Storylines for structuring massive streams of news

Piek Vossen
Faculty of Humanities
VU University Amsterdam
p.t.j.m.vossen@vu.nl

Tommaso Caselli
Faculty of Humanities
VU University Amsterdam
t.caselli@vu.nl

Yiota Kontzopoulou
Faculty of Humanities
VU University Amsterdam
p.kontzopoulou@gmail.com

Abstract

Stories are the most natural ways for people to deal with information about the changing world. They provide an efficient schematic structure to order and relate events according to some explanation. We describe (1) a formal model for representing storylines to handle streams of news and (2) a first implementation of a system that automatically extracts the ingredients of a storyline from news articles according to the model. Our model mimics the basic notions from narratology by adding bridging relations to timelines of events in relation to a climax point. We provide a method for defining the climax score of each event and the bridging relations between them. We generate a JSON structure for any set of news articles to represent the different stories they contain and visualize these stories on a timeline with climax and bridging relations. This visualization helps inspecting the validity of the generated structures.

1 Introduction

News is published as a continuous stream of information in which people reflect on the changes in the world. The information that comes in is often partial, repetitive and, sometimes, contradictory. Human readers of the news trace information on a day to day basis to build up a story over time. When creating this story, they integrate the incoming information with the known, remove duplication, resolve conflicts and order relevant events in time. People also create an explanatory and causal scheme for what happened and relate the actors involved to these schemes.

Obviously, humans are limited in the amount of news that they can digest and integrate in their minds. Even though they may remember very well the main structure of the story, they cannot remember all the details nor the sources from which they obtained the story. Estimates are that on a single working day, millions of news articles are published. Besides the fact that the data is massive, the information is also complex and dynamic. Current search-based solutions and also topic tracking systems (Google trends, Twitter trends, EMM Newsbrief[1], Yahoo news) can point the reader/user to important news but they cannot organize the news as a story as humans tend to do: deduplicating, aggregating, ordering in time, resolving conflicts and providing an explanatory scheme.

In this paper, we present a formal model for representing time series of events as storylines and an implementation to extract data for this model from massive streams of news. Our formal model represents events and participants as instances with pointers to the mentions in the different sources. Furthermore, events are anchored in time and relative to each other, resulting in timelines of events. However, not every timeline is a storyline. We therefore use event relations (*bridging relations*) and event salience to approximate the *fabula*, or plot structure, where the most salient event (the climax of the storyline) is preceded and followed by events that explain it. Our implementation of the storyline extraction module is built on top of an NLP pipeline for processing text that results in a basic timeline structure.

The remainder of this paper is structured as follows. In Section 2, we present the theoretical background based on narratology frameworks which inspired our model described in Section 3. Section 4, then, explains our system for extracting storyline data from news streams according to the model. In Section 5, we report related works and highlight differences and similarities with respect

---

[1] http://emm.newsbrief.eu/NewsBrief/clusteredition/en/latest.html

to our system. Finally, we discuss the status of our work, possible evaluation options and future work in Section 6.

2 What is a story?

Stories are a pervasive phenomenon in human life. They are explanatory models of the world and of its happenings (Bruner, 1990). Our mind constantly struggles to extract meaning from data collected through our senses and, at the same time, tries to make sense out of these data. This continuous search for meaning and meaningful patterns gives rise to stories.

In this paper, we make reference to the narratology framework of Bal (Bal, 1997) to identify the basic concepts which have informed our model. Every story is a mention of a *fabula*, i.e., a sequence of chronologically ordered and logically connected events involving one or more actors. Actors are the agents, not necessarily humans, of a story that perform actions. In Bal's framework "acting" refers both to performing and experiencing an event. Events are defined as transitions from one state to another. Furthermore, every story has a focalizer, a special actor from whom's point of view the story is told. Under this framework, the term "story" is further defined as the particular way or style in which something is told. A story, thus, does not necessarily follow the chronological order of the events and may contain more than one fabula.

