

Extracting Temporal and Causal Relations between Events

Paramita Mirza

Fondazione Bruno Kessler

University of Trento

Trento, Italy

paramita@fbk.eu

Abstract

A notably challenging problem related to event processing is recognizing the relations holding between events in a text, in particular temporal and causal relations. While there has been some research on temporal relations, the aspect of causality between events from a Natural Language Processing (NLP) perspective has hardly been touched. We propose an annotation scheme to cover different types of causality between events, techniques for extracting such relations and an investigation into the connection between temporal and causal relations. In this thesis work we aim to focus especially on the latter, because causality is presumed to have a temporal constraint. We conjecture that injecting this presumption may be beneficial for the recognition of both temporal and causal relations.

1 Introduction

With the rapid growth of information available on the world wide web, especially in the form of unstructured and natural texts, *information extraction* (IE) becomes one of the most prominent fields in NLP research. IE aims to provide ways to automatically extract the available information and store them in a structured representation of knowledge. The stored knowledge can then be useful for many NLP applications, such as question answering, textual entailment, summarization, and focused information retrieval systems.

There are several subtasks within information extraction related to the type of knowledge one wishes to extract from the text, *event extraction* being one of them. Event extraction is considered to be a non-trivial task, due to the fact that mentions of an event in text could be highly varied in terms of sentence construction, and that the attributes describing an event are usually mentioned in several

sentences. However, the most challenging problem in the context of event extraction is identifying the relationship between events.

Events are usually anchored to temporal expressions. The temporal attribute of an event can be used to determine the temporal relationship between events. This information can be useful for the ordering of event sequence in a timeline, e.g. for the better presentation of news or history texts. Moreover, in multi-document summarization of news articles, the relative order of events is important to merge and present information from multiple sources correctly.

A more complex type of relationship between events is causality. Identifying the causal relation between events is an important step in predicting occurrence of future events, and can be very beneficial in risk analysis as well as decision making support.

There is an overlap between causal and temporal relations, since by the definition of causality, the first event (cause) must happen *BEFORE* the second event (effect). We claim that a system for extracting both temporal and causal relations, may benefit from integrating this presumption. The main focus of this research work will be (i) investigating ways to utilize this presumption in building an integrated event relation extraction system, in addition to (ii) exploring ways to develop a robust extraction component for each type of relations (temporal and causal).

2 Background

In NLP, the definition of an event can be varied depending on the target application. In topic detection and tracking (Allan, 2002), the term *event* is used interchangeably with *topic*, which describes something that happens and is usually used to identify a cluster of documents, e.g. *Olympics*, *wars*. On the other hand, information extraction provides finer granularity of event definitions, in which events

are entities that happen/occur within the scope of a document.

There are several annotation frameworks for events and temporal expressions that can be viewed as *event models*,¹ TimeML (Pustejovsky et al., 2003b) and ACE (Consortium, 2005) being the prominent ones.

Both TimeML and ACE define an event as something that happens/occurs or a state that holds true, which can be expressed by a verb, a noun, an adjective, as well as a nominalization either from verbs or adjectives. Consider the following passage annotated with events and temporal expressions (TIMEX). “A *Philippine volcano*, *dormant* EVENT for *six centuries* TIMEX, *exploded* EVENT *last Monday* TIMEX. *During the eruption* EVENT, *lava, rocks and red-hot ash are spewed* EVENT *onto surrounding villages*. *The explosion* EVENT *claimed* EVENT *at least 30 lives*.”

The most important attribute of TimeML that differs from ACE is the separation of the representation of events and temporal expressions from the anchoring or ordering dependencies. Instead of treating a temporal expression as an event argument, TimeML introduces *temporal link* annotations to establish dependencies (temporal relations) between events and temporal expressions (Pustejovsky et al., 2003b). This annotation is important in (i) anchoring an event to a temporal expression (event time-stamping) and (ii) determining the temporal order between events. This distinctive feature of TimeML becomes our main consideration in choosing the event model for our research.

