Textual Entailment w/ Structured Attentions & Composition

Kai Zhao Liang Huang Mingbo Ma
Dept. EECS, Oregon State University

Oregon State UNIVERSITY

Textual Entailment: A Spectrum of Approaches

As of 2016.11
cf. McCarteney & Manning 2007

MacCartney & Manning, 2007
Wang & Manning, 2007
Watanabe et al., 2012
Tian et al., 2014
Filice et al., 2015
…

align: local relation
logic: relation composition

Our Work

align: local relation
tree: relation composition

Rocktäschel et al., 2015
Wang & Jiang., 2015
Bowman et al., 2015

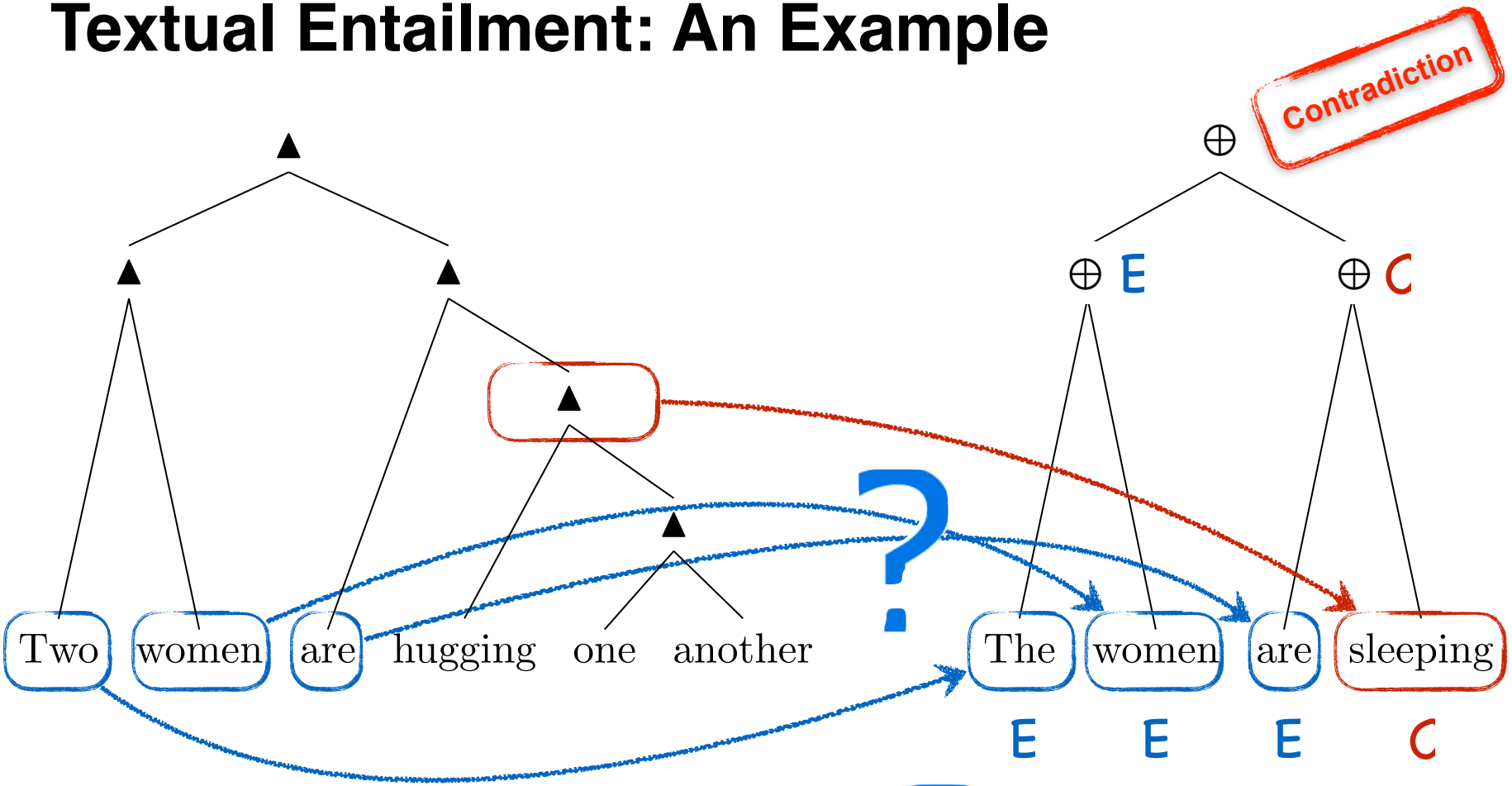
LSTM repr. + **attention** for local inference

mpare RNN
t. repr.

discrete repr.
easy to understand
brittle

continuous repr.
robust
difficult to interpret

Textual Entailment: An Example



Relations:
Entailment
Contradiction
Neutral

COMPOSITION RULES:
Entailment ⊕ Entailment = Entailment
Entailment ⊕ Contradiction = Contradiction
Contradiction ⊕ Contradiction = Neutral

Textual Entailment as Structured Prediction w/ Latent Var.

- For each sentence pair (P, Q) and reference relation y
  - online-fashion: minimize the (stepwise) negative log probability

$$\begin{aligned}\ell &= -\log \Pr(y|P,Q) \\ &= -\log \sum_{\mathbf{A}} \Pr(y, \mathbf{A}|P,Q) \\ &= -\log \sum_{\mathbf{A}} [\Pr(\mathbf{A}|P,Q) \cdot \Pr(y|\mathbf{A},P,Q)]\end{aligned}$$

  - two sub-problems:
    - given an alignment $\mathbf{A} \in \{0,1\}^{|P|\times|Q|}$, calculate entailment bottom-up
    - how to search through all possible latent alignments (exp. many)

$$\mathbf{A} \in \left\{ \quad \cdots \quad \right\}$$

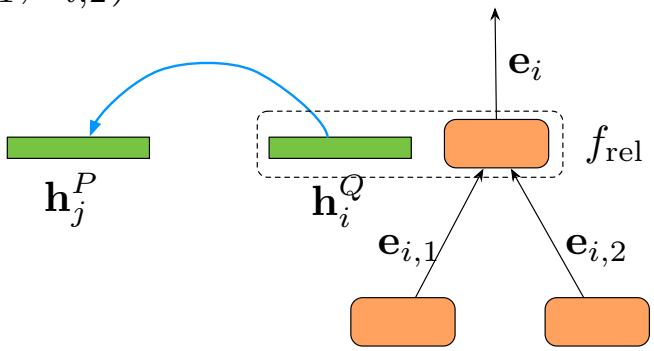
Entailment Calculation

$$\ell = -\log \sum_{\mathbf{A}} [\Pr(\mathbf{A}|P,Q) \cdot \Pr(y|\mathbf{A},P,Q)]$$

- Given an alignment, calculate the entailment bottom-up
  - the entailment relation at hypothesis node i: a continuous vector $\mathbf{e}_i$
  - the meaning representation of the subtree rooted at i is $\mathbf{h}_i^Q$
  - node i is aligned to premise node j, which has meaning repr. $\mathbf{h}_j^P$
  - entailment relations from two children nodes of i: $\mathbf{e}_{i,1}, \mathbf{e}_{i,2}$

$$\mathbf{e}_i = f_{\text{rel}}(\mathbf{h}_i^P, \mathbf{h}_j^Q; \mathbf{e}_{i,1}, \mathbf{e}_{i,2})$$

- Options for f_{rel}
  - Recursive NN (Socher et al., 2013)
  - Convolution NN
  - **Tree LSTM** (Tai et al., 2015)



