Richer Event Description: Integrating event coreference with temporal, causal and bridging annotation

Tim O'Gorman and **Kristin Wright-Bettner** and **Martha Palmer**
Department of Linguistics
University of Colorado Boulder
Boulder CO 80309
`ogormant,kristin.wrightbettner,mpalmer@colorado.edu`

Abstract

There have been a wide range of recent annotated corpora concerning events, either regarding event coreference, the temporal order of events, hierarchical "subevent" structure of events, or causal relationships between events. However, although some believe that these different phenomena will display rich interactions, relatively few corpora annotate all of those layers of annotation in a unified fashion. This paper describes the annotation methodology for the Richer Event Descriptions corpus, which annotates entities, events, times, their coreference and partial coreference relations, and the temporal, causal and subevent relationships between the events. It suggests that such rich annotations of within-document event phenomena can be built with high quality through a multi-stage annotation pipeline, and that the resultant corpus could be useful for systems hoping to transition from the detection of isolated mentions of events toward a richer understanding of events grounded in the temporal, causal, referential and bridging relations that define them.

1 Introduction

Many corpora have been released in the last decade and a half regarding the temporal order of events, the hierarchical "subevent" structure of events, causal relationships between events, or reference between events. However, the lack of large corpora annotated with all of those layers may hinder attempts to train systems that learn to jointly predict different phenomena. Furthermore, the low rates of interannotator agreement within event annotation are an ongoing issue for training and evaluating systems dealing with these phenomena.

The Richer Event Description (RED) corpus presents 95 documents (totaling 54287 tokens) sampled both from news data and casual discussion forum interactions, which contain 8731 events, 1127 temporal expressions (TIMEX3s, section time, and document time labels), and 10320 entity markables. It contains 2390 identity chains, 1863 bridging relations, and 4969 event-event relations encompassing temporal, causal and subevent relations (as well as aspectual ALINK relations and reporting relations), as well as 8731 DOCTIMEREL temporal annotations linking these events to the document time.

The fundamental contribution of the corpus is one in which a wide range of event-event and event coreference relations are annotated in a consistent and integrated manner. By capturing coreference, bridging, temporal, causal and subevent relations in the same annotation, the annotations may provide a more integrated sense of how the events in a particular document relate to each other, and encourage the development of systems that learn rich interactions between systems. Rich interactions between events in a text, moreover, may be useful for a wide range of goals; Liao and Grishman (2010) found that looking at related events within a document could aid ACE-style event detection, and Vossen et al. (2015) discussed the value of combining timelines with bridging and causal relations in the construction of storylines.

This paper covers the details of RED annotation, and illustrates a number of annotation methods used to overcome the challenges of annotating such a rich

inventory. We suggest that the advantages of annotating many different event-event phenomena at once can outweigh those challenges. Our corpus and guidelines will be made publicly available.

2 Related Work

Large-scale corpora for event detection and coreference exist in a number of forms. The original MUC tasks dealt with events and scenarios that fit within a particular ontology, and such ontology-driven event annotations have been extended through the ACE and ERE corpora and through the TAC-KBP evaluations (Humphreys et al. 1997, Bagga and Baldwin 1999, Song et al 2015). Unrestricted event coreference annotations were later developed in OntoNotes (Weischedel et al., 2011) – which annotated event coreference but did not explicitly differentiate events and entities – and in cross-document event corpora such as Lee et al. (2012), Cybulska and Vossen (2014), Minard et al (2016) and Hong et al. (2016).

Corpora for event-event and event-time relations have also been developed, both for temporal information in the TimeML tradition (Pustejovsky et al., 2003; Styler IV et al., 2014; Minard et al., 2016), and causal structure (Bethard, 2007; Mostafazadeh et al., 2016; Mirza et al., 2014; Hong et al., 2016; Dunietz et al., 2015). Subevent relations corpora have also been annotated (Glava and najder, 2014; Hong et al., 2016). In addition to those resources, there are many corpora which do not focus on annotating events directly, but which do annotate causal, temporal, event-structural or coreference relations within limited scopes, such as within the same clause or across adjacent sentences (Banarescu et al., 2013; Carlson et al., 2003; Prasad et al., 2008; Riaz and Girju, 2013; Fillmore and Baker, 2000; Palmer et al., 2005).

