# Optimal and Syntactically-Informed Decoding for Monolingual Phrase-based Alignment

Kapil Thadani and Kathleen McKeown

Columbia University
New York, NY

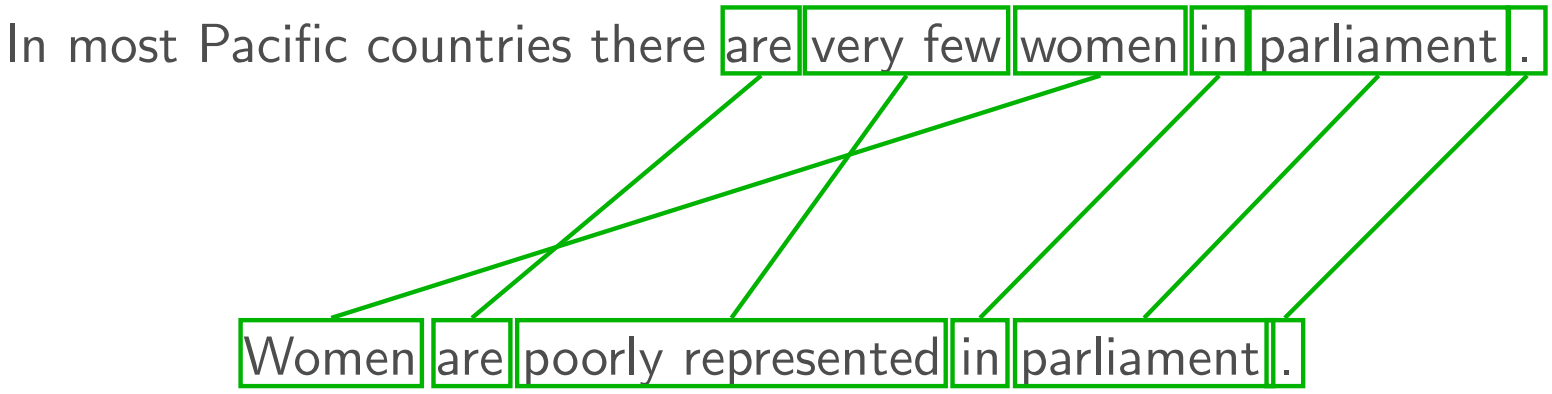
# Phrase-based alignment

In most Pacific countries there are very few women in parliament .

Women are poorly represented in parliament .

MacCartney et al. (2008)

# Phrase-based alignment




MacCartney et al. (2008)

# *Monolingual* alignment

Applications: paraphrasing, entailment, sentence fusion

Differences with MT alignment

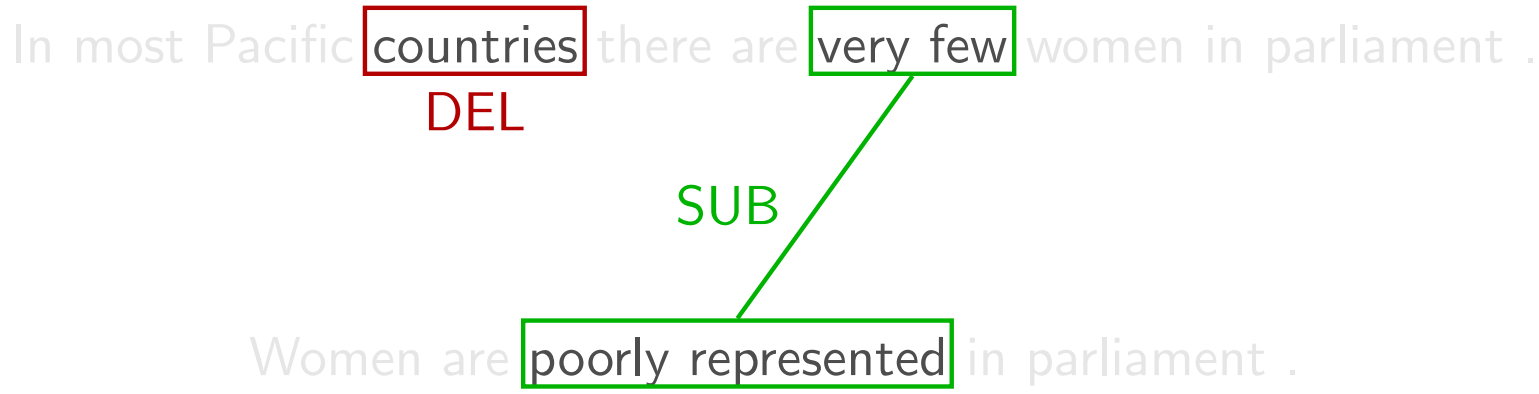
- No semantic equivalence
- Availability of lexical resources (paraphrases, ontologies)

