The Storyline Annotation and Representation Scheme (StaR): A Proposal

Tommaso Caselli and **Piek Vossen**
Vrije Universiteit Amsterdam
De Boelelaan 1105 1081 HV Amsterdam (NL)
`{t.caselli;p.t.j.m.vossen}@vu.nl`

Abstract

This paper illustrates a proposal for the development of an annotation scheme and a corpus for storyline extraction and evaluation from large collections of documents clustered around a topic. The scheme extends existing annotation efforts for event coreference and temporal processing, introducing additional layers and addressing shortcomings. We also show how a storyline can be derived from the annotated data.

1 Introduction

The stream of information is increasing every day posing difficult challenges for the selection and extraction of relevant information. Relevant information can be missed in this vast amount of data, leading to inconsistencies, fragmented reports, or gaps in the extraction and representation of complex stories. Different solutions have been proposed to deal with this problem ranging from the generation of multi-document extractive summaries (Barzilay et al., 1999), to clustering of news with respect to a topic (Swan and Allan, 2000), to the generation of timelines to monitor relevant events in a topic (Shahaf and Guestrin, 2010; Nguyen et al., 2014; Bauer and Teufel, 2015).

In this work, we want to expand on a different approach to select, organize, and represent relevant information from collections of documents clustered around a specific topic. Following Vossen et al. (2015), we adopt the storyline model as a representational device to structure the information, and aim at developing a reference corpus for a quantitative and qualitative evaluation of automatically generated storylines.

A storyline is as a structured index of chronologically ordered events which overcomes representation models based on pure timelines with respect to three aspects: i.) it is able to identify salient events (*climax events*) as the central elements around which a specific topic develops; ii.) it provides an explanatory model for how events connect to each other and contribute to the development of a topic; and iii.) it mimics a pervasive phenomenon in human life, i.e. the use of narrative strategies to organize and make sense of information.

Previous work in storyline generation is limited and in most cases what is labelled as a storyline is a timeline. The main difference is that storylines and narrative structures exhibit some causal and explanatory relation between events and some tension towards a resolution, or climax. We have identified four main contributions (Shahaf et al., 2013; Huang and Huang, 2013; Hu et al., 2014; Laparra et al., 2015) in this area proposing methods to generate storyline datasets. Although each contribution proposes its own definition of storyline, based on the sharing of participants, time and location, one of the commonalities of these works consists of the use of interactions and connections between cross-document topic threads or events which give rise to timelines, i.e. a basic temporal ordering.

Storylines also differ from Narrative Schemas (Chambers and Jurafsky, 2009). Narrative Schemas qualify as sets of partially ordered events with no distinction in relevance or salience

of their elements and, most importantly, with no explanatory power of the ways events are connected together, except for precedence relations. A Narrative Schema looks like an un-prioritized set of events which share some participants, thus leading to the development of entity-centric timelines. Furthermore, the use of entity driven relations (e.g. co-participation) to generate the schemas often result in non-coherent chains of events (Peng and Roth, 2016).

The remainder of this paper will be structured as follows: in Section 2 we will present the main aspects of the storyline model described in (Vossen et al., 2015) and show how these elements have been used to develop a proposal to annotate storylines. Section 3 will report on the preliminary application of the annotation scheme to a corpus presenting insights on the data and interaction between different layers of annotation ranging from event coreference to storyline. Finally, conclusions and future work will be reported in Section 4.

2 Annotating Storylines: A Proposal

The model described in Vossen et al. (2015) is grounded on the narratology framework of Bal (1997) which assumes that every narrative, regardless of the media and content, is a mention of a *fabula*, i.e., a sequence of chronologically ordered and logically connected events involving one or more actors. A *fabula* is a complex structure whose internal components can be decomposed in three main elements: i.) the *rising action(s)*, the event(s) that increases the tension created by a predicament; ii.) the *climax*, the event(s) which creates the maximal level of tension; and iii.) *falling action(s)*, the event(s) which resolve the climax and lower the narrative tension.

These narratological concepts have been translated in the storyline model by providing a definition and a formalization for the following basic components:

- events, participants (actors), locations and time-points (settings);
- the anchoring of events to time and their ordering (a timeline);
- bridging relations: a set of relations between events with explanatory and predictive

value(s).

The proposed annotation scheme aims at grounding these concepts to linguistic elements in document collections. The scheme has been developed to maximize compatibility with existing annotation efforts on event and temporal processing, such as the Richer Event Description (RED) [1], THYME (Styler IV et al., 2014), and TimeML (Pustejovsky et al., 2003a), and event coreference, such as the Event CorefBank+ (ECB+) (Cybulska and Vossen, 2014b).

