

The Circumstantial Event Ontology (CEO)

Roxane Segers and Tommaso Caselli and Piek Vossen

Vrije Universiteit Amsterdam

De Boelelaan 1105 1081 HV Amsterdam (NL)

{r.h.segers,t.caselli;p.t.j.m.vossen}@vu.nl

Abstract

In this paper we describe the ongoing work on the Circumstantial Event Ontology (CEO), a newly developed ontology for calamity events that models semantic circumstantial relations between event classes, where we define circumstantial as explicit and implicit causal relations. The circumstantial relations are defined manually in the ontology for classes of events that involve a change to the same property of a participant. We discuss and contrast two types of circumstantial relations: semantic and episodic circumstantial relations. Further, we describe the meta-model and the current contents of the ontology and outline the future evaluation of the CEO.

1 Introduction

Suppose we read a sentence such as “*Helen was crossing the street; she was hit by a truck*”. As it is clear to most readers, but implicit in this sentence, there must be some relation between “crossing the street” (A) and “being hit by a truck” (B). First, the two events, A and B, share the same participant (“Helen” - “she”) and they stand in a temporal relation of inclusion. However, the interpretation of this sentence as a text, i.e. a unitary message, requires some additional coherence relation between the two events that is not explicitly expressed. In the context of this occurrence, it is normal for a human reader to interpret event B, “hit” as a *consequence* of the event A, “crossing”.

We consider this type of relations between event pairs as a case of a *circumstantial relation*. A circumstantial relation can be best described as a coherence relation between events which allows to interpret and understand their occurrence in the

world in terms of a coherent unitary message. It explains to human readers “why” something happened, without necessarily explaining it. Circumstantial relations are a set of relations which includes temporal, causal, entailment, prevention and contingency relations, among others.

We distinguish two types of circumstantial relations: episodic and semantic. An *episodic* circumstantial relation is a relation that holds between a pair of specific actual event instances in a specific context, where their connection is necessary to understand what is described in a meaningful and coherent way. For instance, the relation between events A and B is a case of an episodic circumstantial relation: A and B may happen independently without implying the other necessarily, but when described in the same context, or circumstance, a connection is created that explains their occurrence as a dependent relation.¹

On the other hand, we define *semantic* circumstantial relations as a relation that holds between event classes (abstracting from actual events), where an event of class C gives rise to another event of class D or vice versa, based on shared properties in the formalization of the classes. For instance: the class “Shooting” has a semantic circumstantial relation with the class “Impacting”, because they both share the property of translocation of an object from location Y to Z. Modeling these relations provides a means to track chains of logically related events and their shared participants within and across documents.

Semantic circumstantial relations thus define possible explanatory sequences of events but not the actual explanatory sequences. Episodic relations define actual circumstantial sequences that fit the semantic model. The Circumstantial Event

¹Of course, not all events can have an explanation. For instance, there is no episodic circumstantial relation that tells us why Helen is crossing the street.

Ontology (CEO), described in this paper, models the semantic relations, based on *shared properties* of the event classes with the intention to support detecting episodic circumstantial relations in texts.

We specify the methodology used in section 2.1. Modeling the relations in an ontology will allow us to 1.) abstract over the different lexical realizations of the same concept (i.e. at an event mention level); 2.) facilitate reasoning between event classes and enrich the extraction of information for event knowledge and event sequences.²

Existing ontologies and models such as SUMO (Niles and Pease, 2001) and FrameNet (Ruppenhofer et al., 2006) do provide explicit causal relations between event classes (SUMO) or preceding and causal relations (FrameNet). These causal relations are strict, meaning that if A happens, then B must happen as well. However, our relations are circumstantial, meaning that some instance of event class C and D can happen independently, but given the circumstance that they coincide, C implies D or D is implied by C. The implication is however not necessary.