Extending the basic framework and focusing on the internal components of the fabula, a kind of universal grammar can be identified which involves the following elements:

- Exposition: the introduction of the actors and the settings (e.g. the location);

- Predicament: it refers to the set of problems or struggles that the actors have to go through. It is composed by three elements: *rising action*, the event(s) that increases the tension created by the predicament, *climax*, the event(s) which creates the maximal level of tension , and, finally, *falling action*, the event(s) which resolve the climax and lower the tension;

- Extrication: it refers to the "end" of the predicament and indicates the ending.

Figure 1: Fabula or Plot structure

Figure 1 is a graphical representation of the internal components of the fabula.

Possible predicaments can be restricted to a closed set of high-level representations (e.g. the actor vs. society; the actor vs. nature; the actor vs himself; the actor vs. another actor), giving rise to recurring units and rules which describe their relations (Propp, 2010).

A further element is the hierarchical nature and the inherent intersection of stories. Multiple stories can be present in a single text and the same event, or set of events, may belong to different stories.

The model allows to focus on each its the components, highlighting different, though connected, aspects: the internal components of the fabula are event-centered; the actors and the focalizer allows access to opinions, sentiments, emotions and world views; and, the medium to the specific genres and styles.

These basic concepts and ingredients apply to every narrative texts, no matter the genre, such as novels, children stories, comic strips. News as a stream of separate articles, however, forms a special type of narrative that tends to focus on climax events on a routine basis (Tuchman, 1973): events with news value need to be published quickly while there may be little information on their rising action(s). At the same time, the falling action(s) and the extrication are not always available, often leading to speculation. Successive news articles may add information to the climax event explaining the rising action(s) towards the climax event and describing any follow up events when time passes.

In the following section we will describe our computational model and how it connects to these basic ingredients.

41

3 A computational model for storylines

Many different stories can be built from the same set of events. The starting point for a story can be a specific entity, a location, an event (Van Den Akker et al., 2011), from which time-ordered series of events spin off through relations that explain the causal nature of their order.

In our model we use the term *storylines* to refer to an abstract structured index of connected events which provides a representation matching the internal components of the fabula (rising action(s), climax, falling action(s) and resolution). On the other hand, we reserve the term *story* for the textual expression of such an abstract structure[2]. Our model, thus, does not represent texts but event data from which different textual representations could be generated. The basic elements of a storyline are:

- A definition of events, participants (actors), locations and time-points (settings)

- Anchoring of events to time

- A timeline (or basic fabula): a set of events ordered for time (chronological order)

- Bridging relations: a set of relations between events with explanatory and predictive value(s) (rising action, climax and falling action)

In the next subsections, we describe how we formalized these ingredients.

3.1 Mentions and instances

As explained in Section 2, a stream of news consists of many separate articles published over time that each give different pieces of information from different temporal perspectives (looking backward or looking forward in time) with partially overlapping information. We therefore first need to make a distinction between mentions of events and the unique instances of events to which these mentions refer. For this, we take the Grounded Annotation Framework (GAF, (Fokkens et al., 2014)) as a starting point. GAF allows to make a formal difference between mentions in texts and instances. Instances are modelled through the Simple Event Model (SEM, (Van Hage et al., 2011).

SEM is an RDF model for capturing event data at an instance level through unique URIs. Following the SEM model, events consist of an action, one or more actors, a place and a time. A textual analysis detects mentions of these instances and their relations, where typically the same instance can be mentioned more than once. GAF connects the representation of these instances in SEM to the mentions in text through a *gaf:denotedBy* relation. Given the following text fragment:

> A380 makes maiden flight to US. March 19, 2007. The Airbus A380, the world's largest passenger plane, was set to land in the United States of America on Monday after a test flight. One of the A380s is flying from Frankfurt to Chicago via New York; the airplane will be carrying about 500 people.

