

Granularity in Natural Language Discourse

Rutu Mulkar-Mehta¹

Information Sciences Institute
University of Southern California
me@rutumulkar.com

February 19, 2011



¹Advisors: Jerry Hobbs (ISI), Eduard Hovy (ISI), Christopher Welty (IBM)

Learning by Reading and its Applications in Intelligent Question-Answering

March 14, 2011 - Paper Submission

April 25, 2011 - Acceptance Notification

May 16, 2011 - Camera ready paper due Website:

<http://www.rutumulkar.com/FAM-LbR-KRAQ-2011.php>

Location: Downtown Barcelona

Motivation

The heart is a muscular organ that pumps blood through the body.

Motivation

The heart is a muscular organ that pumps blood through the body.

Oxygen-poor blood enters the right atrium which is then pumped into the right ventricle.

The blood then moves through the pulmonary artery to the lungs, where the blood is enriched with oxygen.

The oxygen-rich blood is then carried back to the left atrium.

The blood is then pumped to the left ventricle, then the blood is pumped through the aorta and to the rest of the body.

Motivation

The **heart** is a muscular organ that pumps blood through the body.

Oxygen-poor blood enters the **right atrium** which is then pumped into the **right ventricle**.

The blood then moves through the pulmonary artery to the lungs, where the blood is enriched with oxygen.

The oxygen-rich blood is then carried back to the **left atrium**.

The blood is then pumped to the **left ventricle**, then the blood is pumped through the aorta and to the rest of the body.

Motivation

The **heart** is a muscular organ that **pumps** blood through the body.

Oxygen-poor blood **enters** the **right atrium** which is then **pumped** into the **right ventricle**.

The blood then **moves** through the pulmonary artery to the lungs, where the blood is enriched with oxygen.

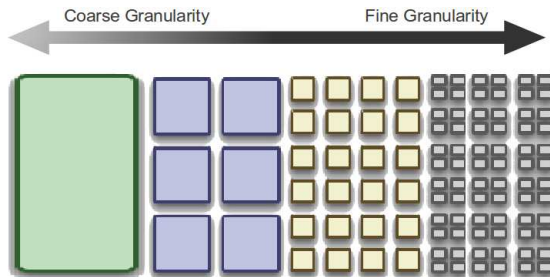
The oxygen-rich blood is then **carried** back to the **left atrium**.

The blood is then **pumped** to the **left ventricle**, then the blood is **pumped** through the aorta and to the rest of the body.

What is Granularity?

Definition

Granularity: the level of detail of description of an event or object.



Domain	Domain elements in decreasing granularity			
Software	Services	Modules	Packages	Classes
Football	Season	Game	Drive	Play
Shopping	Shopping	Travel to Store	Walk	Move a foot
Cleaning	Clean a house	Clean a Room	Dust a cabinet	Dust a shelf

Granular Causality

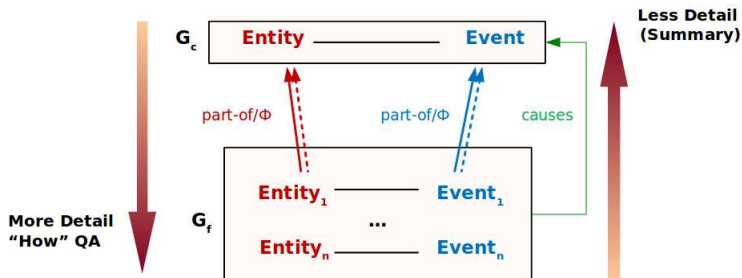
There are 2 types of causality:

- Sequential Causality: *The building collapsed because of the earthquake*
- Granular Causality: *The building collapsed because the roof caved in*

This talk focuses on granular causality, and how identification of granularity structure in text can help infer causal relations.