Approximated Search in the Latent Space

- Stepwise objective

$$\begin{aligned}\ell &= -\log \Pr(y|P,Q) \\ &= -\log \sum_{\mathbf{A}} \Pr(y,\mathbf{A}|P,Q) \\ &= -\log \sum_{\mathbf{A}} [\Pr(\mathbf{A}|P,Q) \cdot \Pr(y|\mathbf{A},P,Q)]\end{aligned}$$

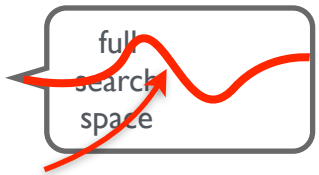
- The structured latent variable $\mathbf{A} \in \{0,1\}^{|P|\times|Q|}$ takes value from an exponentially large space
  - calculating $\sum_{\mathbf{A}} [\Pr(\mathbf{A}|P,Q) \cdot (\ \cdot\ )]$ exactly is intractable
- Approximately search through the space
  - sampling
  - expected alignment

Approximate Search with Sampling

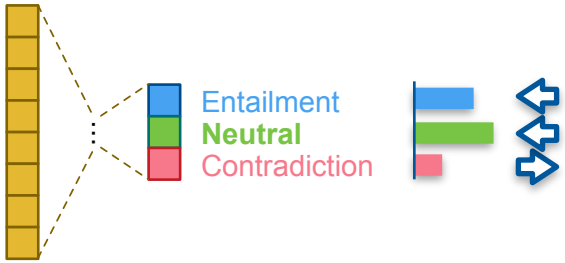
$$\Pr(\mathbf{A}|P,Q) \quad \Pr(y|\mathbf{A},P,Q)$$

Training Iterations

Two women are hugging one another
The women are sleeping



sample an alignment

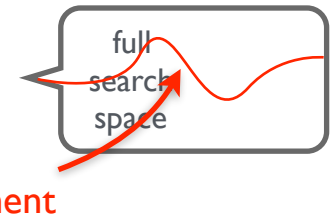


Approximate Search with Sampling

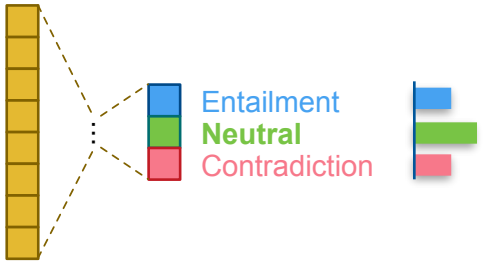
$\Pr(\mathbf{A}|P,Q)$ $\Pr(y|\mathbf{A},P,Q)$

Training Iterations

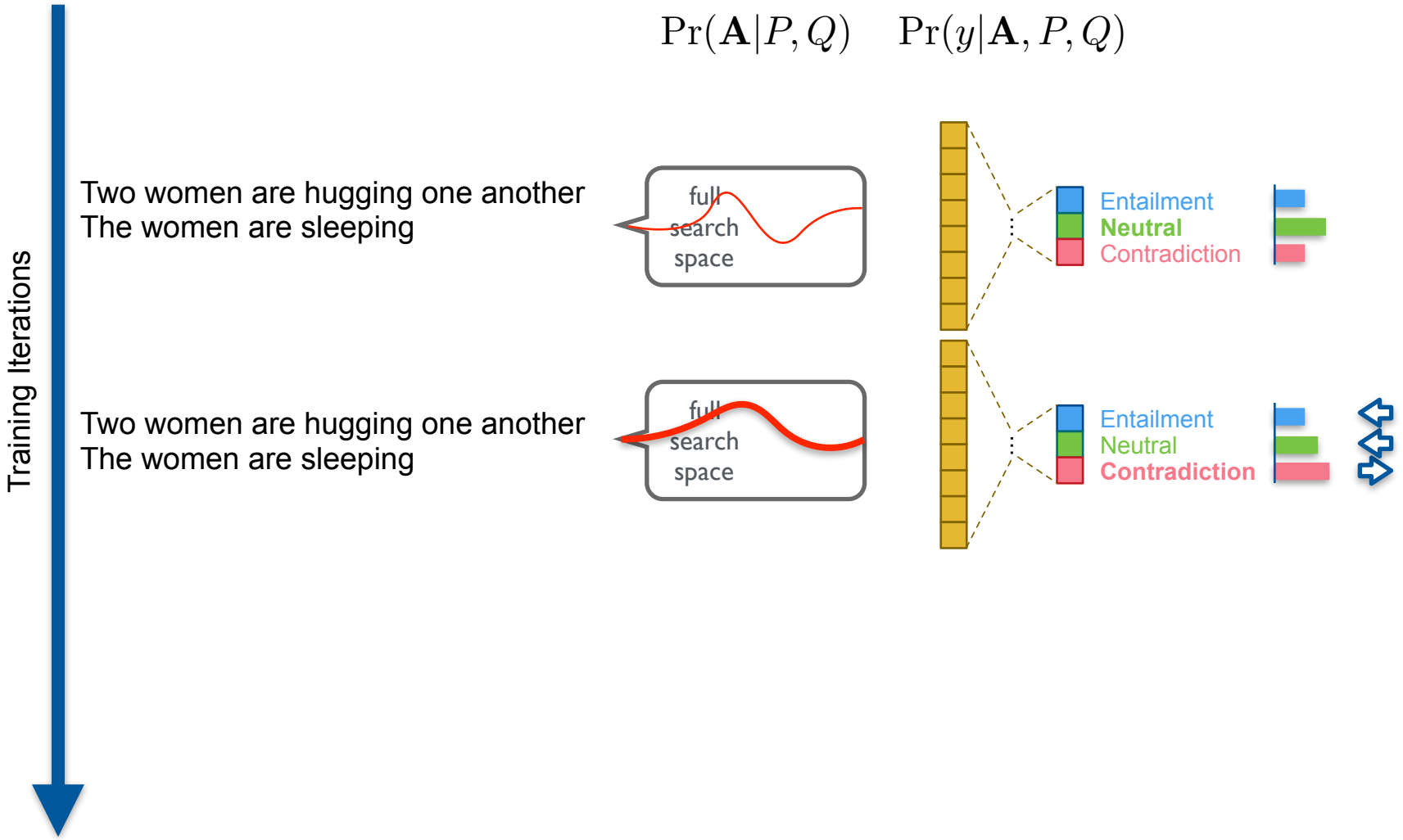
Two women are hugging one another
The women are sleeping



