Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference

# Using Part-Of Relations for Discovering Causality

**Rutu Mulkar-Mehta[1], Christopher Welty[2], Jerry R. Hobbs[1], Eduard Hovy[1]**
me@rutumulkar.com, welty@us.ibm.com, hobbs@isi.edu, hovy@isi.edu
[1]University of Southern California, Information Sciences Institute
[2]IBM Watson Research Center

**Abstract**

Historically, causal markers, syntactic structures and connectives have been the sole identifying features for automatically extracting causal relations in natural language discourse. However various connectives such as "and", prepositions such as "as" and other syntactic structures are highly ambiguous in nature, and it is clear that one cannot solely rely on lexico-syntactic markers for detection of causal phenomenon in discourse. This paper introduces the theory of granularity and describes different approaches to identify granularity in natural language. As causality is often granular in nature (Mazlack 2004), we use granularity relations to discover and infer the presence of causal relations in text. We compare this with causal relations identified using just causal markers. We achieve a precision of 0.91 and a recall of 0.79 using granularity for causal relation detection, as compared to a precision of 0.79 and a recall of 0.44 using pure causal markers for causality detection.

## 1 Introduction

Causality is an important phenomena describing the working of the world. We are forever in the quest to find the cause or purpose of an event. We look for causal relations to explain simple day to day activities to scientific phenomena about the universe or genetics. For instance we question: *why the republicans lost in the elections?*, *why a solar eclipse occurs?*, *why are the streets jammed with traffic at this time of the day?* There have been numerous research initiatives to extract causal structures from discourse. (Girju and Moldovan 2002) have worked on identifying lexico-syntactic features for semi-automatically extracting causal structures from discourse. (Blanco, Castell, and Moldovan 2008) describe techniques to mark the causal relations between a verb phrase and a subordinate clause. (Pechsiri and Kawtrakul 2007) have worked on extracting causal relations using Elementary Discourse Units (EDU's).

All of these works have considered causality as a sequential set of events at the 'same' level. However, causality is often described using a granular structure, where the coarse grained event happens because of a fine grained event. For instance, in *a building collapses because the roof caved in*, the roof is in integral part of the building, and is a sub-event of the entire building collapsing. This paper focuses on granularity, and how such granularity structures can be used to identify causal relations in text.



We use the phenomenon of granularity on a regular basis in our everyday life. For planning and scheduling of important tasks, we often divide or split our tasks into smaller pieces, till each task is easily manageable. For instance, the day-to-day activity of shopping for groceries involves some finer grained events such as *driving to the grocery store*, *carrying a list*, *picking out required items* and *paying the cashier*. Each of these events in turn involve some finer level events. For instance, *driving to the grocery store* involves sub-events like *opening the car door*, *starting the engine*, *planning the route* and *driving to the destination*. The set of fine grained events make up a coarse grained event. When the fine grained events are handled successfully, the coarse grained event is handled successfully. In this sense, granularity decomposition is script or plan decomposition.

In this paper we first introduce our theory of granularity in natural language text (Section 2) along with the variations or options of our theory. We then describe two different approaches that we used to identify presence of granularity structures from text (Section 3). We then describe our experiments for causality identification using causal connectives and using granularity relation indicators (Section 4). Next, we present the analysis of our results, and other issues faced (Section 5). We finally conclude with the conclusions (Section 6).

## 2 Granularity Theory in Natural Language Discourse

We propose a framework or theory for modeling granularity in natural language, in order to represent explicitly the intuitive patterns humans use to shift through various levels of granularity. We focus our theory on event granularities only, not other types of granularity such as temporal and spatial.

