Event Causality in Socio-Political News – Annotation Manual

Fiona Anting Tan
Institute of Data Science, National University of Singapore, Singapore
tan.f@u.nus.edu

Ali Hürriyetoğlu
Koc University, Turkey
ahurriyetoglu@ku.edu.tr

Tommaso Caselli
Rijksuniversiteit Groningen, Netherlands
t.caselli@rug.nl

Nelleke Oostdijk
Radboud University, Netherlands
nelleke.oostdijk@ru.nl

Farhana Ferdousi Liza
University of East Anglia, United Kingdom
f.liza@uea.ac.uk

Tadashi Nomoto
National Institute of Japanese Literature, Japan
nomoto@acm.org

Hansi Hettiarachchi
Birmingham City University, United Kingdom
hansi.hettiarachchi@mail.bcu.ac.uk

Onur Uca
Department of Sociology, Mersin University, Turkey
onuruca@mersin.edu.tr

Iqra Ameer
Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico
iqra@nlp.cic.ipn.mx

Abstract

We build on previously annotated datasets on events from socio-political news by including cause-effect annotations. These include two forms of annotations: (1) Identifying if a cause-effect event type exists in the sentence, and (2) Detecting the text corresponding to the cause-effect-signal spans.

1 Introduction

Causality is a core cognitive concept and appears in many natural language processing (NLP) works that aim to tackle inference and understanding (Jo et al., 2021; Dunietz et al., 2020; Feder et al., 2021). Generally, a causal relation is a semantic relationship between two arguments known as cause and effect, in which the occurrence of one (cause argument) causes the occurrence of the other (effect argument) (Barik et al., 2016).

Understanding the flow of causal events and identifying the causal agents is important in news. For example, in social and political science, researchers might want to observe how event causal graphs change across time in a certain region, perhaps after some policy changes, to understand the impact of the policy.

An existing definition of an event is "any situation that happens, occurs, or holds to be true or false during some time point or time interval." (Pustejovsky et al., 2003). In particular, a cause event refers to a span of text, or argument, that leads to the occurrence of an effect event. An effect event refers to a span of text, or argument, that is a result of the Cause event. A signal is a span of text that explicitly relates two events together. In causal model theory, three main types of causal concepts exists: *Cause, Enable* and *Prevent* (Wolff, 2007).

We believe it is beneficial to create automatic solutions that can identify if a particular sentence contains a causal event, and if so, which parts of it refer to cause, effect and signal (if any). In terms of the meaning of the causality, it is also worthwhile to identify the causal concept involved. Therefore, we propose the annotation of an event causality dataset in the socio-political domain for our task.

Our shared task is a continuation of the workshops Automated Extraction of Socio-political Events from News (AESPEN) in 2020 (Hürriyetoğlu et al., 2020b,a) and Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE) in 2021 (Hürriyetoğlu et al., 2021a,b; Hürriyetoğlu, 2021). We focus on English news in the socio-political and crisis context, obtained from China and South Africa sources. In CASE 2021, Subtask 2 aimed to identify event sentences from sentences without an event trigger. Subtask 4 aimed to mark all word tokens of positive event sentences with its argument[1] and Begin-Inside-Outside (BIO) (Ramshaw and Marcus, 1995) tags. We reused this Subtask 4 dataset, comprising of 869 news documents and 3,559 English sentences, for our event causality task.

The proposed subtasks are:

- **Subtask 1: Causal Event Classification** The first subtask identifies if a given event sentence contains any cause-effect meaning.
- **Subtask 2.1: Cause-Effect Span Detection** This

[1]The arguments used were: Participant, Place, Target, Organizer, Time, and Facility Name.

subtask identifies the spans corresponding to cause and effect per sentence.

- **Subtask 2.2: Signal Span Detection** This subtask identifies the spans corresponding to the signal, or causal connective, per cause and effect relation.
- **Subtask 3: Causal Concept Classification** The third subtask aims to identify the causal concept(s) of a given causal event with the labels: *Cause, Enable, Prevent, Cause-to-End*, *Affect* or none.

With these three subtasks, we hope that automated solutions may be curated for the detection of causal events in news. It is possible for participants to work on a single, multiple or all subtasks concurrently.

2 Related Work

3 Annotation

3.1 Subtask 1: Causal Event Classification

3.1.1 General Definition

We first identified sentences which contain cause-effect relations from ones that do not. We used the definition presented by Penn Discourse TreeBank Version 3 (PDTB-3) for CONTINGENCY, which assigns this label for samples where "one argument provides the reason, explanation or justification for the situation described by the other." (Webber et al., 2019). Specifically, we are the following four Level 2 senses based on definitions from PDTB-3:

- CAUSE: Arg1 and Arg2 are causally influenced but not in a conditional relation. Example connectives are *"because"* and *"since"*.
- PURPOSE: One Arg describes the goal while the other Arg describes the action undertaken to achieve the goal. Example connectives are *"in order to"* and *"so that"*.
- CONDITION: One Arg presents an unrealized situation, that when realized, would lead to the situation described by the other Arg. Example connectives are *"if"* and *"as long as"*.
- NEGATIVE-CONDITION: One Arg presents an unrealized situation, that if it does not occur, would lead to the situation described by the other Arg. Example connectives are *"till"* and *"unless"*.

Since our work is event-based, we did not annotate the Level 2 senses that provided reasons for the speaker to utter a speech, or the hearer to have a belief (+SPEECHACT or +BELIEF types).

Five Tests for Causality For ambiguous examples, we utilized the five tests for causality based on work by Grivaz (2010); Dunietz et al. (2017) to motivate our decisions. Examples (1) to (8) demonstrate how the framework can help justify the expected classification label.

1. **Why:** The example is not causal if the reader is unable to construct a "Why" question regarding the Effect.
2. **Temporal order:** The example is not causal if the Cause does not precede the Effect.
3. **Counterfactual:** The example is not causal if the Effect is equally likely to occur or not occur without the Cause.
4. **Ontological asymmetry:** The example is not causal if the reader can readily swap the Cause and Effect claims in place.
5. **Linguistic:** The example is likely to be causal if it can be rephrased into "X causes Y" or "Due to X, Y."