2.1 STaR: The Storyline Annotation and Representation Scheme

The Storyline Annotation and Representation Scheme (StaR) builds on and extends the ECB+ annotation scheme (Cybulska and Vossen, 2014a). The ECB+ scheme addresses event coreference both at the in- and cross-document levels. Event action coreference is specified as two action mentions which occur/hold true: i.) at the same time; ii.) in the same location; and iii.) with the same actors/participants. Thus, ECB+ data provides access to the first basic elements of the storyline model, i.e., events, participants (actors), locations, and time.

The timeline reconstruction is done by means of a temporal relation tag, `TLINK`, inheriting its semantics from TimeML. Although largely used and adapted to other languages, TimeML-annotated corpora suffer from sparse annotations and poorly connected event/time graphs. For instance, not every event mention is properly anchored to a temporal expression, nor are instructions on *when* annotated ordering relations between events clearly defined. In addition to this, the set of temporal relations adopted by TimeML is very fine-grained, with a total of 13 different values. To overcome these shortcomings of the TimeML annotation, we have designed our guidelines following two principles: i.) each event mention must be anchored to its time of occurrence; ii.) temporal ordering relations must be annotated only when in presence of linguistic evidence, thus limiting inferences. As such, no temporal relation should be annotated on the basis of world knowledge only.

Furthermore, the set of temporal values has been limited to 8 values (`BEFORE`, `AFTER`, `OVERLAP`, `BEFORE_OVERLAP`, `BEGINS_ON`, `ENDS_ON`, `SIMULTANEOUS`, `INCLUDES`). We also annotate temporal relations between events and the Document Creation Time (DCT). The DCT represents a special temporal anchor for actions which expresses a broad temporal dimension (e.g. Present, Past, or Future with respect the time the author created and published the text).

Following the proposal in Cassidy et al. (2014), we also annotate transitive closure relations between pairs of events to develop highly connected event graph. This means that in case of two pairs of events A `BEFORE` B and B `BEFORE` C, we explicitly mark the transitive closure relation A `BEFORE` C.

Finally, we extend the `TLINK` tag with the attribute `contextualModality`, from the RED scheme. It has 4 values: `ACTUAL`, `UNCERTAIN`, `HYPOTHETICAL`, and `GENERIC`. The attribute allows to represent claims of different sources concerning the reality or certainty of a temporal relation. The assignment of the contextual modality values is connected to the factuality profile of the events in the temporal relation but, at the same time, it is assumed to be independent from this latter aspect. The focus is on the factuality of the temporal relation itself. Consider the following examples from two documents about the 2013 Brooklyn riot from ECB+:

1. officers **shot** and killed a 16-year-old Kimani Gray in Brooklyn because he *allegedly* **pointed** a gun at the cops. [ecbplus19_10.xml - sentence 2]
   `TLINK`: pointed `BEFORE` shot - `UNCERTAIN`

2. Gray **pointed** a .38-caliber revolver at the cops before they **opened fire** [ecbplus19_4.xml - sentence 7]
   `TLINK`: pointed `BEFORE` opened fire - `ACTUAL`

ECB+ cross-document event coreference annotation tells us that both mentions of *pointed* are coreferential, as well as *shot* and *opened fire*. The timeline of the events is exactly the same, as expressed by the `BEFORE` relation, due to the presence of evidence such as "because" and "before". However,

the factuality of the `TLINK` is different in the two sources: in example 1. the temporal relation assumes an uncertain value while in example 2. is factual. Modeling these differences is a key element for storyline generation as these disagreements can be used to facilitate the identification of both relevant and interesting information and account for different perspectives on the same topic.

Bridging relations are modeled with a new link tag `PLOT_LINK`. The tag connects the event mentions in a document in order to reconstruct the tripartite structure of the *fabula*: rising actions, climax, and falling action. Two values are associated to the tag: `PRECONDITION`, which marks rising action relations, i.e., events which are circumstantial to, cause or enable another event, and `FALLING_ACTION`, which explicitly mark speculations and consequence relations, i.e. events which are the (anticipated) outcome or the effect of another event. The scheme is silent and neutral with respect to the climax event, i.e. no prior assumption is done. The identification of the climax event, or events, of the topic will emerge from the annotated data and it should correspond to the event(s) that has most incoming `PRECONDITION` (i.e. it is the target element of the relation) and/or outgoing `FALLING_ACTION` (i.e. it is the source element of the relation) links. `PLOT_LINK`s must be grounded on some evidence and not performed on the basis of world knowledge alone. In particular, two event mentions may stand in a `PLOT_LINK` relation if: i.) they share at least one participant (co-participation); or ii.) they stand in a causal or temporal relation; or iii.) if they stand in an entailment relation[2]. Recalling examples 1. and 2., the annotation of the `PLOT_LINK` is as follows:

1a source: *pointed* `PRECONDITION` target: *shot*
   source: *pointed* `PRECONDITION` target: *killed*
   source: *shot* `FALLING_ACTION` target: *killed*

2a source: *pointed* `PRECONDITION` target: *opened fire*

This will result in 3 `PLOT_LINK`s for the event *shot/opened fire* (2 `PRECONDITION` and 1

---

[2] We assume that not all co-participation relation may stand in an entailment relation.