Figure 1 illustrates our theory of granularity. A gran-

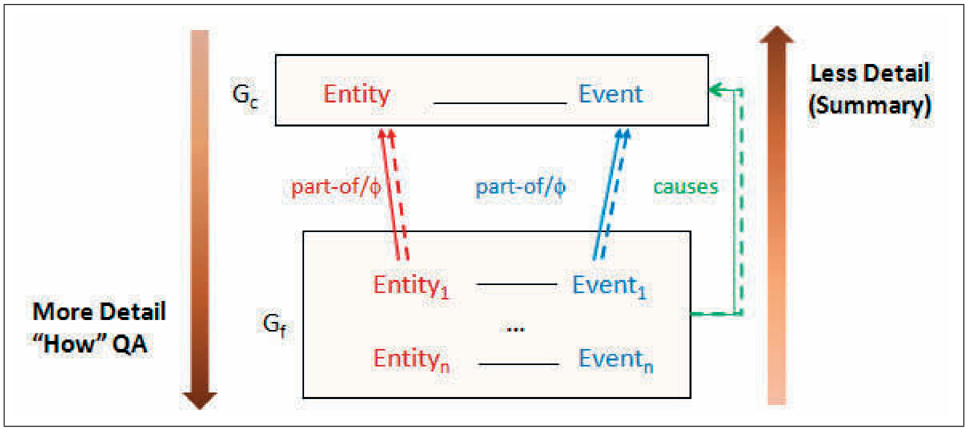


Figure 1: Granularity in Natural Language Descriptions

ularity structure exists only if at least two levels of information are present in text, such that the events in the coarse granularity can be decomposed into the events in the fine granularity and the events in the fine granularity combine together to form at least one segment of the event in the coarse granularity. In Figure 1, $G_c$ represents the phrase or sentence with coarse granularity information and $G_f$ represents a phrase or sentence with fine granularity information. Three types of relations can exist between the objects in coarse and fine granularity: part-whole relationship between entities, part-whole relationship between events, and causal relationship between the fine and coarse granularity. These relations signal a shift in granularity. Instantiating text phrases into this model will produce granularities of text.

This theory of granularity in natural language (Figure 1) can be modified such that only two relations or one relation of the required three relations need be used to instantiate the model, and the other unknown relation(s) can be inferred. These modifications are represented in Figure 1 in the form of dotted lines, where the dotted line is the relation that is interpreted when the other relations (represented as solid lines) are provided. These modifications to the theory of granularity are useful when partial information from a text is mapped to the theory of granularity.

Traditionally causal relations have been extracted using only lexico-syntactic causal markers. Figure 2 shows the sequential structure of causality which is assumed when extracting causal relations. However, causality is often granular in nature, where an event is caused by a set of sub-events, exhibiting a granular structure where the coarse grained events happen because of the fine grained events. In this paper, we emphasize that granular causality is an interesting and unexplored type of causality in natural language. We compare causal structures identified automatically using causal markers, to granular causality identified automatically using part-whole relations. Causality is not always granular, but granularity always contains causal relations.


Figure 2: Sequential Causal relations extracted using lexico-syntactic causal markers

We use this concept with the granularity identification relations to identify causal relations from discourse.

## 3 Approaches for Automatic Granularity Extraction

Section 2 describes the theory of granularity. This section describes the different approaches that can be used to apply this theory to natural language for automatically identifying sentences that have a shift in granularity. We describe two approaches for granularity identification: Shallow surface level approach for mapping part-whole and causal relations from a knowledge base to natural language discourse; Deep semantic approach using an abductive framework for mapping part-whole and causal relations to natural language discourse, with possibilities of incorporating commonsense axioms into the system. Both the approaches follow Algorithm 1.

