6

A Narratology-Based Framework for Storyline Extraction

Piek Vossen, Tommaso Caselli, and Roxane Segers

Abstract. Stories are a pervasive phenomenon of human life. They also represent a cognitive tool to understand and make sense of the world and of its happenings. In this contribution we describe a narratology-based framework for modeling stories as a combination of different data structures and to automatically extract them from news articles. We introduce a distinction among three data structures (timelines, causelines, and storylines) that capture different narratological dimensions, respectively chronological ordering, causal connections, and plot structure. We developed the Circumstantial Event Ontology (CEO) for modeling (implicit) circumstantial relations as well as explicit causal relations and create two benchmark corpora: ECB+/CEO, for causelines, and the Event Storyline Corpus (ESC), for storylines. To test our framework and the difficulty in automatically extract causelines and storylines, we develop a series of reasonable baseline systems.

6.1 Introduction

We experience life as a story developing to the good or to the bad, in which specific events form turning points. We also tell such stories about (other) people; for example, through books, news, and blogs. Stories, as they are commonly understood, are more than *chronologically ordered* sequences of events. In a story, we typically select certain events and not others. One factor that determines this selection is the causal relation that we focus on to *explain* why things happened. Another factor is that we see events from a *perspective* that is reflected in our framing; e.g., focusing on the victim or those responsible. A consequence of this is that stories can be labeled as boring, interesting, exciting, informative, among others. Storytelling creates unique narrative objects in which temporal, explanatory, and perspective relations are mixed. A text thus represents a complex relation between the temporal order of events (a *timeline*); their causal, or explanatory, connections (a *causeline*); and the way in which these relations are framed and reported, taking one event as the turning point; that is, the event with high impact to the framed participants. Other events may be related to this turning point as preceding and explaining (or triggering) its occurrence, or following and being its consequence, and together form a *storyline*. Typically, such turning points are presented as the climax to which stories develop over time, as is proposed in narratological frameworks (Bal, 1997).

Our contributions can be summarized as follows:

- We present a formal model for such storylines, encompassing the notions of timeline, causeline, and storyline as measurable and quantifiable properties of events (Section 6.2).
- We derive two complementary annotation schemes from this model and apply these to news data, resulting in two new benchmark corpora for causelines and storylines (Section 6.3);
- We carry out a set of experiments using various approaches to extract causelines and storylines and evaluate these against the benchmark corpora (Section 6.4).

6.2 A Narratology-Grounded Framework for Storylines Identification

Storytelling is a pervasive human activity dealing with streams of information, selecting the relevant aspects and discarding irrelevant ones. The outcome of the storytelling processes are best defined as *narrative objects*, or simply narratives. We create narratives for reasoning over and understanding changes and to facilitate decision-making processes accordingly (Bruner, 1990; Boyd, 2009; Gottschall, 2012). Narratives qualify as a specific type of discourse only if certain conditions are met (Forster, 1956; Bal, 1997; Mani, 2012), namely:

- they contain a set of events (fictional or real);
- they have participants (i.e., actors) involved in the events;
- they give rise to a chronologically and logically ordered sequence of the events involved;
- they contain a *focalizer* (Bal, 1997), or a perspective from which the narrative is told.

According to these conditions, the single mention of an event (e.g., a killing, an election, or a greeting) is not a narrative per se. The core distinguishing

https://doi.org/10.1017/9781108854221.008 Published online by Cambridge University Press

Figure 6.1 Plot structure representation.

property of narratives from other types of discourse is the combination of *chronologically* and *logically* ordered sequence of events. Narratology frameworks label a chronologically sequence of events as *fabula*, or story, and a logically ordered sequence of events as **plots**. Quoting Forster (1956), a text like "*The king died and then the queen died*" cannot be fully qualified as a narrative because the focus is only on the chronological order of the events (i.e., the *fabula*). On the other hand, a text like "*The king died and then the queen died of grief*" is understood as a narrative because, in addition to the chronological order of the events (i.e., the *fabula*) there is an *explanatory* relation connecting them, in particular, the fact that the queen died *because* of the grief *caused by* the death of the king, her husband. In other words, it instantiates a plot.

Plots make explicit "why" things happen, rather then just telling "how" and "when." The focus is on connecting relevant event sequences in terms of explanatory relations, rather than simply chronological ones, as *fabulae* (stories) do. The logical plot may be realized in various ways in a narrative resulting in a *plot structure*. A plot structure is a complex entity composed by three elements, as illustrated in Figure 6.1.