3.1.2 Refinements

Within the CAUSE sense, we noticed that its sub-senses like CAUSE.REASON and CAUSE.RESULT[2] fit into this framework well. In particular, we were able to construct a question like "Why did `<effect>`?", and answer with "Because (of) `<cause>`." in an extractive manner. However, we required some relaxations in the construction of the "Why" test in order to be able to include other senses into our corpus. We highlight the relaxations as follows:

PURPOSE types We permitted the answer to include additional words like "In order to (achieve the goal of) `<cause>`". In this setting, the cause is a goal, or a justification for the effect action. See Example (3).

CONDITION types We allowed the relaxation of the question to include modal terms, for example: "Why could/would `<effect>` occur?". See Example (7).

CAUSE.NEGRESULT[3] and NEGATIVE-CONDITION types We allowed a further relaxation of the question include "prevent" or "cause-to-end" terms, as in: "Why was `<effect>` prevented?", "Why could be `<effect>` prevented?" or "Why did `<effect>` end?" See Example (8).

3.1.3 Examples

(1) His attackers allegedly drank his blood.

- *Non-Causal*: There is only one event in this sentence.

(2) The protests spread to 15 other towns and resulted in two deaths and the destruction of property.

[2]Quoting the PDTB-3 annotation manual (Webber et al., 2019), "CAUSE.REASON (CAUSE.RESULT) is used when Arg2 (Arg1) gives the reason, explanation or justification, while Arg1 (Arg2) gives its effect."

[3]Quoting the PDTB-3 annotation manual (Webber et al., 2019), "CAUSE.NEGRESULT is used when Arg1 gives the reason, explanation or justification that prevents the effect mentioned in Arg2."

- *Causal*: This sentence contains causal events.

1. **Why:** Why was there "two deaths and the destruction of property"? Because "the protests spread to 15 other towns".

2. **Temporal order:** Protests must spread (Cause) before deaths and destruction (Effect) can occur.

3. **Counterfactual:** Deaths and destruction (Effect) is unlikely to occur if the protests did not spread (NegCause).

4. **Ontological asymmetry:** Using Cause as an Effect to construct a question: Why did "The protests spread to 15 other towns"? Because of deaths and destruction – does not answer the question.

5. **Linguistic:** "The protests spread to 15 other towns" causes "two deaths and the destruction of property".

(3) The three-member Farlam Commission, chaired by retired judge Ian Farlam, was established by President Jacob Zuma to probe into the violence and the deaths of 44 people in wage-related protests.

- *Causal*: This sentence contains causal event of the PURPOSE sense.

1. **Why:** Why was "the three-member Farlam Commision ... established by President Jacob Zuma"? In order "to probe into the violence and the deaths of 44 people in wage-related protests".

2. **Temporal order:** The probe purpose (Cause) must occur before establishing the commission (Effect).

3. **Counterfactual:** Establishing the commission (Effect) is unlikely to occur if the President did not have the goal to probe into the protests (NegCause).

4. **Ontological asymmetry:** Using Cause as an Effect to construct a question: Why did the President "probe into violence and deaths of people in wage-related protests"? Because the Commission was established – does not answer the question.

5. **Linguistic:** "to probe into the violence and the deaths of 44 people in wage-related protests" causes "The three-member Farlam Commission, chaired by retired judge Ian Farlam, was established by President Jacob Zuma".

(4) Chale was allegedly chased by a group of about 30 people and was hacked to death with pangas, axes and spears.

- *Non-Causal*: This sentence exhibits temporal events that have no causal connections. We highlight the tests it fails below:

1. **Why:** Why was Chale "hacked to death with pangas, axes and spears"? Because "Chale was allegedly chased by a group of about 30 people". – This test fails since the Cause does not really explain the reason for why there was the Effect.

2. **Counterfactual:** The counterfactual is "The hacking (Effect) is unlikely to occur if the chasing did not take place (NegCause)." – This test fails since hacking can occur without chasing. In fact, without chasing, Chale might have been hacked earlier.

(5) "I observed the attack on the police, I have no doubt about it," Modiba said during cross-examination.

- *Causal*: This sentence contains a description of an implicit causal event relation within a speech.

1. **Why:** Why did Modiba "have no doubt about it"? Because Modiba "observed the attack on the police".

2. **Temporal order:** Modiba must observe (Cause) before claiming he had no doubts (Effect).

3. **Counterfactual:** Modiba cannot claim that he has no doubts (Effect) if he had not observed the attack (NegCause).

4. **Ontological asymmetry:** Using Cause as an Effect to construct a question: Why was it that Modiba "observed the attack on the police"? Because he had no doubt about it – does not answer the question.

5. **Linguistic:** "I observed the attack on the police" causes "I have no doubt about it".

(6) Both Anoop and Ramanan are also accused in the case related to attack on the Nitta Gelatin office at Panampilly Nagar last year.

- *Non-Causal*: The inclusion of "also" in "also accused" renders that when constructing an answer for the Effect, we need to explain why they are accused in addition to someone else that has already been accused. Therefore, when performing the "Why" test, the answer seems incomplete. Hence, we do not consider this sentence to be causal.

1. **Why:** Why were "both Anoop and Ramanan also accused in the case"? Because of the "attack on the Nitta Gelatin office at Panampilly Nagar last year" – This does not answer the question.

2. **Linguistic:** The “attack on the Nitta Gelatin office at Panampilly Nagar last year” causes “Both Anoop and Ramanan are also accused in the case”. – This does not work.

(7) Striking mineworkers have threatened to halt all mining operations if their employers do not accede to their pay demand.

- *Causal*: This sentence contains CONDITION relations, which we interpret as causal.

1. **Why:** Why could “halt all mining operations" occur? Because “their employers do not accede to their pay demand”.

2. **Temporal order:** Their pay demands must be acceded (Cause) before the halting of mining operations (Effect).

3. **Counterfactual:** The halting of mining operations (Effect) is unlikely to occur if their pay demands were acceded (NegCause).

4. **Ontological asymmetry:** Using Cause as an Effect to construct a question: Why was it that “their employers do not accede to their pay demands"? We cannot find an answer span for this question in the original Effect span.

5. **Linguistic:** “their employers do not accede to their pay demand” causes “halt all mining operations”.

(8) The strike will continue till our demands are conceded.

- *Causal*: This sentence contains NEGATIVE-CONDITION relations, which we interpret as preventive causal type.

1. **Why:** Why could the event that “the strike will continue" be prevented? Because “our demands are conceded”.

2. **Temporal order:** The demands must be conceded (Cause) before preventing the strike from continuing (NegEffect).

3. **Counterfactual:** The strike cannot be stopped from continuing (NegEffect) if the demands are not conceded (NegCause).