### 3.1 Shallow Surface level Techniques

Surface level techniques use simple and shallow syntactic structures such as parse trees to identify entities and events[1]. These events and entities are lexically mapped to part-whole relation lists from background knowledge base to create a granularity structure from the text. For instance, consider the following sentence:

*Elvis Grbac ran 73 yards to complete an 81-yard touchdown play to give the San Francisco 49ers a lead 61 seconds into the game.*

Table 1 shows the part-of and causal relations used as

[1]Nouns are usually entities and verbs usually represent events, with the exception of event nouns

**Algorithm 1** Algorithm for Automatic Discovery of Causal Granularity Structures

```
Obtain part of relations for events (Pev_1,Wev_1 − Pev_n,Wev_n) and entities (Pen_1,Wen_1 − Pen_n,Wen_n)
for all Article A_n do
    Obtain sentences ( S_1...S_m ) in Article A_n
    for all Sentence S_i in A_n do
        for all (Pev_k,Wev_k), k = 1 to x do
            if Pev_k ∈ S_i and Wev_k ∈ S_i then
                for all (Pen_q,Wen_q), q = 1 to y do
                    if Pen_q ∈ S_i and Wen_q ∈ S_i then
                        Inference: S_i contains causal relations between the sentence fragments S_i^1 and S_i^2
                    end if
                end for
            end if
        end for
    end for
end for
Evaluate the Causal Granularity Relations using Annotations
```

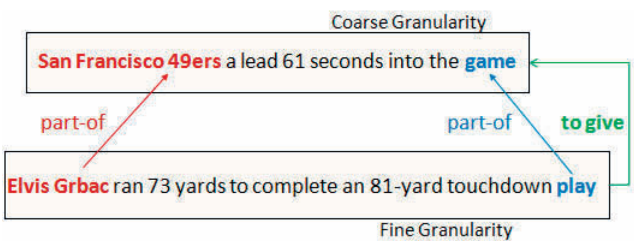


Figure 3: Example of Granularity Structure extraction using shallow techniques

background knowledge. Figure 3 shows the granularity structure extracted from this sentence, by lexically mapping relations from Table 1 and parse structure of the sentence.

## 3.2 Deep Semantic Reasoning

Deep semantic reasoning allows for other forms of reasoning (such as commonsense reasoning) for more detailed analysis and broader coverage to map part-whole relations from the knowledge base for extracting the granularity structure. We use a natural language pipeline that accepts raw texts as input, and converts it into a logical form (Hobbs 1985). We use an abductive inference engine called Mini-TACITUS (Mulkar, Hobbs, and Hovy 2007) that uses the logical form, along with a set of axioms to derive inferences on the input text. This section provides an example of abductive reasoning for granularity identification. For describing the working of the NL pipeline, let us consider the following sentence:

*Chris Kinzer kicked the field goal to give Virginia Tech a victory over North Carolina.*

| Part-Whole relation between entities | |
|---|---|
| **PART** | **WHOLE** |
| Elvis Grbac | San Francisco 49ers |
| **Part-Whole relation between events** | |
| **PART** | **WHOLE** |
| play | game |
| **Causal Connectives** | |
| to give | |

Table 1: Table of Part-Whole and Causal Relations

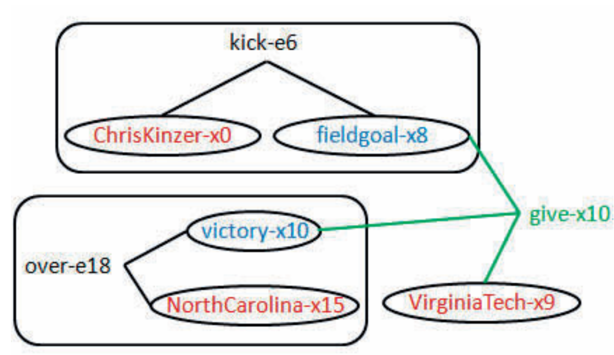


Figure 4: Graphical representation of the Logical Form before Inferencing

In this sentence, we can observe a person level granularity, and an action performed by an individual. There is also a team level granularity and the concept of victory at that granular level. A causal relation exists between these two granularities and is expressed by “to give”.

**Parsing and Logical form conversion**

We use the Charniak parser (Charniak 2000) and then LFToolkit (Rathod and Hobbs 2005) to convert the parse tree into a logical form (Hobbs 1985).

