

# **Cross-training: Improve discourse parser using data generated by another parser**

Apr 12<sup>th</sup>, 2016

Kailang Jiang

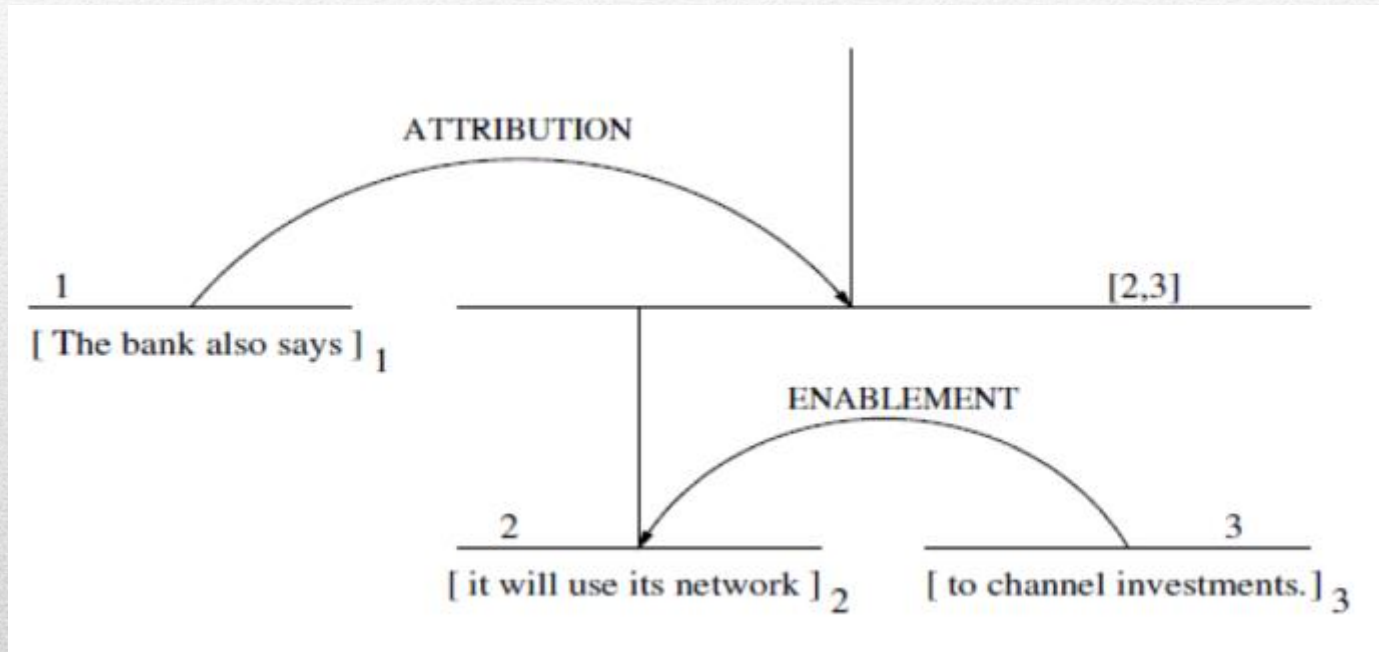
---

- Problem
- Original plan
- Change/Update
- What's done
- In process
- To do

# Overview

---

- Target:
  - Improve existing discourse parsers



# Problem

---

- Problem:
  - Only very few positive training examples for some relations
  - Labeled data are very expensive to get
- Labeled Data:
  - RST-DT (Carlson et al., 2001)
  - PDTB (Prasad et al. 2008)