However, despite the profusion of corpora, only a few of the above resources attempt to provide an integrated annotation of many different event-event relations. Minard et al. (2016) annotated event and entity coreference and temporal relations (as well as semantic roles and cross-document coreference), but omitted both subevent structure and causal relations. Glavas and Snajder (2014) annotated event coreference and subevent relations, but did not capture temporal or causal structure. Hong et al, (2016) annotated a wide inventory of event-event relations, but covered only events within the ERE ontology.

3 Discussion of Annotation

The process of RED annotation is divided into two passes, in order to maximize the quality of event annotations. In the first pass, annotators identify three types of markables: events, temporal expressions, and entities (participants such as people, organizations, objects, and locations). Specific properties of each event are also annotated in this pass, capturing information such as the relation to the document creation time or the modality of the event. Guidelines for these features largely following the Thyme-TimeML specifications (Styler IV et al., 2014), a modification of the ISO-TimeML (Pustejovsky et al., 2003) guidelines designed for clinical text. During that first pass, the entity markables are also annotated with coreference relations and bridging relations.

A second pass occurs only after that first pass is adjudicated, allowing all event-event relations to be labeled over adjudicated events and times. This reduces the propagation of errors from missed events or incorrect events, as the event-event relations and coreference are all annotated between high-quality adjudicated events. It also allows guidelines to be written assuming consistent treatment of event modality, allowing adjudicated modality features to be used when making coreference decisions.

3.1 First pass details and agreement

In many prior annotations such as OntoNotes (Weischedel et al., 2011), markables are only labeled if they participate in coreference chains. In RED annotation, events and entities are annotated regardless of whether they participate in a coreference chain. All occurrences and timeline-relevant states are annotated as events, and entities are annotated according to whether or not they represent an actual discourse referent in the discourse. Such an annotation could easily be adapted to OntoNotes-style annotation (by stripping out the singletons), but adds information that could be very useful for detection of the anaphoricity of mentions, a factor considered to be very useful in coreference resolution

(Harabagiu et al., 2001, Ng and Cardie 2002).

In RED annotation, these entities and events are also labeled using a minimal-span approach in which only the headwords are labeled. This annotation style may reduce the "span match" errors observed by (Kummerfeld and Klein, 2013) in recent systems, and some researchers working on coreference have observed the utility of focusing upon headwords, with (Peng et al., 2015) claiming that "identifying and co-referring mention heads is not only sufficient but is more robust than working with complete mentions" (Peng et al. 2015:1).

Richer Event Description also annotates events and entities with a representation of the polarity and modality of the events and entities in context, making a four-way distinction between AC-TUAL, GENERIC, HEDGED/UNCERTAIN, OR HY-POTHETICAL, and temporal expressions are distinguished into DATE, TIME, DURATION, QUANTI-FIER, PREPOSTEXP and SET, following the Thyme-TimeML annotation of clinical temporal expressions (Styler IV et al., 2014). Figure 1 shows both the accuracy of annotations on these phenomena, and the best performance of systems on a the Tempeval-2016 task, which was on the similarly annotated Thyme data. Modality guidelines were also added to allow annotation of entity modality, primarily to capture reference to generic entities.

A number of additional characteristics of events are annotated (such as intermittence (CONTEXTUAL ASPECT) and whether the event was explicit or implied), but the important additional feature is that of the DocTimeRel, or relationship to document time. Following the methodology of (Pustejovsky and Stubbs, 2011; Styler IV et al., 2014), annotators assume four implicit narrative containers within each document – BEFORE, OVERLAP, BE-FORE/OVERLAP or AFTER document time – and each event is labeled with the best such container. This obviates the necessity of labeling many of the more obvious temporal relations (such as knowing that events in the past happen before events in the future). As can be seen in Table 1, agreement of annotators with the adjudicated gold is very high for such DocTimeRel annotations, and system performance in the clinical domain for this kind of annotation is promising.