FALLING_ACTION), 1 PLOT_LINKs for the event *killed*, and no PLOT_LINK for *pointed*. On the basis of the available data, *shot/opened fire* qualifies as the climax event.

3 The corpus: from ECB+ to ECBStaR

The ECB+ corpus contains 984 news articles and 43 topics, where each topic contains documents reporting on two different "seminal events" or topics. For instance, topic 19 in ECB+ contains two clusters of articles dealing with two different riots (one in Greece and one in Brooklyn). ECB+ annotation is performed through an event-centric approach. Only sentences containing mentions of the target seminal events were annotated, including any other event mention occurring in the same sentence. This has lead to a relevance based annotation: only a subset of the sentences in a document is annotated, i.e., those mentioning the target topic, while the rest is ignored. All event mentions in the selected sentences are annotated, not only those explicitly referring to the target topic. The outcome of this approach resulted in 3,487 annotated event mentions, with 2,050 coreference relations (in- and cross-document).

In Figure 1 we graphically illustrate how storylines can be reconstructed as an outcome of the annotation and interaction between the different layers. The top part of the figure contains sentences from different articles. To simplify the representation, we have only marked events (in bold) and temporal expressions (in italics). The bottom part of the picture contains a representation for event coreferece, timeline, and storyline. by creating unique representations (i.e. instance identifiers).

Each document timeline is merged together into a unique topic/seminal event timeline based on the event coreference data. Timeline representation follows the event-centric annotation of ECB+ but enriches it with time anchoring and ordering relations. In particular: i.) each event instance is associated to its correct time anchor (e.g. "2013-03-09 E1"); ii.) ordering relations based on precedence relations (i.e., BEFORE or AFTER) among instances are represented with numerical indexes (e.g. "1 2013-03-09 E1" -"2 2013-03-09 E6"). In case other types of temporal relations hold between pairs of events, the same numerical index is assigned to
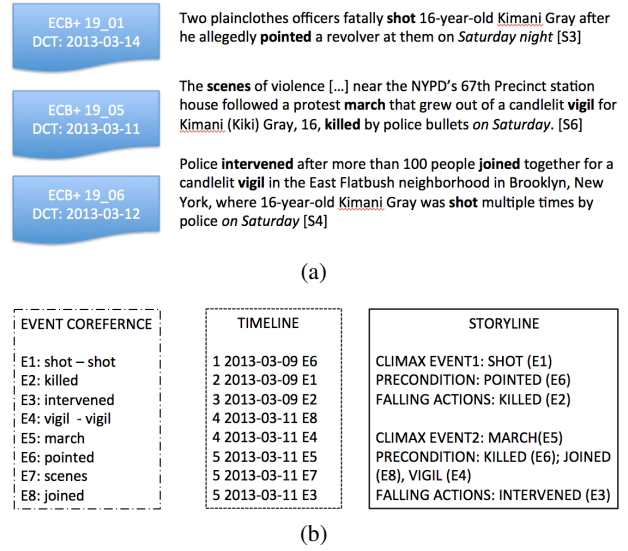

(a)

(b)

Figure 1: Storyline reconstruction and representation.

all events. The specific temporal value is not lost and can be accessed by inspecting the binary relations between the events [3]. Finally, on the basis of the PLOT_LINK annotations, a storyline can be reconstructed. In our specific case, we have identified 2 climax events ("shot" and "march") as both events obtained the same score. Each climax event is then associated to a list PRECONDITIONs and/or FALLING_ACTIONS. The connection between climax events is guaranteed by the timeline data. Furthermore, it is interesting to notice that climax identification is not based on the number of event mentions (i.e., the event with highest number of mentions in the document collection) but on the information derived from PLOT_LINKs.

The annotation effort to develop the ECBStaR corpus is still at an early stage. So far, only 3 seminal events[4] for which the DCT was available have been enriched with TLINKs and PLOT_LINKs, resulting in 33 annotated documents, 3 storylines (one per seminal events), 1,229 TLINKs, 317 PLOT_LINKs, with 223 falling action relations, 46 precondition relations, and 5 climax events. The first interesting data concerns the amount of TLINKs. Storyline annotation aims at creating densely connected temporal relation graphs, to avoid shortcomings of

---

[3] This is not illustrated in the figure to facilitate the reading.

