Event StoryLine Corpus: Annotation Guidelines v0.9

This document contains the annotation guidelines for the development of the ESC v0.9 corpus and the associated annotation guidelines for the creation of a benchmark corpus for storylines.

The storyline extraction task requires the resolution of a set of preliminary subtasks:
1. Selection of relevant sentences
2. Event detection
3. Timex detection and normalization
4. Event participant detection
5. Event coreference: in-document and cross-document
6. Temporal Relation Detection:
    a. event-DCT
    b. event-timex: where the timex corresponds to the temporal anchor of the event, i.e. when the event happened
    c. event-event: temporal ordering relation between events
7. Plot Link Annotation: annotate explanatory relations between event mentions (in-document level)
8. Identification of Climax Events

The development of the ESC v.09 corpus is conducted as an extension of previously annotated data, namely the ECB+ corpus.

Following the ECB+ Annotation Guidelines, the following subtasks are considered completed:

1. *Selection of relevant sentences from the news articles*: in ECB+, only sentences containing reference/relevant reference to the "seminal event" i.e. a specific topic (e.g. the 2013 Brooklyn riots) have been annotated. We keep the annotated sentence as they are - no extension of the already annotated data
2. *Event detection*: ECB+ annotation guidelines for event detection - compatible with TimeML Guidelines - possible changes could be done before releasing the corpus distinguishing between tag names and attributes (minor issue)
3. *Event participant detection*: ECB+ annotation guidelines for participant annotation
4. *Event coreference - in-document and cross-document*: ECB+ annotation guidelines applies

The following subtasks required the development of new annotation guidelines or changes to the ECB+ Guidelines:

Timex detection and normalization

All relevant temporal expression in the document must be annotated and normalized. The TimeML Annotation Guidelines for TIMEX3 annotation applies. Below we report the original TIMEX3 annotation guidelines.

The extent must belong to one of the following grammatical categories:

1. Noun (including Proper Nouns): e.g. *today*, *Thursday*
2. Noun Phrase : e.g. *the morning*, *Friday night*, *the last two years*
3. Adjective: e.g. *current*
4. Adverb: e.g. *recently*
5. Adjective or Adverb Phrase: e.g. *half an hour long*, *two weeks ago*, *nearly a half-hour*

The full extent of a TIMEX3 cannot be a Prepositional Phrase (so that the extent begins with a Preposition) or a clause of any type (for example, the expression cannot start with a subordinating conjunction). Thus, the following are not allowed as TIMEX3 extents: *before Thursday*, *in the morning*, *after the strike ended on Thursday*, *over the last 2 years*. Note though that each of these expressions do include TIMEX3 extents (in bold) that should be marked: *before **Thursday***, *in the **morning***, *after the strike ended on **Thursday***, *over **the last 2 years***.

In terms of its syntactic structure, the full extent of a TIMEX3 should include premodifiers of the time expression (including Determiners). So, for instance, all of the following represent markable TIMEX3s: *that cold day, the next day, late last night, next summer, recent decades, numerous Saturdays, more than a month, no less than 60 days, and just a year ago*.

The full extent of a TIMEX should not include postmodifiers of the time expression, including Prepositional Phrases and dependent clauses. For example, the following are not full extents for TIMEX3, though they do each include markable TIMEXs, highlighted in bold: ***five days** after he came back, **the future** of our peoples, **nearly four decades** of experience, and **months** of renewed hostility*.

1. **ONE TIMEX tag** will be used when there is no intervening token between temporal terms that express values for a single mention of time as in twelve o'clock midnight or where values for units are hierarchically related as in ***Friday evening***. The following examples are all single TIMEX3 tags:

- ***twelve o'clock midnight Friday evening***
- ***8:00 p.m. Friday***
- ***Tuesday the 18th November 1943***
- ***Fall 1998***
- ***this year's summer***

2. **ONE TIMEX tag** will be used when certain prepositions appear within temporal expressions. The best way to identify these special cases is to think about whether the expression can easily be represented as a single temporal value (see the discussion below on the value attribute) as in the following cases:

- ***the second of December summer of 1965***
- ***ten of two***
- ***ten minutes to three half past noon***
- ***eleven in the morning***

3. **TWO TIMEX tags** with no nesting of tags inside one another (meaning the extents do not overlap) will be used in the following situations:

(a) Sequences of two temporal expressions that are ordered one relative to the other. They generally involve the use of temporal prepositions and conjunctions like from, before, after, following, prior to, etc.

- *I'm leaving on vacation **two week] from next Tuesday**.*
- *John left **2 days** before **yesterday**.*

These expressions are referred to as anchored durations and are addressed further in the next section on attributes.