Corpus: MSR RTE2 dataset annotated with human alignments (Brockett, 2007)

- 1600 alignment instances
- Three annotators per instance

# MANLI

MacCartney et al. (2008)




Features $\Phi(e)$ over edits

Learn weights $\mathbf{w}$ using structured perceptron

# Decoding phrase-based alignments

Given $\mathbf{w}$, recover best alignment $E^*$ for any sentence pair

$$E^* = \arg\max_E \sum_{e \in E} \mathbf{w}.\Phi(e)$$

**Challenge:** can't be factored into independent decisions

Need a consistent segmentation over text

e.g., $E$ can't contain both SUB(very few,$*$) and SUB(few,$*$)

MANLI uses stochastic hill-climbing with annealing for stability
**Drawbacks:** approximate, relatively slow

# Exact Decoding

This work

ILP formulation: Boolean variables $x_e$ indicate if $e \in E^*$

$$E^* = \arg\max_x \sum_e x_e \times \mathbf{w}.\Phi(e)$$

Constraint: exactly one edit $e \in E^*$ per token

$$\forall \text{ tokens } t, \quad \sum_{e:\ e \text{ covers } t} x_e = 1$$

Phrase-based MT alignment: DeNero and Klein (2008)

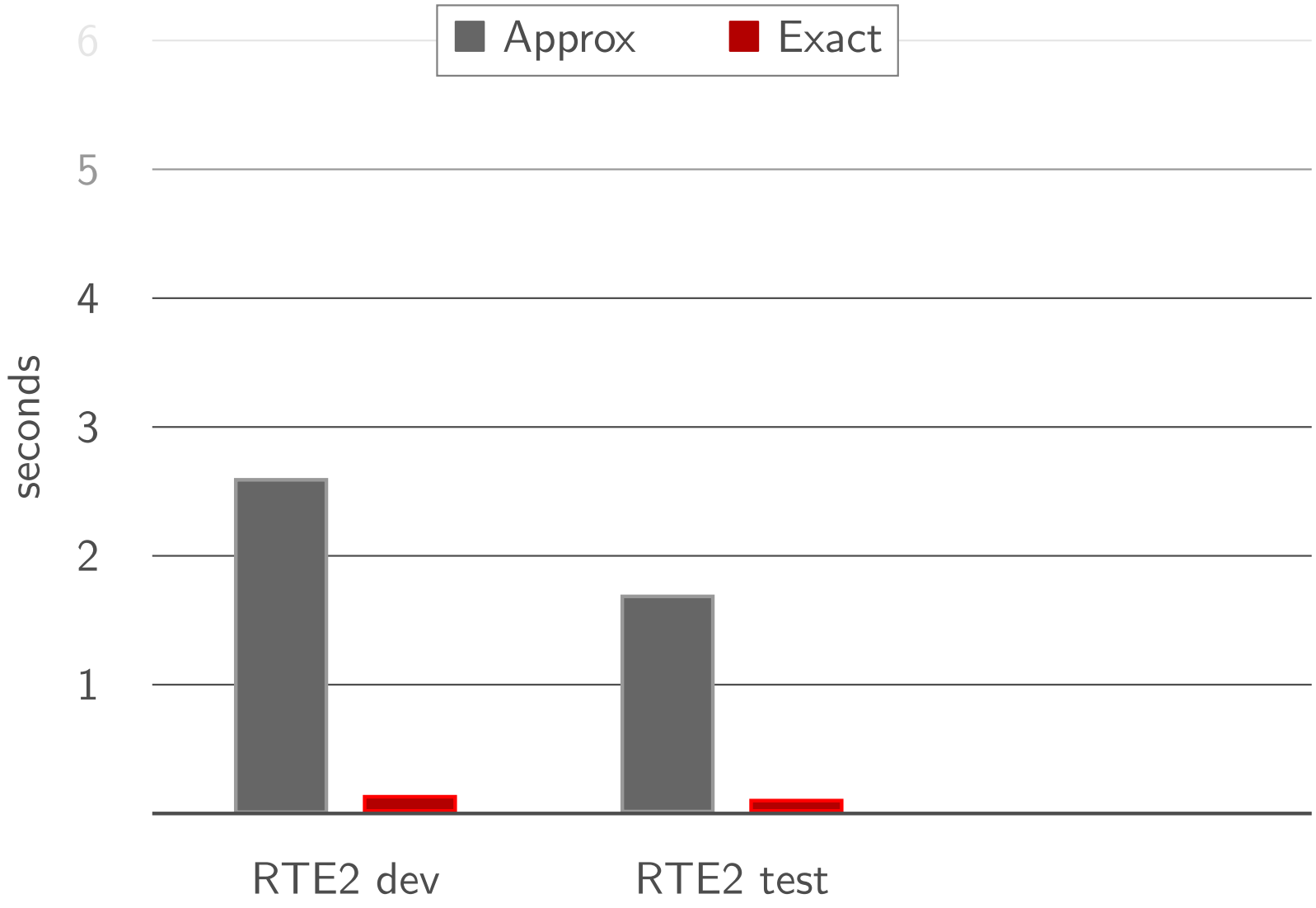
# Performance on RTE2

| System | Data | $P\%$ | $R\%$ | $F_1\%$ | % Perfect |
|---|---|---|---|---|---|
| MANLI (2008) | dev | 83.4 | 85.5 | 84.4 | 21.7 |
| | test | 85.4 | 85.3 | 85.3 | 21.3 |
| MANLI (reimplemented) | dev | 85.3 | 84.8 | 85.0 | 23.8 |
| | test | 87.2 | 86.3 | 86.7 | 24.5 |