4. **Ontological asymmetry:** Using Cause as an Effect to construct a question: Why was it that “our demands are conceded"? Because “the strike could be stopped" – does not answer the question.

5. **Linguistic:** “Our demands are conceded” prevents “the strike will continue”.

For samples with multiple sentences, we annotated each sentence independently, that is, we do not take into account cross-sentence causality[4]. Titles were ignored during annotation. For the subsequent Subtasks 2 and 3, we only retain positive instances for their dataset.

3.2 Subtask 2.1: Cause-Effect Span Detection

We marked out the spans referring to each cause and effect in the context of the event with the <cause> and <effect> tag respectively. It is also possible to use the five checks from Section 3.1 to logically verify the identified Cause-Effect spans. Again, we only focused on annotations within a single sentence.

3.2.1 What is a Cause-Effect span

3.2.1.1 General Principle We defined a cause or effect span as a continuous set of words sufficient for the interpretation of the causal relation meaning. This means that any context modifying or describing the argument relevant to the causal relation was included. Each Cause or Effect span must contain an event, where an event is defined by situations that ‘happen or occur’, or predicates that ‘describe states or circumstances in which something obtains or holds true’ (Pustejovsky et al., 2003).

3.2.1.2 Spans *vs.* Clauses In cases where spans cover entire clauses, our annotations become compatible with PDTB. Examples (9) to (11) are three PDTB examples which convey causal meaning, with their Cause, Effect and Signal spans identified.

(9) A Chemical spokeswoman said <cause>the second-quarter charge was “not material”</cause> and <effect>that no personnel changes were made</effect> <signal>as a result</signal>.[5]

(10) That means <cause>goods could be manufactured closer to customers</cause>, <effect><signal>saving</signal> shipping costs</effect>, he said.[6]

(11) <effect>That’s</effect> <signal>because</signal> <cause>their brokers can require them to sell some shares or put up more cash to enhance the collateral backing their loans</cause>.[7]

[4]Note that in the first release of the dataset, because we obtained our data directly from a previous shared task, there are multiple cases where the example contains multiple sentences. This is a by-product of automatic sentence parsers. We leave such examples as is for now. In the future, such examples will be manually split, so that each sentence will be retained as a separate instance.

[5]Example obtained from PDTB-2 Annotation Manual (Prasad et al., 2008), Example 41, but the Effect span was amended to exclude “and".

[6]Example obtained from PDTB-3 (Webber et al., 2019), WSJ 0317.

[7]Example obtained from PDTB-2 Annotation Manual (Prasad et al., 2008), Example 52.

There were cases where our identified spans did not form a clause, shown in Examples (12) to (14). Examples (13) and (14) demonstrate cases where the Cause spans are nominalizations. Noun phrases as per the span *"The medical crisis in Rajasthan"* shown later in Example (25) were permitted too. Verb phrases were also accepted as per the Effect span of (14).

If the causal relations are explicitly stated (i.e. a connective is present), as per Example (13), we are compatible with the BECauSE 2.0 corpus (Dunietz et al., 2017). One way we differ from BECauSE 2.0 is that we included modals around Signal spans into the Signal, like *"should"* in Example (14).

(12) <cause>The rains</cause> <signal>caused</signal> <effect>the flooding</effect>. [8]

(13) <effect>We are in serious economic trouble</effect> <signal>because of</signal> <cause>inadequate regulation.</cause>.

(14) <cause>The new regulations</cause> <signal>should prevent</signal> <effect>future crises</effect>.[9]

3.2.1.3 Spans contain events As mentioned earlier in General Principle, Cause or Effect spans must contain an event. Therefore, sentences such as Example (15) were excluded. In this sentence, *"John"* is an individual entity, and although he is a causing agent, there is no event in the Cause span.

(15) John caused the fire.

Other examples of an individual entity are pronouns (E.g. *"He"*, *"They"*) and nouns or names of people or institutions (E.g. *"The police"*, *"John"*, *"The Association"*). However, do note that nouns or names of events, policies, or Acts (E.g. *"The Umbrella Movement"*, *"The protest"*) were accepted.

3.2.1.4 Refinements Across our annotation cycles, annotators and curators discussed about examples where there was ambiguity in the exact definition and boundaries of a span. The following paragraphs cover topics where we clarified and aligned the annotation rules further:

Agent, Time and Location Words that describe the Causing or Receiving Agent, Time where event took place and Location of event are treated as relevant contexts of each span. Therefore, according to the General Principle, these words were included into the span. We included words related to the Agent like *"their brokers"* in the Cause span of Example (11) and *"between two shopkeepers"* in the Cause span of Example (16) below. Later in Example (21)(a), Time-related contexts are included in the Effect span.

These "Agent, Time and Location" words were included up to the point where a continuous span is possible, and without changing the intended causal relation meaning. We also only include contexts up to natural stopping points represented by punctuation marks (E.g. *","* and *";"*) for better consistency.

(16) On July 28, <cause>an argument between two shopkeepers</cause> <signal>had led to</signal> <effect> a communal flare-up</effect>.

Modifiers Two common types of modifiers are adverbs and adjectives. Since such descriptions are relevant contexts, we do include them into the spans like the above subparagraph. Some previous Examples are: *"the second-quarter"* ((9)), *"inadequate"* ((13)) and *"On this day"* ((21)(a)).

For examples where the modifiers negate the causal meaning (E.g. *"no", "did not"*), we treat the example as *Non-causal*. They should not appear in Subtask 2. Nevertheless, if present, no Cause-Effect span should be marked. See Example (17).

(17) Curiously, no policeman has died in any of these exchanges of fire.

Subspans: Concurrent Outcomes In English, we use conjunctions like *"and"*, *"or"* or punctuation marks like *","* and *";"* to represent cases where one-or-more or all of the scenarios may occur. In Examples (18) and (19), the Effect spans contain multiple outcomes that can occur concurrently. For simplicity, we will annotate such spans with multiple concurrent outcomes as a single span. If there are multiple signals, as per (19), we highlight all and tie them to the whole Effect span.

(18) But <signal>as</signal> <cause>we work for the people </cause>, <effect>we obliged their request and released the men</effect>," said a Maoist leader who handed over the abducted to the tribal leaders.

(19) Will <cause>the unprecedented protests</cause> <effect><signal>embolden</signal> them to fight for their beliefs in future, or <signal>convince</signal> them that resistance to Bei will is futile</effect>?