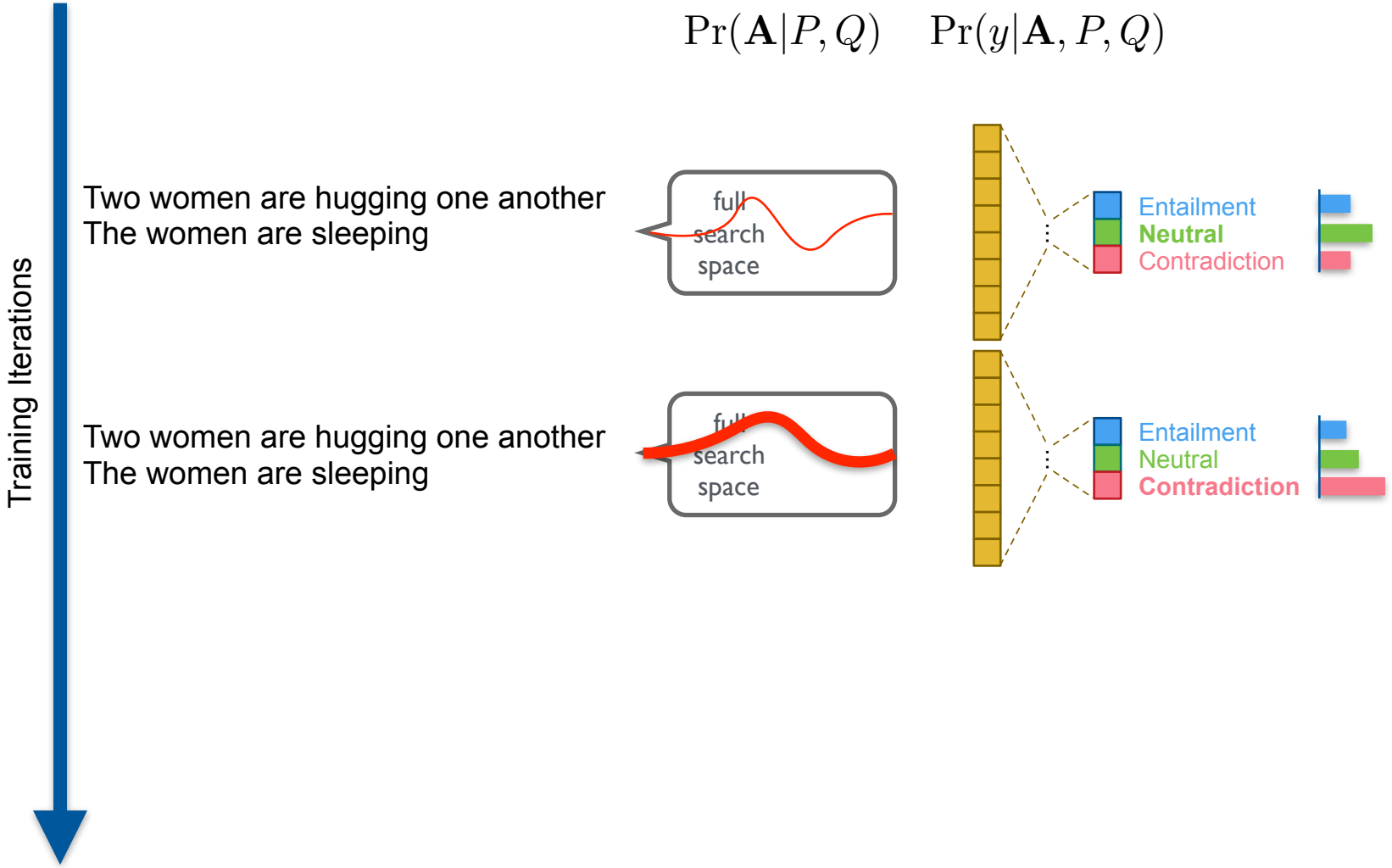
sampled alignment



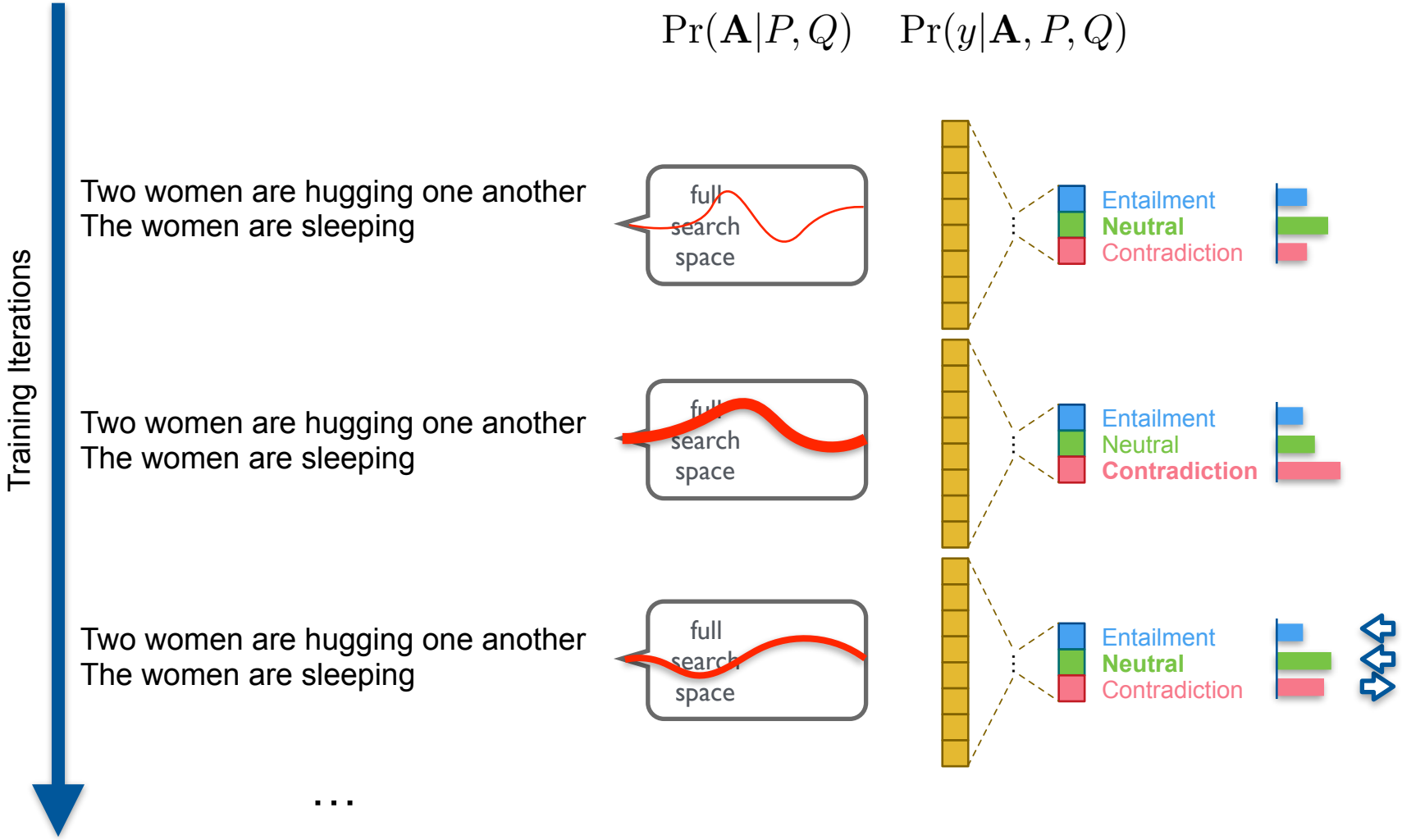
Approximate Search with Sampling



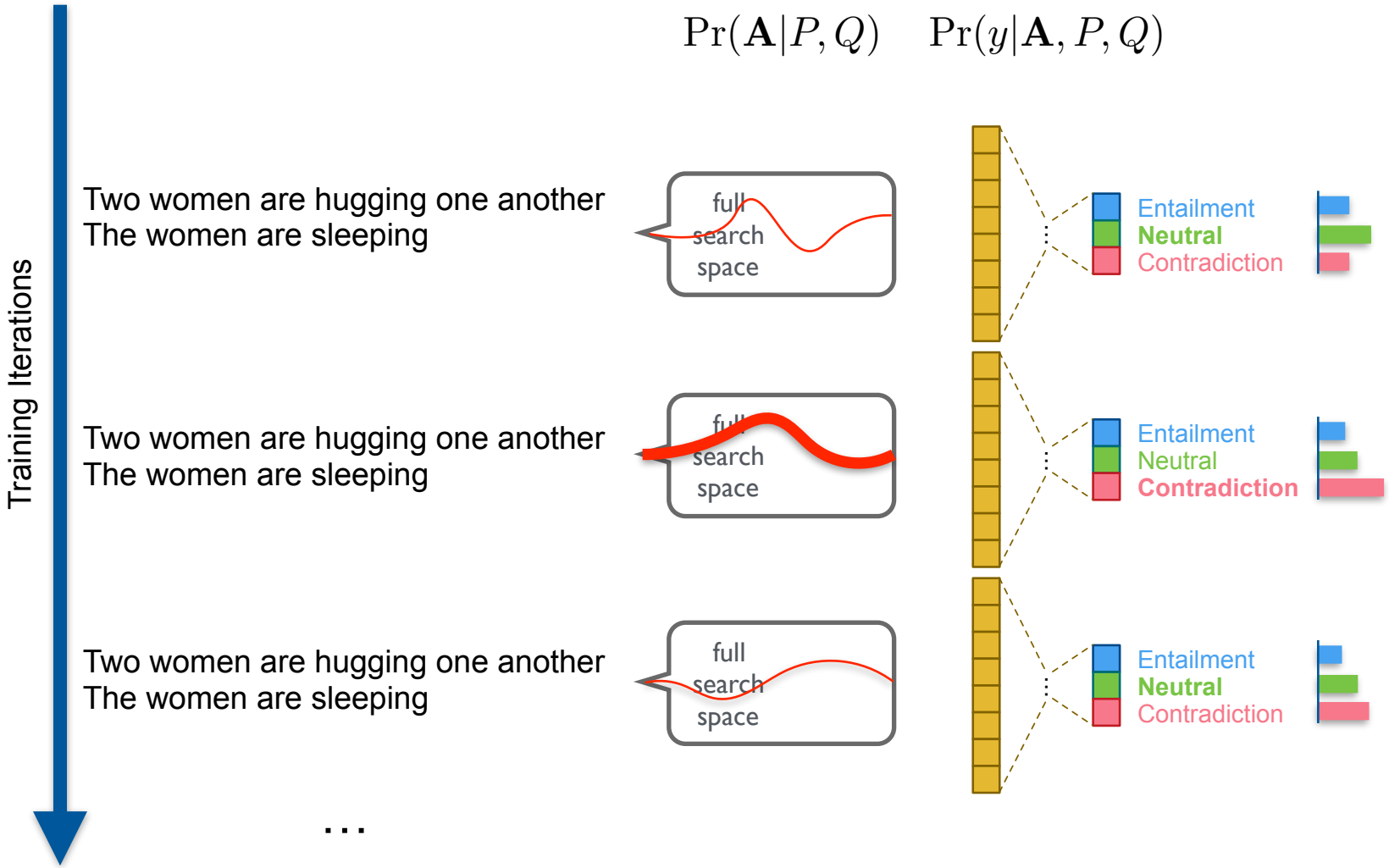
Approximate Search with Sampling



Approximate Search with Sampling

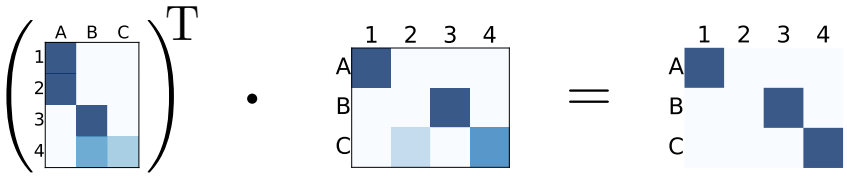


Approximate Search with Sampling



Improving Alignments (Optional)

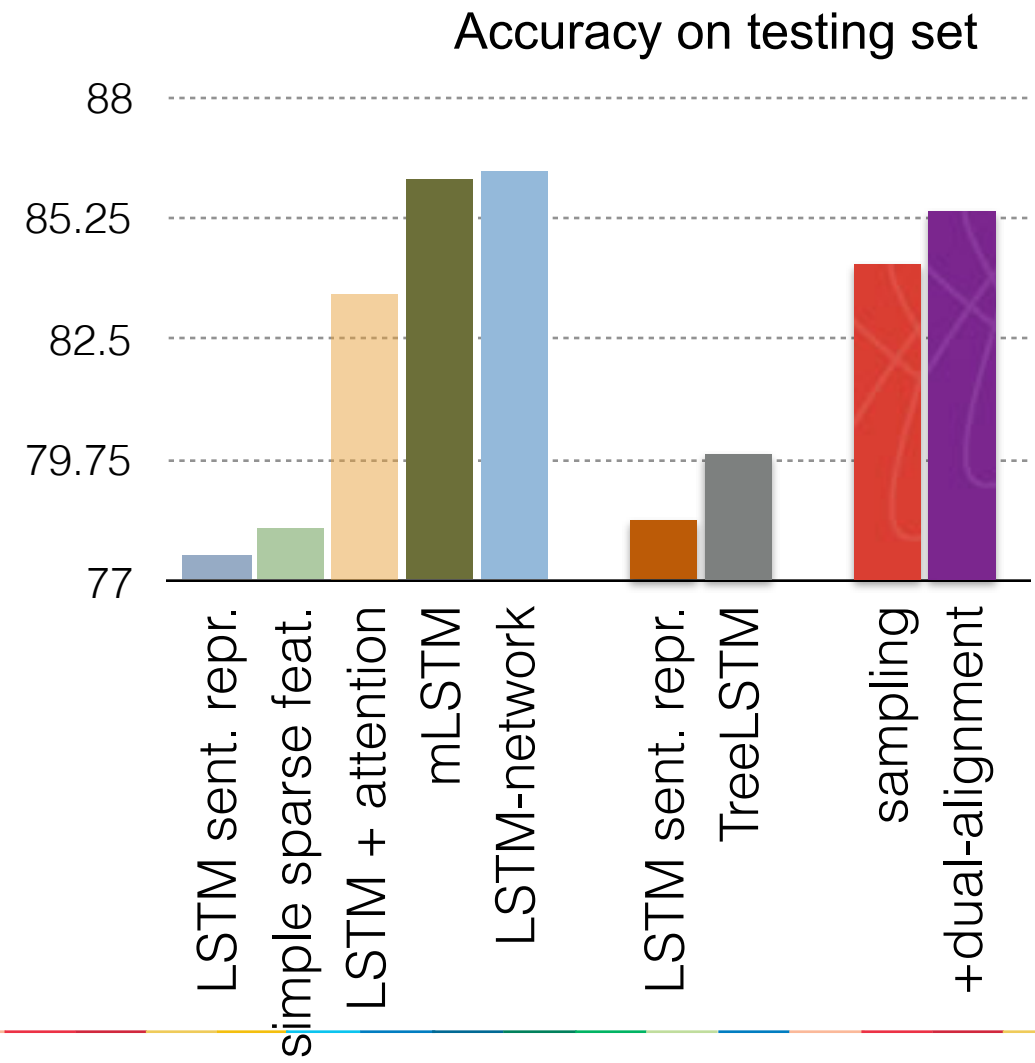
- Improve alignment calculation with some heuristics
  - alignment and relation calculations are separated
- Heuristic: symmetric property
  - if a hypothesis node i is most relevant to a premise node j
  - premise node j should be most relevant to hypo. node i
  - widely used in statistical MT to reduce noice
    - intersection of two directional alignments
    - for alignment probability we use element-wise product



Empirical Evaluation

- Dataset: Stanford Natural Language Inference
  - annotated by mechanical turk
  - 570k sentence pairs in training
  - 10k sentence pairs in development/testing each
  - 3 relations: entailment, neutral, contradiction
- Example sentence pairs
  - entailment
    - Premise: An old man with a package poses in front of an advertisement.
    - Hypothesis: A man poses in front of an ad.
  - contradiction
    - Premise: A statue at a museum that none seems to be looking at.
    - Hypothesis: Tons of people are gathering around the statue.