[4] The seminal events corresponds to the ECB+ data for topics 19, 37, and 41.

previous temporal annotation initiatives based on TimeML which lacked densed connections among the annotated data.

Concerning the PLOT_LINK annotation, the large amount of falling action relations can be explained by taking into account two factors. The first concerns the nature of news data. News is published quickly, and often information about preconditions are lacking or irrelevant. Furthermore, in case of long lasting stories, such as wars, we have access to a limited set of articles concerning a specif sub-event (in our case the bombing in a South Sudan refugee camp - topic 41), thus preventing the identification of rising action relations because this information is considered to be active in the shared knowledge and irrelevant for the current document. The second factor concerns the specific topic which gives rise to the storyline. For instance, natural disaster storylines will barely have mentions of precondition actions as most of the time natural disasters are reported as things that simply occur.

4 Conclusion and Future Work

This paper reports on a proposal of an annotation scheme to develop a reference corpus for storylines from large document collections. The availability of a such a reference corpus will allow a qualitative and quantitative evaluation of storylines, taking into account the two key dimensions: timelines and *fabula*, i.e., the identification of the climax events and the explanatory connections with other events.

One additional insight we hope to achieve with the ECBStaR corpus is the identification of specialized event patterns for different storylines, thus contributing to new models of knowledge template acquisition.

Finally, we are planning to extend the collected data per topic in time. In particular, we aim at harvesting additional documents related to each seminal events to extend the time span of the topic and better monitor the evolution of the story.

References

Mieke Bal. 1997. *Narratology: Introduction to the theory of narrative*. University of Toronto Press.

Regina Barzilay, Kathleen R McKeown, and Michael Elhadad. 1999. Information fusion in the context of multi-document summarization. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 550–557. Association for Computational Linguistics.

Sandro Bauer and Simone Teufel. 2015. A methodology for evaluating timeline generation algorithms based on deep semantic units. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 834–839, Beijing, China, July. Association for Computational Linguistics.

Taylor Cassidy, Bill McDowell, Nathanael Chambers, and Steven Bethard. 2014. An annotation framework for dense event ordering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 501–506, Baltimore, Maryland, June. Association for Computational Linguistics.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 602–610. Association for Computational Linguistics.

Agata Cybulska and Piek Vossen. 2014a. Guidelines for ECB+ annotation of events and their coreference. Technical Report NWR-2014-1, VU University Amsterdam.

Agata Cybulska and Piek Vossen. 2014b. Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26-31.

Po Hu, Min-Lie Huang, and Xiao-Yan Zhu. 2014. Exploring the interactions of storylines from informative news events. *Journal of Computer Science and Technology*, 29(3):502–518.

Lifu Huang and Lian'en Huang. 2013. Optimized event storyline generation based on mixture-event-aspect model. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 726–735, Seattle, Washington, USA, October. Association for Computational Linguistics.

Egoitz Laparra, Itziar Aldabe, and German Rigau. 2015. From timelines to storylines: A preliminary proposal for evaluating narratives. In *Proceedings of the First Workshop on Computing News Storylines*, pages 50–55, Beijing, China, July. Association for Computational Linguistics.

Kiem-Hieu Nguyen, Xavier Tannier, and Veronique Moriceau. 2014. Ranking multidocument event de-

scriptions for building thematic timelines. In *Proceedings of COLING'14*, pages 1208–1217.

Haoruo Peng and Dan Roth. 2016. Two discourse driven language models for semantics. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 290–300, Berlin, Germany, August. Association for Computational Linguistics.

James Pustejovsky, José Castao, Robert Ingria, Roser Saurì, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003a. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Fifth International Workshop on Computational Semantics (IWCS-5)*.

Dafna Shahaf and Carlos Guestrin. 2010. Connecting the dots between news articles. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 623–632. ACM.

Dafna Shahaf, Jaewon Yang, Caroline Suen, Jeff Jacobs, Heidi Wang, and Jure Leskovec. 2013. Information cartography: creating zoomable, large-scale maps of information. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1097–1105. ACM.

William F Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, et al. 2014. Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2:143–154.

Russell Swan and James Allan. 2000. Automatic generation of overview timelines. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 49–56. ACM.

Piek Vossen, Tommaso Caselli, and Yiota Kontzopoulou. 2015. Storylines for structuring massive streams of news. In *Proceedings of the First Workshop on Computing News Storylines*, pages 40–49, Beijing, China, July. Association for Computational Linguistics.