecting SDS's to the scenes they comprise. At present these relational links are all represented via attributes on entities.

Scenes

Scenes are the main element in SceneML. A scene has attributes: `id, time` and `location`. It also includes a list of character sub-elements as there may be more than one character for each scene.

SDSs

Scene description segments (SDSs) are the contiguous sequences of words from the narrative text that comprise the textual realisation of a scene. One SDS cannot belong to more than one scene, but a scene can be composed of multiple SDSs. SDS attributes include: `id` and `scene_id`, i.e the id of the scene that the SDS belongs to.

Time

Time elements used here are the ones developed within ISO-TimeML. Each time annotation includes an `id` attribute and a text segment. Time can also be the time of the storyworld, signalled by the attribute `base`.

Location

We use the location element from ISO-Space. This also includes an `id` attribute that is unique for each location mention and a text span.

Character

Here we use the named entity type `person` from the ACE English Annotation Guidelines for Entities, the only difference being that we allow animals and non-humans as characters if they play the role of characters in the narrative. Person annotations have a unique `id` attribute for each character mention and a text segment.

Narrative Progression Links

Narrative progression links (`nplinks`) link two scenes whose SDSs are textually adjacent. There are different types of nplink depending on the type of temporal relation between the two scenes. SceneML (§3.2) recognises four types of links. `Sequence` links are assigned when the scene change happens because of a change in location or characters, e.g. a character moves to another place, but where the events follow in time directly after those in the linked scene; `Analepsis` links are used when there is a flashback in one scene to a scene in the past. `Prolepsis` links are used when then we are taken forward in time (i.e. flashforward); `Concurrent` links are assigned between two scenes when the transition happens because there is another thread of the story happening at the same time so the transition takes us to different characters and a different place but at the same time.

An extended example of the XML realisation of SceneML annotation can be found in SceneML §4.

3 Methods and Results

3.1 Methods

A small-scale pilot study was carried out to investigate how well-defined our definitions and annotation framework are with respect to scene boundary identification. Two chapters (chapters 3 and 4) of "Bunnies from the Future" [Cor16], a children's story for reading ages 10-13, were selected to be annotated. Three annotators, distinct from the authors, were identified (postgraduate students, non-native speakers of English) in addition to one of the authors. They were given annotation guidelines based on the framework explained in section 2 and were instructed to use the Brat[3] annotation tool to annotate the two chapters following the guidelines, with two simplifying exceptions: (1) annotators were asked not to annotate the `scene` abstract discourse element; (2) they were asked *not* to annotate explicitly any relations. I.e., they were instructed to annotate SDSs and for each SDS they were asked to annotate the first mention of the time and location of and the characters participating in the events described in the SDS.

The first of these simplifications was imposed because the Brat annotation tool, which is a very easy-to-use tool for annotation text spans, does not support the creation of zero-span annotations. That functionality is necessary for creating abstract discourse elements, which can then be linked to spans in the text. We could not