Empirical Evaluation: Quantitative



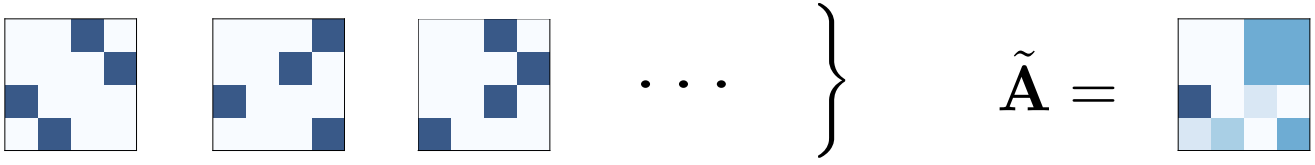
Approximate Search with Expected Alignment

- Review the stepwise objective:
  negative log of expected probability of the relation

$$\begin{aligned}\ell &= -\log \Pr(y|P,Q) \\ &= -\log \sum_{\mathbf{A}} \Pr(y,\mathbf{A}|P,Q) \\ &= -\log \sum_{\mathbf{A}} [\Pr(\mathbf{A}|P,Q)\cdot \Pr(y|\mathbf{A},P,Q)] \\ &= -\log \mathbb{E}_{\Pr(\mathbf{A}|P,Q)}[\Pr(y|\mathbf{A},P,Q)]\end{aligned}$$

  - can be approximated by negative log probability of the relation under expected alignment (Xu et al., 2015)

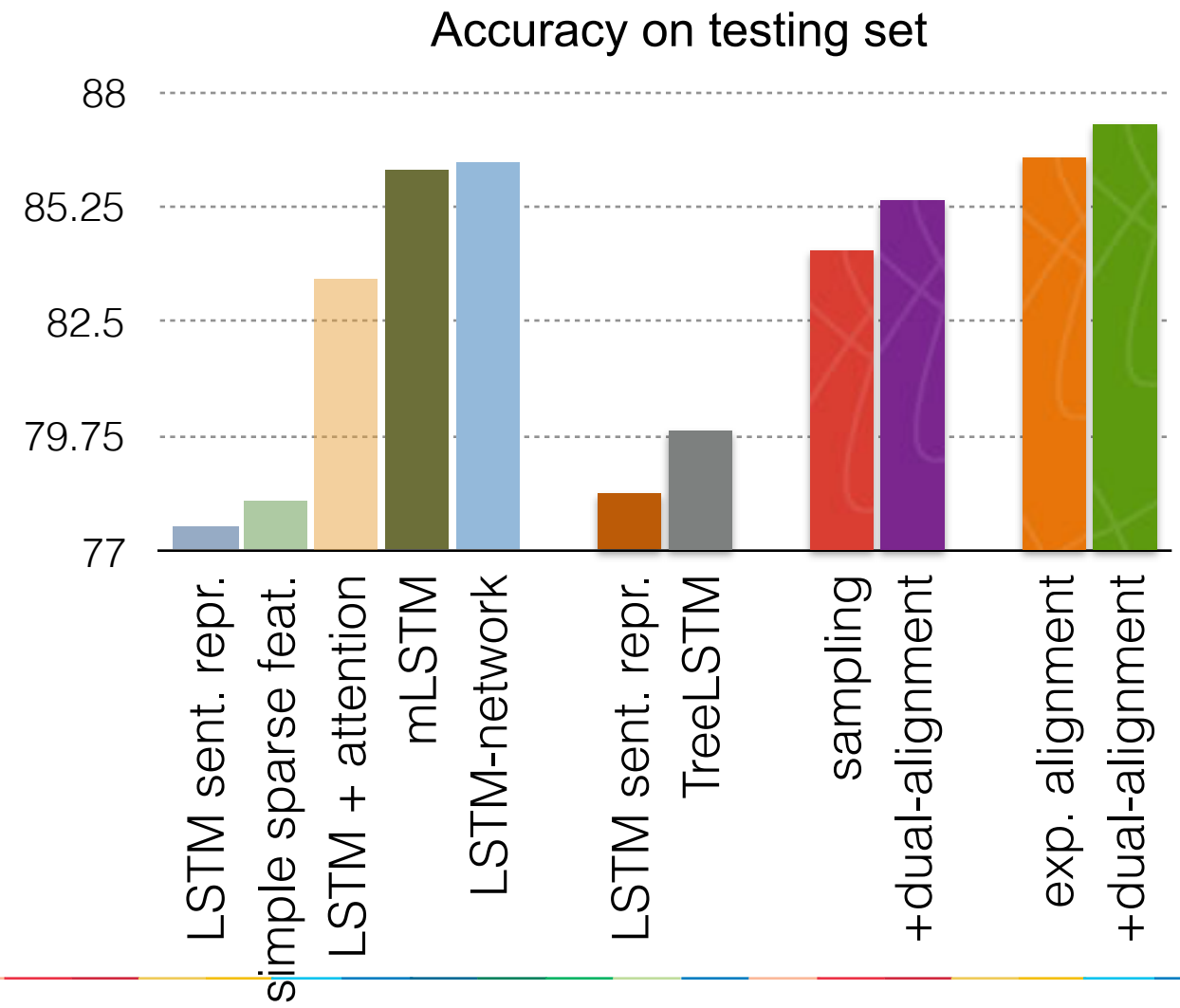
$$\mathbf{A} \in \left\{ \quad \cdots \quad \right\} \qquad \tilde{\mathbf{A}} =$$



    - use one (soft) alignment to approximately represent all alignments

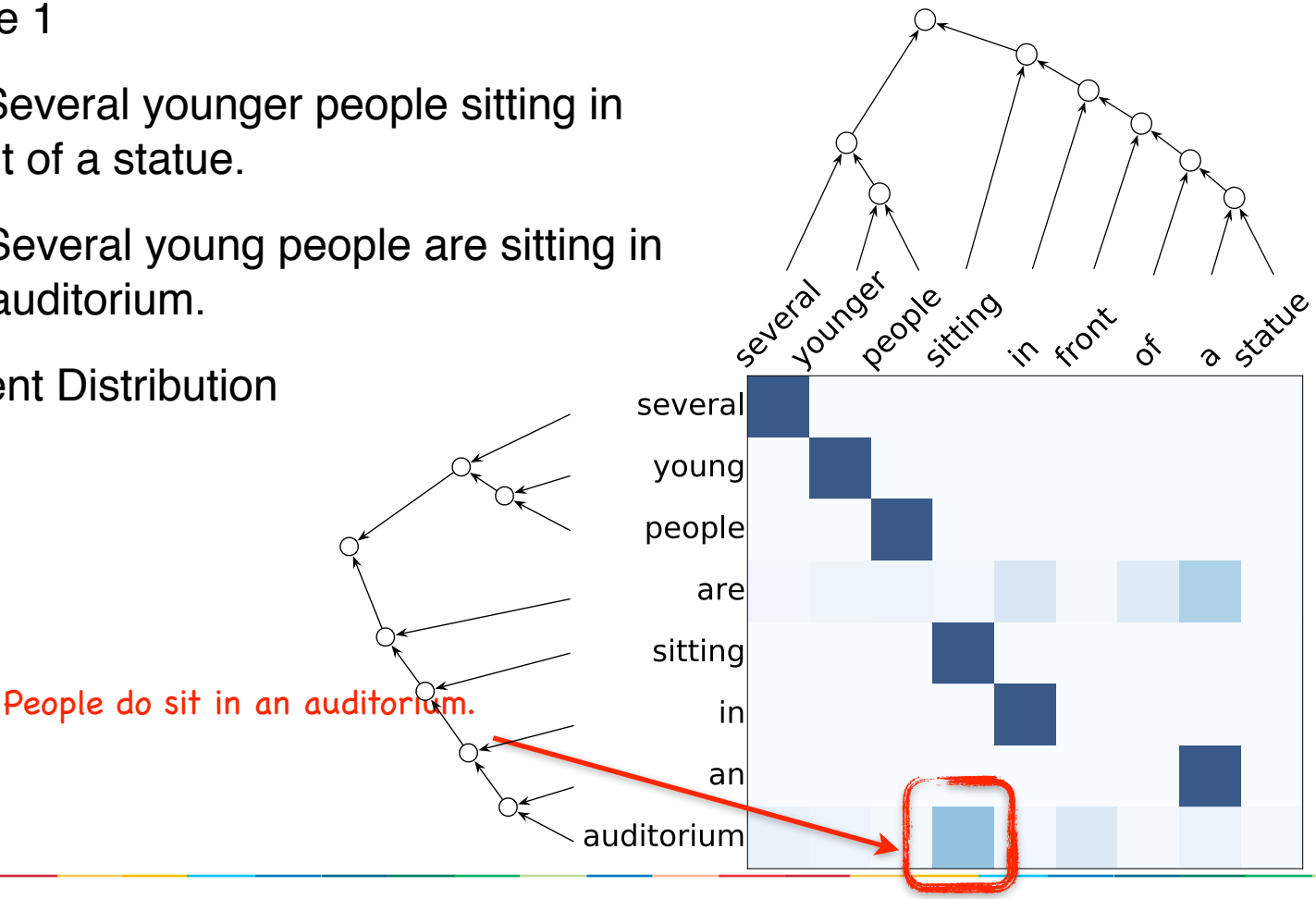
$$\begin{aligned}\ell &\approx -\log \Pr(y|\mathbb{E}_{\Pr(\mathbf{A}|P,Q)}[\mathbf{A}],P,Q) \\ &= -\log \Pr(y|\tilde{\mathbf{A}},P,Q)\end{aligned}$$