We create an RDF representation in SEM with a single instance of a *flying* event through a unique identifier *ev17Flight*. Furthermore, it shows time, place and actor relations to entities identified in DBpedia:

```
:ev17Flight
rdfs:label "maiden flight", "test flight", "flying" ;
gaf:denotedBy
  wikinews:A380_makes_maiden_flight_to_US#char=19,25,
  wikinews:A380_makes_maiden_flight_to_US#char=174,180,
  wikinews:A380_makes_maiden_flight_to_US#char=202,208;
sem:hasTime wikinews:20070319;
sem:hasActor dbp:Airbus_A380, wikinews:500_people;
sem:hasPlace dbp:United_States, dbp:Frankfurt, dbp:Chicago,
 dbp:New_York.
```

The RDF structure provides a unique semantic representation of the event instance through the URI *:ev17Flight*, with `sem:hasActor`, `sem:hasTime` and `sem:hasPlace` relations to the participating entities that are also represented as instances through URIs. The `gaf:denotedBy` relations point to the offset positions in the sources where the event is mentioned. The participants in the event get similar representations with `gaf:denotedBy` relations to their mentions. Events and participants can be mentioned in different sentences and different news articles. Their relations are, however, represented in a single structure, a so-called *event-centric knowledge graph*. As such, GAF provides a natural way for resolving coreference, apply deduplication and aggregate information from different sources. In the above RDF example

3.2 Timelines

Instance representations for events require associating them to time. Such time anchors are minimally required to determine if two mentions of an

event refer to the same event instance. Mentions anchored to different points in time cannot refer to the same event by definition. If no time anchoring is provided, we cannot determine the instance representation of the event and we are forced to ignore the event at the instance level[3]. Event timelines are thus a natural outcome of the model. Timelines are then sequences of event instances anchored to a time expression or relative to each other.

3.3 Towards Storylines

Given a timeline for a specific period of time, we define a storyline S as n-tuples T, E, R such that:

$$\mathbf{T}imepoints = (t_1, t_2, , ..., t_n)$$

$$\mathbf{E}vents = (e_1, e_2, ..., e_n)$$

$$\mathbf{R}elations = (r_1, r_1, ..., r_n)$$

T consists of an ordered set of points in time, E is a set of events and R is a set of bridging relations between these events. Each e in E is related to a t in T. Furthermore, for any pair of events e_i and e_j, where e_i precedes e_j there holds a bridging relation $[r, e_i, e_j]$ in R.

We assume that there is a set of timelines L for every E, which is any possible sequence of events temporally ordered. Not every temporal sequence l of events out of L makes a good storyline. We want to approximate a storyline that people value by defining a function that maximizes the set of bridging relations across different sequences of events l in L. We therefore assume that there is one sequence l that maximizes the values for R and that people will appreciate this sequence as a story. For each l in L, we therefore assume that there is a bridging function B over l that sums the strength of the relations and that the news storyline S is the sequence l with the highest score for B:

$$S(E) = MAX(B(l))$$

$$B(l)) = \sum_{i,j=1}^{n} C(r, e_i, e_j)$$

Our bridging function B sums the connectivity strength C of the bridging relations between all time-ordered pairs of events from the set of

temporally ordered events l. The kind of bridging relation r and the calculation of the connectivity strength C can be filled in in many ways: co-participation, expectation, causality, enablement, and entailment, among others. In our model, we leave open what type of bridging relations people value. This needs to be determined empirically in future research.

The set L for E can be very large. However, narratology models state that stories explain climax events through sequences of preceding and following events. It thus makes sense to consider only those sequences l that include a salient event as a climax and relate the other events to this climax event. Instead of calculating the score B for all l in L, we thus only need to build event sequences around events that are most salient as a climax event and select the other events on the basis of the strength of their bridging relation with that climax. For any climax event e_c, we can therefore define:

$$MAX(B(e_c \epsilon E)) = \max_{i=1}^{n} C(r, e_i, e_c)$$

The climax value for an event can be defined on the basis of salience features, such as:

- prominent position in a source;

- number of mentions;

- strength of sentiment or opinion;

- salience of the involved actors with respect to the source.

