

# *Penn Discourse Treebank Relations and their Potential for Language Generation*

*Kathleen McKeown  
Department of Computer Science  
Columbia University*



**COLUMBIA | ENGINEERING**  
The Fu Foundation School of Engineering and Applied Science



Our work in this area is partly supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20153

# Generation, Summarization and PTDB relations

- PTDB relations have rarely been used for generation and summarization: *WHY?*
- They could replace the use of Rhetorical Structural Theory (RST) relations: *HOW?*
- What is needed to make this happen?

# Outline

- Few examples of how PTDB relations have been used
  - Focused summarization/ open-ended QA (Sasha Blair Goldensohn et al)
  - Summarization (Louis et al)
- RST for language generation
- Speculation on what is needed

# Focused Summarization or Open-ended Question Answering

Blair-Goldensohn, McKeown, Rambow 2007

- Cause and contrast queries
  - Describe [*causes*] of [conviction] in [Edmond Pope convicted for espionage in Russia]
  - Describe [*effects*] of [conviction] in [Edmond Pope convicted for espionage in Russia]
  - Compare [ Korea ] and [ France ] with respect to [ economic outlook]
- PTBD relations improve content selection

# Applying PTDB

- Open-ended QA is a text-to-text generation task
  - Must be able to identify relations in input text
- Blair-Goldensohn et al used PTDB data to develop a classifier to recognize causal and contrastive relations
  - Based on lexical cues
  - Along with topic segmentation, cue word patterns
- Relation/topics represented as lexical pairs

# Application during content selection

- *Goal*: select sentences that express relation in sentence
- *Explicit relations*: Apply patterns using cue words to find sentences
- *Implicit relations*: Use lexical pairs to augment traditional key word matching

# An Example

- Describe [causes] of [conviction] in [Edmond Pope convicted for espionage in Russia]
- Keywords: conviction, edmond, pope, convicted, espionage, russia
- Cue words: accordingly, because, causing
- Relation lexical pairs: juror, confess, inproperli, insuffici, testimoni, statut, wit, limit, mental, dna, prove, ... , arrest, inform, fail



# Relevant sentences

- Extracted by key words and cue words:
  - Pope was convicted for espionage because of eyewitness testimony which proved the case.
- Missed by key words and cue words:
  - Eyewitness testimony proved the prosecution case against Pope.
- Irrelevant but caught by key and cue words:
  - Because of the holiday, the decision on whether to convict Pope will wait until Friday

# Lexical pairs for cause and foci

- PDTB relation lexical pairs link cause and topic
- Find words linked by “cause” to “conviction”
- Relevant:
  - Eyewitness testimony proved the prosecution case against Pope.

# Other lexical pairs for cause

## Focus Term

## RSR terms suggested for cause question

environmental

Hurt, fear, oil, leak, contamin, gasolin, water, land, tanker, threat, lack, cool, coast, 1989

regulation

Interest, supplement, toxic, competit, act, combin, commerc, statut, concern, effect, heavi, contamin, problem, lack

Environmental regulation

Interest, supplement, toxic, hurt, fear, contamin, lack, oil, competit, threat, heavi, pose, concern, leak

# PTDB relations useful for QA content selection

- A relation must be paired with a question (contrast with compare and contrast questions)
- Individual sentences selected: what about relations across sentences?
- Issue: representation of relations as lexical pairs

# Summarization

- Discourse structure should impact summarization (e.g. Sparck Jones 1998)
  - Impact on content selection
  - Impact on summary structure
- Louis and Nenkova experimented with different kinds of discourse features vs. non-discourse
  - Overall structure
  - Sense relations (labels such as cause, contrast, elaboration)

# Discourse Features used

- Overall tree structure of the document
  - RST
  - Depth scores, promotion scores
- Overall graph structured of the document
  - Wolf and Gibson
- Semantic relations between sentences
  - PDTB relations
  - Arg number of specific relations (e.g., arg 1 of implicit contingency)

# Non-discourse features

- Sentence position
- Sentence length
- Paragraph-initial or final document sentence?
- Average, sum and product probabilities of content words in the sentence
- Number of topic signature words in the sentence

# Summarization model

- Extractive, single document summarization
- WSJ news articles
- Classification using learning of sentences as in summary or not



# Data

- Collection of WSJ articles
- Annotated for all three types of discourse features
- RST corpus provides two human created summaries for 150 documents
  - Caveat: Mapped EDUs to sentences
  - Selected 124 documents for which good sentence boundaries
- 15 Summaries from GraphBank

# Classification results

- Structural features outperform other features
  - Graph-based structure better than RST
- Semantic relations give little improvement to content selection; non-discourse better
- Combination of all three give best results

<b>Features used</b>	<b>Acc</b>	<b>P</b>	<b>R</b>	<b>F</b>
structural	78.11	63.38	22.77	33.50
semantic	75.53	44.31	5.04	9.05
non-discourse (ND)	77.25	67.48	11.02	18.95
ND + semantic	77.38	59.38	20.62	30.61
ND + structural	78.51	63.49	26.05	36.94
semantic + structural	77.94	58.39	30.47	40.04
structural + semantic + ND	78.93	61.85	34.42	44.23

Table 1: Accuracy (Acc) and Precision (P), Recall (R) and F-score (F) of important sentences.