- Exposition: the introduction of the actors and the settings (e.g., the location).
- Predicament: the set of problems or struggles that the actors have to go through. It is composed by three elements: (i) *rising action*, the event(s) that increases the tension created by the predicament; (ii) *climax*, the event(s) that create the maximal level of tension; and (iii) *falling action*, the event(s) that resolve the climax and lower the tension. The climax event can be seen as the turning point around which the narrative develops.
- Extrication: which refers to the "end" of the predicament and indicates the ending.

https://doi.org/10.1017/9781108854221.008 Published online by Cambridge University Press

Within narratology, the notions of narrative, *fabula*, plot, and plot structure are closely related and partly overlap. In our model, these notions (*fabula*, plot, and plot structure) find their equivalent in three data structures. In particular, given a narrative document, the *fabula* matches a *timeline* (i.e., a set of temporally anchored and chronologically ordered events); the plot corresponds to a *causeline*, defined as a set of circumstantial relations or, in other words, loose and strict causal relations, explicitly and implicitly expressed; and, finally, the plot structure is converted into a *storyline*, a set of (pairwise) relations between events that expresses the perspective with which the narrative is reported and how the events are connected (i.e., the rising and falling relations) to reach a turning point (i.e., a climax).

These data structures are deeply interrelated to each other, where one entails the other, in particular: a timeline is always entailed by a causeline, and a causeline is always entailed by a storyline. A further aspect to stress is that causelines and storylines are complementary data structures. Causelines makes the "why" between pairs of events explicit, whereas the storylines highlight the relevance of the events to identify a set of preconditions, a turning point, and a set of consequences.

More formally, given a timeline $l \in L$ for a specific period of time, we define a storyline S as n-tuples *T, E, R, P* such that

$$
\begin{aligned}
\mathbf{T}imepoints &= (t_1, t_2, \ldots, t_n)\\
\mathbf{E}vents &= (e_1, e_2, \ldots, e_n)\\
\mathbf{R}elations &= (r_1, r_1, \ldots, r_n)\\
\mathbf{P}erspectives &= (p_1, p_1, \ldots, p_n).
\end{aligned}
$$

T consists of an ordered set of points in time; E is a set of events; R is a set of explanatory relations between events, subsuming causelines; and P is a set of perspectives on the events. Perspectives express the likelihood of an event e to be a climax event. Each e in E is related to a t in T. A storyline, S, can be expressed as a function, F, that finds the most plausible set of plot relations across the timelines $l \in L$ of a text, given timepoints, events, and perspectives, such that

$$
S = argmax(F(l)|T, E, P)
$$

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

https://doi.org/10.1017/9781108854221.008 Published online by Cambridge University Press

The function F sums the connectivity C of explanatory relations in R given a chronologically ordered sequence of events ($l \in L$) with respect to a (candidate) climax event. A storyline will result in that sequence of chronologically ordered events with maximum connectivity score.

To better illustrate the differences and connections between the theoretical notions and the data structures, consider the following example, extracted from the Event Coreference Bank+ Corpus (ECB+; Cybulska and Vossen, 2014). We have marked in bold all event mentions, as defined in ECB+, and illustrate the three data structures:

1. Police **say**$_{e1}$ that on Saturday around 11:30 p.m. Kimani Gray was **standing**$_{e2}$ outside his home with five other young men before **splitting off**$_{e3}$ when he **noticed**$_{e4}$ two plainclothes officers in an unmarked car. After he "**adjusted**$_{e5}$ his waistband and continued to **act**$_{e6}$ in a suspicious manner," officials **say**$_{e7}$ the cops **got out**$_{e8}$ of their car and **approached**$_{e9}$ Gray – who allegedly **turned**$_{e10}$ toward them with a loaded .38-caliber revolver in hand. The 30-year-old sergeant and 26-year-old **fired**$_{e11}$ 11 shots [. . .].

- *timeline*: [NOW] → `includes` → **say**$_{e1}$; **say**$_{e1}$ → `before` → **say**$_{e7}$; **say**$_{e1}$ → `after` → [Saturday around 11:30 p.m.]; [Saturday around 11:30 p.m.] → `includes` → **standing**$_{e2}$; **standing**$_{e2}$ → `before` → **splitting off**$_{e3}$; **splitting off**$_{e3}$ → `simultaneous` → **noticed**$_{e4}$; **adjusted**$_{e5}$ → `before` → **got out**$_{e8}$; **act**$_{e6}$ → `before` → **got out**$_{e8}$; **got out**$_{e8}$ → `before` → **approached**$_{e9}$; **approached**$_{e9}$ → `simultaneous` → **turned**$_{e10}$; **turned**$_{e10}$ → `before` → **fire**$_{e11}$;
- *causelines*: **act**$_{e6}$ → `circumstantial` → **approached**$_{e9}$; **splitting off**$_{e3}$ → `circumstantial` → **noticed**$_{e4}$; **turned**$_{e10}$ → `circumstantial` → **fire**$_{e11}$
- *storyline*: **noticed**$_{e4}$ → `rising_action` → **splitting off**$_{e3}$ → `rising_action` → **adjusted**$_{e5}$ → `rising_action` → **act**$_{e6}$ → `rising_action` → **approached**$_{e9}$ → `rising_action` → **turned**$_{e10}$ → `rising_action` → **fire**$_{e12[climax]}$;