An implementation should thus start from the event with the highest climax score. Next, it can select the preceding event e_l with the strongest value for r. Note that this is not necessarily the event that is most close in time. After that, the event e_l with the strongest connectivity is taken as a new starting point to find any event e_k preceding this event with the highest value for r. This is repeated until there are no preceding events in the timeline l. The result is a sequence of events up to e_c with the strongest values for r. The same process is repeated forward in time starting from e_c and adding e_m with the strongest connectivity value for r, followed by e_n with the strongest connectivity score r to e_m. The result is a sequence of events with local maxima spreading from e_c:

$$...e_k r_{max} e_l r_{max} e_c r_{max} e_m r_{max} e_n \cdots$$

---

[3]In practice, we anchor so-called *timeless* events to the document-creation time by default or speculated events to a future time bucket

This schema models the optimized storyline starting from a climax event. By ranking the events also for their climax score, the climax events will occupy the highest position and the preceding and following events the lower positions approximating the fabula or plot graph shown in Figure 1.

4 Detecting storylines: Preliminary Experiments

In this section we describe a first implementation of our model and its steps for the storyline generation: a.) timeline extraction; b.) climax event identification; c.) rising and falling actions identification.

4.1 Extracting timelines

The timeline extraction is obtained from an NLP pipeline that has been developed in the News-Reader project[4]. The pipeline applies a cascade of modules, ranging from tokenization up to temporal and causal relation extraction, to documents (mention level). Next, it generates a semantic representation of the content in SEM (instance level). The NLP modules generate representations of entities mentioned in the text with possible links to DBpedia URIs, time expressions normalized to dates and a semantic role representation with events and participants linked to FrameNet frames and elements (Baker et al., 1998). Furthermore, coreference relations are created to bind participants and events to instances within each document. The NLP modules interpret mentions in the text, i.e. at single document level. However, given a set of documents or a corpus, these mention based representations are combined resolving cross-document coreference for entities and events, anchoring events to time and aggregating event-participant relations and generating an instance level representation. Details about this process can be found in (Agerri et al., 2014).

The timeline representation anchors events either to a time anchor in the document or to the document publication time. In case a time anchor cannot be determined or inferred, or if the resulting value is too vague (e.g. "PAST_REF"), the event is presented in the timeline but with an under-specified anchor such as XXXX-XX-XX. A natural result of this representation is a time-line of events, as described in (Minard et al.,

---

```
3        2004-XX-XX     1173-3-deal     1173-2-deal
4        2004-03-XX     3307-10-purchased
4        2004-03-XX     3307-2-carry    3307-5-expected   3307-9-expected
4        2004-03-XX     3307-5-sales    3307-8-sales
4        2004-03-XX     3307-8-talks    3307-5-break
4        2004-03-XX     3307-9-flight   3307-4-flight     3307-9-flight
4        2004-03-XX     3307-9-take     3307-9-take       3307-3-take
4        2004-03-XX     4764-9-scheduled
5        2004-10-XX     1173-33-saying  1173-27-saying    1173-38-said
```

Figure 2: Event-centered Timeline

2015). In Figure 2, we show an example of such a timeline constructed from the SemEval 2015 Task 4: TimeLine: Cross-Document Event Ordering[5] data. This representation differs from the Gold data of the task because it is "event-centered". This means that the events are ordered not with respect to a specific actor or entity. Each line corresponds to a time stamped event instance. Lines with multiple events indicate in-document event coreference. The first element of a timeline represents a unique index. Events with under-specified time anchors are put at the beginning of the timeline with index 0. Simultaneous events are associated with the same index. Events here are represented at token level and associated with document id and sentence number.

Although, all events may enter in a timeline, including speech-acts such as *say*, not every sequence of ordered events makes a storyline. The timeline structures are our starting point for extracting a storyline.

4.2 Determining the event salience

Within the set of events in a timeline, we compute for each event its prominence on the basis of the mention sentence number and the number of mentions in the source documents. We currently sum the inverse sentence number of each mention of an event in the source documents:

$$P(e) = \sum_{e_m=1 \to N} (1/S(e_m)).$$

All event instances are then ranked according to the degree of prominence P.