(b) Sequences of two temporal expressions that can be related by a temporal link:

- *I tutored an English student **some Thursdays** in **1998**.*
- *The concert is at **8:00 p.m**. on **Friday**.*
- *The concert is **Friday** at **8:00 p.m**.*

Finally, while postmodifiers that introduce a related event are never included in the TIMEX3 extent, adverbial postmodifiers are included in the TIMEX3 extent as in:

- **the best second quarter ever**
- **three years ago**

Each of these examples are treated as single TIMEX3 tags.

Special attention needs to be given to range expressions such as *from 1992 through 1995* and *August 6-8*. Since these expressions contain calendar information that cannot be fully represented with one TIMEX, they are annotated as TWO TIMEX tags, the extent of which does not include any temporal signals. The two TIMEX3 tags that are created for *from 1992 through 1995* are [1992] and [1995]. For *August 6-8*, the extents are [August 6] and [8].

4. **Non-consuming Temporal Expressions:** Some cases require the use of an additional temporal expression that does not explicitly refer to a particular span of text. These are called *non-consuming temporal expressions*. They are generally introduced by way of **anchored durations** (e.g. *2 days before yesterday*) or **range expressions** (e.g. *August 6-8*). All events must be anchored their proper temporal expressions via a TLINK. Consider the following example:

- *John is leaving **two weeks** from **today**.*

There are at least two temporal expressions in this sentence: *two weeks* and *today*. There is also one event, leaving. Looking ahead, we will want to anchor this event to a particular date, namely the date that is two weeks from today. In order to represent this implicit date, we create a non-consuming TIMEX. The new tag will have all of the TIMEX attributes as described in the next section, but it will not wrap around any text.

In extending the ECB+ annotation for temporal expression we will not adopt the full TimeML attributes for temporal expression annotation. We report only the relevant attributes below, namely:

- `Type`
- `Value`
- `anchorTimeId`
- `beginPoint` and `endPoint`

Given that ECB+ associate the timex type to the tag name (e.g. TIME_DATE; TIME_DURATION) etc. the TimeML `type` attribute is not used. Nevertheless, we report the guidelines for its annotation in order to facilitate the selection of the right TIME tag from the ECB+ Guidelines.

`Type`. Each TIMEX is assigned one of the following types: DATE (`<TIME_DATE>` tag in ECB+), TIME (`<TIME_TIME_OF_DAY>` tag in ECB+), DURATION (`<TIME_DURATION>` tag in ECB+), or SET (`<TIME_REPETITION>` tag in ECB+). The format of the value attribute is determined by the type of TIMEX3. For instance, a DURATION must have a value that begins with the letter 'P' since durations represent a period of time. This will be elaborated on below in the value section. In addition, some attributes are used specifically with certain types of temporal expressions.

The following examples illustrate each possible type value. TIMEX3 markable expression is in bold face:

DATE: The expression describes a calendar time.

- *Mr. Smith left **Friday, October 1, 1999***
    the second of December
    yesterday
    *in **October of 1963***
    *in **the summer of 1964***
    *on **Tuesday 18th***
    *in **November 1943***
    this year's summer
    two weeks** from **next Tuesday
    last week
    late Tuesday

DATE can also be the value for the type attribute of each of the two TIMEX3 markable expressions constituting a range. e.g.,:

- *John left between **Monday** and **Wednesday***

TIME: The expression refers to a time of the day, even if in a very indefinite way (as in the two last examples below):

- *Mr. Smith left **ten minutes to three***
  *at **twenty after twelve***
  *at **half past noon***
  *at **eleven in the morning***
  *at **9 a.m. Friday, October 1, 1999***
  the morning of January 31
  late last night
  last night

As before, TIME can also be the type value for each of two TIMEX3 markable expressions that together refer to a temporal range (e.g., *Mr. Smith left between **8:00 a.m**. and **10:00 a.m.***).

DURATION: The expression describes a duration. This value is assigned to explicit durations like the following:

- *Mr. Smith stayed **2 months** in Boston*
  48 hours
  three weeks
  ***20 days** in July*
  ***3 hours** last Monday.*

As a rule, if any specific calendar information is supplied in the temporal expression, then the type of the TIMEX3 must be either DATE or TIME. Some annotators may be tempted to make something like 1985 a DURATION if the context suggests that an event holds throughout that year. However, since it is required that a DURATION have a particular format in the value attribute, such a temporal expression as the one described here must be a DATE.

SET: The expression describes a set of times.

- *John swims **twice a week**.*
  every 2 days.