[3] http://brat.nlplab.org/

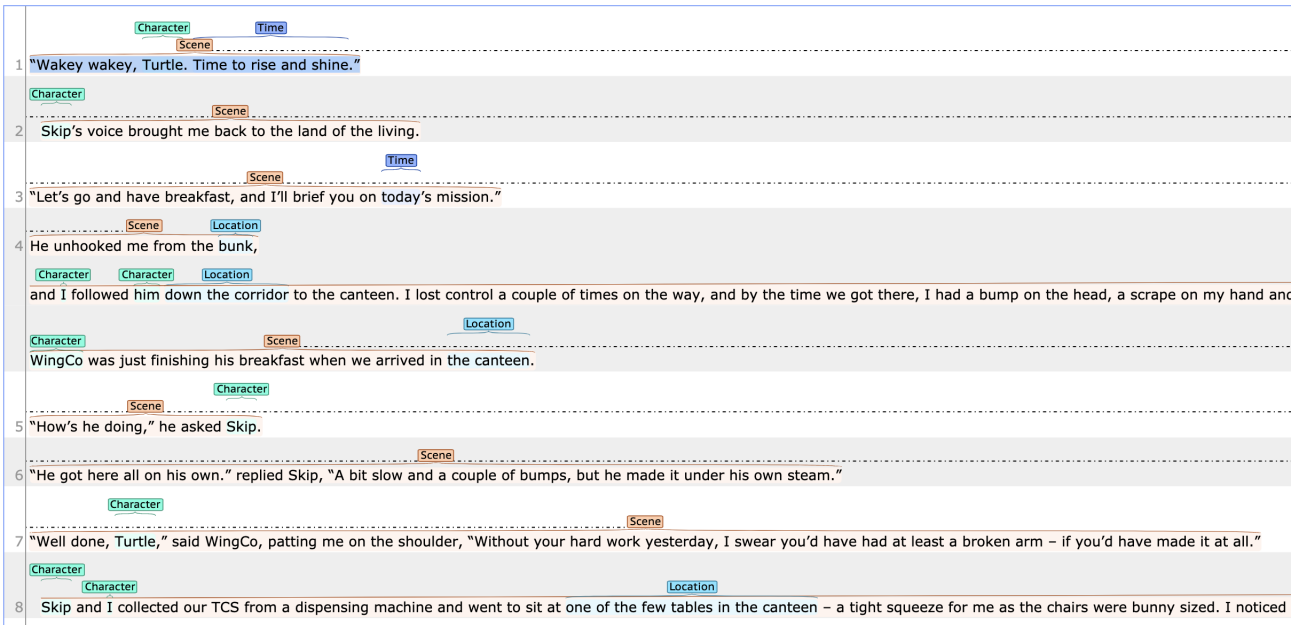

Figure 1: Screenshot showing an example of the pilot annotation using Brat.

find any tool that would allow this and was equally easy-to-use, and did not have the resources to create our own. The problem can be circumvented by linking all SDSs in one scene together with a "same-scene-as" relational link, thus extensionally specifying a scene as the set of its SDSs.

The second simplification was imposed for several reasons. First, for this first annotation exercise we were primarily concerned with determining whether or not annotators could accurately identify and agree on SDS boundaries. The problem of determining what the narrative links between scenes are is very much a secondary problem compared to this. As it turns out, all but one of the scenes in the chosen texts consist of a single SDS and follow each other in temporal sequence. Single-SDS scenes are likely a characteristic of this level of children's stories and one reason why we chose such stories for this pilot, though we did not deliberately chose chapters because they had few scenes with multiple SDSs. Second, given that virtually all scenes consist of a single SDS, the need to link same-scene SDSs is low. This led us to ignore it altogether for this exercise, though clearly it will need to be addressed when dealing with more complex narrative structure. Finally, given the absence of explicit scene elements in the annotation, the times, characters and locations needed to be annotated in each SDS participating in a scene. This has the disadvantage of forcing the annotator to annotate these things multiple times per scene, when there are multiple SDSs per scene, but means that (1) the relations linking the scene elements time, character and location to the scene can be inferred from simple presence of annotated strings of these types within an SDS and (2) allows the annotator, and subsequently analyst, to validate that distinct scenes involve a distinction in at least one of time, location and characters (in fact identity of these elements could be used as another method to indirectly link multiple SDSs in a single scene together, though annotator error could make this unreliable).

Figure 1 shows an example annotation from chapter 4; entity segments are highlighted.

3.2 Results

Tables 1, 2 and 3 show the results of the annotation exercise. Table 1 shows information on the numbers of entity mentions annotated by each annotator and the averages of these numbers. Note that the chapters differ in size: chapter 3 is 124 sentences long (2756 words) and chapter 4 is only 65 (1775 words). The columns *SDS, Character, Location* and *Time* indicate the number of entity mentions annotated for each category by each annotator. Two averages are computed: (1) entity mentions per chapter averaged over annotators (2) average entity mentions per *SDS* for each annotator averaged across all annotators.

Table 2 shows inter-annotator agreement results. Pair-wise inter-annotator agreement results are computed using Cohen's kappa. Averaged inter-annotator agreement results between all annotators are computed using Fleiss's kappa. $\kappa 1$ refers to the kappa score for segments of type *SDS*. For SDSs the kappa score is computed by

Table 1: Statistics about the annotations. *SDS, Character, Time* and *Location* columns refer to the number of segments marked as each entity type per chapter for each annotator. Averages of these numbers by chapter/by *SDS* are also shown.

	Chapter 3				**Chapter 4**			
	SDS	**Char**	**Time**	**Loc**	**SDS**	**Char**	**Time**	**Loc**
Ann1	4	19	2	5	5	14	4	6
Ann2	5	21	3	3	6	28	4	6
Ann3	13	30	8	12	10	28	4	10
Ann4	8	21	0	5	9	19	1	6
Av/Chapter	7.5	22.75	3.25	6.25	7.5	22.25	3.25	7.25
Av/SDS	–	3.67	0.4	0.78	–	3.33	0.55	1.11

Table 2: Inter-annotator agreement results. $\kappa 1$ refers to the kappa score for *SDS*, $\kappa 2$ refers to kappa score for all other entities together.