We implemented a greedy algorithm in which the most prominent event will become the climax event[6]. Next, we determine the events with the strongest bridging relation preceding and following the climax event in an iterative way until there are no preceding and following events with a bridging relation. Once an event is added to a storyline it cannot be added to another storyline. For

---

all remaining events (not connected to the event with the highest climax score), we select again the event with the highest climax score of the remaining events and repeat the above process. Remaining events thus can create parallel storylines although with a lower score. When descending the climax scores, we ultimately are left with events with low climax score that are not added to any storyline and do not constitute storylines themselves.

For determining the value of the bridging relations we use various features and resources, where we make a distinction between structural and implicit relations:

- Structural relations:
  - co-participation;
  - explicit causal relations;
  - explicit temporal relations;

- Implicit relations:
  - expectation based on corpus co-occurrence data;
  - causal WordNet relation;
  - frame relatedness in FrameNet;
  - proximity of mentions;
  - entailment;
  - enablement.

Our system can use any of the above relations and resources. However, in the current version, we have limited ourselves to co-participation and FrameNet frame relations. Co-participation is the case when two events share at least one participant URI which has a PropBank relation *A0*, *A1* or *A2*. The participant does not need to have the same relation in the two events. Events are related to FrameNet frames if there is any relation between their frames in FrameNet up to a distance of 3.

In the Appendix A, we show an example of a larger storyline extracted from the corpus used in the SemEval 2015 Timeline task. The storyline is created from a climax event [`"purchase"`] involving Airbus with a score of *61*. The climax event is marked with `C` at the beginning of the line. Notice that the climax event of this storyline is also reported in Figure 2, illustrating the event-centered timeline (`[4 2004-03-XX 3307-10-purchased]`). After connecting the other events, they are sorted according to their time anchor. Each line in Appendix A is a unique

event instance (between square brackets) anchored in time, preceded by the climax score and followed by major actors involved[7]. We can see that all events reflect the commercial struggle between *Airbus* and *Boeing* and some role played by governments.

In Figure 3, we visualize the extracted storylines ordered per climax event. Every row in the visualization is a storyline grouped per climax event, ordered by the climax score. The label and weight of the climax event is reported in the vertical axis together with the label of the first participant with an *A1* Propbank role, which is considered to be most informative. Within a single row each dot presents an event in time. The size of the dot represents the climax score. Currently, the bridging relations are not scored. A bridging relation is either present or absent. If there is no bridging relation, the event is not included in the storyline. When clicking on a target storyline a pop up windows open showing the storyline events ordered in time (see Figure 4). Since we present events at the instance level across different mentions, we provide a semantic class grouping these mentions based on WordNet which is shown on the first line. Thus the climax event "purchase" is represented with the label more general label "buy" that represents a hypernym synset. If a storyline is well structured, the temporal order and climax weights mimic the fabula internal structure, as in this case. We expect that events close to the climax have larger dots than more distant events in time[8].

Stories can be selected per target entity through the drop-down menu on top of the graph. In the Figure 3, all stories concerning Airbus are marked in red.

Comparing the storyline representation with the timeline (see Figure 2) some differences can be easily observed. In a storyline, events are ordered in time and per climax weight. The selection of events in the storyline is motivated by the bridging relations which exclude non-relevant events, such as *say*.

We used the visualization to inspect the results. We observed that some events were missed because of metonymic relations between partici-

---

[7]We manually cleaned and reduced the actors for reasons of space.

[8]In future work we will combine the prominence with a score for the strength of the bridging and reflect this in the size.

pants, e.g. *Airbus* and *Airbus_380* are not considered as standing in a co-participation relation by our system because they have different URIs. In other cases, we see more or less the opposite: a storyline reporting on journeys by *Boeing* is interrupted by a plane crash from *Airbus* due to overgenerated bridging relations.

What is the optimal combination of features still needs to be determined empirically. For this we need a data set, which we will discuss in the next subsection.

4.3 Benchmarking and evaluation

In this phase we are not yet able to provide an extensive evaluation of the system.

Evaluation methods for storylines are not trivial. Most importantly, they cannot be evaluated with respect to standard measures such as Precision and Recall. In this section, we describe and propose a set of evaluation methods to be used as a standard reference method for this kind of tasks.