`Value`. The value attribute details what temporal information is contained in the TIMEX3. This value is given in an extended ISO 8601 format. The following examples, from previous sections, partially illustrate the use of the value attribute for times of the day, dates, durations, and sets. Note though that these examples may not cover every possible temporal expression you may come across. If you are uncertain of the format of a particular TIMEX3 value, please refer to the TIDES guidelines for TIMEX2, available at:

- http://timeml.org/site/terqas/readings/MTRAnnotationGuide_1_2.pdf
- http://www.timeml.org/publications/timeMLdocs/annguide_1.2.pdf

anchorTimeID. This attribute is used when the value of a TIMEX3 can only be established by making reference to some other TIMEX3 in the text. The ID of this anchoring temporal expression is given in the `anchorTimeID` attribute.

An anchor time is always outside the scope of the TIMEX in question, as shown in the following example. Please note the use of t0 as a common anchorTimeID. In the examples, this ID refers to the document creation time (DCT).

- *Mary left on **Thursday** and John arrived **the day** after.*

```
Mary left on
<TIME_DATE tid="t1" value="1998-WXX-4" anchorTimeID="t0">
Thursday
</TIME_DATE>
and John arrived
<TIME_DATE tid="t2" value="1998-WXX-5" anchorTimeID="t1">
the day
</TIMEX3>
after.
```

beginPoint and **endPoint**. These attributes are used when a *duration* is anchored by another time expression, as well as for range expressions. If only one of these points is provided in the text, the annotator can create an empty TIMEX3 to represent the missing point. In some ways, the `beginPoint` and `endPoint` attributes are similar to `anchorTimeID`. The values stored in these attributes can be used by temporal functions to compute the missing point and create a tag for it. The following example shows how this might occur:

```
John begins teaching
<TIME_DURATION tid="t1" value="P1W"
beginPoint="t2" endPoint="t3">
one week
</TIME_DURATION>
from
<TIME_DATE tid="t2" value="XXXX-9-15">
September 15
</TIMEX3>
<TIME_DATE tid="t3" value="XXXX-9-22"
anchorTimeID="t1" />
```
[EMPTY TIMEX]

Finally, if a TIMEX3 corresponds to the DCT, it must be marked by checking the associated attribute.

In ECB+ it can be the case that timex3 have been annotated as part of an event mention extent. In these cases, the annotator have to mark the temporal expression with a tag of its own (this means that we allow multiple annotations on the tokens overlapping over different text expressions). In addition to this, the created temporal expression can be used to anchor

the event mentions in time. In the following examples, we underline the event extent and mark in bold the timex3 expression, plus resolve the anchoring relations between events and timexes. Notice that the event mention "2004 tsunami" contains a timex3 as its pre-modifier which has to be marked and normalized, and which will function as temporal anchor for the event "2004 tsunami"

- *6 . 1 - magnitude quake strikes Indonesia's Aceh province in region hit by **2004 tsunami***
  2004 `CONTAINS` *2004* tsunami
  2004 `CONTAINS` hit

Pay attention to cases where the possible timex3 is a proper noun (e.g. Christmas, Boxing Day, May Day, and similar) and it is associated with an event noun e.g. *the Boxing Day tsunami*; *the Easter flu attack*. In these cases, check if the expression is an entry in the English Wikipedia; e.g. the Boxing Day tsunami - https://en.wikipedia.org/wiki/2004_Indian_Ocean_earthquake_and_tsunami .
If you can find an entry in Wikipedia, then you do not mark the eligible temporal expression as a timex3. On the other hand, if there is no entry in Wikipedia, then you can mark and normalize the eligible temporal expression.

N.B. in case of a proper noun event which contains a numerical temporal expression in it (e.g. the 9/11 attacks; the 13th of November Paris attacks), you have to mark the temporal expressions and create relevant temporal anchoring links.

Temporal Relation Detection
Temporal Relation Detection aims at making explicit:

- The temporal anchoring relation between an event and a temporal expression
- The ordering relations between pairs of events

The task is not trivial and will be splitted in the following subtasks:

- TLINK between event and DCT
- TLINK between event and relevant temporal expressions (temporal anchoring)
- TLINK between pairs of events (event ordering)

Currently, four different approaches to temporal relation annotation can be identified:
- Standard TimeML: the original TimeML annotation guidelines are pretty poor in term of instructions on when annotate a temporal relation. They loosely instruct the annotators to mark all relevant TLINK between pairs of events, pairs of event-timex3; pairs of events-dct. It adopts a very fine grained range of temporal values, including a non-temporal values (`IDENTITY`) to annotated coreferential relations between events.
- TimeML TimeBank-Dense: the annotation aims at creating a dense time-graph for events and temporal expressions relations. The set of temporal relation values has been reduced to 6 (`BEFORE, AFTER, INCLUDES, IS_INCLUDED, SIMULTANEOUS, VAGUE`). Annotators were constrained to mark up i.) all relations between event pairs

in the same sentence and in the following sentence; ii.) all relations between event-timex pairs in the same sentence and in the following sentence; iii.) relations between events and the DCT, in particular, each event mention has a TLINK with the DCT. The annotation tool used infer also annotations based on transitive closure. The VAGUE value is used to mark-up both cases where a TLINK is not really existing and cases where the value of the TLINK is not easy to identify. The annotated dataset has by far more TLINK values than any other existing annotated corpus with TLINKs but the most frequent TLINK value is VAGUE.