# Problem

---

- RST-DT

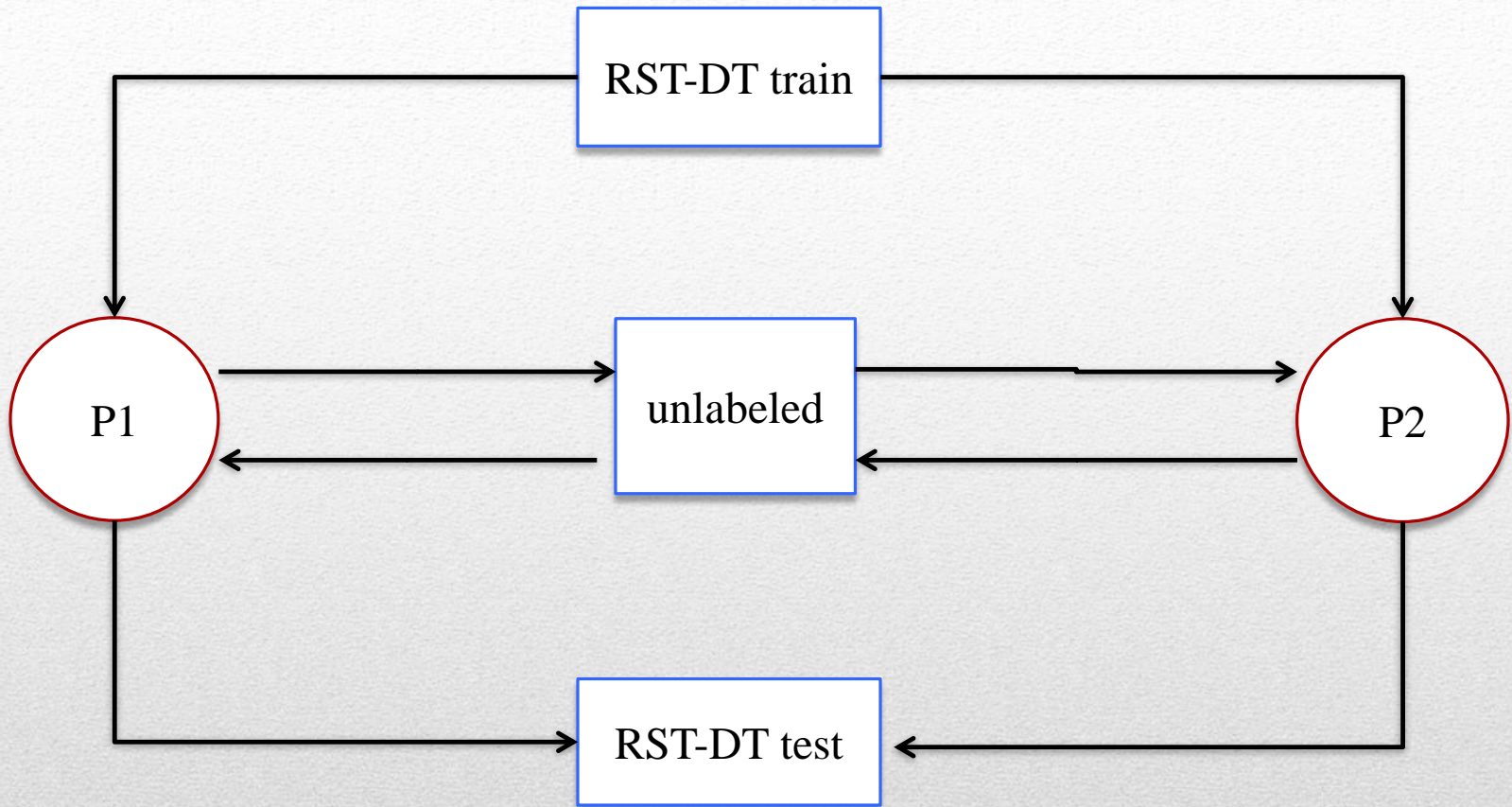
Elaboration	Joint	Attribution	Same-Unit	Contrast	Explanation
Background	Temporal	Cause	Enablement	Comparison	Evaluation
Topic-Change	Topic-Comment	Condition	Textual Organization	Manner-Means	Summary

- PDT

Level-1	Temporal	Contingency	Comparison	Expansion
	Synchrony	Cause	Contrast	Conjunction
	Asynchronous	Pragmatic Cause	Pragmatic Contrast	Instantiation
Level-2		Condition	Concession	Restatement
		Pragmatic Condition	Pragmatic Concession	Alternative
				Exception
				List

# Problem

---



# Original Plan

---

- Analyze how following factors affect parsing results
  - Quality of enriched training data
  - Size of enriched training data
  - Explicit/Implicit enriched training datas
- Analyze the impact of different factors on
  - Frequent/Infrequent relations

# **Change/Update**

---

- Read relevant paper
- Understanding the code of one parser

# **What's done**

---



- Experiments on this parser
  - Including analyzing the effect of different factors
  - Including performance on different relations

**In process**

---

- Read the paper and understand the code of the other parser
- Perform similar experiments on the other parser

# To do

---