Future researchers who are interested to study more granular singular events can consider augmenting the dataset by splitting such spans into subspans by using rule-based search for "and" or "or" keywords within spans.

[8] Examples obtained from TimeML Annotation Guidelines Version 1.2.1 (Saurı et al., 2006), but the Effect span was amended to become a noun phrase to be in line with our rules. The original example from TimeML marks only "flooding" as the Effect.

[9] Example obtained from (Dunietz et al., 2017).

Subspans: Consecutive Outcomes In cases where a span could be split into subparts that describe a series of Cause and Effect events, annotators were instructed to annotate them as separate spans. This is because we are interested to recreate a storyline, and such examples are important to test if models can truly identify a series of events that lead to one another. We provide two positive and one negative example below:

In Example (20), we can view the series of causal events as *"an altercation"* caused *"a youth ... was allegedly severely injured in a thrashing ..."* in (b), and the latter event then caused *"the clashes erupted"* in (a). Instead of combining both the Cause and Effect in (b) as a single Cause span for (a), we asked annotators to mark them separately as shown.

(20) (a) <effect>The clashes erupted</effect> after <cause>a youth belonging to a minority Muslim sect was allegedly severely injured in a thrashing by a youth from the majority sect</cause> following an altercation.

(b) The clashes erupted after <effect>a youth belonging to a minority Muslim sect was allegedly severely injured in a thrashing by a youth from the majority sect</effect> <signal>following</signal> <cause>an altercation</cause>.

In Example (21), we can view the series of causal events as: Motivated to be *"against pass law legislation"* led to the motivation *"to protest"* in (b), and the latter motivation instigated the event *"... around 20,000 women participated in a national march"* in (a).

(21) (a) <effect>On this day in 1956 around 20,000 women participated in a national march</effect> <cause><signal>to</signal> protest</cause> against pass law legislation...

(b) On this day in 1956 around 20,000 women participated in a national march <effect>to protest</effect> <cause>against pass law legislation</cause> ...

We urge annotators to revert to the five checks for causality when in doubt to split subspans or not. Even though some sentences may contain words that tend to be signals (E.g. *"to"*), it might not mean subspans are warranted. In Example (22), if we attempt to split the Effect span into two, as per *"We will continue"* and *"to boycott the work in all courts"*, the Why check would fail[10]. This is because *"to boycott the work ..."* here serves as a description of what they are continuing, and not a Motivation/Goal/Cause.

(22) <effect>We will continue to boycott the work in all courts</effect> <signal>till</signal> <cause>the District Magistrate looks into our demands</cause>," Samal said.

To conclude this section, our approach to annotate consecutive events is to mark each event span separately. In an example with three sequential events, where $A \rightarrow B \rightarrow C$, we annotated two relations: The first relation marks $A \rightarrow B$, while the second relation marks $B \rightarrow C$. For the second relation, one may argue that A can be a relevant context to event B. According to the General Principle, relevant contexts of each span should be included. Therefore, in post-processing, we would also provide alternative annotations that join consecutive annotations, such that the second relation marks $AB \rightarrow C$.

Subspans: Concurrent and Consecutive Outcomes In some instances, we had both concurrent and consecutive outcomes. The rules described in the previous two paragraph segments still apply. Earlier in (19), we demonstrated how to annotate the Effect span with multiple concurrent outcomes as single span. However, if we notice, the first half of the Effect span contains another relation which can be annotated as per (19)(b). Both relations were annotated for (19).

(19) (b) Will the unprecedented protests <effect>embolden them to fight</effect> <signal>for</signal> <cause>their beliefs in future</cause> ...

Speech Markers For reported speech, we did not annotate across the speech boundary. This is demonstrated by Example (18), where we did not annotate *"a Maoist leader who handed over the abducted to the tribal leaders"* as another possible Effect span. *"A Chemical spokeswoman said"* and *"he said"* were also words outside of the speech boundary that were excluded from spans in Examples (9) and (10) respectively. Our rationale is that we are currently focused on single sentences, and texts that cross the speech boundary tend to also involve crossing a sentence boundary.

Titles & Punctuation Marks Our sentences were obtained from a previous shared task and some examples contained texts that included the news' title. These titles were typically in a format where the location of the news is capitalized, followed by the title name, and finally the time of report. *"ADILABAD : TRS condemns arrests March 11 , 2011 00:00 IST"* is one such example. Annotators were told to avoid annotating titles.

Punctuation marks at the start or end of any span, like *"."* or *"?"*, were excluded. Punctuation marks were only included if it was in the middle of the span, like the Effect of Example (19).

Demonstrative Pronouns Demonstrative pronouns were specially accepted as spans. Demonstrative pronouns are pronouns used to refer to a specific entity, which might be within or in another sentence. The

[10] **Why:** Why "we will continue"? We expect answers like "Because the District Magistrate did not look into our demand". This cannot be answered by "To boycott the work".

pronouns *"That"* and *"This"* in (11), (23) and (24) are some examples.

(23) <cause>This</cause> <signal>resulted into</signal> <effect> a clash between the members of the two communities</effect>.

(24) <effect>This</effect> was <cause><signal>to</signal> thwart economic well-being of the population</cause>.

Demonstrative pronouns might refer to an event, but we cannot confirm without further information about surrounding sentences. We acknowledge that there is a chance whereby these pronouns are directed at non-events, which would violate our criteria for a Cause or Effect span (as per Sections 3.2.1.1 and 3.2.1.3). However, given the possibility that these pronouns could refer to events, for better coverage, we included such examples at our current stage of annotation. When we move to multiple sentences in the future, we will retrieve them again for inter-sentence co-reference resolution.

3.2.1.5 Dealing with disjoint spans: To ensure consistency, all our examples will contain only one continuous Cause span, and one continuous Effect span. Thus, if any span is represented by multiple clauses while adhering to the General Principle, we link them up into one continuous span. This is demonstrated by (25)(c) below, where the key Effect spans are two-parts – "*The medical crisis*" and "*deepened on Friday*" – and therefore, we combined them together as one span.

For cases where linking up key subspans to form our argument span either (1) overlaps the other argument's span or (2) changes the intended meaning of the causal relationship, we remove these examples for now to avoid confusion in modelling. We do not have such examples to show yet.