# Disappointing News

- Compared models using discourse structure to graph-based models
  - Graph is induced using lexical similarity between sentences
  - Page-rank metrics determine salient sentences (Erkan and Radev 2004, Mihalcea and Tarau 2005)
- Graph model based on lexical similarity even more helpful than discourse
  - 53% F-score vs 42% (RST) or 48% (GB)

# PDTB not useful for summarization?

- Extractive, single document summarization of news
  - Baselines for news summarization hard to beat
- Abstractive
  - Selecting parts of sentences, rewriting sentences
- Louis and Nenkova suggest it could be useful for organizing summaries
  - Readability of summaries still an issue

# Rhetorical Structure Theory and Language Generation

- Used for content selection
- Connects intentions with RST relations
- Uses planning to reason about how RST relations can produce desired effects
- Selection of RST relations causes selection of content

# Example (Moore and Paris)

*Remove the cover. You'll need the Phillips screwdriver. It's in the top drawer of the toolbox. Do you have it?*

- Intention: increase the user's ability to perform a task -
  - enablement relation
  - Tell the user what tool to use
- Intention: help the user find the tool -
  - circumstance relation
  - Tell the location of the tool

# Other Uses of Relations

- Map combinations of relations to recipes, much like schema
- Once planned, full structure is available
  - Select connectives
  - Map to sentence structure
- Without intention
  - Map combinations of information to likely relations
  - Aggregation commonly used for comparison
    - *Canines and felines have backbones*



# New work at Columbia: Explanation of Prediction

- Task: prediction of technical emergence from scientific journals
  - A machine learning approach
- Explain why system predicted what it did
  - Machine learning output as input
- Exploring discourse relations to
  - Aggregate input
  - Overall discourse structure

,

RNAi : 2006-2010 : CQ1

---

## Was there a community of practice around RNAi during 2006-2010?

The answer is **YES**, with a confidence of **72%**

---

### Topic Summary

RNA interference (RNAi) is an RNA-dependent gene silencing process within living cells. The selective and robust effect of RNAi on gene expression makes it a valuable research tool, both in cell culture and in living organisms.

### Justification and evidence for answer

Many indicators suggest a positive answer to the CQ, especially within the [Coauthorship Graph](#), [Time Series](#) and [Funding](#) groups (see below).

#### Coauthorship Graph

[Click for detailed view](#)

The [coauthorship graph](#) for RNAi spans 520 authors, and it has the properties of a small-world network which is typical of real-world communities. It is a [fully connected network](#) with a high [clustering coefficient](#) as well as a reasonably high [power law coefficient](#).

Coauthorship graph indicators are the most powerful when determining the answer for CQ1, and in this case their values strongly point in the

RNAi : 2006-2010 : CQ1

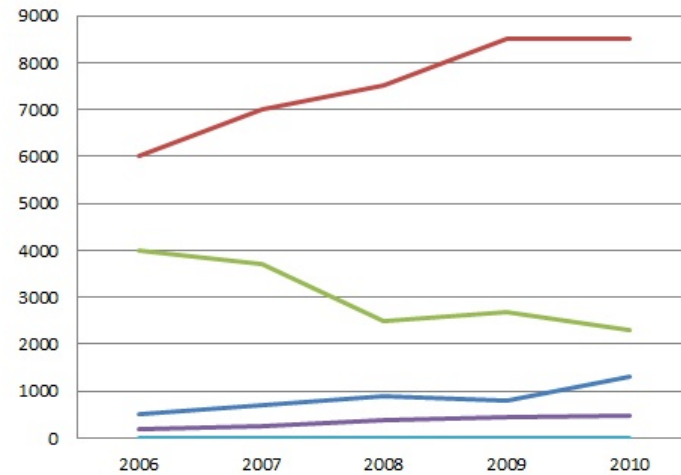
## Time series

A time series is a sequence of data points measured at successive time instants spaced at uniform time intervals (in our case, years). Time series analysis looks at the way that various functions of the RDG behave over time.

The time series for number of papers and number of unique authors have very high slopes, indicating a lively community of practice which evolves over time. The number of in-citations and out-citations also progress visibly over time.

Time series indicators are very important when determining the answer for CQ1. In this case, they clearly suggest a positive answer.

Time series: RNAi



	2006	2007	2008	2009	2010
papers	500	700	900	800	1300
in-citations	6000	7000	7500	8500	8500
out-citations	4000	3700	2500	2700	2300

# Position

*Lack of work on relations in the field makes it hard for people to adopt*

- *Papers:* showing how relations could be used for generation
- *Explicit comparison:* Advantage of PDTB relations over RST
- *More data:* Some relations not well covered.

# Discussion

- Why should language generation and summarization researchers work with PDTB relations over RST?
- Is there a better way to use the relations than the small amount of work already there?

# Discussion

- Language Generator
  - Are RST relations used for generation because they always have been?
  - Are they more useful because they provide overall structure which informs mapping to sentences?
  - Is there more data available for RST?
  - Given definitions of PDTB relations, are they suitable for selecting content?

# Discussion

- Summarization
  - Would abstraction make more use of relations and structure than extraction?
  - Would ordering of sentences be helped by semantic relations?

# Discussion

- Question Answering
  - Are relations more obviously applicable because they place constraints on content?
    - (when defined as lexical pairs)
  - Could each relation be considered a question and thus relevant when that question asked?
  - Should relations be combined when answering a question?