An Unsupervised Approach to Recognizing Discourse Relations

Marcu and Echihabi 2002

The Problem

- Discourse relations obviously exist.
- Most sentences require discourse context to be understood.

Example Pairs

1) a. Such standards would preclude arms sales to states like Libya, which is also currently subject to a U.N. Embargo.

b. *But* states like Rwanda before its present crisis would still be able to legally buy arms.

2) a. South Africa can afford to forgo the sales of guns and grenades

b. *because* it actually makes most of its profits from the sale of expensive, high-technology systems like laser-designated missiles, aircraft electronic warfare systems, tactical radios, anti-radiation bombs, and battlefield mobility systems.

Existing Theories

- Lots of theories with fine-grained distinctions
- Not computationally tractable
  - *Require general world knowledge*
  - *Semantic interpretation*
  - *Intentions and* ***illocutions***

Linguistic Side Note

- **Illocutionary Act -** The *intended* meaning of an utterance.
- **Locutionary Act -** The *literal* surface meaning of an utterance.
- **Perlocutionary Act –** The *effect* the meaning of an utterance on the listener.
- *Ex. "Don't play with fire."*

  Loc. - Surface meaning

  Illoc. - Warning

  Perloc. - Persuaded to comply

First Experiment

- They hypothesize that they can determine discourse relation labels without semantic interpretation via word pairs.
- For example, a CONTRAST relation:
  - John is *good* in math and sciences.
  - Paul *fails* almost every class he takes.
- This idea extends to pairs like *embargo* and *legally*
- Can't use WordNet antonym relations because they don't cover relations like *embargo* and *legally*.

Relation Mappings

CONTRAST	CAUSE-EXPLANATION-EVIDENCE	ELABORATION	CONDITION
ANTITHESIS (M&T) CONCESSION (M&T) OTHERWISE (M&T) CONTRAST (M&T) VIOLATED EXPECTATION (Ho) (CAUSAL | ADDITIVE) - (SEMANTIC | PRAGMATIC) - NEGATIVE (K&S)	EVIDENCE (M&T) VOLITIONAL-CAUSE (M&T) NONVOLITIONAL-CAUSE (M&T) VOLITIONAL-RESULT (M&T) NONVOLITIONAL-RESULT (M&T) EXPLANATION (Ho) RESULT (A&L) EXPLANATION (A&L) CAUSAL - (SEMANTIC | PRAGMATIC) - POSITIVE (K&S)	ELABORATION (M&T) EXPANSION (Ho) EXEMPLIFICATION (Ho) ELABORATION (A&L)	CONDITION (M&T)

Table 1: Relation definitions as union of definitions proposed by other researchers (M&T – (Mann and Thompson, 1988); Ho – (Hobbs, 1990); A&L – (Lascarides and Asher, 1993); K&S – (Knott and Sanders, 1998)).

- Conflate fine-grained categories from several DRTs into four labels

Text Spans → Word Pairs

- Cartesian product defined over two text spans (pairwise combinations of words)
- Goal is to maximize probability of relation label given pair of text spans:

$$argmax_{(r_k)} P\left(r_k | W_{1,} W_2\right)$$

- By assuming word pairs are independent:

$$P\left(W_{1,} W_2 | r_k\right) = \prod_{((w_i, w_j) \in W_{1,} W_2)} P\left(\left(w_i, w_j\right) | r_k\right)$$

- $P((w_i, w_j) \mid r_k)$ computed using Maximum Likelihood Estimators smoothed using Laplace method *(allows non-zero probability to be assigned to word pairs that do not occur in training data)*

Pairwise Relation Classification

- Trained classifier with the word pairs.
- Cue phrases were removed to avoid teaching the classifier to distinguish between patterns used for creating training data (“but”, “because”, etc.)
- Test using corpus of 5000 examples labeled with one target relation and 5000 with another target relation to establish 50% baseline.

Pairwise Classifier Results

	CONTRAST	CEV	COND	ELAB	NO-REL-SAME-TEXT	NO-REL-DIFF-TEXTS
CONTRAST	-	87	74	82	64	64
CEV			76	93	75	74
COND				89	69	71
ELAB					76	75
NO-REL-SAME-TEXT						64

Table 3: Performances of classifiers trained on the Raw corpus. The baseline in all cases is 50%.

- Nice results, but they noticed that the learning curve flattened out around 2,000,000 examples.
- Suspected noise in the training data was the culprit

Experiment 2: Operation Cleanup

- They hypothesize that only using a subset of significant word pairs would improve results
- Create training set from BLIPP corpus (the parsed corpus)
- Extracted only the nouns, verbs, and cue phrases
- Repeated process from Experiment 1