Previous work on the encoding of semantic relations between event pairs has focused on specific subsets of circumstantial relations. For instance, one example is the encoding of the entailment relations in WordNet (Fellbaum, 1998). With respect to the WordNet approach in this work, we abstract from various event types (i.e. lexical items) and do not depend on relations defined at a synset level by formalizing event knowledge and relations in an ontology. Another related approach are narrative chains as described in (Chambers and Jurafsky, 2010) that provide chains of various event mentions. However, the relation between these mentions is not specified explicitly but based on co-occurrence of participants and a basic precedence relation. Manual inspection of these chains revealed that dissimilar relations are implied within these chains, varying from temporal ordering, to episodic, up to causal. The Penn Discourse Tree-Bank (PDTB) (Prasad et al., 2007) annotates contingency relations, of which causal relations are a subclass. In PDTB, the focus of the annotation is between two Abstract Objects (called Arg1 and Arg2), corresponding to discourse units, rather than event mentions. The contingency relation is

annotated either in presence of an explicit connective, i.e. a lexical item, connecting the two abstract objects or implicitly by adjacency in discourse. In our approach, contingency relations are one of the possible values which express circumstantial relations, and, most importantly, they are independent of the presence of connectives or adjacency in discourse but grounded on (shared) properties of events.

A resource such as the CEO is envisioned to be of added value for several NLP tasks such as script mining, question answering, information extraction and textual entailment, among others. Furthermore, the explicitly defined relations between events can be of help in reconstructing storylines (Vossen et al., 2015), (van den Akker et al., 2010) and improve the coherence of the narrative chain models (Chambers and Jurafsky, 2010).

The remainder of this paper is organized as follows: in section 2 we describe the meta model and the development of CEO; in section 3 we report on plans and current work to evaluate CEO; in section 4 we conclude with final remarks and future work.

2 The Circumstantial Event Ontology

The CEO builds upon an existing event ontology called the Event and Implied Situation Ontology (ESO) (Segers et al., 2016). ESO is designed to run over the output of Semantic Role Labeling systems by making explicit the ontological type of the predicative element and the situation that holds before, during and after the predicate. Each so called pre-, post- and during situation consists of a set of properties and roles that define what holds true. For instance, as can be seen in Figure 2 the pre- and post-situations of the event class “Damaging” define:

- that something is in a “relatively plus” state (pre-situation);
- that this something is in a “relatively less” state, i.e. it underwent a loss or a negative change, relatively to the state before the damaging (“+”) (post-situation);
- that some object is in a state ‘damaged’ after the event (post-situation);
- that something has some damage which has some negative effect on some activity (post-situation).

²CEO will be made publicly available with a CC-BY-SA license.

pre situation:	damaging-undergoer damaging-state-1	inState hasRelativeValue	damaging-state-1 "+"
post situation:	damaging-undergoer damaging-state-2 damaging-undergoer damaging-undergoer damaging-damage	inState hasRelativeValue isDamaged hasDamage hasNegativeEffectOn	damaging-state-2 "." true damaging-damage activity

Figure 1: The ESO assertions for the class Damaging

ESO allows to track chains of states and changes over time, whether explicitly reported or inferred. However, ESO does not provide any explicit definition on what event class logically precedes or follows some other event class, i.e. the pre-, post- and during situations provide only descriptions of properties of the participants of the event in analysis. CEO aims at extending ESO, by further developing the event hierarchy, the expressiveness of the pre-, post-, and during situations, and, finally, the definition of the circumstantial semantic relations between the classes.

2.1 The CEO Meta Model

CEO is an OWL2 ontology, still under development, which currently consists of 250 event classes, 65 roles, and 58 unique properties that model the pre-, post- and during situations of the event classes.

The CEO meta model fully adopts and extends the ESO model (Segers et al., 2016). The reasons to reuse and extend it are twofold: 1) The ESO classes and roles are mapped to FrameNet, therefore we can rely on existing SRL techniques and models to instantiate CEO (Björkelund et al., 2009; de Lacalle et al., 2016); 2) ESO provides a model that defines what situation, or state, is true before and after an event, thereby already providing the initial hooks to define the circumstantial semantic relations. Event classes are connected by checking if a shared property holds in one of the following conditions:

- between a post-situation of class X and the pre-situation of class Y;
- between the post-situation of class X and the during situation of class Y;
- between the during situation of class X and the pre-situation of class Y.

Figure 2.1 illustrates this approach and the CEO meta model. In the Figure, the class "Damaging"

has a post-situation where is stated that some object is damaged (X isDamaged true). For the static class "BeingDamaged", the same statement is defined as a during situation, meaning that during the state "BeingDamaged", some object is in a damaged state. As such, both classes are tied together, based on a shared property. Further, the role of the entity that undergoes the change (here: X) is mapped to several FrameNet frame elements while the class, e.g. "Damaging", is mapped to both a SUMO class and FrameNet frames.