The simplified logical form for the above sentence is shown as follows:

*ChrisKinzer-nn'(e4,x0) & kick-vb'(e6,x0,x8) & fieldgoal-nn'(e13,x8) & give-vb'(e10,x8,x9,x10) & VirginiaTech-nn'(e16,x9) & a'(e19,x10,e15) & victory-nn'(e15,x10) & over-in'(e18,x10,x15) & NorthCarolina-nn'(e20,x15)*

The graphical version of this logical form is shown in Figure 4. The logical form is the input to the Mini-TACITUS system for abductive inference.

**Abductive Inference Engine: Mini-TACITUS**

Mini-TACITUS [2](Mulkar, Hobbs, and Hovy 2007) attempts to find the best possible explanation for the content of the sentence, given a set of general axioms as the knowledge base. A small set of axioms is shown in Table 2. Axioms

[2] *http://www.rutumulkar.com/download/TACITUS/tacitus.php*

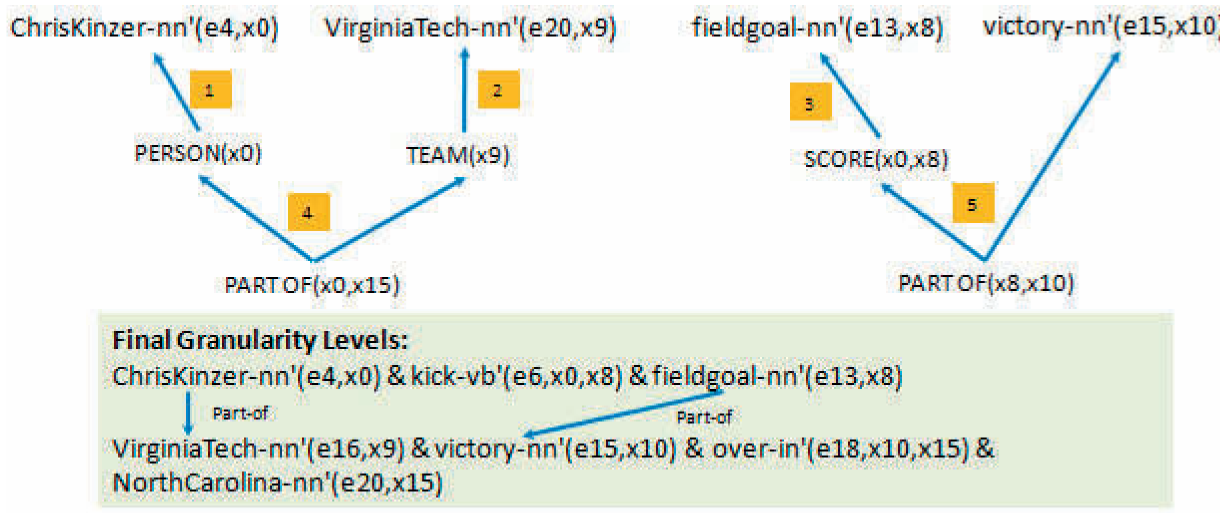


Figure 5: Deep Inferencing for Granularity Level Identification. The numbers next to the arrows represent the axiom numbers from Table2.

| | |
|---|---|
| 1 | PERSON(x1) → ChrisKinzer-nn'(e1,x1) |
| 2 | TEAM(x1) → VirginiaTech-nn'(e1,x1) |
| 3 | SCORE(x1,x2) → fieldgoal-nn'(e1,x1) |
| **Part-Whole relation between events** | |
| 4 | PART-OF(x0,x3) → PERSON(x0) & TEAM(x3) |
| **Part-Whole relation between entities** | |
| 5 | PART-OF(x1,x2) → SCORE(x1,x2) & victory-nn'(e1,x2) |

Table 2: Sample axioms for Mini-TACITUS

1, 2 and 3 are instantiation axioms. Axiom 1 can be read as, *if x1 is a person, one possibility is that the x1 is Chris Kinzer*. Axiom 4 says that *if there is a person and a team, the person might be a part of the team*[3]. Similarly Axiom 5 says that *if there is a scoring event and a victory event, scoring might be a part of the victory*.