Moreover, TimeML is the annotation framework used in TempEval-3², the most recent shared task on temporal and event processing. The ultimate goal of this evaluation campaign is the automatic identification of temporal expressions, events, and temporal relations within a text (UzZaman et al., 2012).

The main tasks defined in TempEval-3 include: the automatic extraction of *TimeML entities*, i.e. temporal expressions and events, and the end-to-end automatic extraction of both TimeML entities and temporal links/relations. The result of TempEval-3 reported by UzZaman et al. (2013)

¹There are other event models based on web ontology (RDFS+OWL) such as LODE (Shaw et al., 2009), SEM (van Hage et al., 2011) and DOLCE (Gangemi et al., 2002), which encode knowledge about events as triples. Such models can be seen as ways to store the extracted knowledge to perform the reasoning on.

²<http://www.cs.york.ac.uk/semeval-2013/task1/>

shows that even though the performances of systems for extracting TimeML entities are quite good (>80% F-score), the overall performance of end-to-end event extraction systems suffers from the low performance of the temporal relation extraction system. The state-of-the-art performance on the temporal relation extraction task yields only around 36% F-score. This becomes the main reason of focusing our research on the extraction of event relations.

3 Research Problem

We consider two types of event relations to be extracted from text, which are *temporal relations* and *causal relations*. Causal relations are related to temporal relations since there is a temporal constraint in causality, i.e. the cause must precede the effect. Considering this presumption, and the assumption that there are good enough systems to extract temporal expressions and events, we define two main problems that will be addressed in this research work:

1. Given a text annotated with entities (temporal expressions and events), how to automatically extract temporal and causal relations between them.
2. Given the temporal constraint of causality, how to utilize the interaction between temporal relations and causal relations for building an integrated event relation extraction system for both types of relations.

4 Research Methodology

There are several aspects of the mentioned problems that will become our guidelines in continuing our research in this topic. The following sections will give a more detailed description of these aspects including the arising challenges, some preliminary results to address the challenges and our future research directions.

4.1 Temporal Relation Extraction

As previously mentioned, we consider the TimeML annotation framework because it explicitly encodes the temporal links between entities (events and temporal expressions) in a text. In TimeML, each temporal link has a temporal relation type assigned to it. There are 14 types of temporal relations specified in TimeML version 1.2.1 (Saurí et al., 2006),

which are defined based on Allen’s interval algebra (Allen, 1983), as illustrated in Table 1.

	<i>a</i> is <i>BEFORE</i> <i>b</i> <i>b</i> is <i>AFTER</i> <i>a</i>
	<i>a</i> is <i>IBEFORE</i> <i>b</i> <i>b</i> is <i>IAFTER</i> <i>a</i>
	<i>a</i> <i>BEGINS</i> <i>b</i> <i>b</i> is <i>BEGUN_BY</i> <i>a</i>
	<i>a</i> <i>ENDS</i> <i>b</i> <i>b</i> is <i>ENDED_BY</i> <i>a</i>
	<i>a</i> is <i>DURING</i> <i>b</i> <i>b</i> is <i>DURING_INV</i> <i>a</i>
	<i>a</i> <i>INCLUDES</i> <i>b</i> <i>b</i> IS <i>INCLUDED</i> in <i>a</i>
	<i>a</i> is <i>SIMULTANEOUS</i> with <i>b</i>
	<i>a</i> is <i>IDENTITY</i> with <i>b</i>

Table 1: Temporal relations in TimeML annotation

Recalling the low performances of currently available systems on the temporal relation extraction task, including the state-of-the-art systems according to TempEval-3, it is still insufficient to use the existing temporal relation extraction systems to support real world applications, such as creating event timelines and temporally-based question answering. Therefore, as the first step we take as an objective finding ways to improve the current state-of-the-art performance on temporal relation extraction task.