Theory of Causal Granularity in Natural Language Discourse

The following is a visual description of the theory of granularity in Natural Language Text



Real Examples of Granularity in Natural Language Descriptions

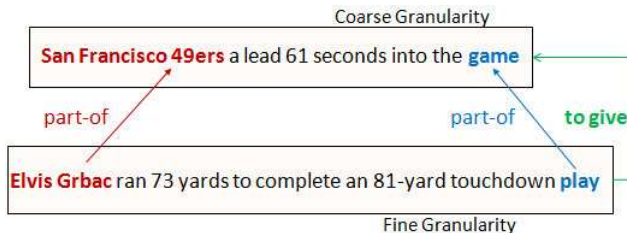
Consider the following sentence:

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.

Real Examples of Granularity in Natural Language Descriptions

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.



Real Examples of Granularity in Natural language Descriptions

How did the WTC *collapse*?

As the joists on one or two of the most heavily burned floors *gave way* and the outer box columns began to *bow outward*, the floors above them also *fell*. The floor below could not support the roughly 45,000 t of ten floors (or more) above *crashing down* on these angle clips. This started the domino effect that *caused* the building to *collapse* within ten seconds

Understanding Granularity Relations

Part-of Relations (Winston et al. 1987)	
Category	Example
Component-Integral	<i>pedal - bike</i>
Member-Collection	<i>ship - fleet</i>
Portion-Mass	<i>slice - pie</i>
Stuff-Object	<i>steel - car</i>
Feature-Activity	<i>pay - shop</i>
Place-Area	<i>LA - USA</i>

Understanding Granularity Relations

Part-of Relations (Winston et al. 1987)	
Category	Example
Component-Integral	<i>pedal - bike</i>
Member-Collection	<i>ship - fleet</i>
Portion-Mass	<i>slice - pie</i>
Stuff-Object	<i>steel - car</i>
Feature-Activity	<i>pay - shop</i>
Place-Area	<i>LA - USA</i>

Causal Relations (Girju et al. 2002)		
Category	Type	Example
Causal Connectives	Prepositional	<i>because of, thanks to, due to</i>
	Adverbial	<i>for this reason, the result that</i>
	Clause links	<i>because, since, for</i>
Causation Verbs		<i>kill, melt</i>
		<i>poison, hang</i>
Conditionals		<i>If S1 then S2.</i>

Evaluation of Causal Granularity Theory

Sure! You can propose a fancy theory!
but is it real?

Annotation Study for Granularity Theory Evaluation

The following are the details for the Annotation Study conducted:

- **Articles:** 37 articles from three domains
(Travel (confluence.org), Timebank (LDC2006T08), Games (Wikipedia.org))
- **Annotators:** 5 People from Mechanical Turk

Annotation Study for Granularity Theory Evaluation

The following are the details for the Annotation Study conducted:

- **Articles:** 37 articles from three domains
(Travel (confluence.org), Timebank (LDC2006T08), Games (Wikipedia.org))
- **Annotators:** 5 People from Mechanical Turk
- **Annotation Guidelines:**
 - 1 Is one paragraph a subevent of the other paragraph?
 - 2 Did one paragraph cause the other paragraph?
 - 3 Is one paragraph less detailed and the other paragraph more detailed?
 - 4 Did one paragraph happen after the other paragraph?

Example Task

Consider the following two paragraphs:

- 1: *Officials in California are warning residents of dangerous and unpredictable landslides.*
- 2: *Experts say the ground is so saturated it cannot absorb any more water. So that means soil will fall off in chunks and destroy anything in its path.*

Example Task

Consider the following two paragraphs:

1: *Officials in California are warning residents of dangerous and unpredictable landslides.*

2: *Experts say the ground is so saturated it cannot absorb any more water. So that means soil will fall off in chunks and destroy anything in its path.*

Agreement:

para2 is a subevent of para1

para2 causes para1

para2 is more detailed than para1

para2 happens before or at the same time as para1.

Example Task

Consider the following two paragraphs:

1: *Officials in California are warning residents of dangerous and unpredictable landslides.*

2: *Experts say the ground is so saturated it cannot absorb any more water. So that means soil will fall off in chunks and destroy anything in its path.*

Agreement:

para2 is a subevent of para1

para2 causes para1

para2 is more detailed than para1

para2 happens before or at the same time as para1.