The evaluation of a storyline must be based, at least, on two aspects: informativeness and interest. A good storyline is a storyline which interest the user, provides all relevant and necessary information with respect to a target entity, and it is coherent. We envisage two types of evaluation: direct and indirect. Direct evaluation necessarily needs human interaction. This can be achieved in two methods: using experts and using crowdsourcing techniques.

Experts can evaluate the data provided with the storylines with respect to a set of reference documents and check the informativeness and coherence parameters. Following (Xu et al., 2013), two types of questions can be addressed at the *micro-level* and at the *macro-level* of knowledge. Both evaluation types address the quality of the generated storylines. The former addresses the efficiency of the storylines in retrieving the information while the latter addresses the quality of the storylines with respect to a certain topic (e.g. the commercial "war" between Boeing and Airbus). Concerning metrics, micro-knowledge can be measured by the time the users need to gather the information, while the macro-knowledge can be measured as the text proportion, i.e. how many sentences of the source documents composing the storyline are used to write a short summary.

Crowdsourcing can be used to evaluate the storylines by means of simplified tasks. One task can ask the crowd to identify salient events in a corpus and then validate if the identified events correlate with the climax events of the storylines.

Indirect evaluation can be based on a cross-document Summarization tasks. The ideal situation is the one in which the storyline contains the most salient and related events. These sets of data can be used either to recover the sentences in a collection of documents and generate an extractive summary (story) or used to produce an abstractive summary. Summarization measures such as ROUGE can then be used to evaluate the quality of summaries and, indirectly, of the storylines (Nguyen et al., 2014; Huang and Huang, 2013; Erkan and Radev, 2004).

5 Related Works

Previous work on storyline extraction is extensive and ranges from (computational) model proposals to full systems. An additional element which distinguishes these works concerns the type of datasets, i.e., fictitious or news documents, used or referred to for the storyline generation or modelization. Although such differences are less relevant for the development of models, they are important for the development of systems. Furthermore, the task of storyline extraction is multidisciplinary, concerning different fields such as Multi Document Summarization, Temporal Processing, Topic Detection and Tracking. What follows is a selection of previous works which we consider more strictly related to our work.

Chambers and Jurafsky (Chambers and Jurafsky, 2009) extended previous work on the identification of "event narrative chains", i.e., sets of partially ordered events that involve the same shared participant. They propose an unsupervised method to learn *narrative schemas*, i.e. coherent sequences of events whose arguments are filled with participants' semantic roles. The approach can be applied to all text types. The validity of the extracted narrative schemas (event and associated participants) have been evaluated against FrameNet and on a narrative cloze task: a variation of the cloze task defined by (Taylor, 1953). The narrative schema proposed perform much better than the simpler narrative chains, achieving an improvement of 10.1%.

McIntyre and Lapata (McIntyre and Lapata, 2009) developed a data-driven system for short children's stories generation based on co-

occurrence frequencies learned from a training corpus. They generate story structure in the form of a tree, where each node is a sentence assigned with a score based on the mutual information metric as proposed in (Lin, 1998). The story generator traverses the tree and generates the story by selecting the nodes with the highest scores. Evaluation was carried out by asking to 21 human judges to rank the generated stories with respect to three parameters: Fluency, Coherence and Interest. The results have shown that the story generated by the system outperforms other versions of the system which rely on deterministic approaches. One relevant result from this work is the scoring of the tree nodes and the consequent generation of the story based on these scoring which aims at capturing the internal elements of the fabula.

Nguyen et al. (Nguyen et al., 2014) developed a system for thematic timeline generation from news articles. A thematic timeline is a set of ranked time anchored events based on a general-domain topic provided by a user query. The authors developed a two-step approach inter-cluster ranking algorithm which aims at selecting salient and non-redundant events. The topic timeline is built from time clustered events, i.e. all events occurring at a specific date and relevant with respect to the user query. The dates are ranked by salience on the basis of their occurrences with respect to the topic related events retrieved by the query. On top of this temporal cluster, events are ranked per salience and relevance. Event salience is obtained as the average of the term frequency on a date, while event relevance is a vector based similarity between the query and the time clustered document. A re-ranking function is used to eliminate redundant information and provide the final thematic timeline. The timeline thus obtained have been evaluated against Gold standard thematic timelines generated by journalists with respect two parameters: the dates and the content. As for the dates, the evaluation aims at comparing that the dates selected as relevant and salient for a certain topic are also those which occur in the Gold data. Mean Average Precision has been used as a metrics with the system scoring 77.83. The content evaluation determines if the selected events also occur in the Gold data. For this evaluation the ROUGE metric has been used, assuming that the generated timeline and the Gold data are summaries. The system scored Precision 31.23 and Recall 26.63 outper-