3.2.1.6 Overlapping spans As a consequence of the rules from the previous paragraph, for each relation identified, the Cause and Effect spans do not overlap. Signals may overlap within the Cause and Effect arguments and this phenomenon does occur frequently. This is true especially for AltLex connectives, and is consistent with PDTB-3's annotation scheme for argument spans and connectives.[11]

3.2.1.7 Exclusions Any sentence described by any of the following conditions were excluded from annotations:

1. No Cause-Effect relation exists: Although *Non-causal* examples were excluded from Subtask 2, this exclusion condition holds as a general rule.
2. Correlational: Some relationship exists, but it is unclear what the direction of the relationship is. (E.g. "Smoking is linked to cancer").
3. Pragmatic and Inference uses of Causality: In some cases, causal markers are used to express pragmatic causes or to indicate evidence. (E.g. "Let's do this later because we are on a tight schedule." and "The car was driven recently, because the hood is hot."[12])
4. Argument does not contain any event: This is explained earlier with Example (15).
5. Argument requires discontinuous spans: This is explained earlier with in Paragraph "Dealing with disjoint spans:".
6. Cause and Effect spans overlap: This is explained earlier with in Paragraph "Overlapping spans".

3.2.2 Accommodating multiple relations

Each sentence may have multiple cause-effect relations, and we will annotate them all. This means that we will have duplicated sentences in our dataset.

An example sentence[13] with three possible annotations for cause-effect relations is shown below:

(25) (a) <cause>The medical crisis in Rajasthan</cause> due to an indefinite strike by doctors, <effect>which <signal>has crippled</signal> health services and <signal>left</signal> over a dozen patients dead</effect>, deepened on Friday with 5,000 doctors submitting their resignations.

(b) <effect>The medical crisis in Rajasthan</effect> <signal>due to</signal> <cause>an indefinite strike by doctors</cause>, which has crippled health services and left over a dozen patients dead, deepened on Friday with 5,000 doctors submitting their resignations.

(c) <effect>The medical crisis in Rajasthan due to an indefinite strike by doctors, which has crippled health services and left over a dozen patients dead, deepened on Friday</effect> <signal>with</signal> <cause>5,000 doctors submitting their resignations</cause>.

As previously mentioned, *"The medical crisis in Rajasthan"* is a noun phrase. In all three examples, the Cause span is a noun phrase. In (b), the Effect span is a verb phrase with a coordinated structure. It contains two concurrent Effect events which have been combined into one span, as required by Subspan rules in Section 3.2.1.4 Refinements. Meanwhile, (c) has an independent clause for Effect.

[11] "PDTB-3 annotators have been permitted to include material from both Arg1 and Arg2 in the AltLex expression" (Webber et al., 2019).

[12] Examples obtained or adapted from BECAUSE Annotation Guide V3 (Dunietz et al., 2017).

[13] Example obtained from AESPEN 2020 shared task dataset, or from the article at The Indian Express.

3.2.2.1 Multi-relation Annotation Firstly, our annotation tool must allow for the submission of multiple relations per sentence. Additionally, for easier cross-checking of agreement across annotators, after annotations are submitted, we sort the annotations per example sequentially (along the sentence) by Cause first, followed by Effect, then Signal, if tied. For instance, Example (25)(a) or ID "0001-01" has the first occuring Cause span at the very beginning of the sentence, whereas as we move down to Example (25)(c) or ID "0001-03", the Cause span is located at the end of the sentence.

3.2.2.2 Multi-relation Evaluation To evaluate sentences with multiple relations, our example ID of the each relation for the same sentence will have a similar prefix. For example, we will label Examples (25)(a) to (c) as "0001-01", "0001-02", "0001-03" and "0001-04" respectively. This is to allow participants to submit predicted annotations in any order during evaluation. Our evaluation code will group submissions by sentence, and return the highest score from all possible combinations of prediction and actual sequence labels.

3.2.2.3 Overlapping annotations As our data source is news articles which often contain complex flow of events, there are many cases where the cause of one relation overlaps with the Effect of another relation, vice versa. For instance, when considering Examples (25)(a) and (b) together, the flow of events is as such, where the right arrow ($\rightarrow$) could be interpreted as "*led to*" or "*caused*": "*an indefinite strike by doctors*" $\rightarrow$ "*the medical crisis in Rajasthan* $\rightarrow$ "*crippled health services*" | "*left over a dozen patients dead*". The middle phrase, *"The medical crisis in Rajasthan"*, is an Effect in Example (25)(b) and a Cause in Example (25)(a). To better align the phrases for future co-reference resolution or event chains construction work, we aim to keep to the same span length, especially for sentences with multiple relations.

Exceptions to this rule include cases when the emphasis or meaning of the span is slightly different. For instance, compare Example (25)(b) versus (c). In Example (25)(c), it is necessary to include the words "*deepened on Friday*" in the Effect span to convey the meaning that the crisis was not only ongoing, but that it worsened because doctors submitted their resignations.

3.2.3 Contrast with TimeML format

The TimeML format was introduced by Pustejovsky et al. (2003, 2005); Boguraev et al. (2007) and adopted by UzZaman et al. (2013) in the creation of TempEval-3 corpus that continued the research on temporal expression, event, and relation extraction tasks (Verhagen et al., 2009; Pustejovsky and Verhagen, 2009). Subsequently, Mirza et al. (2014) added annotations to identify causal event relations on the TempEval-3 corpus, and this corpus is called CausalTimeBank[14]. For TimeML formats, events are expressed by verbs, nominalizations, adjectives, predicative clauses, or prepositional phrases (Saurı et al., 2006). They only mark key elements instead of the whole relevant phrase[15]. Extracting an example sentence from one of the documents in CausalTimeBank[16], notice that all events are annotated by the head word instead of the full phrase.

(26) (a) **Sentence:** Iraq `<EVENT eid="e222" class="REPORTING">` said`</EVENT>` it `<EVENT eid="e223" class="OCCURRENCE">`invaded`</EVENT>` Kuwait `<C-SIGNAL cid="c224">`because of`</C-SIGNAL>` `<EVENT eid="e225" class="OCCURRENCE">`disputes`</EVENT>` over oil and money.

(b) **Causal Link:** `<CLINK lid="l141" c-signalID="c224" eventInstanceID="ei225" relatedToEventInstance="ei223"/>`

(c) **Reformatted Sentence:** Iraq said it `<effect>`invaded`</effect>` Kuwait `<signal>`because of`</signal>` `<cause>` disputes`</cause>` over oil and money .