Verbatim comments from Annotators:

.. "Without the rain or loose soil, landslide warnings don't happen" ..

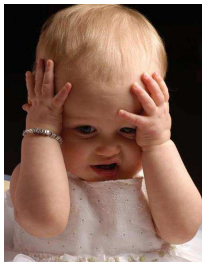
.. "Para2 will result in Para1" ..

Annotation Evaluation

Annotation Agreement:

Two annotators were considered to be in agreement if they agreed with questions 1, 3 and 4.

“Causality” was a confusing feature



1: I wanted to visit the confluence point located in the extreme southwest of Hunan Province.

2: To get to the confluence, I caught the Hong Kong-to-Shanghai intercity train on Friday afternoon.

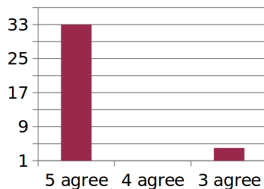
Does 1 cause 2 or 2 cause 1?

Does goal cause event or events cause goal?

All the annotators agreed that a sub-event explains how an event happens, or a sub-event causes an event. We counted this in lieu of our causality question (2)

Agreement Analysis

Annotator ID	A2-K9	A2-CA	A2-GH	A3-CN	A3-FM	Average
A2-K9		0.82	0.76	1	0.76	0.833314
A2-CA	0.82		0.93	0.82	0.93	0.876835
A2-GH	0.76	0.93		0.76	1	0.861683
A3-CN	1	0.82	0.76		0.76	0.833314
A3-FM	0.76	0.93	1	0.76		0.861683



The average pairwise Kappa (Cohen 1960): 0.85

Example case of disagreement

1: Some 1,500 ethnic Albanians marched Sunday in downtown Istanbul, burning Serbian flags.

2: The police barred the crowd from reaching the consulate, but allowed them to demonstrate nearby.

Yes! Granularity Shift: “demonstrations” are a part of a “march”

No! Granularity Shift: “demonstrations” happened after the “march”



Ok! But I need more convincing!

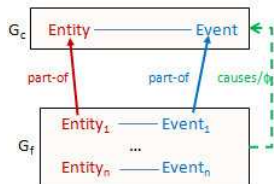
Algorithm

- 1: Obtain part of relations for events ($Pev_1, Wev_1 - Pev_n, Wev_n$) and entities ($Pen_1, Wen_1 - Pen_n, Wen_n$)
- 2: **for all** Article A_n **do**
- 3: Obtain sentences ($S_1...S_m$) in Article A_n
- 4: **for all** Sentence S_i in A_n **do**
- 5: **for all** (Pev_k, Wev_k), $k = 1$ to x **do**
- 6: **if** $Pev_k \in S_i$ and $Wev_k \in S_i$ **then**
- 7: **for all** (Pen_q, Wen_q), $q = 1$ to y **do**
- 8: **if** $Pen_q \in S_i$ and $Wen_q \in S_i$ **then**
- 9: Inference: S_i contains causal relations between the sentence fragments S_i^1 and S_i^2
- 10: **end if**
- 11: **end for**
- 12: **end if**
- 13: **end for**
- 14: **end for**
- 15: **end for**
- 16: Evaluate the Causal Granularity Relations using Annotations

Experiments

1 Objective

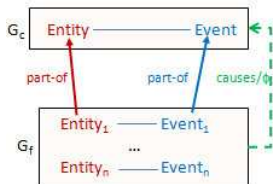
1 Causality detection using granularity shift identification



Experiments

1 Objective

1 Causality detection using granularity shift identification



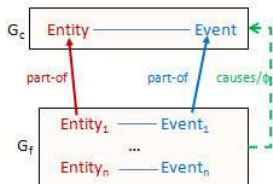
2 Causal relation detection using causal markers