Figure 5 shows the process of abduction using the Table 2 for inference. Axioms 1, 2 and 3 are applied first, to infer *Person*, *Team* and *Score*. Axioms 4 and 5 introduce *part-of* relations between the *Person and Team* and *score and victory* respectively. The final set of granularity levels obtained from the sentence after inferencing, with the two granularity levels connected by *part-of* relations. This process becomes more complex as the axiom numbers increase.

## 4 Experimental Details

The objective was to compare the performance of causality relation detection using only granularity features as opposed to using pure causal connectives/indicators. We performed the experiments by implementing the following systems:

[3]The propositions person, team, score etc. are written in CAPS as they are domain theory propositions and not propositions from the text. Case sensitivity is one way to distinguish between these propositions

1. Causality detection using granularity shift identification
   (a) Surface level techniques
   (b) Deep semantic reasoning
2. Causal relation detection using causal markers
   (a) Domain independent causal markers
   (b) Domain dependent causal markers

From here on, we will refer to each of these systems by their corresponding number.

### 4.1 Corpus details

We performed our experiments on a part of the LDC - *New York Times Annotated corpus* (LDC2008T19A) that describes football games. We selected the first 31 articles describing football games. There were a total of 544 sentences in this corpus. Section 5 describes the reasons why a larger corpus could not be used.

### 4.2 Background knowledge-base

For experiments 1(a) and 1(b) a background knowledge-base containing part-whole relations was required. We use meronymic relations described in (Winston, Chaffin, and Herrmann 1987) as the part whole relations between events and between entities in our granularity model. A detailed description for this selection can be found in (Mulkar-Mehta, Hobbs, and Hovy 2011). (Winston, Chaffin, and Herrmann 1987) introduce six types of part-whole relationships of which the Feature-Activity (*paying* is part of *shopping*) type relation is used as the part-whole relation for events, and the rest are part-whole relations for entities. As most of the entity part-whole relations in football are player-team relations (member-collection from (Winston, Chaffin, and Herrmann 1987)), we extracted the list of team-player relations from *http://databasefootball.com*. Example relation instances include – *William Floyd is a*

*Player*, *San Francisco 49ers is a Team*, *Player is part of a Team*. For events, there does not exist a dictionary of part-whole relations that can be extracted from the web. To get event part-whole relations, 10 articles were studied (∼100 sentences), and a list containing 95 domain specific sub event relationships was manually developed. Example relation instances include – *touchdown is part of game*, *game is part of a season*.

For experiments 2(a) and 2(b) a background knowledge base containing domain specific and domain independent causal cue words was required. The domain independent causal cue words were obtained from the causal QA system developed by (Prager, Chu-Carroll, and Czuba 2004) for TREC QA. The domain dependent causal cue words were extracted by studying 10 articles (∼100 sentences) and manually creating lists of football domain specific causal cue words such as - *gave*, *lead to*, *set up* and so on.

### 4.3 Gold Standard

A gold standard was created on the union of all the sentences that were marked to have a positive causality by any of the above systems. All the sentences in the dataset were independently annotated by 2 annotators. Each annotator was asked to judge whether the given sentence contained a causal relation, and was asked to mark the causal cue words in the sentence. The Kappa agreement (Cohen 1960) between the two annotators was 0.86. The annotations from the primary annotator were taken as the gold standard for evaluation.

## 5 Results and Analysis

The precision and recall of systems 1(a), 1(b) and 2(b) are shown in Figure 6. System 2(a) which used the domain independent causal cue words performed extremely poorly, with a precision of 0.25. This showed that causal markers are very domain specific, and differ largely with the domain selected. For instance, the causal markers in a biomedical domain such as (Mulkar-Mehta et al. 2009) are completely disjoint from the causal markers in the football domain.