Empirical Evaluation: Quantitative



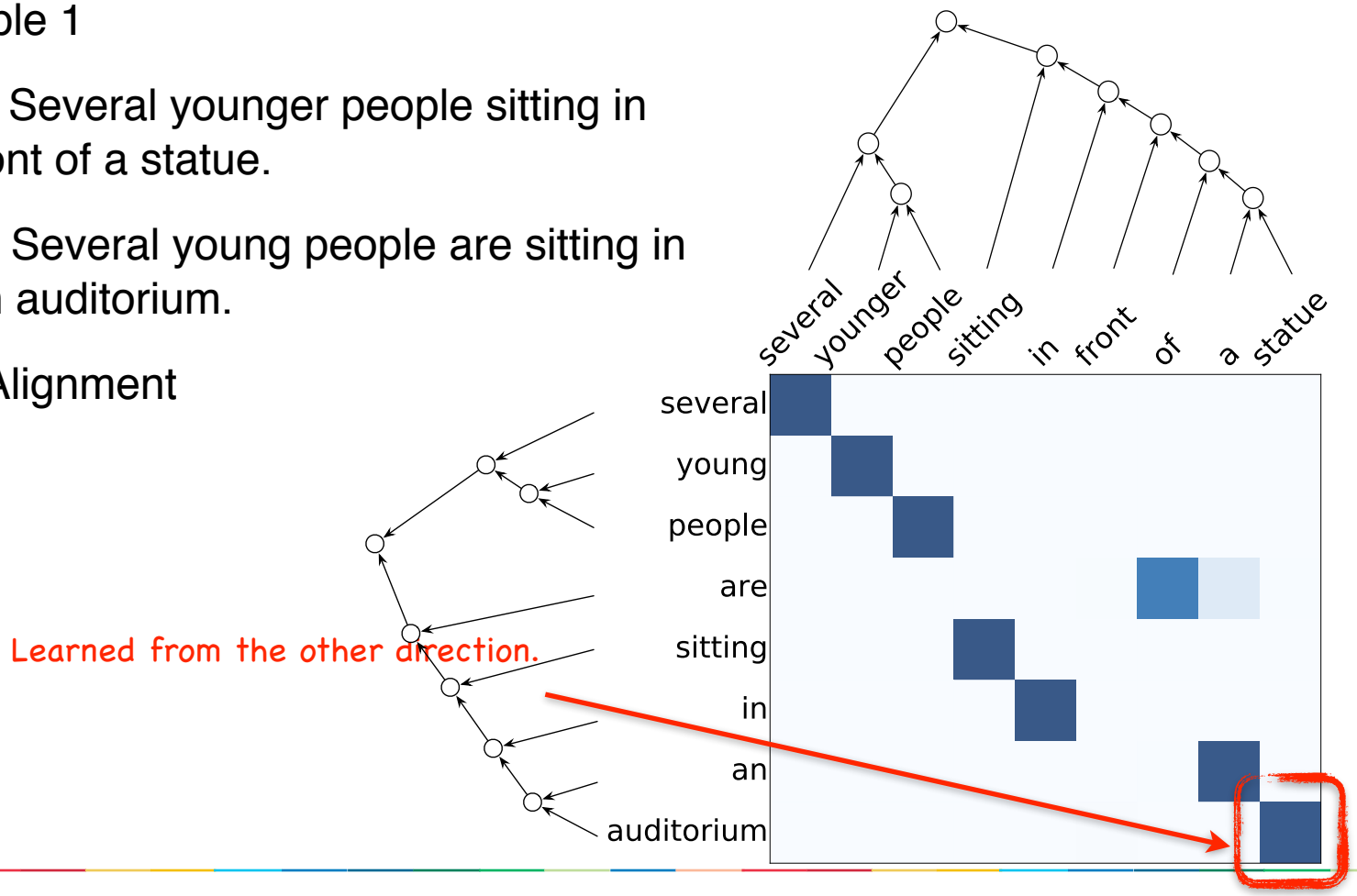
Empirical Evaluation: Qualitative

- Example 1
  - P: Several younger people sitting in front of a statue.
  - H: Several young people are sitting in an auditorium.
- Alignment Distribution