Coreference in the first pass is done between

Markable	ITA	gold	TempEval2016
Entity	85.9	92.8	
Event	86.1	93.0	90.3
Timex3	70.8	84.9	79.5
Features	**ITA**		
Timex3 class	91.9	96.5	77.2
Entity Modality	92.3	96.5	85.1
Event Polarity	95.2	98.3	88.7
Event Modality	72.9	91.5	85.5
Event DocTimeRel	84.4	92.0	75.6

Table 1: Agreement F1 for Eventy, Event and TIMEX3 detection. Scores for features only measured when annotators agree that an event exists. Highest reported scores on Tempeval-2016 (a corpus annotated with similar event guidelines) are reported to give an approximation of system performance

all entities in the document, alongside annotation of apposition relations and three bridging relations. The bridging relations are important for capturing a range of anaphora phenomena that are not strict identity relationships (Clark, 1975; Poesio et al., 1997). SET/MEMBER was a label used both for set-subset and set-member relationships, PART/WHOLE captured relationships between entities that physically composed a larger whole, and a general BRIDGING relation was used for any class of bridging that did not fit into other categories, such as events of differing modality, allegations of identity (such as links between "the murderer" and a particular suspect).

The fact that this annotation explicitly labels modality and polarity features can have important consequences for coreference and bridging annotation. Even annotations which do not annotate genericity, such as OntoNotes coreference (Weischedel et al., 2011), have very specific rules about how they are annotated (in the case of OntoNotes, generics noun phrases are only linked to pronouns referring to them, or in specific headline constructions). This means that annotator behavior is dependent upon a separate decision (whether or not a markable is generic) that is never explicitly annotated. RED explicitly annotates modality, and constrains IDEN-TITY relations to only apply to be between elements with the same modality and polarity, and providing bridging relations to capture relations that do not pass this strict definition of identity. We evaluate entity coreference scores using the reference imple-

49

mentation of a variety of scoring metrics that was provided in (Pradhan et al., 2014), which are shown in Table 2. All agreement numbers are scored on a 55-document subset of the corpus sampled from discussion fora and newswire documents.

	muc	bcub	ceafe	conll f1
entity IAA	75.3	68.5	67.6	70.4
vs gold	80.06	85.0	79.5	81.8

Table 2: Agreement scores between annotators and agreement with gold, for Entity coreference

Table 3 shows the scores of entity-entity and event-event bridging relations.

Entity	ITA	gold
set/member	21.5	46.5
whole/part	25.8	56.3
bridging	7.1	25.6
apposition	51.2	67.6

Table 3: Agreement F1 scores on other entity coreference relations (Apposition, Set/Subset, Part/Whole and Bridging), both between annotators and when compared to adjudicated gold

We can also note the subset of this corpus used for agreement calculation was also annotated within the rich ERE paradigm (Song et al., 2015), which allows an inter-schema comparison of the overlap between a defined ontology of "relevant events" and the annotations presented here. 86.3% of all ERE Event mentions have strictly the same span as an Event annotated in RED, and that number grows to 89.5% when accommodating partial span matches. The missing 10.5% is largely due to markables which an annotation such as RED views as merely "entities" – rather than events – in RED annotations, as in the examples. This highlights the level to which corpora disagree on how to handle events that are entailed by entity mentions:

(1) These **nominees** have dedicated their careers to serving the public good, (ERE Personnel.nominate event)

(2) MILITANT SAYS HE IS BEHIND **FATAL** NIGER ATTACK (ERE Life.Die event)

The reverse is quite different, with only 25% of RED events having correlates in ERE, which are largely simply events that do not fit into the ontology used in (Song et al., 2015).