- Rich Event Description (RED): It adopts a different annotation philosophy with respect to TimeML/TimeBank and TimeBank-Dense: Temporal Relations - especially ordering relations - between events are annotated only when "*there is explicit grammatical evidence for the ordering – most obviously, prepositions and connectives like "before", or things like "and then" – do not infer anything about the relationships between them*". They extend the annotation of temporal information with the temporal/narrative containment relations: CONTAIN and CONTAINS-SUBEVENT. The narrative containment relation marks an "EVENT [which] is completely temporally contained within the temporal span of the EVENT or TIMEX3 it is related to. In other words, the contained event occurs entirely within the temporal bounds of the event[/timex] it is contained within. This relation is most often used to mark when an EVENT is contained entirely by a narrative container." They use 10 different temporal relations, which can be reduced to 5 TLINK values (BEFORE, OVERLAP, BEGINS_ON, ENDS_ON, SIMULTANEOUS), 2 values for narrative containers (CONTAINS and CONTAINS-SUBEVENT) and other values for causal/precondition relations.

- Temporal Anchoring: the annotation associates the correct temporal anchor to all event mentions, regardless of their position in the text/document. The event anchors are annotated as attributes of the event mentions. It is fundamental the distinction between durative and punctual/non-durative event mentions. Durations express the anchoring via a *beginPoint* and *endPoint* attribute. Relations with the DCT are annotated when the DCT is the most informative time anchor. It extends the annotations of the TimeBank-Dense corpus.

For the Storyline annotation task, we will follow - with minor adaptations - the RED Guidelines. Version 0.9 of the dataset DOES NOT CONTAIN TLINKs between event pairs.

TLINK event-DCT

This subtask aims at connecting *each semantically full event to the DCT*. This allows to position in time an event with respect to a meta-textual information.
Semantically full events: all events which express an happening or a situation in the world. We explicitly exclude the following subtypes of events from the annotation:
- Aspectual events (start/end)
- Events contributing to the factuality profiling of another event (e.g. try to + V) and which are not cognitive events (think/believe/seem etc.)
- Cause event mentions (e.g. cause, lead, result, facilitate, help to + V, produce etc.)

- Events which introduce a reported speech, a direct speech, or a mixed reported speech: only the events in the content portion of the construction will have TLINK relations the DCT

The annotators have to manually create the link between the DCT and the target events (directionality: DCT - event mention).
Once the link is created, the annotator has to select one of the following values:
- `BEFORE`: the DCT is preceding the event:
  - *[DCT: 2016-11-21] On the 4th of Dec. Italy will **vote** on a referendum*
  - *2016-11-21[DCT]* `BEFORE` *vote*
- `AFTER`: it is the opposite of the BEFORE relation. The event is ordered as one following the DCT.
- `CONTAINS`: corresponds to the temporal containment relation between event and timex. The event is temporally contained in the DCT.
- `OVERLAP`: "a single temporal relation that encompasses all the different notions of two things happening at the same time, but is less specific than CONTAINS or SIMULTANEOUS." This can refer to cases where an event has started before the DCT and it is still ongoing at DCT (or presented as ongoing); or any other cases of temporal overlap between the event and the DCT.

TLINK event-timex3

This subtask aims at connecting *each semantically full event to its temporal anchor or time extension/measure*.
Semantically full events: all events which express an happening or a situation in the world. We explicitly exclude the following subtypes of events from the annotation:
- Aspectual events (start/end)
- Events contributing to the factuality profiling of another event (e.g. try to + V) and which are not cognitive events (think/believe/seem etc.)
- Cause event mentions (e.g. cause, lead, result, facilitate, help to + V, produce etc.)
- Events which introduce a reported speech, a direct speech, or a mixed reported speech: only the events in the content portion of the construction will have TLINK relations timexes

Temporal anchoring is performed by answering to the following question: "*When did X happen?*". Temporal anchors will be marked with the value `CONTAINS`. The directionality of temporal anchor annotation is from the TIMEX to the EVENT.