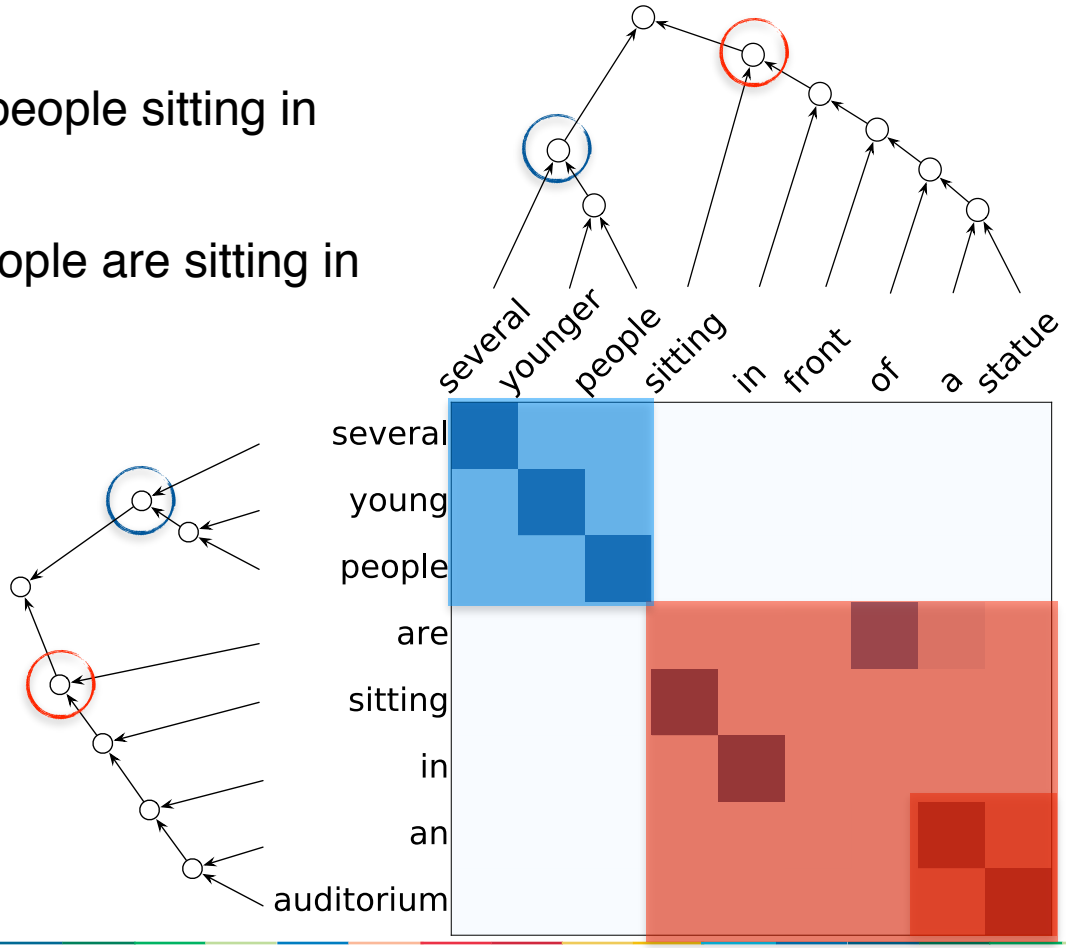
Empirical Evaluation: Qualitative

- Example 1
  - P: Several younger people sitting in front of a statue.
  - H: Several young people are sitting in an auditorium.
- Dual-Alignment



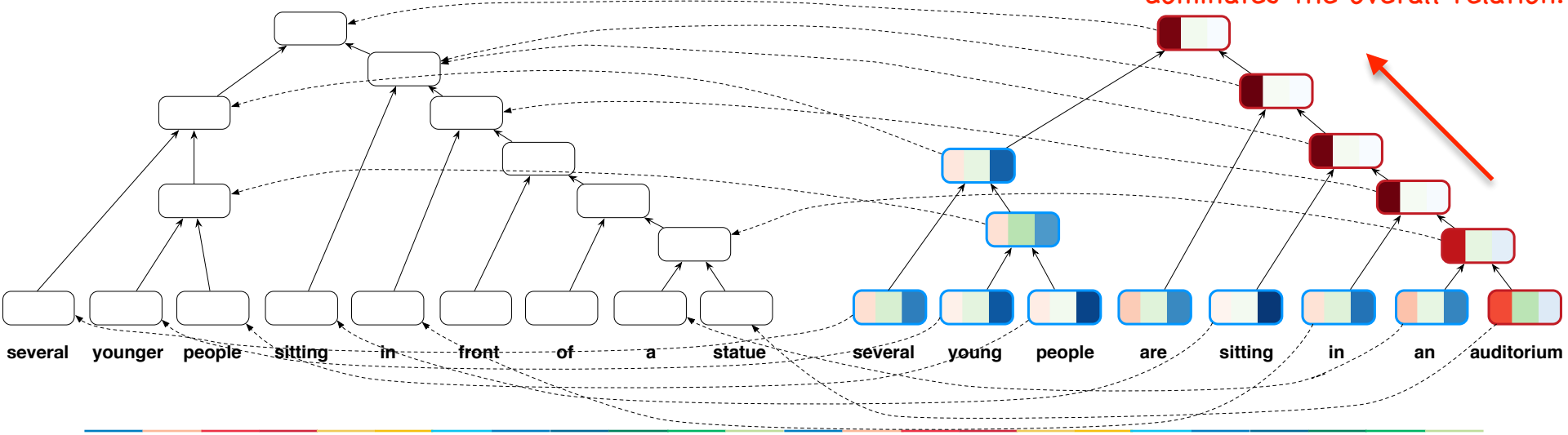
Empirical Evaluation: Qualitative

- Example 1
  - P: Several younger people sitting in front of a statue.
  - H: Several young people are sitting in an auditorium.
- Dual-Alignment
  - tree-level



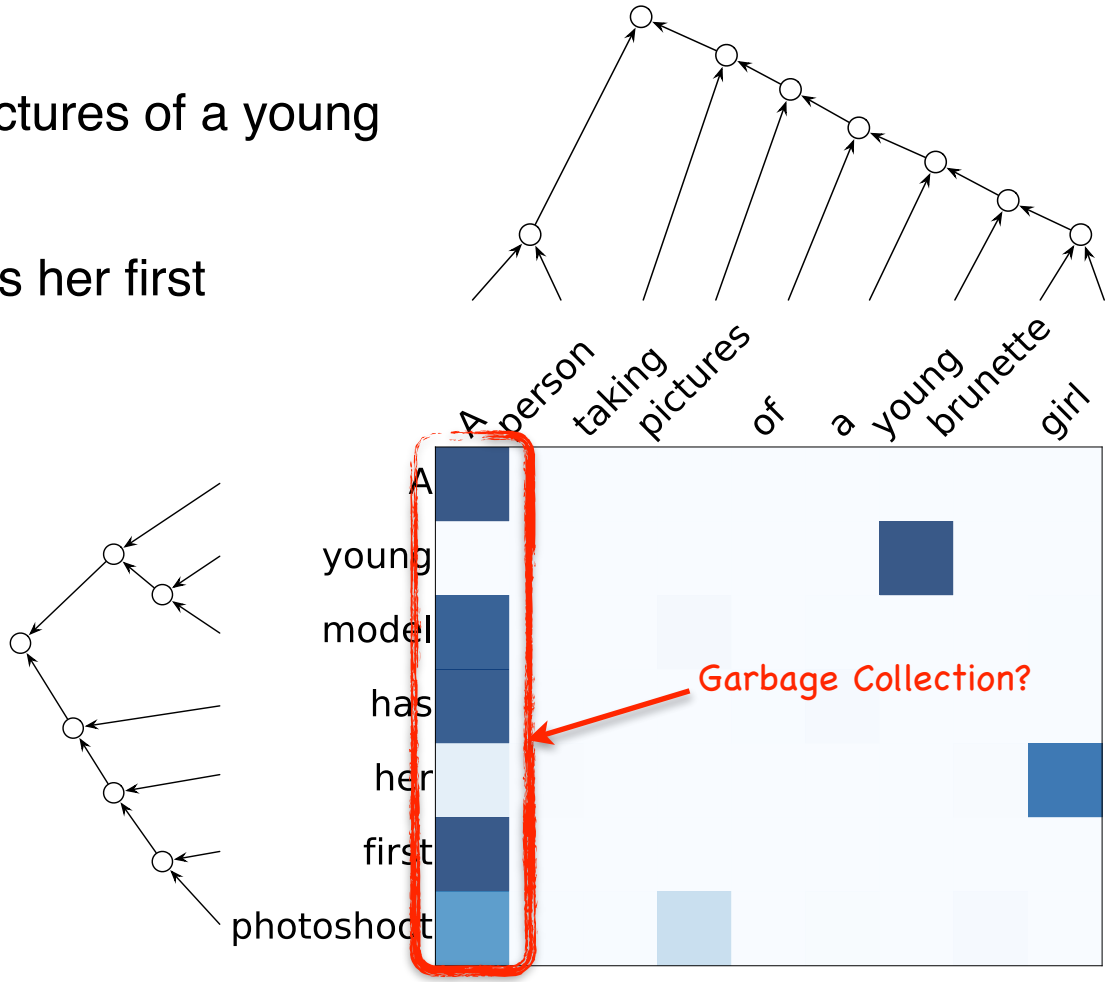
Empirical Evaluation: Qualitative

- Example 1
  - P: Several younger people sitting in front of a statue.
  - H: Several young people are sitting in an auditorium.
- Entailment Composition