# Performance on RTE2

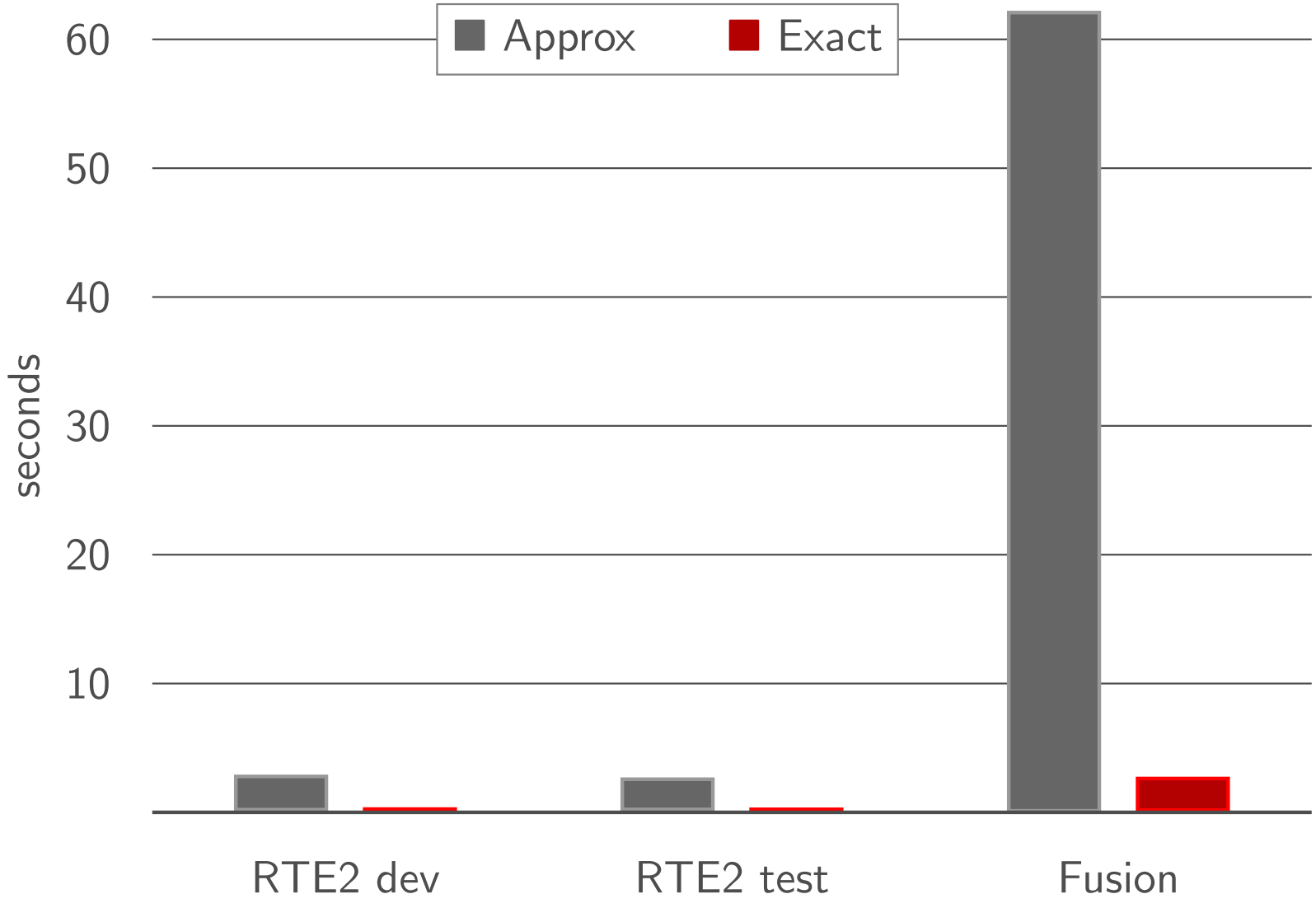
| System | Data | $P\%$ | $R\%$ | $F_1\%$ | $\%$ Perfect |
|---|---|---|---|---|---|
| MANLI (2008) | dev | 83.4 | 85.5 | 84.4 | 21.7 |
| | test | 85.4 | 85.3 | 85.3 | 21.3 |
| MANLI (reimplemented) | dev | 85.3 | 84.8 | 85.0 | 23.8 |
| | test | 87.2 | 86.3 | 86.7 | 24.5 |
| with exact decoding | dev | 85.7 | 84.7 | 85.2 | 24.6 |
| | test | 87.8 | 86.1 | 86.8 | 24.8 |

# Average runtime



# Exact Decoding

Average runtime



# Function words

Sunday's earthquake was felt in the south Indian city of Madras on the mainland , as well as other parts of south India . The Naval meteorological office in Port Blair said it was the second biggest aftershock after the Dec. 26 earthquake .

The city of Madras is located in Southern India .