In case of durative events, we will mark the following information:
- Begin point: use the TLINK value `BEGINS_ON` - the link directionality is from the EVENT to the relevant timex marking the beginning of the event.
- End point: use the TLINK value `ENDS_ON` - the link directionality is from the EVENT to the relevant timex marking the end of the event.
- In case a durative event is fully temporally contained within a DURATION timex, mark the relations with `CONTAINS`, else use the `OVERLAP` value

The annotator has to mark each events to its relevant temporal anchor, regardless of the position of the timex3 in the document (e.g. the relevant temporal anchor can be in the same sentence or not). The following rules apply:
- Check if the relevant temporal anchors is a timex3 in the same sentence, if so, create the appropriate link between the timex3 and the event;
- If the relevant temporal anchor is not a timex3 in the same sentence, first check if it is realised by an empty timex3 associated with DURATIONs; if so, create the appropriate link between the timex3 and the event;
- If the relevant temporal anchor is not a timex3 in the same sentence, and it is not an empty timex3, check for the relevant anchor in the previous sentence (or sentences); the first sentence that contains a mention of the relevant time anchor has to be selected as the correct one;
- If the relevant temporal anchor is not a timex3 in the same sentence, it is not an empty timex3, and it is not in a previous sentence, check the following sentences; the first sentence that contains a mention of the relevant time anchor has to be selected as the correct one
- If the temporal anchor is not stated in the document, the annotators have to create an empty timex3 markable, trying to narrow down the exact temporal anchor as precisely as possible. Consider the following example: in the document there is no mention of when exactly the earthquake happened. The DCT of the document is a time of day expression, including hours and minutes. This means that event mentions in the document are either before or after the DCT. The DCT does not CONTAINS any event. Nevertheless, we can infer, or narrow down, a more general time expressions, e.g. *July 03 2013*, which can be used to answer the questions "When did the event happened?". In practical terms, when such a case occurs in a text, the annotator have to create an empty timex3 (in this case, an empty timex3 to account for *July 03 2013*) and use this timex as possible temporal anchor.

  - *[July 03 , 2013 6 : 22AM] - DCT*
    An earthquake killed six children and left 14 others trapped.

    2013-07-03 [empty timex3]
    2013-07-03 CONTAINS earthquake
    2013-07-03 CONTAINS killed
    2013-07-03 CONTAINS trapped

- If neither the temporal anchor nor the DCT are stated in the document, then:
  - If the DCT can be retrieved from the article URL, create an empty timex3, mark it as DCT and complete DCT-event annotation;
  - If the DCT cannot be retrieved from the article URL, create an empty timex3, mark it as DCT, assign it the value "PAST_REF", and complete DCT-event annotation;
  - If the DCT and the temporal anchor are the same, then there is no need to create an extra empty timex3, the timex3-event annotations are not to be performed. **N.B.** This applies only in this case, i.e. when there is NO DCT nor temporal anchor in the documents. It differs from cases where the DCT is present but the temporal anchors is missing in the document (see example

above "*[July 03 , 2013 6 : 22AM] (DCT) An earthquake killed six children and left 14 others trapped.*")

Temporal anchoring can hold between a punctual temporal expressions and events. In this case, we cannot use the `CONTAINS` value, `SIMULTANEOUS` applies:

- *Earthquake Of Magnitude 6 . 1 Strikes Indonesia's Aceh Province Killing One on* **July 02 2013 7 : 41 AM**
  July 02 2013 7 : 41 AM `SIMULTANEOUS` *Strikes*
  July 02 2013 7 : 41 AM `SIMULTANEOUS` *Earthquake*

Answering the questions "When did **the event strikes** happened?" and "When did **the event earthquake** happened" lead to the same answer: *July 02 2013 7 : 41 AM .*

TLINK Values for EVENT-TIMEX3 pairs

The following values will be used for temporal relation between EVENT-TIMEX3 pairs:
- `SIMULTANEOUS`: relations between event and timex3 of type TIME-OF-DAY if expressing a punctual value (i.e.. if the value is expressed as an hour) must be marked with the SIMULTANEOUS value.
- `BEGINS-ON`: "BEGINS-ON signals that the EVENT begins on the timex3 it is related to. This type of TLINK will only occur with EVENTs which have a non-trivial temporal span. Relations with punctual EVENTs will usually be marked with BEFORE instead" [RED Guidelines]
- `BEGUN-ON`: the inverse of the preceding relation, BEGINS-ON
- `ENDS-ON`: "ENDS-ON signals that the EVENT ends on the timex3 it is related to. As with BEGINS-ON, this type of TLINK will only occur with EVENTs which have a non-trivial temporal span. Relations with punctual EVENTs will usually be marked with BEFORE instead." [RED Guidelines]
- `ENDED-ON`: the inverse of the preceding relation, ENDS-ON
- `CONTAINS`: corresponds to the temporal containment relation event-timex3. The relations expresses full temporal containment between one element and the other:
  - *The* **ceremony** *is on* **22nd of March**
    - 22nd of March CONTAINS ceremony
  It roughly corresponds to the TimeML INCLUDES relation
- `IS-CONTAINED`:the inverse of the preceding relation, CONTAINS. It corresponds to the TimeML IS_INCLUDED relation
- `OVERLAP`: "a single temporal relation that encompasses all the different notions of two things happening at the same time, but is less specific than CONTAINS or SIMULTANEOUS. This can refer to two nearly simultaneous events, an EVENT that occurs during a timex3 (but where containment is not entirely sure). OVERLAP is used for linking TIMEX3s of type SET with other EVENTs." [Red Guidelines]