In example 1, the timeline reflects the temporal order and anchoring of all events mentioned in the narrative. Each event mention is associated with a temporal expression (e.g., [NOW]; [Saturday around 11:30 p.m.]) or directly temporally related to each other. Every event mention is present, even those that do not contribute to the plot or the plot structure (e.g., **say**$_{e1}$ and **say**$_{e7}$). On the other hand, the multiple causelines retain only the events that express a circumstantial relation; i.e., those events for which we can identify some

https://doi.org/10.1017/9781108854221.008 Published online by Cambridge University Press

loose causal relation. Finally, the storyline extends (as in this case) or inherits causelines and makes explicit additional explanatory relations (using rising actions only in this case) that may lead to a (candidate) climax event (i.e., **fire**$e_1$2). Each rising action in example 1 provides a meaningful (explanatory) connection (i.e., an explanation) between event pairs, contributing to understanding why the chronologically following event occurred and how the sequence of events led to the occurrence of the climax event (**fire**$e_1$2).

If after the main event **fire**$e_1$2 the narrative mentions consequences, such as dying and protests, these will be related as falling actions, making it a turning point of the climax.[1]

Being able to properly distinguish and reconstruct these three data structures is essential to extracting and identifying storylines from documents in a principled way. Furthermore, this framework calls for the development of appropriate solutions, according to the specific data structure that is targeted. In the remainder of this chapter we will show how the framework has been translated with respect to data sets and corpora, focusing on causelines and storylines. As for timelines, we adopt an established annotation framework based on ISO-TimeML (Pustejovsky et al., 2010) and its extensions (Cassidy et al., 2014; O'Gorman et al., 2016).[2]

6.3 From Theory to Data: Annotating Causelines and Storylines

We translated our framework into two different annotation schemes, one for causelines and the other for storylines, and applied it to news data. The annotation process has been used as an empirical validation of the components of the model, resulting in two new benchmark corpora.

6.3.1 Modeling Causelines

Causelines address a specific type of explanatory relations in narratives, namely, reasons why events happened. Causality is the most clear semantic relation that may provide such explanations. However, causality is also

[1] For this topic, other news articles that were published later report on the riots following the shooting, in which case these riots are the climax and firing at the boy is a rising event, and damage, looting, and arrests are the consequences of the riots.
[2] For a detailed overview about timeline extraction, see Chapter 4.

a debated relation, and causes often remain vague or implicit in stories (see Chapter 5 for an overview on causality and automatic approaches to causal relation extraction.). Building on Ikuta et al. (2014), we therefore use the umbrella term *circumstantiality* to capture a broad range of weak and strong relations, including causality, enablement, prevention, and entailment. These relations differ from strict causal relations in that the consequence is not logically necessary but (culturally or empirically) expected, made possible or explained by backward presupposition; e.g., *you crossed the streets and therefore you were hit by a truck*. As a sum, circumstantial relations make explicit why one event enabled the next event, to facilitate the understanding of a narrative. Causelines are then simply sequences of event mentions connected by circumstantial relations.

Previous work has modeled circumstantial relations in very different ways. A first notable contribution is The Penn Discourse Treebank (PDTB; Prasad et al., 2008). In their framework, contingency relations give rise to a hierarchy, including causality, enablement, and condition. Annotation is conducted at the level of discourse segments and not the level of event mentions, which unfortunately makes it impossible to know the actual pairs of events involved. Chklovski and Pantel (2004) and Hu et al. (2013, 2017) targeted contingency relations directly for pairs of events. Although they used different methods (e.g., lexical patterns vs. statistical association measures, such as pointwise mutual information or causal potential), they still only identified general contingency relations between verbs, ignoring the specific context of occurrence in a document. These systems produce databases of pairs of events (i.e., verbs) that stand in a contingency relation. Other initiatives such as CaTeRS (Mostafazadeh et al., 2016), Causal TimeBank (Mirza et al., 2014), and BeCauSE 2.0 (Dunietz et al., 2017) target only a specific subset of relations; i.e., only causal relations expressed through certain causal connectives. A different approach and take on causality is addressed in the richer event description (RED) corpus (Ikuta et al., 2014; O'Gorman et al., 2016), where annotators are free to make causal inferences even in absence of particular connectives.

In our work, we extend previous approaches in two major aspects. First, we target "circumstantial" relations between pairs of events, thus requiring event mentions in documents to be our minimal annotation units. Secondly, we extend the annotations to also include other types of relations such as such as enablement, prevention, and entailment, even in the absence of explicit textual markers.