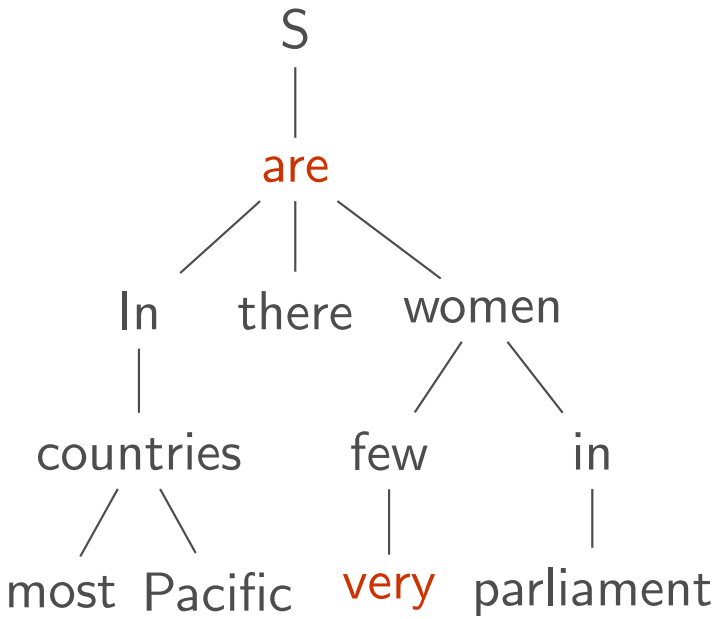
# Function words

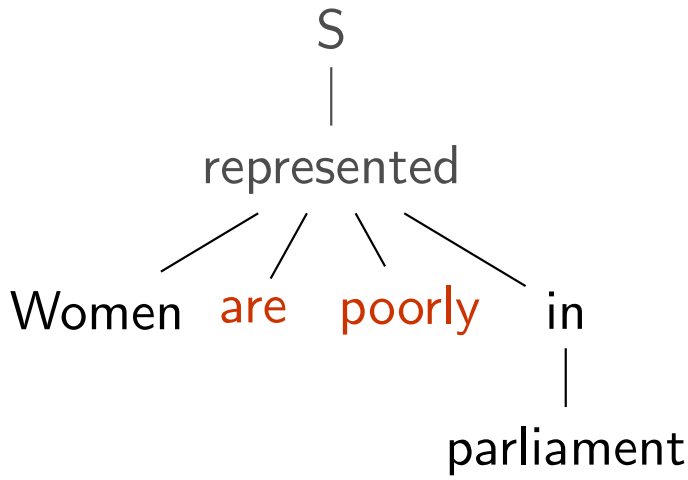
Sunday's earthquake was felt in the south Indian city of Madras on the mainland , as well as other parts of south India . The Naval meteorological office in Port Blair said it was the second biggest aftershock after the Dec. 26 earthquake .

The city of Madras is located in Southern India .

# Syntactically-informed features?

Unclear how to encode syntactic dependency information in features




SUB (are very, are poorly)

# Syntactically-informed *constraints*

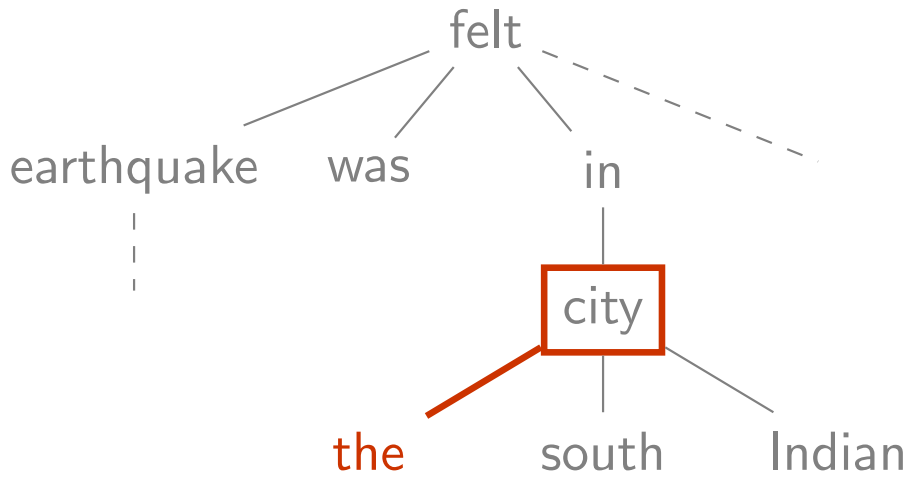
Constraint: Boolean variables $y_t$ to indicate tokens covered by SUB edits

$$\forall \text{ tokens } t, \quad y_t \; - \sum_{\substack{e\text{: } e \text{ is SUB} \\ e \text{ covers } t}} x_e \; = \; 0$$