3.2 Event Coreference and Event Bridging

After the adjudication of event and entity markables, event coreference is done alongside the annotation of other event-event relations and event bridging relations. Annotating upon events after a markable adjudication pass is intended to increase consistency in how events are annotated. The annotation of event coreference alongside bridging, temporal, causal and subevent relations adds a different kind of consistency; because annotators cannot relate two event mentions in multiple ways, boundaries differentiating phenomena such as subevent and coreference phenomena are strictly defined, and guidelines are necessarily structured to make those boundaries clear-cut.

Table 4 shows coreference and bridging performance of the event annotations done in this second pass of annotation.

	muc	bcub	ceafe	conll f1
event IAA	68.2	65.1	63.2	65.5
vs gold	84.5	82.8	82.3	83.2
Event	F1 IAA	vs gold		
Set/subset	25.1	64.6		
bridging	5.8	51.9		

Table 4: Agreement scores between annotators and agreement with gold for event coreference

3.3 Temporal and Subevent Annotation

This annotation followed recent work in the TimeML tradition (Pustejovsky and Stubbs, 2011; Styler IV et al., 2014) in focusing upon informative temporal annotations, primarily through two kinds of temporal "containers". The first kind of container is the the relationship that each event has with the time of document creation (the DOCTIMEREL feature annotated in the first pass). The second source of "narrative container" annotation is a focus in the annotations on capturing temporal structure using CONTAINS (INCLUDES, in ISO-TimeML) relations between events and on capturing event-time relationships. This focus on temporal annotations can be measured directly – 40.7% of RED temporal relations are one of the two types of CONTAINS relations, whereas the equivalent relations in TimeBank 1.2 take up only 35% of the relations (using counts reported in (D'Souza and Ng, 2013)).

RED annotation expands upon that narrative container approach by adding subevent annotation. As with causal relations, it is noted that subevent relations also carry temporal information, and therefore they are captured by subtyping CONTAINS relations into two subtypes; purely temporal containment (CONTAINS), and a CONTAINS-SUBEVENT relation, which requires that the contained event be both spatiotemporally contained and also a subevent, being a part of the script or event structure of the larger event. When annotators agreed that two events were linked by some kind of CONTAINS relation, they agreed about distinction between CONTAINS and CONTAINS-SUBEVENT 90.2% of the time.

An outcome of this focus upon annotating both DOCTIMEREL, CONTAINS, and CONTAINS-SUBEVENT relations is a great deal of hierarchical temporal structure in a document, from which one may be able to infer the temporal relationship between two events purely through the temporal relationship of their narrative containers. RED expands upon that by adding event coreference, so that one may make temporal inference not just over a particular event mention, but all mentions of the same event. If one particular "chant" is part of a larger "protest" event, and the annotator knows that some mention of that "protest" is BEFORE a "speech" that instigates it, then the relationship between the "chant" and the "speech" can be viewed by annotators as inferrable, and therefore does not need to be annotated. RED guidelines furthermore limit BEFORE and OVERLAP relations to contexts in which the relation is perceived by an annotator to be explicitly expressed in the context. Section details more nuanced agreement results of temporal annotation.

3.4 Causal Annotation

Causation has often been divided into CAUSE, ENABLE and PREVENT, as outlined in Hobbs (2005) and Wolff (2007), and implemented in Mirza et al. (2014) and Mostafazadeh et al. (2016). RED annotation, based on preliminary studies of causal annotation in (Ikuta et al., 2014), adopted a two-way distinction between CAUSES and PRECONDITION similar to the distinction often made between "Cause" and "enable". RED represents "prevent" relations simply through polarity (being the cause or precondition for a negated event), which does re-

quire that all prevented events have a negated polarity. These CAUSES and PRECONDITION labels have been noted to generally combine with temporal information, and therefore annotators annotate causality with one of four fused labels: BEFORE/CAUSES, OVERLAP/CAUSES, BEFORE/PRECONDITION, and OVERLAP/PRECONDITION. This distinction has similarly been suggested in (Mostafazadeh et al., 2016), and bears practical similarity to the decisions in Hong et al. (2016) to allow multiple labels between two events, or the layered annotation of Mirza et al. (2014) on top of temporal structure.