We believe that in social-political causal extraction, the context of the event is as important as the actual agent and action. Therefore, in order to extract spans in context, we would have annotated the CausalTimeBank example sentence as follows, where the underlined texts indicate the TimeML spans:

(27) Iraq said `<effect>`it invaded Kuwait`</effect>` `<signal>`because of`</signal>` `<cause>`disputes over oil and money`</cause>`.

However, if time permits, it might be beneficial for the community if we could annotate both formats so that researchers can create models that work across annotation schemes.

3.2.4 A Proposal: `Minimum` vs `Maximum` spans:

The exact boundaries of an argument might be subjective to the reader. For example, in Examples (25)(a) and (b) above, the event span represented by *"The medical crisis in Rajasthan"* can also be more succinctly referred to as *"crisis"*. Alternatively, some readers might refer

[14] https://hlt-nlp.fbk.eu/technologies/causal-timebank

[15] For example, the TimeML annotation guidelines states that: "If the event is a nominalization which appears as the head of the noun phrase along with other elements (specifiers, complements, modifiers), only the head element will be marked.". Examples and further explanations are available in their guide.

[16] Example obtained from CausalTimeBank, file "AP900815-0044.tml".

to the same span as *"The medical crisis"*, *"medical crisis"* or *"crisis in Rajasthan"*. Furthermore, it might be unnecessarily harsh to categorize models that identify the head word(s) but missed out a few boundary words as one that mispredicts.

Hence, there are benefits in acknowledging the maximum and minimum boundaries necessary for constructing our cause-effect arguments. We recommend admitting two annotations per span type: (1) `Minimum` and (2) `Maximum` span. In Examples (25) and (27), we highlighted `Maximum` span annotations and underlined `Minimum` spans. `Maximum` spans are in line with general principle introduced in earlier in Section 3.2.1. On the other hand, `Minimum` spans aim to be aligned with annotations rules from TimeML, briefly introduced in Section 3.2.3, and further defined in the next section. By default, if not explicitly mentioned, we refer to spans of the `Maximum` kind. In evaluation, systems should include the `Minimum` span, and the best systems should match up to the `Maximum` spans. However, given the current time constraints, we defer the annotation of `Minimum` spans to future work.

3.3 Subtask 2.2: Signal Span Detection

We marked out the spans referring to the causal connective in the context of the causal event with the `<signal>` tag. Examples with signals have been presented throughout in Section 3.2.

3.3.1 Type of signals

Although we did not differentiate the types of discourse signals involved in our corpus, in this section, we briefly introduce the types available to improve understanding of what are causal signals.

- **Explicit signals** have an explicit connective. In our work, common explicit causal signals are verbs (E.g. *"cause", "compel", "aid", "lead"*), prepositions (E.g. *"to", "following"*), conjunctions (E.g. *"because"*) that relate a causing event with an effect event.
- **Alternative lexicalizations (AltLex)** is an open-class of markers with no syntactic limitations. Signals presented in Examples (25)(b) to (c) and (28) demonstrate some usages. Some example AltLex expressions from PDTB-3 include *"making", "easing", "that is why"* and *"attributed that to"*.
- **Implicit signals** must be inferred. In cases where the signal is not obvious or available, we do not mark any signal word. See Example (29).

(28) "`<cause>`The fit is `<signal>`so good`</signal></cause>`, `<effect>`we see this as a time of opportunity`</effect>`", he said.

(29) `<cause>`South Korean consumer prices rose 5% in the first 10 months of this year according to the Bank of Korea and the Economic Planning Board`</cause>`, `<signal>`*(Implicit=thus,thereby)*`</signal>` `<effect>`matching the government's target for the entire year`</effect>`. [17]

3.3.2 What is a signal span

3.3.2.1 Examples from other works For annotators' reference, we compiled a list[18] of causal signals from PDTB-3 (Webber et al., 2019) and Girju and Moldovan (2002). For PDTB-3, we extracted only explicit connectives with CONTINGENCY.CAUSE sense for 5 or more occurrences. For the latter source, we extracted only low ambiguity and high frequency terms for reference.

3.3.2.2 Tips for identifying signal spans To verify if your signal span is a causal connective for AltLexes, a useful trick is to check if it can be replaced by one of the Causal Concept Verbs which we will introduce in Subtask 3: *Cause*, *Enable*, *Prevent*, *Affect*, or *Cause-to-End*. For example (25)(b), we can rephrase the two-concurrent claims to be *"The medical crisis in Rajasthan, which has ~~crippled~~prevented health services"* and *"The medical crisis in Rajasthan ~~left~~caused over a dozen patients to die."*.

3.3.2.3 Temporal signals Temporal signals like *"in"*, *"during"*, *"when"* and *"after"* which describes transitions of time are not causal signals. The markers that indicate concurrent outcomes described earlier in Section 3.2.1.4 (E.g. *"and", "or", ",", ";"*) are also not deemed as causal. This is true, even if it seems possible to use rules from 3.3.2.2 to replace *"and continued"* with *"causing"* in Example (30). This is because *"and continued"* describes a continuity in time. The causal element between the Cause and Effect span is implicitly inferred by the reader. For such examples, annotators marked the Cause and Effect spans without the temporal signal. Example (31) is another demonstration.

(30) However, `<cause>`trade unions refused to accept it`</cause>` and `<signal>`*(Implicit=thus)*`</signal>` `<effect>`continued their strike`</effect>`.

(31) `<effect>`Two militants has also been killed`</effect>` `<signal>`*(Implicit=due_to)*`</signal>` in `<cause>`the fire fight`</cause>`.

3.3.2.4 PURPOSE signals PURPOSE signals convey the meaning of a Goal/Motivation (Cause) leading to an 'Action to Goal' (Effect). For reference, some PURPOSE signals accepted by PDTB-3 are: *"(in order) to be", "(in order) to", "in", "for", "so (as/that)", "with the goal (of)", "by", "for the purpose of", "for that purpose"*.

[17] Examples obtained from PDTB-3(Webber et al., 2019).

[18] https://docs.google.com/spreadsheets/d/1S29j8IWGUzTHWaqb7rMBrolqx_v4SZiX6cyug0rV-_U/edit?usp=sharing

A common PURPOSE signal is *"to"*, short for "in order to". However, do note that there are many usages of *"to"* that do not convey PURPOSE meaning, for example in "He cannot wait to start the movie.". Check if *"to"* can be rephrased as *"in order to"* before including it as a signal.