Only need to consider relationship between tokens within a sentence

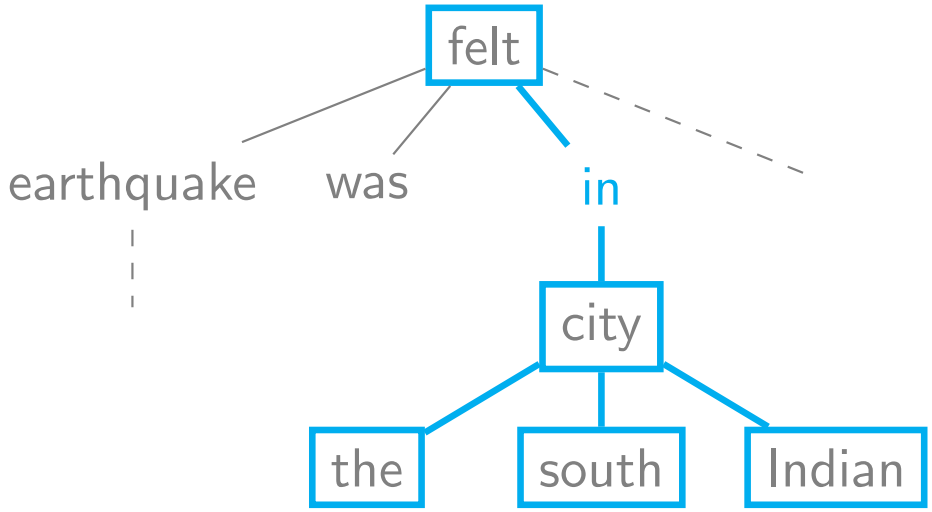
# Syntactically-informed constraints

(1) **Modifier constraints:** determiners, conjunctions, modals



# Syntactically-informed constraints

(2) **Lineage constraints:** prepositions, particles



# Performance on RTE2

| System | Data | $P\%$ | $R\%$ | $F_1\%$ | $\%$ Perfect |
|---|---|---|---|---|---|
| MANLI (reimplemented) | dev | 85.3 | 84.8 | 85.0 | 23.8 |
| | test | 87.2 | 86.3 | 86.7 | 24.5 |

# Performance on RTE2

| System | Data | $P\%$ | $R\%$ | $F_1\%$ | $\%$ Perfect |
|---|---|---|---|---|---|
| MANLI (reimplemented) | dev | 85.3 | 84.8 | 85.0 | 23.8 |
| | test | 87.2 | 86.3 | 86.7 | 24.5 |
| exact + modifier constraints | dev | **86.8** | 84.5 | **85.6** | 25.3 |
| | test | **88.8** | 85.7 | **87.2** | 29.9 |

# Performance on RTE2

| System | Data | $P\%$ | $R\%$ | $F_1\%$ | % Perfect |
|---|---|---|---|---|---|
| MANLI (reimplemented) | dev | 85.3 | 84.8 | 85.0 | 23.8 |
| | test | 87.2 | 86.3 | 86.7 | 24.5 |
| exact + modifier constraints | dev | **86.8** | 84.5 | **85.6** | 25.3 |
| | test | **88.8** | 85.7 | **87.2** | 29.9 |
| exact + lineage constraints | dev | **86.1** | 84.6 | **85.3** | 24.5 |
| | test | **88.2** | 86.4 | **87.3** | 27.6 |

# Performance on RTE2

| System | Data | $P\%$ | $R\%$ | $F_1\%$ | % Perfect |
|---|---|---|---|---|---|
| MANLI (reimplemented) | dev | 85.3 | 84.8 | 85.0 | 23.8 |
| | test | 87.2 | 86.3 | 86.7 | 24.5 |
| exact + modifier constraints | dev | **86.8** | 84.5 | **85.6** | 25.3 |
| | test | **88.8** | 85.7 | **87.2** | 29.9 |
| exact + lineage constraints | dev | **86.1** | 84.6 | **85.3** | 24.5 |
| | test | **88.2** | 86.4 | **87.3** | 27.6 |
| exact + both | dev | **87.1** | 84.4 | **85.8** | 25.4 |
| | test | **89.5** | 86.2 | **87.8** | 33.0 |

# Summary

- Speed gains via ILP decoding
  - Exact solutions
  - 20x runtime improvement over search-based decoding
- Syntactic constraints for controlling function-word mismatch
  - Defined at the token level
  - Significant gains in precision
  - Large increase in number of perfect alignments