For relating the classes we investigate two options. Either we leave the relation between the classes implicit and track possible paths connecting the classes based on the shared properties. Another possibility is that we define explicit relations between the classes. For the latter case, we propose to define two properties: 1.) "hasCircumstantialPreEvent" (HCPRE), which expresses that an event class (e.g. "Shooting") is elicited by another one (e.g. "BeingArmed"); and 2.) "hasCircumstantialPostEvent" (HCPoE) which expresses that an event class (e.g. "Shooting") elicits another one ("Impacting"). Both properties are modeled as a non-inverse property of each other and as non-propagational. This implies that the relation only holds between two event classes and does not inherit to any of its subclasses. Also, if there is a "hasCircumstantialPostEvent" property between event class A and B, this does not imply that there is a relation from B back to A, unless specified otherwise. However, at this moment the pre-, post- and during situations, which are used to connect the classes, do not provide the information to determine the directionality of the HCPRE and HCPoE relations.

Figure 2.1 illustrates a chaining of calamity events and their relations. On the left, we show the event classes and on the right the pre-, post- and during situations. Note that we do not show the subclass hierarchy here, but only the binding of a subset of event classes based on shared properties. For instance, the class "Shooting" has a HCPoE relation to "Impacting", while the class "BeingArmed" has a HCPRE relation to "HavingAPurpose".

2.2 Building the CEO

CEO is designed to capture chains of events in newswire, more specifically calamity events. We define a calamity event as any event where some

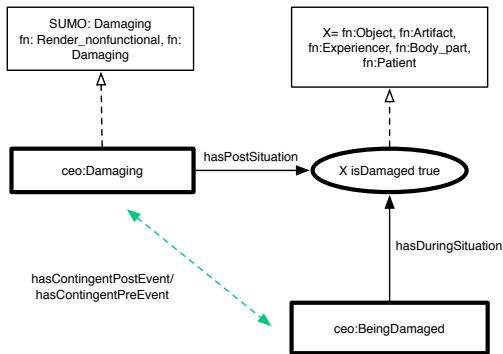


Figure 2: The meta model of CEO and the mappings to the external resources FrameNet (fn) and SUMO at class and at role levels.

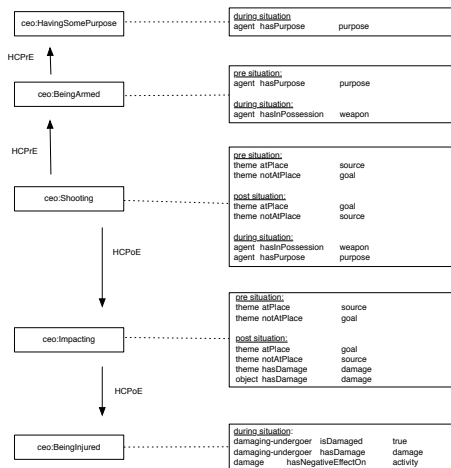


Figure 3: Explicit chaining of event classes (left) and their shared properties in the pre, post and during situation (right).

situation turns from relatively positive to some relatively negative state due to some changes in the world. Event classes that define processes are also modeled in CEO, where some agent tries to improve some situation in reaction to some calamity, i.e. going from a relatively negative situation back to a relatively positive situation. Examples of calamity event classes are “CyberAttack” and “Earthquake”. Examples of event classes where an attempt to some improvement of a situation is made are “Repairing” and “Evacuation”.

ESO already provides us with some event classes for calamities, though the coverage is rather limited as ESO was designed for the economic-financial domain. As such, we massively extended the hierarchy from the initial 63 event classes in ESO to the 250 event classes currently in CEO. To the best of our knowledge, no formal ontology specific for calamities and the inter-event relations exist. Some thesauri such as the IPTC³ contain terms for calamities but these are not formalized and provide few relations. Therefore, we decided to define a new model, reusing existing resources as much as possible.

As an input for the calamity classes defined in CEO, we partially were able to reuse Chamber’s narrative chains (Chambers and Jurafsky, 2010) for as far as these pertained to calamities of some sort. This selection was made manually, based on at least three calamity events per event chain. Further, we manually selected FrameNet frames that capture calamity events. We used the SUMO ontology as a backbone for modeling our initial list of verbs and frames. Finally, we defined SKOS mappings from each CEO event class to FrameNet and SUMO.⁴ thus providing the opportunity to use CEO on SRL labeled text as well as to find the vocabulary expressing calamities by means of the lexical units mapped to frames in FrameNet and the mappings to Princeton WordNet that are defined in SUMO.