The common approach towards temporal relation extraction is dividing the task into two sub-tasks: *identifying the pairs* of entities having a temporal link and *determining the relation types*. The problem of identifying the entity pairs is usually simplified. In TempEval-3, the possible pairs of entities that can have a temporal link are defined as (i) main events of consecutive sentences, (ii) pairs of events in the same sentence, (iii) an event and a time expression in the same sentence, and (iv) an event and the document creation time (UzZaman et al., 2013). The problem of determining the label of a given temporal link is usually regarded as a classification problem. Given an ordered pair of entities (e_1 , e_2) that could be either *event-event*, *event-timex* or *timex-timex* pair, the classifier has to assign a certain label representing the temporal relation type.

We focus on the latter subtask of classifying temporal relation types, assuming that the links between events and time expressions are already established. Several recent works have tried to address this complex multi-class classification task by using sophisticated features based on deep pars-

ing, semantic role labelling and discourse parsing (D’Souza and Ng, 2013; Laokulrat et al., 2013). In Mirza and Tonelli (2014) we have shown that a simpler approach, based on lexico-syntactic features, can achieve comparable results.

A classification model is trained for each category of entity pair, i.e. event-event, event-timex and timex-timex, as suggested in several previous works (Mani et al., 2006; Chambers, 2013). However, because there are very few examples of timex-timex pairs in the training corpus, it is not possible to train a classifier for these particular pairs. Moreover, they only add up to 3.2% of the total number of extracted entity pairs; therefore, we decided to disregard these pairs.

We follow the guidelines and the dataset provided by the organizers of TempEval-3 so that we can compare our system with other systems participating in the challenge. The TBAQ-cleaned corpus is the training set provided for the task, consisting of the TimeBank (Pustejovsky et al., 2003a) and the AQUAINT corpora. It contains around 100K words in total, with 11K words annotated as events (UzZaman et al., 2013).

Simple Feature Set. We implement a number of features including the commonly used ones (UzZaman et al., 2013), which take into account morpho-syntactic information on events and time expressions, their textual context and their attributes. Other features rely on semantic information such as typical event durations and explicit temporal connective types. However, we avoid complex processing of data. Such semantic information is based on external lists of lexical items and on the output of the *addDiscourse* tagger (Pitler and Nenkova, 2009). We build our classification models using the Support Vector Machine (SVM) implementation provided by YamCha³.

We perform feature engineering in order to select from our initial set of features only those that improve the accuracy of the classifiers. This allows us to select the best classification models for both event-event pairs and event-timex pairs.

Inverse Relations and Closure. Motivated by the finding of Mani et al. (2006) that bootstrapping the training data through a *temporal closure* method results in quite significant improvements, we investigate the effect of enriching the training data with *inverse relations* and *closure-based inferred*

³<http://chasen.org/~taku/software/yamcha/>

relations.

However, we adopt a simpler approach to obtain the closure graph of temporal relations, by applying the *transitive closure* only within the same relation type, e.g. $e_1 \text{ BEFORE } e_2 \wedge e_2 \text{ BEFORE } e_3 \rightarrow e_1 \text{ BEFORE } e_3$. It produces only a subset of the relations produced by the *temporal closure* (Verhagen, 2005; Gerevini et al., 1995). The problem of finding the transitive closure of a directed acyclic graph can be reduced to a boolean matrix multiplication (Fischer and Meyer, 1971).

Training data	event-event	event-timex
TBAQ	48.28%	73.82%
TBAQ-I	47.77%	74.45%
TBAQ-IC	46.39%	74.45%

Table 2: Classifier accuracies with different training data: TBAQ (TimeBank+AQUAINT), TBAQ-I (TBAQ+inverse relations) and TBAQ-IC (TBAQ+inverse relations+transitive closure).

Evaluation and Analysis. Our test data is the newly annotated TempEval-3-platinum evaluation corpus provided by TempEval-3 organizers, so that we can compare our system with other systems participating in the task. First, to investigate the effect of enriching the training data with inverse relations and transitive closure, we evaluate the system performance trained with different datasets, as shown in Table 2. A randomization test between the best performing classifier and the others shows that by extending the training data with inverse relations and transitive closure, the improvement are not significant. Applying inverse relations and transitive closure extends the number of training instances but makes the already skewed dataset more imbalanced, thus it does not result in a significant improvement.