This annotation aims towards logical definitions for cause and preconditions outlined in Ikuta et al (2014). This defines CAUSES as being true "if, according to the writer, the particular EVENT Y was inevitable given the particular EVENT X.", and PRECONDITION as being true when, "had the particular EVENT X not happened, the particular EVENT Y would not have happened.". Following (Bethard et al., 2008; Bethard, 2007; Prasad et al., 2008), those logical definitions were supplemented by guidelines for particular contexts, and for paraphrasing with particular implicit connectives, and case-by-case guideline for specific problematic frames, to handle edge cases which where challenging for classification by logical definition alone. Table 1 illustrates an example in which all four relations are illustrated:

*The **ouster** of Morsi and the subsequent **suppression** of the Brotherhood has **enraged** the groups members and led to a spate of scapegoating **attacks** by Muslim extremists*
ouster BEFORE/CAUSES **enrage**
ouster BEFORE/PRECONDITION **attacks**
suppression OVERLAP/CAUSE **enrage**
suppression OVERLAP/PRECONDITION **attacks**

Figure 1: An examples with the four causal types

3.5 Annotation Example

To give a summarizing sense of the output of the kind of annotation, we illustrate the culmination of the different layers of annotation with two sentences from the corpus. Figure 2 illustrates the relations annotated during the first pass, and which elements would be annotated as entities. Figure 3 illustrates

the "events" which would be annotated in that first pass (which would also receive modality, polarity, relationship to document time, etc.) and the event-event relations annotated in a second pass.
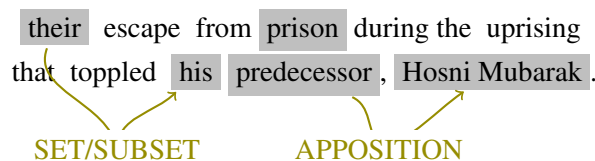

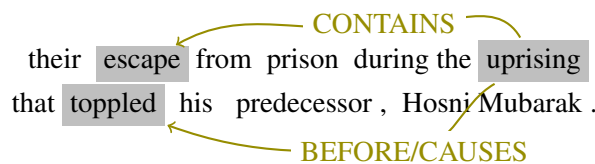**Figure 2:** Example of entity annotation

Figure 3: Example of event annotation pass

4 Annotation Analysis

4.1 Temporal Evaluation

We examine the relation agreement scores between annotators, and between annotators and the gold adjudicated data. While one might evaluate each relationship type – such as BEFORE/CAUSES – as an independent relation, that makes it difficult to compare this relationship annotation to prior endeavors, which have been focused upon temporal annotation tasks. Table 5 therefore also lists what the relation agreement would be if one were to remove causal and subevent relations (for example, treating BEFORE/CAUSES as BEFORE and CONTAINS-SUBEVENT as CONTAINS).

Those temporal relations can also be measured using the temporal closure evaluation method of Uzzaman et al. (2011), which proposes applying temporal closure to the reference (or in this case, gold) annotations when evaluating calculating precision, and apply temporal closure on the annotator annotations when calculating recall.

There have been suggestions for adapting the idea of closure to encompass making inference about other relations, as well. Glavas et al. (2014), for example, suggest that the subevent relation is transitive and should be measured with closure.

### 4.2 Locality and Density of Relation Annotations

Actual comparison of event corpora is made complicated by the wide variance in how many events (or relation-bearing predicates) are annotated per sentence, and how many relations are explicitly annotated, how many implicit relations are inferable, and the distance that is allowed when one makes an annotation. Annotation schemes such as Propbank (Palmer et al., 2005), FrameNet (Fillmore and Baker, 2000), Preposition annotation (Litkowski and Hargraves, 2005; Srikumar and Roth, 2013; Schneider et al., 2015) or AMR (Banarescu et al., 2013) have captured large quantities of temporal and causal relationships, but largely do so within very limited distances from a predicate. Other annotations such as PDTB (Prasad et al 2008) or RST (Carlson et al., 2003) may also capture relations, but are limited to adjacent sentences or adjacency pairs within rhetorical structure. However, there are plenty of contexts where an event may be clearly within a causal chain or an event-subevent relationship outside of such limited scopes.