Unlike NEGRESULT, which is related to the Causal Concept Verbs *Prevent* or *Cause-to-End*, there is no such thing as a NEGPURPOSE sense. Therefore, even if the Motivation contains Causal Concept Verbs (E.g. *"prevent"* (37), *"thwart"* (24)) that seem like signal words, we do not mark them as signals. Our plan is to then identify *Prevent* or *Cause-to-End* semantics later in Subtask 3.

There are also cases where PURPOSE types are implicitly expressed. In (21)(b), we can rewrite the expression in (b) to (c) below. Again, notice that *"against"* is not a signal.

(21) (b) ...<effect>to protest</effect> <cause>against pass law legislation</cause>...

(c) ...<effect>to protest</effect> <signal>*(Implicit=in_order_to_be)*</signal> <cause>against pass law legislation</cause>...

3.3.2.5 Description *vs.* Cause Some connectives are challenging to determine if they convey causal meaning (E.g. *"for"*, *"about"*). When in doubt, revert to the five linguistic checks.

3.3.3 When are signals part of cause-effect spans

In this section, we provide more guidelines and examples where signal words are excluded from cause-effect spans, and when they are included. The rule-of-thumb we employed was to exclude signals from cause or effect spans unless the signals are a necessary part of the arguments' meaning.

3.3.3.1 Conjunctions & adverbial connectors A conjunction is used to connect words, phrases, and clauses. Signals that are conjunctions were excluded from cause-effect spans. Some examples of causal conjunctions are *"because"*, *"as"*, *"since"*, *"if...then"*, *"until"*, *"till"*, and *"so"*. Conjunctive adverbial phrase used to indicate cause-and-effect relationships, like *"as a result"*, are synonymous with conjunctive adverbs, like *"therefore"*. *"hence"*, *"consequently"* and *"accordingly"*. Therefore, conjunctive adverbial phrases should be excluded from cause-effect spans too. Refer to the following examples and (9), (11), (13) and (18).

(32) <effect>Spokesman Keith Khoza said they had decided to March to Prime Media</effect> <signal>because</signal> <cause>the cartoon had raised various concerns</cause>.

(33) <effect>They blocked access to the villages</effect> <signal>by</signal> <cause>cutting down trees</cause>.

(34) <signal>If</signal> <cause>it does not act</cause>, <effect>the protests will continue</effect>.

3.3.3.2 Prepositions A preposition is used prior to a noun, pronoun, or noun phrase to indicate time, place, location, direction or spatial relationships. Preposition-based signals were excluded from cause-effect spans, unless they are PURPOSE type (discussed in the next paragraph). Some examples of prepositions are words like *"for"*, *"to"*, *"till"*, *"until"*, *"following"*, *"from"*, and *"after"*. (19)(b), (20)(b) and (22) reflects some sentences with prepositions as signals. There are some multi-word prepositions like *"because of"* and *"due to"*, demonstrated in Examples (27) and (25)(b) respectively. Notice that some signals can be either a conjunction or preposition, depending on whether the words that come after it are noun-based or not, like *"till"*. The following are a few more examples for illustration.

(35) <effect>Heavy police presence was seen in the area</effect> <signal>due to</signal> <cause>the protest</cause>.

(36) <effect>Two men have been charged</effect> <signal>for</signal> <cause>their murders</cause>.

There is one exception where we included the signal word into the cause-effect span: If the example is a PURPOSE type, and the preposition is right before the Motivation/Goal/Cause span. This exception allows us to distinguish between a goal and an actual realized event when disambiguating the Cause span. In Example (37), *"to"* is a signal span that ties the the Goal (Cause) span to the 'Action to Goal' (Effect).

(37) The DGP said <effect>steps have been taken</effect> <cause><signal>to</signal> prevent the escalation of violence</cause>.

- **Why:** Why was it that "steps have been taken"?
  - (a) Cause span includes signal: Because "to prevent the escalation of violence"
  - (b) Cause span excludes signal: Because "prevent the escalation of violence"

Imagine performing the Why-check and responding to the question with either Answer (a) or (b). Answer (b) is misleading because it suggests that the prevention of the escalation has occurred. By including prepositions that reflects that the Cause span is a Goal-type into the Cause span, like *"(in order) to"*, we avoid such misleading interpretations. More Examples are available in (21)(a) and (24).

3.3.3.3 Verbs Verbs are used to describe actions, events, or states. We categorized the verb-based signals

into two types: (1) Verbs that are completely replaceable by *"cause (by)"* or *"prevent (by)"*, (2) Verbs where causal meaning is only one part of its definition.

In the first case, such verb signals are excluded from the Cause or Effect spans because they do not provide additional information to the arguments. In other words, including or excluding such signals in the Cause or Effect spans do not change the meaning of the spans at all. Refer to Examples (12), (14), (23) and (38) to (40) below.

(38) .... for <effect>all damages</effect> <signal>caused during</signal> <cause>the fire</cause>.

(39) ... <cause>the recent happenings in Orissa </cause> <signal>is causing</signal> <effect>great concern</effect>...

(40) <cause>The incident </cause> <signal>has sparked</signal> <effect>protests across New Delhi ...</effect>

Nevertheless, there are exceptions where we included the verb-based signals event if they are of the first type. The rationale to do so is that incomplete arguments arose if the signals were excluded. We noticed two such situations where this exception was needed: (a) when there are multiple signals describing concurrent events in Examples (41) and (42), and (b) when relative pronouns (E.g. *"who"* and *"which"*) are at the start of the argument in Examples (43) and (44).

(41) ... <cause>the FeesMustFall protests </cause> <effect><signal>that caused</signal> unrest and <signal>turned</signal> violent towards the end of the year</effect>.

(42) <cause>The incident </cause> <effect><signal>triggered</signal> outrage across the country and <signal>was condemned</signal> as despicable ... </effect>

(43) <effect>The 1967 riots</effect>, <cause>which were <signal>triggered by</signal> a labour dispute in San Po Kong</cause>...

(44) <cause>Durban University of Technology (DUT) students went on strike early this month</cause> <effect>which <signal>led to</signal> the institution being closed for over a week</effect>.

For the second type of verb-based signals, they are included in the Cause or Effect span because they describe the Cause or Effect event itself. For Examples (45) to (48), if the verb-based signals were removed from the Cause or Effect span, the Cause-Effect relation no longer has the same meaning. Earlier Examples (10), (19), and (25)(a) have similar traits too.