	Ann2				Ann3				Ann4			
	Ch3		Ch4		Ch3		Ch4		Ch3		Ch4	
	$\kappa 1$	$\kappa 2$	$\kappa 1$	$\kappa 2$	$\kappa 1$	$\kappa 2$	$\kappa 1$	$\kappa 2$	$\kappa 1$	$\kappa 2$	$\kappa 1$	$\kappa 2$
Ann1	0.60	0.54	0.33	0.27	0.15	0.27	0.20	0.30	0.23	0.19	0.10	0.21
Ann2					0.27	0.20	0.42	0.39	0.19	0.24	0.35	0.30
Ann3									**0.72**	0.33	**0.95**	0.52
Average $\kappa 1$ Ch3: 0.36					Average $\kappa 2$ Ch3: 0.29							
Average $\kappa 1$ Ch4: 0.41					Average $\kappa 2$ Ch4: 0.34							

considering each sentence as a potential candidate for a scene segment boundary. So, each sentence is represented by either a 1 or 0, 1 if the sentence either contains a scene segment boundary or is preceded or followed by one, 0 otherwise. $\kappa 2$ refers to the kappa score computed for all other entity types (*Time, Character* and *Location*). Here we treat the problem of recognising these three entity types as a token classification problem, following the widely used named entity recognition approach of IOB tagging (see, e.g., [JM09]). Each word is tagged as either *Time, Character, Location* or *Outside*, where the *Outside* tag is given to words that are not part of any entity mention[4].

Table 3 shows percentage agreement results between each annotator pair for each entity type. These are computed by creating a confusion matrix of token entity type labels (including type *Outside*) for each annotator pair and dividing each value on the diagonal in this confusion matrix by the corresponding row total.

3.3 Discussion

Here we discuss disagreement between the annotators with a view to determining how our annotation guidelines and/or processes should be improved and whether or not there are any underlying conceptual problems with our approach.

In the general the Cohen's kappa scores show what is commonly interpreted as fair (0.21–0.40) to moderate (0.41–0.60) agreement with a few cases of slight and substantial agreement around the edges. However, as the qualitative interpretation of kappa scores is contentious, we use these scores primarily as a diagnostic tool, highlighting areas of relative agreement and disagreement. Looking the scores overall, several observations can be made. First, the annotator pair (Ann3,Ann4) agree much more than any other annotator pair. Second, generally and on average, $\kappa 1$ scores are higher than $\kappa 2$ scores, i.e. agreement on SDSs is higher than agreement on entities.

Regarding the percentage agreement results on entity annotation in Table 3, it can be seen that for most cases, *Character* and *Time* entities have higher agreement results than *Location* entities (agreement for *Outside* tags will always be high given the unbalanced nature of the data, i.e. most tokens are outside of any entity

[4]For simplicity we do not include a 'B' tag, as instances of contiguous distinct entity mentions of the same type are extremely rare.

Table 3: Percentage agreement results between annotator pairs for each entity type by token. Here *O* refers to the *Outside* tag.

	Ann2				Ann3				Ann4			
	Character	**Location**	**O**	**Time**	**Character**	**Location**	**O**	**Time**	**Character**	**Location**	**O**	**Time**
Ann1	0.24	0.15	0.99	0.8	0.28	0.15	0.99	0.26	0.33	0.05	0.99	0
Ann2					0.36	0.09	0.99	0.33	0.48	0.05	0.98	**1**
Ann3									**0.73**	**0.88**	0.98	**1**

mention). Note that these figures are heavily dependent on agreement in SDS annotation. If one annotator annotates two SDSs where another annotates just one, the first annotator will have twice the number of time, location and character annotations, since these entity types are to be annotated for each SDS. Hence low scores are to expected where agreement in the number of SDSs is not high. Indeed we can see that for annotators 3 and 4, where SDS agreement is higher than for any other annotator pair, agreement on entities is also very much higher than for any other annotator pairing.

Analysis of the annotations reveals two underlying causes of the observed disagreement: (1) lack of understanding of the guidelines and task and (2) lack of clarity or specificity in the guidelines. These are often not easy to distinguish.

Lack of understanding

It emerged in questioning following the annotation exercise that some annotators (Ann1 and Ann2) relied only on the authors' verbal explanation of the guidelines and task and had either not read the guidelines at all or had not read them carefully. E.g., in some places we find two distinct *location* entities tagged in the same *SDS*, in clear contradiction of the guidelines, suggesting either the annotator did not pay enough attention or simply did not understand. Although our annotators are all good speakers of English and studying at the postgraduate level in well-respected English-speaking universities, being a non-native speaker of English led to misinterpretations of some sentences or expressions in the text that obviously caused mistakes in annotation For example, one annotator labelled *earth* as a scene *location* in the idiomatic expression *how on earth had he ...* where clearly it is not. In another case, in the example sentence *Sorry, old chap, had an attack of the wobbles. Dashed embarrassing*, the word *Dashed* was annotated as a character.