To capture such a wide range of (possibly implicit) circumstantial relations, we make use of the Circumstantial Event Ontology (CEO).[3] CEO (Segers et al., 2018) models circumstantial relations through *shared properties* of the event classes. A circumstantial relation is defined as a relation that holds between event classes, where an event of class *A* may give rise to another event of class *B* if properties resulting from the happening of *A* form preconditions to enable the happening of *B*. For instance, the class “ceo:Shooting” allows for a semantic circumstantial relation with the class “ceo:Impacting,” because they both share the property of translocation of an object from location X to Y, the former as the outcome of the event (postcondition) and the latter as a condition to take place (precondition). Modeling these event properties provides a means to chain logically related events and their shared participants within and across documents. With respect to existing ontologies (SUMO; Niles and Pease, 2001) and lexicons (FrameNet; Baker et al., 1998; and WordNet; Fellbaum, 1998), CEO formalizes event knowledge and relations at the most abstract level rather than for specific lexical items and concepts. The axioms in CEO do not define (lexical) concepts exhaustively but capture only the minimally implied properties that reflect the change.

CEO consists of 223 event classes, of which 189 are fully modeled with pre, during, and postsituations. We defined 92 binary properties and 29 unary properties. In total, 189 unique situation rules were defined that consist of 192 binary situation rule assertions and 264 unary rule assertions. All classes are mapped to FrameNet frames (265 mappings) and SUMO classes (195 mappings) and the CEO roles to FrameNet elements (265 mappings).

We enriched the ECB+ corpus, originally annotated for event coreference (Cybulska and Vossen, 2014), with circumstantial relations, resulting in the ECB+/CEO corpus.[4] We extracted 22 CEO-compliant topics (508 articles) that cover calamities such as earthquakes, murders, hijacks, and arsons and automatically added event mentions by means of a compliant state-of-the art system (Caselli and Morante, 2018).[5] After this, two trained annotators completed the markup of the circumstantial relations between events, as well as coreference relations for all automatically generated event mentions not yet covered in the ECB+ corpus.[6] Annotators were asked to connect pairs of calamity events with a circumstantial link if one event could be used to explain the occurrence of the other.

[3] github.com/cltl/CEO-Ontology
[4] For more details, readers are referred to the work of Segers et al. (2018)
[5] The system achieves 0.8187 F1-score for event trigger detection on the TempeEval-3 test set.
[6] If two event mentions are coreferential, they also denote the same concept and share the same participants, time, and location.

In total, 3,038 new event instances expressing calamities were annotated, with 3,448 new event coreference sets and 2,244 circumstantial links. On average, every ECB+/CEO article contains about 7 new coreference sets and about 5 different circumstantial relations. Inter-annotator agreement has been measured with Cohen's kappa on a subset of 21 articles, reaching $\kappa = 0.76$.

6.3.2 Modeling Storylines

Models for understanding narratives mainly focused on fictional texts (Rumelhart, 1975; Lehnert, 1981; Mani, 2012). Research on storyline extraction from nonfictional data is limited. The majority of this work models storylines as a topic clustering task of documents over time (Allan, 2002; Becker et al., 2011; Kawamae, 2011; Binh Tran et al., 2013; Aggarwal, 2014). Shahaf and Guestrin's (2010) seminal work, on the other hand, was the first to argue that clustering is not sufficient and emphasizes that coherence is the defining aspect of storylines. They proposed to generate storylines as coherent chains of documents. However, their storylines do not reflect the unfolding of a specific event in the world but rather structure a document collection as two connecting dots; i.e., its beginning and end documents. Other work targets the problem as a temporal summarization task (Huang and Huang, 2013) or as a joint distribution over locations, entities (organizations and persons), keywords (event mentions), and a set of topics (Zhou et al., 2016). Chambers and Jurafsky (2009, 2010) introduced a new representation format, called narrative event chains, based on unsupervised learning from news stories. Their assumption is that narrative coherence is reflected in events sharing coreferring protagonists. Each chain is essentially an entity-centric timeline rather than a storyline.

Following our model, we created the Event Storyline Corpus (ESC; Caselli and Vossen, 2016, 2017). Similar to ECB+/CEO, we built on top of ECB+ and added storyline interpretations as an extra annotation layer, called a *plot link*. Plot link relations are used to connect only events that actively contribute to the expression of a storyline matching the internal components of the plot structure: rising action(s), climax, and falling action(s).

Plot relations are directional relations, involving a source, S_e, and a target event, T_e. Their annotation is conducted in a two-step approach: first, identify all eligible pairs of events that can express a plot relation and then classify each relation with one of the following values: `rising_action` or `falling_action`.

We did not explicitly annotate the climax event. We assume that climax events can be indirectly derived from the annotated pairs as those events toward which most rising and falling events point, either directly or indirectly. They

https://doi.org/10.1017/9781108854221.008 Published online by Cambridge University Press

should come out as the natural turning points, where the direction of the plot changes from rising to falling.

Plot links are compatible with other discourse annotation initiatives, namely, PDTB. In many cases a plot relation can be mapped to other categories such as "backgound," "narration," or "reason." However, plot links differ from such frameworks because they mainly aim at capturing the perspective of a narrative to connect to (one or multiple) climax events. Secondly, the binary values of the relation values reduce the complexity of the annotation process. Finally, plot relations highlight event–event relations at a microlevel of the discourse dimension.