3 Evaluation

The CEO will be evaluated against a benchmark corpus to determine precision and recall for both the classes and the semantic circumstantial relations. For this, we plug the CEO into an existing NLP pipeline for text annotation and analysis (Vossen et al., 2016) For this, we are cur-

³<https://iptc.org/>

⁴<https://www.w3.org/2004/02/skos/>

rently annotating part of the ECB+ corpus (Cybulska and Vossen, 2014). We selected 24 topics that describe a calamity event. In our annotation, we only use the existing event mention annotations and add new mentions if they realize an event calamity class. In addition to this, the annotators define co-reference sets among event mentions and the semantic circumstantial relations. As such, we can evaluate what events are captured by our ontology and what relations can be successfully reconstructed. For the annotation, we use the CAT annotation tool (Bartalesi Lenzi et al., 2012). Additionally, we are designing a Question-Answering task, where systems will have to provide answers to questions "why" a certain event has taken place rather than factoid questions by providing the most relevant and direct preceding event that can be seen as an explanation.

4 Conclusion and Future Work

We have described current ongoing work on an event ontology that captures calamity events in newswire and the semantic circumstantial relations that hold between event classes, based on shared properties in the pre-, post- or during situations defined for each class. Future work includes the further development of the ontology with a focus on defining the circumstantial semantic relations between the classes and an extension of the expressivity of the pre-, post- and during situations of the event classes. Further, we will evaluate the added value of our model both intrinsically, against a manually annotated corpus, and extrinsically, by means of a QA task.

5 Acknowledgements

This research was funded by the NWO Spinoza Prize project "Understanding Language by Machines.

References

V. Bartalesi Lenzi, G. Moretti, and R. Sprugnoli. 2012. Cat: the celct annotation tool. In *LREC*. pages 333–338.

A. Björkelund, L. Hafdell, and P. Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of CoNLL-2009*. Boulder, CO, USA.

N. Chambers and D. Jurafsky. 2010. A database of narrative schemas. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2010)*.

A. Cybulska and P. Vossen. 2014. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*. Reykjavik, Iceland. http://www.lrec-conf.org/proceedings/lrec2014/pdf/840_Paper.pdf.

M. Lopez de Lacalle, E. Laparra, I. Aldabe, and G. Rigau. 2016. A multilingual predicate matrix. In *Proceedings of Language Resources and Evaluation Conference (LREC 2016)*.

C. Fellbaum. 1998. *WordNet: an electronic lexical database*. MIT Press.

I. Niles and A. Pease. 2001. Towards a standard upper ontology. In *Proceedings of FOIS-Volume 2001*. ACM.

R. Prasad, E. Miltsakaki, N. Dinesh, A. Lee, A. Joshi, L. Robaldo, and Bonnie L Webber. 2007. The penn discourse treebank 2.0 annotation manual .

J. Ruppenhofer, M. Ellsworth, M. Petruck, C.R. Johnson, and J. Scheffczyk. 2006. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California.

R. Segers, M. Rospocher, P. Vossen, E. Laparra, G. Rigau, and A.L. Minard. 2016. The event and implied situation ontology: Application and evaluation. In *Proceedings of Language Resources and Evaluation Conference (LREC 2016)*.

C. van den Akker, L.M. Aroyo, A.K. Cybulska, M.G.J. van Erp, P. Gorgels, L. Hollink, C. Jager, S. Legne, L. van der Meij, J. Oomen, J. van Ossensbruggen, G. Schreiber, R. Segers, P.T.J.M. Vossen, and B. Wielinga. 2010. *Historical Event-based Access to Museum Collections*, CEUR-WS (online), pages 1–9.

P. Vossen, R. Agerri, I. Aldabe, A. Cybulska, M. van Erp, A. Fokkens, E. Laparra, A. Minard, A. Palmero Aprosio, G. Rigau, M. Rospocher, and R. Segers. 2016. *Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news*. *Special Issue Knowledge-Based Systems, Elsevier* <https://doi.org/dx.doi.org/10.1016/j.knosys.2016.07.013>.

P. Vossen, T. Caselli, and Y. Kontzopoulou. 2015. Storylines for structuring massive streams of news. In *Proceedings of the 1st Workshop on Computing News StoryLines (CNews 2015) at the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)*. Beijing, China.