Plot Link Annotation

The PLOT LINK annotation task aims at connecting via explanatory relations pairs of events. The question that PLOT LINKs answer is "*why EVENT/STATE X happened/is happening/will happen (held/is holding/will hold as true)*?" The answer to this question will point to other

event mentions in the text/document. The PLOT LINK task aims at reconstructing the plot structure of a topic - i.e. a collection of ordered documents in time and from different sources (or of a single document) - following the tripartite notion of *plot structure* developed in narratology.

Two possible values for the PLOT LINK task have been identified:

1. `RISING_ACTION`: it is used to annotate relations between events, which correspond to the rising actions in the plot structure. They are events, which contributes to the creation of a predicament, or tension in the story. For the annotation efforts, they could be seen as the *causes*, the *reasons* which facilitated the happening of the event to which they are connected. Notice that causal relation are special cases of precondition relations. The annotator should paraphrase the relation between the two events in this way "*event A is the precondition for event B*"

    a. The **earthquake** prompted a tsunami **alert**
       earthquake *RISING_ACTION* **alert**

    b. The police **shot** the boy who was **pointing** a gun at them
       pointing *RISING_ACTION* **shot**

2. `FALLING_ACTION`: it is used to annotate relations between events, which correspond to the falling actions in the plot structure. They are events, which contributes to the resolution of a predicament, or tension in the story. For the annotation efforts, they could be seen as the *effects*, the *consequences* which facilitated the happening of the event to which they are connected. Notice that causal relation are special cases of precondition relations. The annotator should paraphrase the relation between the two events in this way "*event A is the falling action of event B*"

    a. The tsunami **alert** arrived three minutes after the **earthquake**
       alert *FALLING_ACTION* **earthquake**

RISING_ACTION and FALLING_ACTION are asymmetrical, and one is the inverse of the other, i.e. *event A* (pointing a gun) *is the precondition of event B* (shooting) **vs**. *event B* (shooting) *is the falling action of event A* (pointing a gun). None of the relations is transitive. This requires to mark all PLOT_LINKs which can be identified.

The annotator should prefer the order of presentation of the events in the document in order to mark up these relations e.g.:

- The **earthquake killed** 14 and **left** hundred **trapped** in **collapsed** buildings.
  earthquake *RISING_ACTION* **killed**
  earthquake *RISING_ACTION* **collapsed**
  trapped *FALLING_ACTION* **collapsed (collapsed** *RISING_ACTION* **trapped)**

Annotators are free to identify the pairs of events which may stand in a PLOT_LINK relation. We are not going to create a predefined set of pairs of events which may stand in a plot link,

as this will require to create a really large graph with all possible event pairs (all same sentence pairs and all cross-sentence pairs).

To better guide the annotation process, annotators may exclude from the PLOT_LINK annotation:

- Aspectual events (start/end)
- Events contributing to the factuality profiling of another event (e.g. try to + V) and which are not cognitive events (think/believe/seem etc.)
- Cause event mentions (e.g. cause, lead, result, facilitate, help to + V, produce etc.)
- Events which introduce a reported speech, a direct speech, or a mixed reported speech.

Additional rules:
1. If possible, create a pair of events standing in PLOT_LINK relation using event mentions in the same sentence; if this is not possible, look for the nearest source (or target) element of the PLOT_LINK pairs in the previous (or following) sentences.
    a. For cross-sentence annotation: try to limit the PLOT_LINK annotation between events in subsequent sentences (i.e. the following sentence with respect to the target events). In case, multiple events can be identified as possible RISING and FALLING actions wrt the target events, check if one of the conditions apply and follow the associated rule(s):
        i. Not marking the PLOT_LINK relation may result in missing information, and thus in an incomplete and inconsistent story:
            - ANNOTATE the PLOT_LINK
2. When possible, avoid using pronouns to create PLOT_LINK relations;
3. Annotated all possible PLOT_LINKs for which you have a confidence. In case of doubts between which pairs of events should a PLOT_LINK be annotated, annotate all pairs which make sense. As illustrated in the following examples: what is the rising action (or falling action) of the events **killing** and **injuring**? The earthquake? The flattening of buildings and the landslides? In these cases, all pairs which make sense must be annotated.