The ESC v1.2 corpus has been realized by merging together experts and crowdsourced annotation (Caselli and Inel, 2018). It is composed of 258 documents grouped into 22 topics/stories, 6,315 event mentions, and 6,383 plot links. `rising_action` relations amount to 3,160, and 3,225 are the `falling_action` relations. Expert annotation was conducted by two trained annotators and spanned both intra- and intersentence levels (interannotator agreement with Cohen's kappa on a subset of 44 articles is $\kappa = 0.62$). The evaluation of the crowd data (i.e., spammer removal and data quality) has been conducted by applying the CrowdTruth disagreement-aware methodology (Aroyo and Welty, 2014, 2015). The manually annotated data (both experts and crowd) at mention level can be expanded using the event coreference chains in ECB+, thus increasing the number of plot relations.

6.4 Validating Causelines and Extracting Storylines

To obtain insight into our claims about causelines and storylines, we ran a set of experiments using different systems on both ECB+/CEO and ESC v1.2.

As a baseline system (bl), we simply assume that events in the sequential order of mentioning in a text also express a circumstantial and/or a plot relation.

To use CEO, we developed the CEO-Pathfinder module[7] that reads the ontology and a CEO lexicon. The lexicon was derived from the ECB+/CEO annotations and has 655 words manually mapped unambiguously to a CEO class. The CEO-Pathfinder module looks up every mention of an event from the corpus in the lexicon and obtains its corresponding CEO class. Next, it compares the pre-, during, and post-situation assertions of pairs of CEO classes to see whether there is a match such that a post- or during situation assertion

[7] https://github.com/cltl/ceopathfinder

https://doi.org/10.1017/9781108854221.008 Published online by Cambridge University Press

of *ceo:Class1* is the pre-situation assertion of *ceo:Class2*. We check the events in both orders. If a single property matches, we assume that there is a valid circumstantial relation and/or a plot relation. The current implementation does not check whether the situations apply to the corresponding referents of the participants as defined in the ontology.

The CEO-Pathfinder has the option to assume a hidden event between pairs to bridge pre- and postconditions. For example, *shooting* may require *hitting* as an event before explaining *injuring*. To consider the pair *shooting* and *injuring* to be circumstantial, the CEO-Pathfinder scans the ontology for any potential event that may follow *shooting* to enable *injuring*. We experimented with two levels of intermediate events: (i) one intermediate (hidden) event (*ceo1*) and (ii) two intermediate (hidden) events (*ceo2*).

To better assess the performance of CEO, we compare it with three other variants. First, we reuse FBK-PRO (Mirza and Tonelli, 2014) as integrated in the NewsReader platform (Vossen et al., 2016) and extract all explicitly marked causal relations. Secondly, we use *narrative chains* (Chambers and Jurafsky, 2009) as a resource to determine circumstantial and plot relations. For any event pair in the evaluation data, we create a relation if the pairs occur in that order in any narrative chain. Finally, we implement a system that exploits causal relations from FrameNet (Baker et al., 1998). For each pair of events it checks whether any of the associated FrameNet frames stand in a *Causative_of* or *Inchoative_of* relation according to FrameNet 1.7. We use the FrameNet lexicon to obtain all possible frames associated with the words listed in the evaluation data as events.

Summing up the different systems:

- **bl**: Baseline system that assumes that the order of mentioning two events in the text reflects a circumstantial/plot relation.
- **ceo**: Two events only have a circumstantial/plot relation if there is a property match across post-, during, and preconditions.
- **ceoi1**: The same as **ceo** except that it may assume another event as a hidden bridge between two events.
- **ceoi2**: The same as **ceo** except that there can be two other events as a hidden bridge.
- **fbkcl**: Events are connected through a circumstantial/plot relation if there is an explicit causal relation detected by the FBK-PRO system.
- **fn**: Events are connected through a circumstantial/plot relation if there is an explicit causal relation in FrameNet.
- **nc**: Events are connected through a circumstantial/plot relation if they are listed together in a narrative chain.

Table 6.1. *Causelines: P, R, and F1 scores on ECB+/CEO*

	Same sentence only			Full text		
System	**P**	**R**	**F1**	**P**	**R**	**F1**
bl	0.775	**0.799**	**0.786**	0.248	**1.000**	0.397
ceo	0.881	0.107	0.191	0.576	0.300	0.394
ceoi1	0.872	0.208	0.337	0.485	0.421	0.451
ceoi2	0.874	0.284	0.429	0.442	0.488	**0.464**
fn	0.879	0.022	0.043	0.647	0.110	0.188
fbkcl	**0.898**	0.027	0.052	**0.794**	0.038	0.073
nc	0.760	0.016	0.031	0.468	0.044	0.080

Results for causelines are illustrated in Table 6.1. We provide the precision (P), recall (R), and harmonic mean (F1) for pairs of mentions observed in the same sentence and across the complete document. In the latter case, all mention pairs are candidates, which we expect to give the highest recall and lowest precision. For the events in the same sentence, we expect the highest precision and lowest recall. For FrameNet and explicit causal relations detected by FBK-PRO, we expect high precision and low recall as well. All scores have been computed by taking into account automatic expansion of the annotated data to include event coreference.