We then train our classifiers for event-event pairs and event-timex pairs by exploiting the best feature combination and using the best reported dataset for each classifier as the training data. The two classifiers are part of our temporal classification system called *TRelPro*.

Compared with the performances of other systems participating in TempEval-3 reported in UzZaman et al. (2013), *TRelPro* is the best performing system both in terms of precision and of recall. The result of our system using simpler features confirms the finding reported in UzZaman et al. (2013), that a system using basic morpho-syntactic features is hard to beat with systems using more

complex semantic features, if not used properly.

System	F1	Precision	Recall
TRelPro	58.48%	58.80%	58.17%
UTTime-1, 4	56.45%	55.58%	57.35%
UTTime-3, 5	54.70%	53.85%	55.58%
UTTime-2	54.26%	53.20%	55.36%
NavyTime-1	46.83%	46.59%	47.07%
NavyTime-2	43.92%	43.65%	44.20%
JU-CSE	34.77%	35.07%	34.48%

Table 3: TempEval-3 evaluation on the classification of temporal relation types

4.2 Causal Relation Extraction

Unlike the *temporal order* that has a clear definition, there is no consensus in the NLP community on how to define *causality*. Causality is not a linguistic notion, meaning that although language can be used to express causality, causality exists as a psychological tool for understanding the world independently of language (van de Koot and Neeleman, 2012). There have been several attempts in the psychology field to model causality, including the *counterfactual model* (Lewis, 1973), *probabilistic contrast model* (Cheng and Novick, 1991; Cheng and Novick, 1992) and the *dynamics model* (Wolff and Song, 2003; Wolff et al., 2005; Wolff, 2007), which is based on Talmy’s *force dynamic* account of causality (Talmy, 1985; Talmy, 1988).

In information extraction, modelling causality is only the first step in order to have guidelines to recognize causal relations in a text. In order to have an automatic extraction system for causal relations (particularly using a data-driven approach) and most importantly to evaluate the performance of the developed extraction system, it is important that a language resource annotated with causality is available.

Even though there are several corpora annotated with causality, e.g. Penn Discourse Treebank (PDTB) (Prasad et al., 2007) and PropBank (Palmer et al., 2005),⁴ we are not aware of any standard benchmarking corpus for evaluating event causality extraction, as it is available for temporal relations in TimeML. This motivates us to create a language resource annotated with both temporal and causal relations in a unified annotation scheme, for the main purpose of investigating the interaction between both types of relations. It becomes the objective of the second stage of our research, in

⁴PDTB annotates causality between related clauses, while PropBank annotates causality between a verb and its cause clause.

addition to building an automatic extraction system for event causality using the developed corpus.

In Mirza et al. (2014) we have proposed annotation guidelines for causality between events, based on the TimeML definition of events, which considers all types of actions (punctual and durative) and states as events. Parallel to the <TLINK> tag in TimeML for temporal relations, we introduced the <CLINK> tag to signify a causal link. We also introduced the notion of causal signals through the <C-SIGNAL> tag, parallel to the <SIGNAL> tag in TimeML indicating temporal cues.

C-SIGNAL. C-SIGNAL is used to mark-up textual elements signalling the presence of causal relations, which include all causal uses of: *prepositions* (e.g. because of, as a result of, due to), *conjunctions* (e.g. because, since, so that), *adverbial connectors* (e.g. so, therefore, thus) and *clause-integrated expressions* (e.g. the reason why, the result is, that is why).

CLINK. A CLINK is a directional relation where the causing event is the *source* (indicated with s in the examples) and the caused event is the *target* (indicated with T). The annotation of CLINKs also includes the *c-signalID* attribute, with the value of the ID of C-SIGNAL marking the causal relation (if available).