Experiments

1 Objective

1 Causality detection using granularity shift identification



2 Causal relation detection using causal markers



2 Corpus: 31 articles of LDC - *New York Times Annotated corpus* (LDC2008T19A) that describes football games.

3 Sentences: 544 sentences

Experiment 1: Using Part of Relations to Discover Causality

Entities	
PART	WHOLE
Elvis Grbac	SF 49ers
William Floyd	SF 49ers

Events	
PART	WHOLE
play	game
touchdown	game

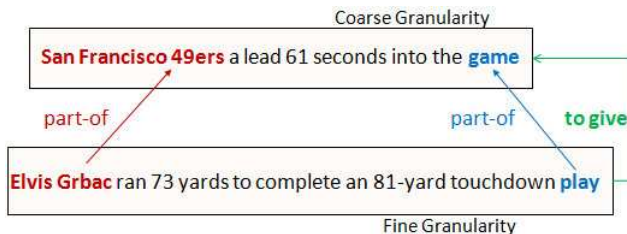
Elvis Grbac ran 73 yards to complete an 81-yard touchdown play to give the San Francisco 49ers a lead 61 seconds into the game.

Experiment 1: Using Part of Relations to Discover Causality

Entities	
PART	WHOLE
Elvis Grbac	SF 49ers
William Floyd	SF 49ers

Events	
PART	WHOLE
play	game
touchdown	game

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.



Experiment 2: Using Causal connectives to Discover Causality

Cause
lead to
give
cause
drive

John Carney's conversion kick
gave the San Diego Chargers
their first lead.

Experiment 2: Using Causal connectives to Discover Causality

Cause
lead to
give
cause
drive

John Carney's conversion kick
gave the San Diego Chargers
their first lead.

John Carney's conversion kick
(CAUSES)
the San Diego Chargers their first lead.

Gold Standard

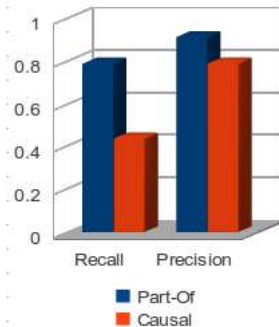
A gold standard was created on the union of all the sentences that were marked to have a positive causality by any of the above systems.

Annotators: 2 annotators.

Guidelines:

- Does the sentence contain a causal relation?
- Mark the causal cue words in the sentence

Kappa Agreement (Cohen 1960): 0.86.



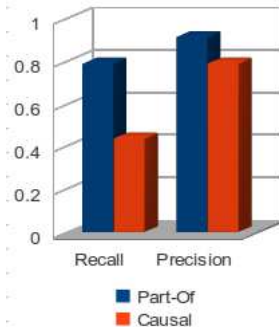
Prepositions are inherently ambiguous in nature:

*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

Here causality is conveyed by the word “behind”.



Prepositions are inherently ambiguous in nature:

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*

Here causality is conveyed by the word “behind”.

Data Collection

Part of Relations

- Part-of Relations for Entities: team-player relations from <http://databasefootball.com>
- *William Ali Floyd is part of San Francisco 49ers*
- Part of Relations for Events: 10 articles were studied (~100 sentences), 95 sub event relationships developed.
- *touchdown is part of game*

Data Collection

Part of Relations

- Part-of Relations for Entities: team-player relations from <http://databasefootball.com>
- *William Ali Floyd is part of San Francisco 49ers*
- Part of Relations for Events: 10 articles were studied (~100 sentences), 95 sub event relationships developed.
- *touchdown is part of game*

Causal Cue Words

- Domain independent causal cue words: Obtained from the causal QA system developed by Prager et al. 2004 for TREC QA
- *because, cause*
- Domain dependent causal cue words: Obtained by studying 10 articles (~100 sentences).
- *gave, lead to, set up*

Future Work

- Automatic Extraction of Part-Of Relations (Entities and Events)
- Using Granularity structures for answering “How” style questions
- Coreference problem for named entities
- Differentiating between Granularity and Causal Granularity

Conclusion

- Proposed a theory of Causal Granularity
- Human Experiment to judge the features of Causal Granularity
- Use Part-of Relations to extract Causal Relations

Thank you!