Not surprising, the results reflect our predictions. The baseline obtains the best results when building circumstantial relations using only events in the same sentence, although the highest precision is provided by using FBK-PRO (*0.898*), followed by CEO without hidden events and FrameNet. Narrative chains (nc) lag behind. On the other hand the CEO approaches have 10 times higher recall than the other non-baseline approaches. Also remarkable, CEO with intermediate hidden events (ceo1 and ceo2) hardly suffer in precision but double the recall.

Recall gets maximized at the cost of precision when event pairs are extended beyond the sentence level. In the full-text scenario, the baseline has 100% recall but the other approaches also increase in recall, showing that a substantial number of circumstantial relations holds across events mentioned in different sentences. In the full-text scenario, CEO (ceo2) gives the best F1 score (*0.464*). Because CEO can still gain from recall by improving the lexical coverage while there is no potential gain for the baseline, we expect the CEO approach to exceed the baseline in the future.

Table 6.2 reports the scores for storylines with CEO-Pathfinder, using P, R, and F1. Similar to causelines, we have expanded the set of event pairs using

Table 6.2. *Storylines: P, R, and F1 scores on ESC v1.2 test data*

	Same sentence only			Full test		
Models	**P**	**R**	**F1**	**P**	**R**	**F1**
bl	0.276	**0.327**	**0.292**	0.174	**0.984**	**0.291**
ceo	0.313	0.006	0.011	0.181	0.021	0.038
ceo1	0.302	0.013	0.024	0.190	0.047	0.075
ceo2	0.349	0.021	0.040	0.204	0.072	0.107
fn	0.563	0.002	0.004	0.328	0.008	0.017
fbkcl	**0.774**	0.005	0.010	**0.512**	0.009	0.018
nc	0.104	0.002	0.003	0.111	0.003	0.006

available annotations for event coreference. According to our model, we expect that all systems should results in lower scores (in absolute terms) when applied to the storyline data set.

As the results show, all approaches, including the baseline (bl), have lower scores, in general. FBK-PRO remains the system that still obtains the best precision but with extremely low recall. Recall is actually the weakest aspect of storyline extraction. We interpret this as cues concerning (i) lack of coverage of the lexicons used in CEO-Pathfinder (i.e., CEO lexicon, FrameNet lexicon, and narrative chains) and, most importantly, (ii) the different nature of relations with respect to causelines. Although we assume that storylines entail causelines, storylines are built on rising and falling relations, which are not captured by causelines.

We compared the annotated events pairs with the lexicons to obtain more insight into their coverage. The lexicon annotated with CEO classes has a coverage of 77.71%, whereas FrameNet has a coverage of 45.46% and narrative chains 13.13%, where the latter lexicons do not include inflected forms. Early experiments with embeddings show that lexical coverage of the CEO lexicon can easily be expanded to 86.79% considering the top 30 most similar words.

6.5 Conclusion

In this chapter we present a formal model for storylines based on ideas from narratology. We show how theoretical notions such as story, plot, and plot structure can be translated into three corresponding data structures, namely, timelines, causelines, and storylines.

Our work is ongoing and there are a number of directions we plan to follow. Currently, the two corpora are released separately. Because both annotations are based on ECB+, it makes sense to combine them and release them as one corpus. To detect circumstantial relations, we can also combine the resources that are available: CEO, FrameNet, SUMO, WordNet, and narrative chains. This will increase the recall and, through voting, possibly also the precision. Because recall seems to be the main bottleneck, it is worthwhile to use unsupervised methods to expand the resources; e.g., through word embeddings, clustering, and topic detection. If unsupervised methods can also provide training data, e.g., finding news articles on similar events, we could experiment with end-to-end systems that do not rely on gold annotations of events, participants, time expressions, coreference relations. Such end-to-end systems are more challenging not only because of possible error propagation but also because it is more difficult to decide which events are relevant and which ones are not. So far we addressed mostly *semantic* relations between events, but news articles also contain *episodic* relations that are specific for that particular incident; e.g., the fact that it was raining that day. Knowledge-based approaches cannot predict episodic relations. We may have to learn these from large quantities of news articles on similar incidents. Finally, we experimented so far only with extracting the causeline and storyline from each single document separately. The ECB+ corpus lends itself to cross-document extraction as well. A future work will be to combine the causelines and storylines from separate documents into a single causeline and storyline across documents that report on the same incident.

References

Aggarwal, Charu C. 2014. Mining Text and Social Streams: A Review. *SIGKDD Explorations*, **15**(2), 9–19.

Allan, James. 2002. Introduction to Topic Detection and Tracking. Pages 1–16 of: Allan, James (ed.), *Topic Detection and Tracking*. Boston: Springer.

Aroyo, Lora, and Welty, Chris. 2014. The Three Sides of CrowdTruth. *Journal of Human Computation*, **1**, 31–34.