(45) <cause>The move</cause> <effect><signal>angered</signal> residents who had been protesting in front of the court ... </effect>

- **Linguistic:** "The move" "angered residents" $\neq$ "The move" caused "residents".

  (a) The definition of "angered" is "to make someone angry". It contains elements of causal meaning but cannot be completely replaced by the word "caused".

(46) <cause><signal>Irked over</signal> the arrests</cause>, <effect>the protestors staged dharna</effect>.

(47) <effect>They were on strike for about 65 days</effect> <cause><signal>protesting</signal> the low wages</cause>.

(48) The most visible impact of <cause>the strike</cause> was on transport sector, <effect><signal>affecting</signal> bus and train services</effect>.

3.3.3.4 Phrases Although signals are typically comprised of few words, there are no actual restrictions on how many words are permitted in a signal span. Phrases that are signals typically coincide with AltLex signals and/or conjunctive adverbial phrases. These signals can overlap wholly, in-part, or not at all with Cause and Effect spans. In Example (49), *"so intense that"* it is an extended form of *"so...that"* conjunctions, and is similar to Example (28). A few more examples are provided.

(49) <cause>The blast was <signal>so intense</cause> that</signal> <effect>the motorbike split into pieces</effect>.

(50) ...<cause>a similar attack last December in New Delhi </cause> <signal>that sparked</signal><effect> nationwide protests</effect>.

(51) <effect>The relay fast</effect> is <cause><signal>a way of</signal> supporting Anna from Pune</cause>.

(52) ... <cause>only stringent punishment such as death sentence</cause> <effect><signal>could act as a deterrent</signal> against such 'heinous crime'</effect>.

Acknowledgements

References

Biswanath Barik, Erwin Marsi, and Pinar Öztürk. 2016. Event causality extraction from natural science literature. *Res. Comput. Sci.*, 117:97–107.

Branimir Boguraev, James Pustejovsky, Rie Kubota Ando, and Marc Verhagen. 2007. Timebank evolution as a community resource for timeml parsing. *Lang. Resour. Evaluation*, 41(1):91–115.

Jesse Dunietz, Greg Burnham, Akash Bharadwaj, Owen Rambow, Jennifer Chu-Carroll, and Dave Ferrucci. 2020. To test machine comprehension, start by defining comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7839–7859, Online. Association for Computational Linguistics.

Jesse Dunietz, Lori Levin, and Jaime Carbonell. 2017. The BECauSE corpus 2.0: Annotating causality and overlapping relations. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 95–104, Valencia, Spain. Association for Computational Linguistics.

Amir Feder, Katherine A Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E Roberts, et al. 2021. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *arXiv preprint arXiv:2109.00725*.

Roxana Girju and Dan I. Moldovan. 2002. Text mining for causal relations. In *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference, May 14-16, 2002, Pensacola Beach, Florida, USA*, pages 360–364. AAAI Press.

Cécile Grivaz. 2010. Human judgements on causation in French texts. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Ali Hürriyetoğlu, editor. 2021. *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*. Association for Computational Linguistics, Online.

Ali Hürriyetoğlu, Osman Mutlu, Erdem Yörük, Farhana Ferdousi Liza, Ritesh Kumar, and Shyam Ratan. 2021a. Multilingual protest news detection - shared task 1, CASE 2021. In *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*, pages 79–91, Online. Association for Computational Linguistics.

Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, Jakub Piskorski, Reyyan Yeniterzi, Deniz Yuret, and Aline Villavicencio. 2021b. Challenges and applications of automated extraction of socio-political events from text (CASE 2021): Workshop and shared task report. In *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*, pages 1–9, Online. Association for Computational Linguistics.

Ali Hürriyetoğlu, Erdem Yörük, Vanni Zavarella, and Hristo Tanev, editors. 2020a. *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020*. European Language Resources Association (ELRA), Marseille, France.

Ali Hürriyetoğlu, Vanni Zavarella, Hristo Tanev, Erdem Yörük, Ali Safaya, and Osman Mutlu. 2020b. Automated extraction of socio-political events from news (AESPEN): Workshop and shared task report. In *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020*, pages 1–6, Marseille, France. European Language Resources Association (ELRA).

Yohan Jo, Seojin Bang, Chris Reed, and Eduard H. Hovy. 2021. Classifying argumentative relations using logical mechanisms and argumentation schemes. *Trans. Assoc. Comput. Linguistics*, 9:721–739.

Paramita Mirza, Rachele Sprugnoli, Sara Tonelli, and Manuela Speranza. 2014. Annotating causality in the TempEval-3 corpus. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, pages 10–19, Gothenburg, Sweden. Association for Computational Linguistics.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K. Joshi, and Bonnie L. Webber. 2008. The penn discourse treebank 2.0. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco*. European Language Resources Association.

James Pustejovsky, José M. Castaño, Robert Ingria, Roser Saurí, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. Timeml: Robust specification of event and temporal expressions in text. In *New Directions in Question Answering, Papers from 2003 AAAI Spring Symposium, Stanford University, Stanford, CA, USA*, pages 28–34. AAAI Press.

James Pustejovsky, Jessica Littman, and Roser Saurí. 2005. Arguments in timeml: Events and entities. In *Annotating, Extracting and Reasoning about Time and Events, International Seminar, Dagstuhl Castle, Germany, April 10-15, 2005. Revised Papers*, volume 4795 of *Lecture Notes in Computer Science*, pages 107–126. Springer.

James Pustejovsky and Marc Verhagen. 2009. SemEval-2010 task 13: Evaluating events, time expressions, and temporal relations (TempEval-2). In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, pages 112–116, Boulder, Colorado. Association for Computational Linguistics.

Lance Ramshaw and Mitch Marcus. 1995. Text chunking using transformation-based learning. In *Third Workshop on Very Large Corpora*.

Roser Saurı, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. 2006. Timeml annotation guidelines version 1.2. 1.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. SemEval-2013 task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 1–9, Atlanta, Georgia, USA. Association for Computational Linguistics.

Marc Verhagen, Robert J. Gaizauskas, Frank Schilder, Mark Hepple, Jessica Moszkowicz, and James Pustejovsky. 2009. The tempeval challenge: identifying temporal relations in text. *Lang. Resour. Evaluation*, 43(2):161–179.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. The penn discourse treebank 3.0 annotation manual. *Philadelphia, University of Pennsylvania*.

Phillip Wolff. 2007. Representing causation. *Journal of experimental psychology: General*, 136(1):82.