Figure 4 shows the distance (in sentences) between events with various kinds of relations. One may see that CONTAINS-SUBEVENT relations have much more long-distance relations than terms. This is largely due to the nature of "subevent" relations, which can be annotated across many sentences, because the kind of world knowledge used to mark those relations is not reliant upon local words or constructions.

One may see that the RED annotation is far more dense, in having many event-event relations per event, and has a longer tail of long-distance relations for causal, contains and subevent relations. Indeed, roughly 18% of event chains have two or more relations (temporal, causal, subevent, or bridging) to other events or times. Figure 5 shows the distributions involved, and illustrates the natural idea that event coreference increases the number of event-event relations seen per event chain, and therefore the amount of contextual information about each event.

		inter-annotator			annotator against gold		
Relation	count	F1	Temp. only	+closure	F1	temp. only	+closure
before	609	23.2			60.9		
before/causes	260	22.8	41.0	42.1	62.2	70.1	70.8
before/precondition	492	24.4			59.9		
overlap	346	10.0			45.7		
overlap/causes	221	26.2	20.6	22.2	59.8	54.4	55.2
overlap/precondition	174	4.9			46.7		
contains	983	64.0	53.0	54.4	81.1	76.9	77.7
contains-subevent	729	25.8			66.7		
begins-on	209	18.0	18.0	18.4	64.4	64.4	65.0
ends-on	138	28.3	28.3	28.3	69.2	69.2	69.2
simultaneous	57	0	0	0	43.5	43.5	43.5

Table 5: Agreement on event-event relations, and total corpus counts of each relation
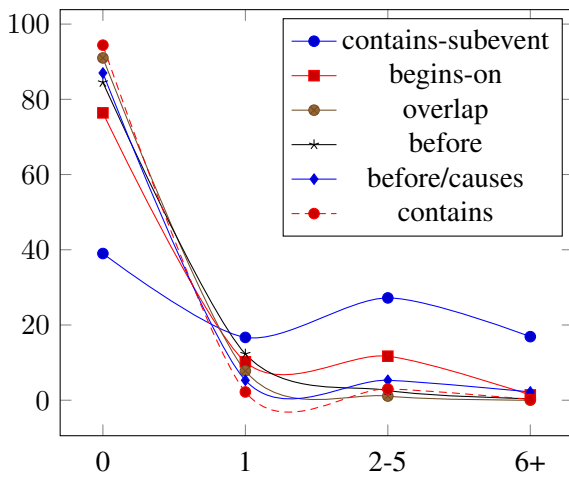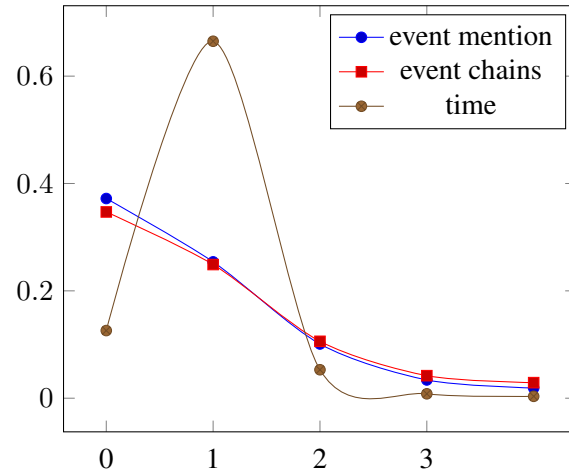

Figure 4: Distance between events, with 0 denoting within-sentence relations.