Aroyo, Lora, and Welty, Chris. 2015. Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation. *AI Magazine*, **36**(1), 15–24.

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

Bal, Mieke. 1997. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press.

Becker, Hila, Naaman, Mor, and Gravano, Luis. 2011. Beyond Trending Topics: Real-World Event Identification on Twitter. *ICWSM*, **11**(2011), 438–441.

Binh Tran, Giang, Alrifai, Mohammad, and Quoc Nguyen, Dat. 2013. Predicting Relevant News Events for Timeline Summaries. Pages 91–92 of: *Proceedings of the 22nd International Conference on World Wide Web*. ACM.

Boyd, Brian. 2009. *On the Origin of Stories*. Harvard University Press.

Bruner, Jerome S. 1990. *Acts of Meaning*. Vol. 3. Harvard University Press.

Caselli, Tommaso, and Inel, Oana. 2018. Crowdsourcing StoryLines: Harnessing the Crowd for Causal Relation Annotation. Pages 44–54 of: Caselli, Tommaso, Miller, Ben, van Erp, Marieke, et al. (eds.), *Proceedings of the 2018 Events and Stories in the News Workshop (EventStory 2018)*. Santa Fe, NM: Association for Computational Linguistics.

Caselli, Tommaso, and Morante, Roser. 2018. Systems' Agreements and Disagreements in Temporal Processing: An Extensive Error Analysis of the TempEval-3 Task. In: Calzolari, Nicoletta, Choukri, Khalid, Cieri, Christopher, et al. (eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association.

Caselli, Tommaso, and Vossen, Piek. 2016. The Storyline Annotation and Representation Scheme (StaR): A Proposal. Pages 67–72 of: Caselli, Tommaso, Miller, Ben, van Erp, Marieke, Vossen, Piek, and Caswell, David (eds.), *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*. Austin, TX: Association for Computational Linguistics.

Caselli, Tommaso, and Vossen, Piek. 2017. The Event StoryLine Corpus: A New Benchmark for Causal and Temporal Relation Extraction. Pages 77–86 of: Caselli, Tommaso, Miller, Ben, van Erp, Marieke, et al. (eds.), *Proceedings of the Events and Stories in the News Workshop*. Vancouver: Association for Computational Linguistics.

Cassidy, Taylor, McDowell, Bill, Chambers, Nathanael, and Bethard, Steven. 2014. An Annotation Framework for Dense Event Ordering. Pages 501–506 of: Toutanova, Kristina, and Wu, Hua (eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: Volume 2. Short Papers*. Baltimore, MD: Association for Computational Linguistics.

Chambers, N., and Jurafsky, D. 2010. A Database of Narrative Schemas. In: Calzolari, Nicoletta, Choukri, Khalid, Maegaard, Bente, et al. (eds.), *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2010)*.

Chambers, Nathanael, and Jurafsky, Dan. 2009. Unsupervised Learning of Narrative Schemas and Their Participants. Pages 602–610 of: Su, Keh-Yih, Su, Jian, Wiebe, Janyce, and Li, Haizhou (eds.), *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*. Association for Computational Linguistics.

Chklovski, Timothy, and Pantel, Patrick. 2004. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. Pages 33–40 of: Lin, Dekang, and Wu, Dekai (eds.), *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain: Association for Computational Linguistics.

Cybulska, Agata, and Vossen, Piek. 2014. Using a Sledgehammer to Crack a Nut? Lexical Diversity and Event Coreference Resolution. In: Calzolari, Nicoletta, Choukri, Khalid, Declerck, Thierry, et al. (eds.), *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*.

Dunietz, Jesse, Levin, Lori, and Carbonell, Jaime. 2017. The BECauSE Corpus 2.0: Annotating Causality and Overlapping Relations. Pages 95–104 of: Schneider, Nathan, and Xue, Nianwen (eds.), *Proceedings of the 11th Linguistic Annotation Workshop*. Valencia, Spain: Association for Computational Linguistics.

Fellbaum, Christiane (ed.). 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Finlayson, Mark A., Cremisini, Andreas, and Ocal, Mustafa. 2021. Extracting and Aligning Timelines. Pages 91–109 of: Caselli, Tommaso, Palmer, Martha, Hovy, Eduard, and Vossen, Piek (eds), *Computational Analysis of Storylines: Making Sense of Events*. Cambridge University Press.

Forster, Edward Morgan. 1956. *Aspects of the Novel*. Harcourt.

Gottschall, Jonathan. 2012. *The Storytelling Animal: How Stories Make Us Human*. Houghton Mifflin Harcourt.

Hu, Zhichao, Rahimtoroghi, Elahe, Munishkina, Larissa, Swanson, Reid, and Walker, Marilyn A. 2013. Unsupervised Induction of Contingent Event Pairs from Film Scenes. Pages 369–379 of: Yarowsky, David, Baldwin, Timothy, Korhonen, Anna, Livescu, Karen, and Bethard, Steven (eds.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, WA: Association for Computational Linguistics.