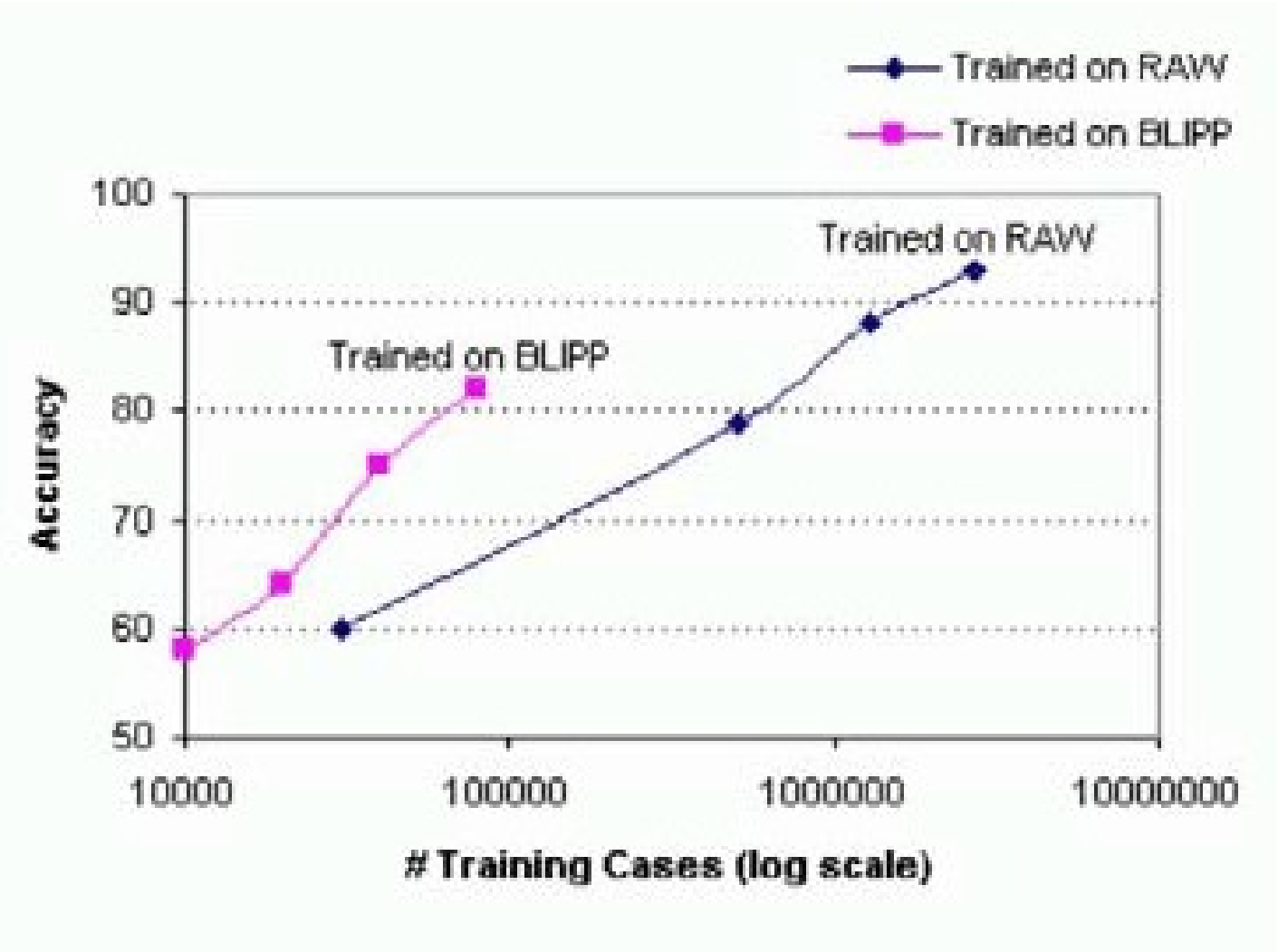
Results

	CONTRAST	CEV	COND	ELAB	NO-REL-SAME-TEXT	NO-REL-DIFF-TEXTS
CONTRAST	-	62	58	78	64	72
CEV			69	82	64	68
COND				78	63	65
ELAB					78	78
NO-REL-SAME-TEXT						66

Table 4: Performances of classifiers trained on the BLIPP corpus. The baseline in all cases is 50%.

- Similar results achieved with 100,000 examples
- Suggests performance increase by using subset of more informative word pairs

Learning Curves



Experiment 3 – Established Theory (RST)

- The goal was to show whether classifiers trained in this way can be used with an established theory
- Retrained classifier on RAW corpus, but left cue words in place
- Tested on RST Corpus from Carlson et al. (2001)

Results

	CONTR	CEV	COND	ELAB
# test cases	238	307	125	1761
CONTR	—	**63** 56	**80** 65	**64** 88
CEV			**87** 71	**76** 85
COND				**87** 93

Table 5: Performances of Raw-trained classifiers on manually labeled RST relations that hold between elementary discourse units. Performance results are shown in bold; baselines are shown in normal fonts.

- Second row indicates number of examples in RST data
- Indicates that these classifiers work well for certain RST relations
- Weaker results are the result of ill-defined categories: *"...result is consistent with the discourse model proposed byKnott et al. (2001), who suggest that ELABORATION relations are too ill-defined to be part of any discourse theory."*

In Conclusion...

- Discourse relations exist
- We still don't know exactly what their nature, number, and taxonomy looks like
- It may be possible to develop automated techniques for defining empirically justified discourse relations