Lack of clarity and detail in the guidelines

Some annotators included definite articles in the annotated entity mention, e.g. *the stone ages* was annotated as a time by one annotator and *stone ages* by another (to assess the effect of such minor variation in annotation we re-measured inter-annotator agreement after removing all stop words and found that results slightly improved). Some *time* entities were annotated as the time of a scene while in fact they just reference other times. E.g., in *Do you not have good fabrics in the future?* the word *future* was annotated as the time of the scene. Regarding tagging characters, confusion arose as to whether to annotate the fullest form of the character mention, the first mention, or every mention. All of these small divergences need to be eliminated by more fully and explicitly addressing them in the guidelines.

Two deeper issues were detected with respect to scene boundary determination and account for most of the variation between annotators with respect to SDS boundary placement. One is to do with "scene transition segments", typically short phrases or a sentence or two that indicate a character is moving from one scene to another. For example: *and soon I emerged back into the corridor looking like a new man.* Should this text be annotated as belonging to the preceding or succeeding scene, or is it a scene in its own right or part of no scene, but a new "scene transition" element that should be added to the annotation scheme? The other issue is to do with granularity of scene segmentation. E.g. if one minor character leaves a scene does this imply a new scene, given our definition of scene as "a unit of a story in which the elements of time, location, and main characters are constant."? Or should this be viewed as too minor a change to count as a scene change?

Again, if, for example, a character goes into a dressing room off another space and his changing clothes in the dressing room is described while he continues talking to another character outside the dressing room ([Cor16], Chp. 4), is this a significant enough change of location to constitute a scene change? Further work needs to be done to articulate a clear position with respect to these edge cases.

4 Conclusion

We have presented a pilot annotation experiment in which annotators are asked to use SceneML to annotate several chapters in a children's story. Our aim was to assess the viability of the scene segmentation task and the adequacy of our guidelines. Results show that the task is feasible, but suggest several changes need to be made to the annotation process and guidelines to improve inter-annotator agreement. First, annotators should be better trained, and filtered from the annotator pool if their English language understanding or understanding of the task is too weak. This can be assessed by asking them to do a trial annotation exercise, which is reviewed by expert annotators before allowing them to proceed with real annotation. Second, the guidelines should be refined to reduce confusion, addressing the specific issues identified in the preceding section. With these changes we are convinced agreement between annotators on the task can be substantially increased.

Once agreement between annotators has been assured at a higher level, our planned future work includes a number of activities. First, we will extend the annotation to include all SceneML elements, including narrative progression links and explicit linking of SDSs. Second, we plan to apply the scheme to a much wider range of texts, including historical and contemporary fiction for adults, as well as biography. Non-textual narrative genres, such as film, will also be considered. Third, we plan to investigate automating the process of scene annotation after getting sufficient manually annotated data to train a predictive model, using supervised machine learning techniques. Other potential future activities include investigating the application of the scheme to other languages. There is no reason in principle why SceneML should be restricted to the English language, though entity annotation guidelines for target languages need to be available. Application to other languages would be welcome and would serve to confirm the universality of the approach.

References

[AGKK11] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, and Krishnaram Kenthapadi. Enriching textbooks with images. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 1847–1856, New York, NY, USA, 2011. ACM.

[CL02] Charles B. Callaway and James C. Lester. Narrative prose generation. *Artif. Intell.*, 139(2):213–252, August 2002.

[Cor16] J. Corcoran. *Bunnies from the Future*. www.freekidsbooks.org, 2016.

[DFUL17] Bo Dai, Sanja Fidler, Raquel Urtasun, and Dahua Lin. Towards diverse and natural image descriptions via a conditional gan. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2970–2979, 2017.

[FL10] Yansong Feng and Mirella Lapata. Topic models for image annotation and text illustration. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 831–839. Association for Computational Linguistics, 2010.

[GA19] Robert Gaizauskas and Tarfah Alrashid. Sceneml: A proposal for annotating scenes in narrative text. In *Workshop on Interoperable Semantic Annotation (ISA-15)*, page 13, 2019.

[JM09] Daniel Jurafsky and James H. Martin. *Speech and Language Processing (2nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2009.

[JWL04] Dhiraj Joshi, James Z Wang, and Jia Li. The story picturing engine: finding elite images to illustrate a story using mutual reinforcement. In *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, pages 119–126. ACM, 2004.

[KPD+11] Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C Berg, and Tamara L Berg. Baby talk: Understanding and generating image descriptions. In *Proceedings of the 24th CVPR*. Citeseer, 2011.

[Sch10] Wolf Schmid. *Narratology: An introduction*. Walter de Gruyter, Berlin, 2010.

[YTDIA11] Yezhou Yang, Ching Lik Teo, Hal Daumé III, and Yiannis Aloimonos. Corpus-guided sentence generation of natural images. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 444–454. Association for Computational Linguistics, 2011.

[ZKZ+15] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, pages 19–27, Washington, DC, USA, 2015. IEEE Computer Society.