Hu, Zhichao, Rahimtoroghi, Elahe, and Walker, Marilyn. 2017. Inference of Fine-Grained Event Causality from Blogs and Films. Pages 52–58 of: Caselli, Tommaso, Miller, Ben, van Erp, Marieke, et al. (eds.), *Proceedings of the Events and Stories in the News Workshop*. Vancouver: Association for Computational Linguistics.

Huang, Lifu, and Huang, Lian'en. 2013. Optimized Event Storyline Generation Based on Mixture-Event-Aspect Model. Pages 726–735 of: Yarowsky, David, Baldwin, Timothy, Korhonen, Anna, Livescu, Karen, and Bethard, Steven (eds.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, WA: Association for Computational Linguistics.

Ikuta, Rei, Styler, Will, Hamang, Mariah, O'Gorman, Tim, and Palmer, Martha. 2014. Challenges of Adding Causation to Richer Event Descriptions. Pages 12–20 of: Mitamura, Teruko, Hovy, Eduard, and Palmer, Martha (eds.), *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. Baltimore, MD: Association for Computational Linguistics.

Kawamae, Noriaki. 2011. Trend Analysis Model: Trend Consists of Temporal Words, Topics, and Timestamps. Pages 317–326 of: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. ACM.

Lehnert, Wendy G. 1981. Plot Units and Narrative Summarization. *Cognitive Science*, **5**(4), 293–331.

Mani, Inderjeet. 2012. Computational Modeling of Narrative. *Synthesis Lectures on Human Language Technologies*, **5**(3), 1–142.

Mirza, Paramita, Sprugnoli, Rachele, Tonelli, Sara, and Speranza, Manuela. 2014. Annotating Causality in the TempEval-3 Corpus. Pages 10–19 of: Kolomiyets, Oleksandr, Moens, Marie-Francine, Palmer, Martha, Pustejovsky, James, and Bethard, Steven (eds.), *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*. Gothenburg, Sweden: Association for Computational Linguistics.

Mirza, Paramita, and Tonelli, Sara. 2014. An Analysis of Causality between Events and Its Relation to Temporal Information. Pages 2097–2106 of: Tsujii, Junichi, and Hajic, Jan (eds.), *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Mostafazadeh, Nasrin, Grealish, Alyson, Chambers, Nathanael, Allen, James, and Vanderwende, Lucy. 2016. CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures. Pages 51–61 of: Palmer, Martha, Hovy, Ed, Mitamura, Teruko, and O'Gorman, Tim (eds.), *Proceedings of the Fourth Workshop on Events*. San Diego, CA: Association for Computational Linguistics.

Niles, Ian, and Pease, Adam. 2001. Towards a Standard Upper Ontology. Pages 2–9 of: *Proceedings of the International Conference on Formal Ontology in Information Systems. Volume 2001*. FOIS '01. New York: Association for Computing Machinery.

O'Gorman, Tim, Wright-Bettner, Kristin, and Palmer, Martha. 2016. Richer Event Description: Integrating Event Coreference with Temporal, Causal and Bridging Annotation. Pages 47–56 of: Caselli, Tommaso, Miller, Ben, van Erp, Marieke, Vossen, Piek, and Caswell, David (eds.), *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*. Austin, TX: Association for Computational Linguistics.

Prasad, Rashmi, Dinesh, Nikhil, Lee, Alan, et al. 2008. The Penn Discourse TreeBank 2.0. In: Calzolari, Nicoletta, Choukri, Khalid, Maegaard, Bente, et al. (eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association.

Pustejovsky, James, Lee, Kiyong, Bunt, Harry, and Romary, Laurent. 2010. ISO-TimeML: An International Standard for Semantic Annotation. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association.

Rumelhart, David E. 1975. Notes on a Schema for Stories. Pages 211–236 of: Bobrow, Daniel G., and Collins, Allan (eds.), *Representation and Understanding*. San Diego, CA: Morgan Kaufmann.

Segers, Roxane, Caselli, Tommaso, and Vossen, Piek. 2018. The Circumstantial Event Ontology (CEO) and ECB+/CEO: An Ontology and Corpus for Implicit Causal Relations between Events. Pages 4585–4592 of: Calzolari, Nicoletta, Choukri, Khalid, Cieri, Christopher, et al. (eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association.

Shahaf, Dafna, and Guestrin, Carlos. 2010. Connecting the Dots between News Articles. Pages 623–632 of: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.

Vossen, Piek, Agerri, Rodrigo, Aldabe, Itziar, et al. 2016. NewsReader: Using Knowledge Resources in a Cross-Lingual Reading Machine to Generate More Knowledge from Massive Streams of News. *Knowledge-Based Systems*, **110**, 60–85.

Zhou, Deyu, Xu, Haiyang, Dai, Xin-Yu, and He, Yulan. 2016. Unsupervised Storyline Extraction from News Articles. Pages 3014–3021 of: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. New York: AAAI Press.