Wolff (2007) has shown that the dynamics model covers three main types of causal concepts, i.e. CAUSE, ENABLE and PREVENT. The model has been tested by linking it with natural language, Wolff and Song (2003) show that the three causal concepts can be lexicalized as verbs : (i) CAUSE-type verbs, e.g. *cause, prompt, force*; (ii) ENABLE-type verbs, e.g. *allow, enable, help*; and (iii) PREVENT-type verbs, e.g. *block, prevent, restrain*. Its connection with natural language becomes the main reason of basing our annotation guidelines for causality on the dynamics model.

We limit the annotation of CLINKs to the presence of an explicit causal construction linking two events, which can be one of the following cases:

1. Expressions containing **affect verbs** (*affect, influence, determine, change*), e.g. *Ogun ACN crisis* s **influences** the *launch* T of the All Progressive Congress.
2. Expressions containing **link verbs** (*link, lead, depend on*), e.g. *An earthquake* T in North America was **linked to a tsunami** s in Japan.
3. **Basic construction of causative verbs**, e.g.

The purchase s **caused** the *creation* T of the current building.

4. **Periphrastic construction of causative verbs**, e.g. *The blast* s **caused** the *boat to heel* T violently, where the causative verb (*caused*) takes an embedded verb (*heel*) expressing a particular result.
5. Expressions containing **causative conjunctions and prepositions**, which are annotated as C-SIGNALs.

Note that for causative verbs we consider sets of verbs from all types of causal concepts including CAUSE-type, ENABLE-type and PREVENT-type verbs.

Manual Annotation of Event Causality. Having the annotation guidelines, we are about to complete the annotation of event causality. We have annotated a subset of training corpus from TempEval-3 used in the temporal relation extraction, i.e. TimeBank. The agreement reached by two annotators on a subset of 5 documents is 0.844 Dice's coefficient on C-SIGNALs (micro-average over markables) and 0.73 on CLINKs.

After completing causality annotation, the next step will be to build an automatic extraction system for causal relations. We will consider to use a supervised learning approach, as well as the similar features employed for temporal relation classification task, in addition to lexical information (e.g. WordNet (Fellbaum, 1998), VerbOcean (Chklovski and Pantel, 2004)) and the existing causal signals.

4.3 Integrated Event Relation Extraction

During the last stage of our research work, we will investigate the interaction between temporal and causal relations, given the temporal constraint of causality. The ultimate goal is to build an integrated event relation extraction system, that is capable of automatically extracting both temporal and causal relations from text.

Few works have investigated the interaction between these two types of relations. The corpus analysis conducted by Bethard et al. (2008) shows that although it is expected that almost every causal relation would have an underlying before relation, in reality, 32% of causal relations in the corpus are not accompanied by underlying before relations. One of the possible causes is that the considered event pairs are conjoined event pairs under the ambiguous *and* conjunctive.

Consider the sentence “*The walls were shaking* τ *because of the earthquake* ς .” Looking at the explicit causal mark *because*, there is a causal relation between the events *shaking* and *earthquake*. However, according to Allen’s interval algebra or the TimeML annotation framework we cannot say that event *earthquake* occurred *BEFORE* the event *shaking*, because both events happen almost at the same time (could be *SIMULTANEOUS*), and in both frameworks there is no overlap in *BEFORE* relations. During our manual annotation process, we encountered the case where the cause event happens after the effect, as in “*Some analysts questioned* τ *how much of an impact the retirement package will have, because few jobs will end* ς *up being eliminated.*” Further investigations are needed to address this issue.

Rink et al. (2010) makes use of manually annotated temporal relation types as a feature to build a classification model for causal relations between events. This results in 57.9% of F1-Score, 15% improvement of performance in comparison with the system without the additional feature of temporal relations. The significant increase of performance proves that the temporal relations between causal events have a significant role in discovering causal relations. On the other hand, a brief analysis into our preliminary result on temporal relation extraction shows that there is a possibility to employ causality to improve the temporal relation classification of event-event pairs, specifically to reduce the number of *false positives* and *false negatives* of *BEFORE* and *AFTER* relations scored by the system. Our hypothesis is that temporal and causal relations can be of mutual benefit to the extraction of each other.