Figure 5: Relations per event, time, or event chain

4.3   Error Analysis of False Positive annotations

An ongoing issue with this annotation is whether annotators agree on which markables should be related. To explore these errors, we did a manual error analysis of relations discarded during adjudication. We randomly sampled 60 instances from the six relations constituting 87% of the errors (the two CONTAINS relation, the two PRECONDITION relations, and BEFORE and OVERLAP), and clustered them into kinds of issues, listed below:

Presupposition(12) :   A particular edge case in RED causal relations are instances where annotators don't infer much of a causal link between events, but where event 2 definitionally assumes event 1.   One might have to get married once in order for later marriages to be called a "remarriage", for example, but many would hesistate to say that the marriage had a BEFORE/PRECONDITION relation to "remarriage".

Modality(8) :   Annotators disagreeing about relations that were ruled out due to different modalities or differing polarity of the events involved. We assume that most such errors are corrected in adjudication.

Idiomatic(8) :   Annotators differing either in the exact interpretation of a complex temporal expression, or regarding the temporal structure implied by a particular multiword expression.

Containment(6) :   Annotators who agree regarding whether two events are part of a larger struc-

ture of event-subevent relations or containment relations, but disagreed upon which event to attach to. Such relations are usually agreement under temporal closure.

Inferrable(5) : Temporal OVERLAP or BEFORE relations inferrable through document time, and which therefore did not require annotation.

Resultatives(4) : Interpretations of whether the temporal spans of events such as "injured" or "encouraged" refer to the state of being injured/encouraged or to the precipitating event.

Other(17)

5 Conclusion

This paper presents a set of guidelines for annotating causality, temporal relations, subevent relations, coreference and bridging coreference, and presents evaluations of the quality of these annotations. While the individual kinds of phenomena annotated in this corpus have been studied before, such relations have not been annotated together in the same datasets.

We also note that the details of this annotation are similar to other recently developed corpora, perhaps signaling that parallel work in this area may be trending towards a consensus. One such point can be seen in similar treatments of how causal and temporal links are annotated. Both this work and the CaTeRS corpus (Mostafazadeh et al., 2016) adopt very similar treatments of causality, in which the temporal and causal links are joined together into links such as BEFORE/CAUSES. Work such as Mira et al. (2014), while annotating causal relations separately from TimeBank temporal links, have focused upon learning the relationships between causal and temporal structure.

It is hoped that such a richly annotated corpus can provide the opportunity for joint learning that may not be viable with existing corpora. The described corpus will be released, and the guidelines are publicly available. While it remains the case that no singular corpus has become a standardized benchmark for the development of many of these relations, we hope that the current work may help move the community further towards general annotation and prediction of event coreference, representation and

event-event relations, and that it may shed light upon the utility of annotating many kinds of event phenomena over the same corpus.

References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proc. of the Linguistic Annotation Workshop and Interoperability with Discourse*.

Steven Bethard, William J Corvey, Sara Klingenstein, and James H Martin. 2008. Building a Corpus of Temporal-Causal Structure. In *LREC*.

Steven John Bethard. 2007. *Finding event, temporal and causal structure in text: A machine learning approach.*

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue*, pages 85–112. Springer.

Herbert H Clark. 1975. Bridging. In *Proceedings of the 1975 workshop on Theoretical issues in natural language processing*, pages 169–174. Association for Computational Linguistics.

Agata Cybulska and Piek Vossen. 2014. Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution. In *LREC*, pages 4545–4552.

Jennifer D'Souza and Vincent Ng. 2013. Classifying temporal relations with rich linguistic knowledge. In *HLT-NAACL*, pages 918–927.

Jesse Dunietz, Lori Levin, and Jaime Carbonell. 2015. Annotating Causal Language Using Corpus Lexicography of Constructions. In *The 9th Linguistic Annotation Workshop held in conjuncion with NAACL 2015*, page 188.

Charles J. Fillmore and Collin F. Baker. 2000. FrameNet: Frame Semantics Meets the Corpus. In *Poster presentation, 74th Annual Meeting of the Linguistic Society of America*, January.

Goran Glava and Jan najder. 2014. Constructing Coherent Event Hierarchies from News Stories. *TextGraphs-9*, page 34.