    a. A powerful **earthquake** in Indonesia's Aceh province **flattened** buildings and sparked **landslides** Tuesday, **killing** at least five people and **injuring** dozens
    *earthquake RISING_ACTION **flattended***
    *earthquake RISING_ACTION **landslides***
    *earthquake RISING_ACTION **killing***
    *earthquake RISING_ACTION **injuring***
    *flattened RISING_ACTION **killing***
    *flattened RISING_ACTION **injuring***
    *landslides RISING_ACTION **killing***
    *landslides RISING_ACTION **injuring***

4. If a pair of events share at least one participant, check if a PLOT_LINK can be created between them:
    a. *Many people were **injured** and it is difficult to **evacuate** them.*
    *injured RISING_ACTION **evacuate** (shared participant: many people-them)*

5. Events introducing a reported or direct speech (ACTION_REPORTING/NEG_ACTION_REPORTING) are always excluded from PLOT_LINK relations even when the event trigger is a speech act verb (declare, condemn, sentence, vote, etc.). These event triggers are best annotated when opinions about events are annotated. For a comprehensive list and description of Speech Act predicates in English see Vanderveken "SEMANTIC ANALYSIS OF ENGLISH PERFORMATIVE VERBS"

6. In case between two (or more) event mentions exist a CONTAINS (IS_CONTAINED) temporal relation or a subevent relation, the following may rule apply:
   a. The event container (i.e. the event anchor of the CONTAINS/IS_CONTAINED) must be interpreted as the outcome/explanation/consequence of the contained event:
      i. *the plane had circled before launching **two bombing raids**$_{event\_container}$. Five bombs were **dropped**, of which four exploded.*
         - **two bombing raids** FALLING_ACTION **dropped**
      ii. *Five bombs were **dropped** in **two bombing raids**$_{event\_container}$*
         - ***dropped*** RISING_ACTION ***two bombing raids***
   b. Should the event container be mentioned in subsequent sentences with respect to the events it contains, you have to create the necessary PLOT_LINK relations following the order of presentation of the events in the document; e.g.:
      i. *A convicted child molester who was supposedly confined to a wheelchair **overpowered** two prison guards today , **handcuffed** them , **stole** their weapons and **walked** off **wearing** one of their uniforms. [...] $_{sentence\ 3}$ Arcade Joseph Comeaux Jr . , 49 , is serving a life sentence for several charges – including indecency with a child – and was en route from the Estelle Unit in Huntsville , Texas , to the Stiles Unit in Beaumont , Texas , when he **escaped**$_{event\_container}$ just after 9 a . m $_{-\ sentence\ 4}$*
         - ***overpowered*** RISING_ACTION ***escape***
         - ***handcuffed*** RISING_ACTION ***escape***
         - ***stole*** RISING_ACTION ***escape***
         - ***walked*** RISING_ACTION ***escape***
         - ***wearing*** RISING_ACTION ***escape***

7. To avoid the overgeneration of PLOT_LINKs, the following test apply:
   a. If an event, Y, which has already been associated to a PLOT_LINK, i.e. it is part of an event pair (X-Y), is NECESSARY in order to answer the "why" question of an additional event pair with an event component of a previous pair (e.g. X-Z), then DO NOT ANNOTATE the PLOT_LINK between the events in the target pairs; e.g.:
      i. *He initially was **imprisoned** on February 1979 on three 10 - year **sentences** for **rape** of a child , aggravated **rape** of a child , and **burglary***
         - ***imprisoned**$_X$ - **sentences**$_Y$ (why he was **imprisoned**? Because of the **sentences**) = OK pair*

- **imprisoned$_X$ - rape$_Z$** *(why he was **imprisoned$_X$**? Because of the <u>**sentences$_Y$**</u> for **rape$_Z$**) = KO pair*
- **sentences$_Y$ - rape$_Z$***(why the **sentences**? Because of the **rape**) = <u>OK pair</u>*

NOTE: this applies according to the actual textual occurrences of events in the document; e.g.:

> ii.   *He initially was **imprisoned** on February 1979 for **rape** of a child , aggravated **rape** of a child , and **burglary***
>> - **imprisoned$_X$ - rape$_Y$** *(why he was **imprisoned$_X$**? Because of the **rape$_Y$**) = <u>OK pair</u>*