forming baseline systems based on date frequency only and a version of the system without the re-ranking function.

6 Conclusion and Future Work

We presented a computational model for identifying storylines starting from timelines. The model is based on narratology frameworks which have proven valid in the analysis of different types of text genres. A key concept in our model is the climax event. This notion is a relative one: each event has a climax score whose weight depends on the number of mentions and the prominence of each mention. Individual scores are normalized with respect to a data set to the maximum score. Next, storylines are built from climax events through bridging relations. In the current version of the system, we have limited the set of bridging relations to co-participation and FrameNet frame relations. Both relations are not trivial and pose some questions on how to best implement them. In particular, the notion of co-participation needs to be better defined. Possible solutions for this issue may come from previous works such as (Chambers and Jurafsky, 2009).

The set of proposed bridging relations requires further refinements both in terms of definitions and on their implementation. In particular, the big question is how to find the right balance between lexicographic approaches and machine learning techniques for identifying complex relations such as causations, enablement and entailment.

The preliminary results are encouraging although still far from perfect. Evaluation of the extracted storyline is still an open issue which has been only discussed in a theoretical way in this contribution. Methods for evaluating this type of data are necessary as the increasing amount of information suggests that approaches for extracting and aggregating information are needed.

The model proposed is very generic, but its implementation is dependent on a specific text type, news articles, and exploit intrinsic characteristics of these type of data. An adaptation to other text genres, such as fictitious works, is envisaged but it will require careful analyses of the characteristics of these data.

Acknowledgements

the NWO Spinoza Prize project Understanding Language by Machines (sub-track 3).

References

Rodrigo Agerri, Itziar Aldabe, Zuhaitz Beloki, Egoitz Laparra, Maddalen Lopez de Lacalle, German Rigau, Aitor Soroa, Antske Fokkens, Ruben Izquierdo, Marieke van Erp, Piek Vossen, Christian Girardi, and Anne-Lyse Minard. 2014. Event detection, version 2. NewsReader Deliverable 4.2.2.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 17th International Conference on Computational Linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

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

Jerome S Bruner. 1990. *Acts of meaning*. Harvard University Press.

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*, pages 602–610, Suntec, Singapore, August. Association for Computational Linguistics.

Günes Erkan and Dragomir R Radev. 2004. Lexrank: graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, pages 457–479.

Antske Fokkens, Aitor Soroa, Zuhaitz Beloki, Niels Ockeloen, German Rigau, Willem Robert van Hage, and Piek Vossen. 2014. NAF and GAF: Linking linguistic annotations. In *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, page 9.

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.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2*, pages 768–774. Association for Computational Linguistics.

Neil McIntyre and Mirella Lapata. 2009. Learning to tell tales: A data-driven approach to story generation. 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*, pages 217–225, Singapore.

Anne-Lyse Minard, Manuela Speranza, Eneko Agirre, Itziar Aldabe, Marieke van Erp, Bernardo Magnini, German Rigau, and Ruben Urizar. 2015. SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 777–785, Denver, Colorado, June. Association for Computational Linguistics.

Kiem-Hieu Nguyen, Xavier Tannier, and Veronique Moriceau. 2014. Ranking multidocument event descriptions for building thematic timelines. In *Proceedings of COLING'14*, pages 1208–1217.

Vladimir Propp. 2010. *Morphology of the Folktale*. University of Texas Press.

Wilson L Taylor. 1953. "cloze procedure": a new tool for measuring readability. *Journalism quarterly*.