Both the granular causality identification systems (systems 1(a) and 1(b)) outperformed the pure causal extraction systems (systems 2(a) and 2(b)) with a precision and recall of 0.92 and 0.79 respectively for the surface level system 1(a) and precision and recall of 0.89 and 0.60 respectively for the deep semantic system 1(b). The precision and recall of the system 2(b) using domain causal markers was 0.79 and 0.44 respectively. System 2(b) had a lower recall than 1(a) and 1(b) as these systems marked sentences which were missed by 2(b). Following are a few examples:

*Derek Loville added a 19-yard touchdown catch and a one-yard touchdown run in the second quarter as the San Francisco 49ers rolled to a 31-7 half-time edge*

Here causality is represented by the word "as".

*The Miami Dolphins went ahead 21-6 at halftime behind three touchdown throws by Dan Marino, who found Keith Jackson twice and gave seldom-used Mike Williams his first touchdown in four seasons with the Miami Dolphins*

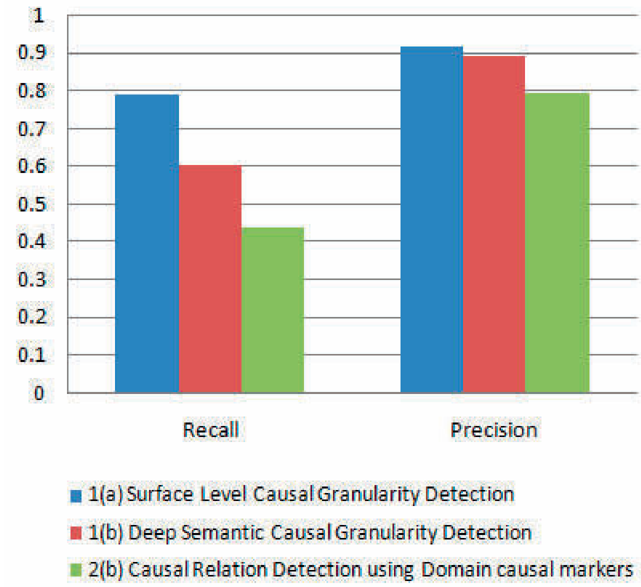


Figure 6: Accuracy for detection of causality

Here causality is conveyed by the word "behind".

There were two major reasons for low precision of system 2(b). First, the parser sometimes incorrectly parsed "lead" as a verb, when it was a noun in the sentence. Such sentences were incorrectly identified to have a causal relation. The following sentence is an example:

*Nogh seized a 10-0 lead in the first quarter on Jeff Reed's 45-yard field goal and a three-yard Jerome Bettis touchdown run but the New York Jets answered on Doug Brien's 42-yard field goal and a 75-yard punt return touchdown by Santana Moss.*

Another reason for low precision of system 2(b) was that the domain specific causal verbs often had other meanings besides causality.

### 5.1 Limitations of the Reasoning Engine

It was unexpected that the deep semantic reasoning system 1(b) would be outperformed by the surface level system 1(a). Upon inspection, we found the limitations of the reasoning engine (Mini-TACITUS) (Mulkar, Hobbs, and Hovy 2007) which were blocking inferencing. One of the major problems we faced with Mini-TACITUS was that the system did not allow backchaining on the same proposition more than once. This created issues in cases where a sentence contained mentions of a team and multiple players in the team. Mini-TACITUS was able to apply only one of the axioms, and often marked the sentence to not have a granularity relation. Another issue with Mini-TACITUS was that it did not support loopy axioms. This meant that one could not write axioms such as $a \rightarrow b$ and $b\rightarrow a$ in the same axiom set.

### 5.2 Issues with Mapping Part-Whole relations

Mapping part-whole relations to the entity mentions in the discourse proved to be a very complex task. Although we were able to extract the part-whole relations for entities from *http://databasefootball.com*, the entities extracted from the website were all in their canonical form i.e. all the players and teams were mentioned by their full name. However this is rarely the case in discourse mentions of player or team names. For instance, the entity *William Ali Floyd*, was mentioned as *William Floyd* or *Floyd* in discourse. In some scenarios entities had a preferred name which was often different from their given names. For instance the entity *David Michael Brown* has a preferred name of *Dave Brown*. This made part-whole mapping more challenging as current co-reference solutions are not usable for this task. Similar was the case for team names. For instance *San Francisco 49ers*, *49ers* and *San Francisco* are all used interchangeably.