Goran Glava, Jan Snajder, Parisa Kordjamshidi, and Marie-Francine Moens. 2014. HiEve: A Corpus for Extracting Event Hierarchies from News Stories. In *Proceedings of 9th language resources and evaluation conference*.

Jerry R Hobbs. 2005. Toward a useful concept of causality for lexical semantics. *Journal of Semantics*, 22(2):181–209.

Yu Hong, Tongtao Zhang, Sharone Horowit-Hendler, Heng Ji, Tim O'Gorman, and Martha Palmer. 2016. Building a Cross-document Event-Event Relation Corpus. In *Proceedings of LREC 2016*.

Rei Ikuta, William F. Styler IV, Mariah Hamang, Tim OGorman, and Martha Palmer. 2014. Challenges of

Adding Causation to Richer Event Descriptions. *Proceedings of the 2nd Workshop on Events*, page 12.

Jonathan K Kummerfeld and Dan Klein. 2013. Error-driven analysis of challenges in coreference resolution. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Wash*, pages 18–21.

Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. 2012. Joint entity and event coreference resolution across documents. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 489–500. Association for Computational Linguistics.

Shasha Liao and Ralph Grishman. 2010. Using document level cross-event inference to improve event extraction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 789–797. Association for Computational Linguistics.

Kenneth C Litkowski and Orin Hargraves. 2005. The preposition project. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 171–179.

Anne-Lyse Minard, Manuela Speranza, Ruben Urizar, Begona Altuna, Marieke van Erp, Anneleen Schoen, and Chantal van Son. 2016. MEANTIME, the News-Reader multilingual event and time corpus. *Proceedings of LREC2016*.

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.

Nasrin Mostafazadeh, Alyson Grealish, Nathanael Chambers, James Allen, and Lucy Vanderwende. 2016. CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures. In *Proceedings of the 4th Workshop on Events*.

M. Palmer, P. Kingsbury, and D. Gildea. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Haoruo Peng, Kai-Wei Chang, and Dan Roth. 2015. A joint framework for coreference resolution and mention head detection. *CoNLL 2015*, 51:12.

Massimo Poesio, Renata Vieira, and Simone Teufel. 1997. Resolving bridging references in unrestricted text. In *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, pages 1–6. Association for Computational Linguistics.

Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, and Michael Strube. 2014.

Scoring coreference partitions of predicted mentions: A reference implementation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)*, pages 30–35.

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 *LREC*. Citeseer.

J. Pustejovsky and A. Stubbs. 2011. Increasing informativeness in temporal annotation. In *ACL HLT 2011*, page 152.

James Pustejovsky, Jos M Castano, Robert Ingria, Roser Sauri, Robert J Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R Radev. 2003. TimeML: Robust specification of event and temporal expressions in text. *New directions in question answering*, 3:28–34.

Mehwish Riaz and Roxana Girju. 2013. Toward a better understanding of causality between verbal events: Extraction and analysis of the causal power of verb-verb associations. In *Proceedings of the annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*.

Nathan Schneider, Vivek Srikumar, Jena D Hwang, and Martha Palmer. 2015. A hierarchy with, of, and for preposition supersenses. In *The 9th Linguistic Annotation Workshop held in conjuncion with NAACL 2015*, page 112.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. 2015. From light to rich ere: annotation of entities, relations, and events. In *Proceedings of the 3rd Workshop on EVENTS at the NAACL-HLT*, pages 89–98.

Vivek Srikumar and Dan Roth. 2013. Modeling semantic relations expressed by prepositions. *Transactions of the Association for Computational Linguistics*, 1:231–242.

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

Naushad UzZaman and James F Allen. 2011. Temporal evaluation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 351–356. Association for Computational Linguistics.

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

Ralph Weischedel, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin, Sameer Pradhan, Lance Ramshaw, and Nianwen Xue. 2011. OntoNotes: A large training corpus for enhanced processing. *Handbook of Natural Language Processing and Machine Translation. Springer*.