Gaye Tuchman. 1973. Making news by doing work: Routinizing the unexpected. *American journal of Sociology*, pages 110–131.

Chiel Van Den Akker, Susan Legêne, Marieke Van Erp, Lora Aroyo, Roxane Segers, Lourens van Der Meij, Jacco Van Ossenbruggen, Guus Schreiber, Bob Wielinga, Johan Oomen, et al. 2011. Digital hermeneutics: Agora and the online understanding of cultural heritage. In *Proceedings of the 3rd International Web Science Conference*, page 10. ACM.

Willem Robert Van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. 2011. Design and use of the simple event model (sem). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2):128–136.

Shize Xu, Shanshan Wang, and Yan Zhang. 2013. Summarizing complex events: a cross-modal solution of storylines extraction and reconstruction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1281–1291, Seattle, Washington, USA, October. Association for Computational Linguistics.

Appendix A: Storyline Example

```
:Airbus

29      20040101 ["manufacturer","factory","manufacture"] :Boeing:European_aircraft_manufacturer_%2C_Airbus_%2C:Airbus
3       20041001 ["charge","level","kill"] :United_States_Department_of_Defense:the_deal
[C]61   20040301 ["purchase"] :People_s_Republic_of_China:Airbus_aircraft
23      20050613 ["win"] :European_aircraft_manufacturer_%2C_Airbus_%2C:Boeing
6       20050613 ["aid","assistance","assist"] :Airbus:Boeing:for_the_new_aircraft
1       20050613 ["spark"] :Airbus
15      20061005 ["compensate"] :Airbus:of_its_new_superjumbo_A380_%27s
22      20070228 ["cut","have","reduction","make"] :Airbus:the_company
39      20070319 ["supply","provide","resource","supplier","fund","tube"] :European_Aeronautic_Defence_and_Space_Company_EADS_N.V.
                                                                          :Airbus:United_States_Department_of_Defense
21      20070319 ["carry","carrier"]:the_airplane:Airbus_will
12      20070609 ["jet"]:Airbus:Airbus_A320_family
3       20070708 ["write","letter"]:Airbus:Boeing
21      20080201 ["ink","contract","sign"]:Royal_Air_Force:Airbus
13      20090730 ["lead","give","offer"] :France:Airbus
4       20041124 ["personnel","employee"] :Airbus:Former_military_personnel
20      20141213 ["carry","flight","fly"] :The_plane:Airbus
```
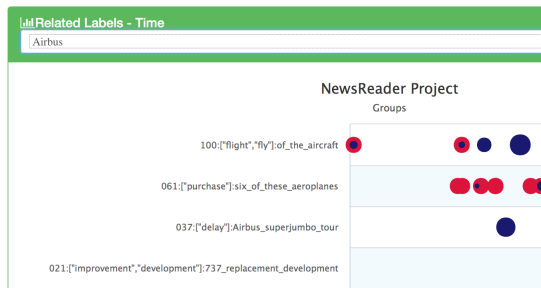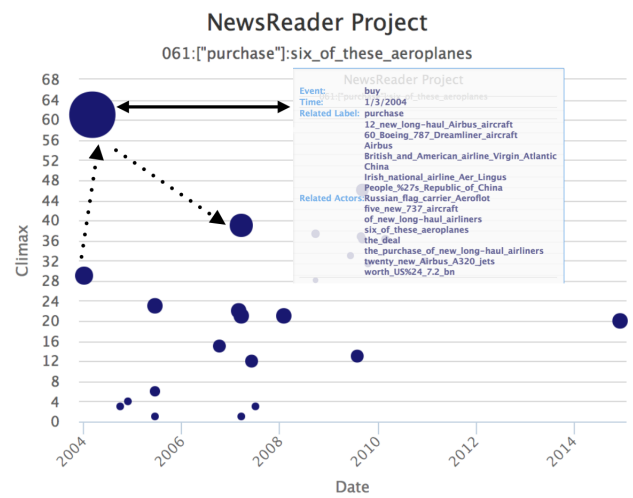

Figure 3: Airbus storylines order per climax event

Figure 4: Airbus storyline for climax event [61] "purchase"

49