8. In presence of one of the following constructions:
    a. Copulative predicates + ADJ$_{EVENT}$ (e.g. *be, seem, appear; etc.*):
        i.   *The students <u>seems</u> **exhaused***
    b. Inchoative predicates + ADJ$_{EVENT}$ (e.g. *become, turn (into), etc.*):
        i.   *The protest <u>turned</u> **ugly***
    c. Aspectual predicates + ADJ$_{EVENT}$ (e.g. *start, continue, keep, etc.*):
        i.   *The family <u>kept</u> **hopeful***
    d. Creation predicates + ADJ$_{EVENT}$ | NOUN$_{EVENT}$ (e.g. *create, give rise (to), etc.*):
        i.   *100 angry mourners clashed with cops, tossing bottles and <u>creating</u> a **melee***
    e. Causative predicates + ADJ$_{EVENT}$ (e.g. *cause, make, etc.*):
        i.   *He <u>made</u> them **happy***

Only the ADJ$_{EVENT}$| NOUN$_{EVENT}$ (i.e. the predicative complement) can enter a PLOT_LINK relation, the verbal element of the construction is excluded:

- *100 angry mourners **clashed** with cops, **tossing** bottles and <u>creating</u> a **melee***
    - **clashed - melee** *:* PLOT_LINK
    - **tossing - melee** *:* PLOT_LINK

9. Grammatical events, i.e. events which are necessary for the sentence to be grammatical but which are semantically dependent on another event and that: i.) do not introduce any change of state which is not already introduced by the content event; ii.) do not have a time span outside the content event, are NEVER part of a PLOT_LINK:
    a. *The **<u>earthquake</u> hit** Indonesia and **<u>killed</u>** 100*
        i.   **earthquake - killed**
        ii.   NO PLOT_LINK with "**hit**" (grammatical event)

10. In case of events realized by complex NPs of the form:
    a. "*NOUN$_{EVENT}$ NOUN$_{EVENT}$*": only the NP head can enter a PLOT_LINK:
        i.   *The scenes of violence on blocks near the NYPD ' s 67th Precinct station house followed a **protest march***
            - Only **march** can enter a PLOT_LINK
    b. "*NOUN$_{EVENT}$ of NOUN$_{EVENT}$*": the following test applies:

i. Check if the element to be put in the "why" question (or answer) for the candidate event pair is composed only by the first "$NOUN_{EVENT}$" or by the entire NP "$NOUN_{EVENT}$ of $NOUN_{EVENT}$":
  - If only by the first "$NOUN_{EVENT}$" is sufficient to formulate a grammatical and meaningful question (or answer), then only the first "$NOUN_{EVENT}$" must be used in the PLOT_LINK;
  - If the entire NP "$NOUN_{EVENT}$ of $NOUN_{EVENT}$" is necessary to formulate a grammatical and meaningful question (or answer), then only the "*of $NOUN_{EVENT}$*" (i.e. the second component of the complex NP) must be used in the PLOT_LINK:
    a. *The **scenes** of **violence** on blocks near the NYPD's 67th Precinct station house followed a protest **march***
      i. ***scenes - march*** : *why "scenes? because of the "march"* : NO PLOT_LINK
      ii. ***violence - march*** : *why "scenes of violence"? because of the "march"* : PLOT_LINK

11. Events realized by adjectives functioning as modifiers of nominal entities (e.g. ADJ_{EVENT} - $NOUN_{ENTITY}$) such as "**enraged** protesters", "**unexploded** bomb" and similar, ARE EXCLUDED from PLOT_LINKs - the semantic heads such NPs are ENTITIES, not events.

12. Final constructions between two events A-B:
    a. If event B is the goal to be achieved, then event A stands in a PRECONDITION relation with B
    b. If A is the goal to be achieved, then event A stands in a FALLING_ACTION relation with B

Identification of Climax Events

NOTE: This annotation layer is not available in v0.9 of the dataset

The identification of the climax event, in this stage of the annotation, will be done by computing the incoming and outgoing PLOT LINKs associated with each *event instance* per topic.

Nevertheless, once the document has been annotated completely, the annotators have to select one or more climax event mentions per document. The annotation must be done only once, excluding possible coreferential link - we will take care of this automatically. Once the eligible climax event (or events) has been identified, the annotator(s) has to mark the attribute `CLIMAX_EVENT` in the corresponding event tag (or tags).

The eligible `CLIMAX_EVENT(s)` must be annotated either in the Title or in the first sentence of the article. To identify the `CLIMAX_EVENT(s,` read the entire article and try to understand what is the "most salient event".

References

- ECB+ Annotation Guidelines http://www.newsreader-project.eu/files/2013/01/NWR-2014-1.pdf
- ESC v.09Annotation Scheme proposal: http://aclweb.org/anthology/W/W16/W16-5708.pdf