Taking into account different classification frameworks and possible configurations for the integrated system, for example, cascading the temporal and causal relation extraction systems, or one system for both relation types in one pass, we will explore the possibilities and evaluate their performances. Furthermore, there is a possibility to exploit a global optimization algorithm, as explored by Chambers and Jurafsky (2008) and Do et al. (2012), to improve the performance of a pairwise classification system.

One possible classification algorithm under our considerations, which can be used for extracting both temporal and causal relations in one pass, is General Conditional Random Fields (CRFs).

General CRFs allow us to train a classification model with arbitrary graphical structure, e.g. a two-dimensional CRF can be used to perform both noun phrase chunking and PoS tagging at the same time. And its skip-chain mechanism allows us to create a *chain of entity pairs*, which may improve the classification performance.

5 Conclusion

Event extraction has become one of the most investigated tasks of information extraction, since it is the key to many applications in natural language processing such as personalized news systems, question answering and document summarization. The extraction of relations that hold between events is one of the subtasks within event extraction gaining more attention in the recent years, given the beneficial and promising applications.

We have presented a research plan covering the topic of automatic extraction of two event relation types, i.e. temporal and causal relations, from natural language texts. While there has been a clearly defined framework for temporal relation extraction task, namely TempEval-3, there is none for causal relation extraction. Furthermore, since causality has a temporal constraint, we are interested in investigating the interaction between temporal and causal relations, in the context of events.

We propose a three-stage approach to cover this research topic. The first stage includes improving the state-of-the-art performance on temporal relation extraction. During the second stage we propose an annotation scheme to create a corpus for causal relations, based on the established annotation framework for events and temporal relations, namely TimeML. The created language resource will then be used to build the automatic extraction system for causal relations, and also to provide the benchmarking evaluation corpus. Finally, the last stage includes investigating the interaction between temporal and causal relations, in order to build an integrated system for event relation extraction, which is the ultimate goal of this research work.

Acknowledgments

The research leading to this paper was partially supported by the European Union’s 7th Framework Programme via the NewsReader Project (ICT-316404). We also thank Google for travel and conference support for this paper.

References

- James Allan, editor. 2002. *Topic Detection and Tracking: Event-based Information Organization*. Kluwer Academic Publishers, Norwell, MA, USA.
- James F. Allen. 1983. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, November.
- Steven Bethard, William Corvey, Sara Klengenstein, and James H. Martin. 2008. Building a corpus of temporal-causal structure. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).
- Nathanael Chambers and Dan Jurafsky. 2008. Jointly combining implicit constraints improves temporal ordering. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, pages 698–706, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Nate Chambers. 2013. Navytime: Event and time ordering from raw text. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 73–77, Atlanta, Georgia, USA, June. Association for Computational Linguistics.
- Patricia W. Cheng and Laura R. Novick. 1991. Causes versus enabling conditions. *Cognition*, 40(1-2):83 – 120.
- Patricia W. Cheng and Laura R. Novick. 1992. Covariation in natural causal induction. *Psychological Review*, 99(2):365–382.
- Timothy Chklovski and Patrick Pantel. 2004. Verbocean: Mining the web for fine-grained semantic verb relations. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 33–40, Barcelona, Spain, July. Association for Computational Linguistics.
- Linguistic Data Consortium, 2005. *ACE (Automatic Content Extraction) English Annotation Guidelines for Events*.
- Quang Xuan Do, Wei Lu, and Dan Roth. 2012. Joint inference for event timeline construction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL '12*, pages 677–687, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jennifer D'Souza and Vincent Ng. 2013. Classifying temporal relations with rich linguistic knowledge. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 918–927.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
- Michael J. Fischer and Albert R. Meyer. 1971. Boolean matrix multiplication and transitive closure. In *SWAT (FOCS)*, pages 129–131. IEEE Computer Society.
- Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. 2002. Sweetening ontologies with dolce. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, EKAW '02*, pages 166–181, London, UK, UK. Springer-Verlag.
- Alfonso Gerevini, Lenhart Schubert, and Stephanie Schaeffer. 1995. The temporal reasoning tools timegraph i-ii. *International Journal of Artificial Intelligence Tools*, 4(1-2):281–299.
- Natsuda Laokulrat, Makoto Miwa, Yoshimasa Tsuruoka, and Takashi Chikayama. 2013. Uttime: Temporal relation classification using deep syntactic features. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 88–92, Atlanta, Georgia, USA, June. Association for Computational Linguistics.
- David Lewis. 1973. Causation. *The Journal of Philosophy*, 70(17):pp. 556–567.
- Inderjeet Mani, Marc Verhagen, Ben Wellner, Chong Min Lee, and James Pustejovsky. 2006. Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, ACL-44*, pages 753–760, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Paramita Mirza and Sara Tonelli. 2014. Classifying temporal relations with simple features. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*.
- 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)*.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106, March.
- Emily Pitler and Ani Nenkova. 2009. Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, ACLShort '09*, pages 13–16, Stroudsburg, PA, USA. Association for Computational Linguistics.

- Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo, and Bonnie L Webber. 2007. The penn discourse treebank 2.0 annotation manual. Technical report.
- J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, et al. 2003a. The timebank corpus. In *Corpus Linguistics*, volume 2003, page 40.
- James Pustejovsky, José Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003b. TimeML: Robust specification of event and temporal expressions in text. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*.
- Bryan Rink, Cosmin Adrian Bejan, and Sanda M. Harabagiu. 2010. Learning textual graph patterns to detect causal event relations. In *FLAIRS Conference*.
- Roser Saurí, Jessica Littman, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky, 2006. *TimeML Annotation Guidelines, Version 1.2.1*.
- Ryan Shaw, Raphaël Troncy, and Lynda Hardman. 2009. Lode: Linking open descriptions of events. In *Proceedings of the 4th Asian Conference on The Semantic Web, ASWC '09*, pages 153–167, Berlin, Heidelberg. Springer-Verlag.
- Leonard Talmy. 1985. Force dynamics in language and thought. *Chicago Linguistic Society*, 21:293337.
- Leonard Talmy. 1988. Force dynamics in language and cognition. *Cognitive Science*, 12(1):49–100.
- Naushad UzZaman, Hector Llorens, James F. Allen, Leon Derczynski, Marc Verhagen, and James Pustejovsky. 2012. Tempeval-3: Evaluating events, time expressions, and temporal relations. *CoRR*, abs/1206.5333.
- 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 *Proceedings of the Seventh International Workshop on Semantic Evaluation, SemEval '13*, pages 1–9, Atlanta, Georgia, USA, June. Association for Computational Linguistics.
- H van de Koot and A Neeleman, 2012. *The Theta System: Argument Structure at the Interface*, chapter The Linguistic Expression of Causation, pages 20 – 51. Oxford University Press: Oxford.
- Willem Robert van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. 2011. Design and use of the simple event model (sem). *Journal of Web Semantics*, 9(2):128–136.
- Marc Verhagen. 2005. Temporal closure in an annotation environment. *Language Resources and Evaluation*, 39(2-3):211–241.
- Phillip Wolff and Grace Song. 2003. Models of causation and the semantics of causal verbs. *Cognitive Psychology*, 47(3):276 – 332.
- Phillip Wolff, Bianca Klettke, Tatyana Ventura, and Grace Song, 2005. *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin. APA decade of behavior series*, chapter Expressing Causation in English and Other Languages, pages 29–48. Washington, DC, US: American Psychological Association, xx, 316 pp.
- Phillip Wolff. 2007. Representing causation. *Journal of Experiment Psychology: General*, 136:82–111.