One of the solutions for team names was to mark the location and team name as a team and then apply co-reference and select the largest fragment as the canonical form for the team. However, the canonical form is not always mentioned in the article. For instance consider the following paragraph:

*San Francisco moved ahead 7-3 on Floyd's two-yard touchdown run 11:19 into the game. Merton Hanks returned an interception 31 yards to set up a 36-yard Doug Brien field goal that gave the 49ers a 23-3 lead.*

This sequence of sentences mentions *San Francisco* and *49ers* but not the canonical form of the entity *San Francisco 49ers*. In order to simplify the downstream process, we manually converted all the proper names into their canonical form. As a result, we are not currently able to scale this to a larger corpus.

## 6 Conclusions

In this paper we describe a theory of granularity as it occurs in natural language text. We use this theory for identification of sentences containing causal relations. We compare this with a system that identifies causal relations using causal markers only. Our granularity based system outperforms the causal markers based system in precision as well as recall. This provides strong evidence that causality is not always sequential in nature and can often have a granular structure, which is a theory that has been largely overlooked.

## Acknowledgements

This research was supported by the Defense Advanced Research Projects Agency (DARPA) Machine Reading Program under Air Force Research Laboratory (AFRL) prime contract no. FA8750-09-C-0172. Any opinions, findings, and conclusion or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the DARPA, AFRL, ONR, or the US government.

## References

Blanco, E.; Castell, N.; and Moldovan, D. 2008. Causal Relation Extraction. *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)* 310–313.

Charniak, E. 2000. A Maximum Entropy Inspired Parser. *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* 132 – 139.

Cohen, J. 1960. A coefficientof agreement for nominal scales. *Educational and Psychological Measurement*.

Girju, R., and Moldovan, D. 2002. Mining Answers for Causation. *Proceedings of American Association of Artificial Intelligence* 15–25.

Hobbs, J. R. 1985. Ontological Promiscuity. *Proceedings of the 23rd annual meeting on Association for Computational Linguistics* 60–69.

Mazlack, L. 2004. Granular causality speculations. *IEEE Annual Meeting of the Fuzzy Information, 2004. Processing NAFIPS '04.* 690–695.

Mulkar, R.; Hobbs, J. R.; and Hovy, E. 2007. Learning from Reading Syntactically Complex Biology Texts. *Proceedings of the AAAI Spring Symposium Commonsense'07. Stanford University, CA*.

Mulkar-Mehta, R.; Hobbs, J. R.; Liu, C.-C.; and Zhou, X. J. 2009. Discovering Causal and Temporal Relations in Biomedical Texts Recognizing Causal and Temporal Relations :. *Proceedings of the AAAI Spring Symposium, Standord CA*.

Mulkar-Mehta, R.; Hobbs, J. R.; and Hovy, E. 2011. Granularity in Natural Language Discourse. *International Conference on Computational Semantics, Oxford, UK* 360—-364.

Pechsiri, C., and Kawtrakul, A. 2007. Mining Causality from Texts for Question Answering System. *IEICE Transactions on Information and Systems* E90-D(10):1523–1533.

Prager, J. M.; Chu-Carroll, J.; and Czuba, K. 2004. A Multi-Strategy, Multi-Question Approach to Question Answering. *In New Directions in Question-Answering, ed. M. Maybury. Menlo Park, CA: AAAI Press.*

Rathod, N., and Hobbs, J. R. 2005. LFToolkit.

Winston, M. E.; Chaffin, R.; and Herrmann, D. 1987. A Taxonomy of Part-Whole Relations. *Cognitive Science* 11(4):417–444.