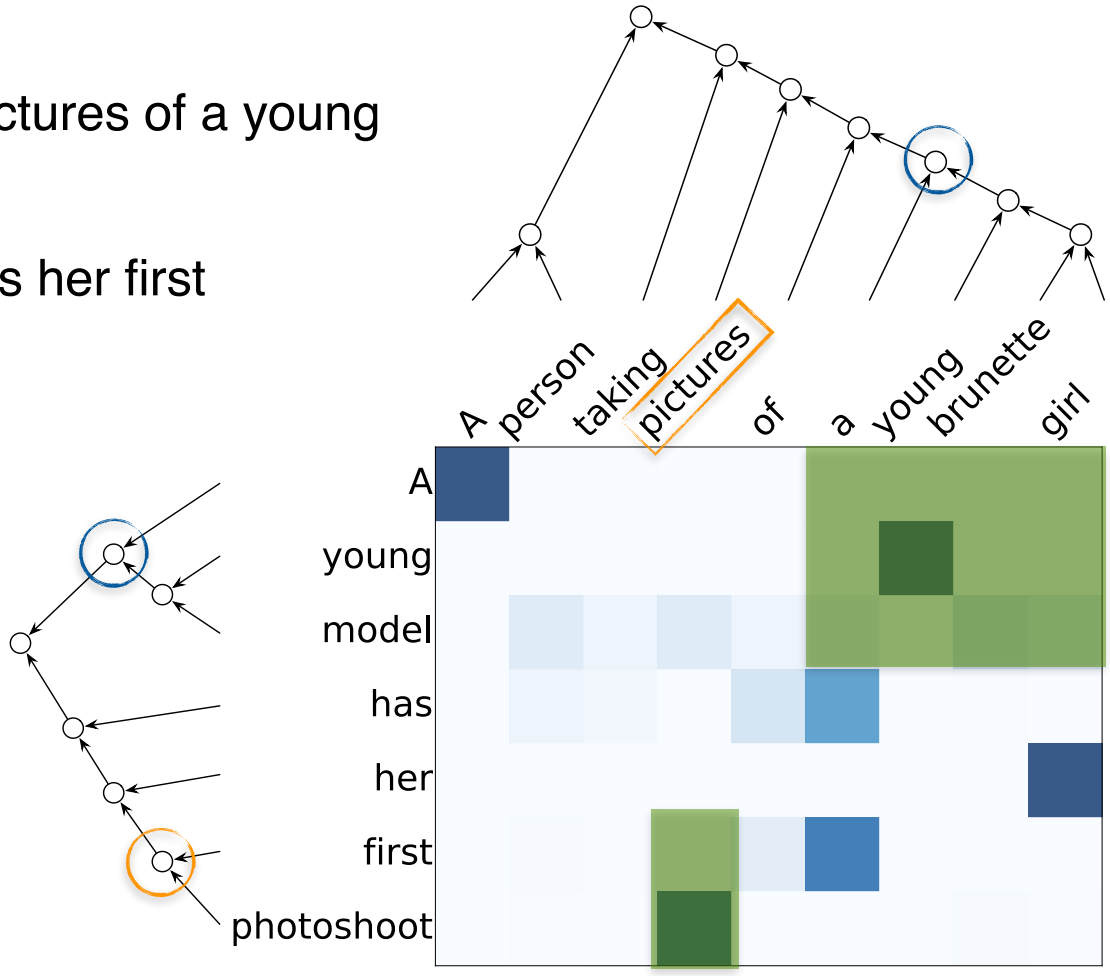
Empirical Evaluation: Qualitative

- Example
  - P: A person taking pictures of a young brunette girl.
  - H: A young model has her first photoshoot.
- Alignment Distribution



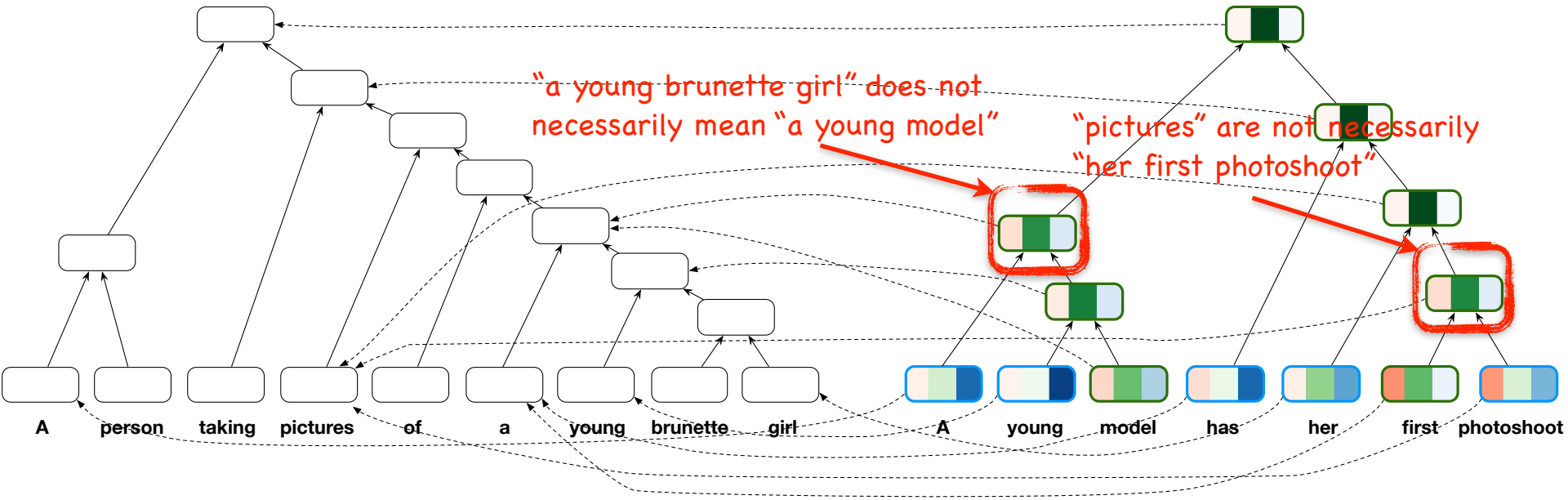
Empirical Evaluation: Qualitative

- Example
  - P: A person taking pictures of a young brunette girl.
  - H: A young model has her first photoshoot.
- Dual-Alignment



Empirical Evaluation: Qualitative

- Example
  - P: A person taking pictures of a young brunette girl.
  - H: A young model has her first photoshoot.
- Entailment composition



Discussion

- Some difficult examples: we need to modify our framework to handle these
  - More variations in syntax
    - The boy loves the girl.
    - The girl loves the boy.
    - The boy loved the girl.
  - Some are more subtle
    - A stuffed animal on the couch.
    - An animal on the couch.
    - A stuffed toy on the couch.
- Traditional Datasets: Pascal RTE, FraCas

Conclusion

- A soft version of natural logic
- Textual entailment problem as a latent variable structured prediction problem
- Approximate the exponentially large latent space with
  - sampling
  - expected alignment
- Align from two directions to improve alignment
- Learn an RNN for entailment relation composition
- Easy to interpret:
  - what is the entailment relation at each hypothesis tree node
  - how these entailment relations are composed

However

- Textual entailment is still a difficult problem
- For example, our approach does not handle following cases:

Entailment Neutral Contradiction

Premise: A man is waiting in front of a red light.

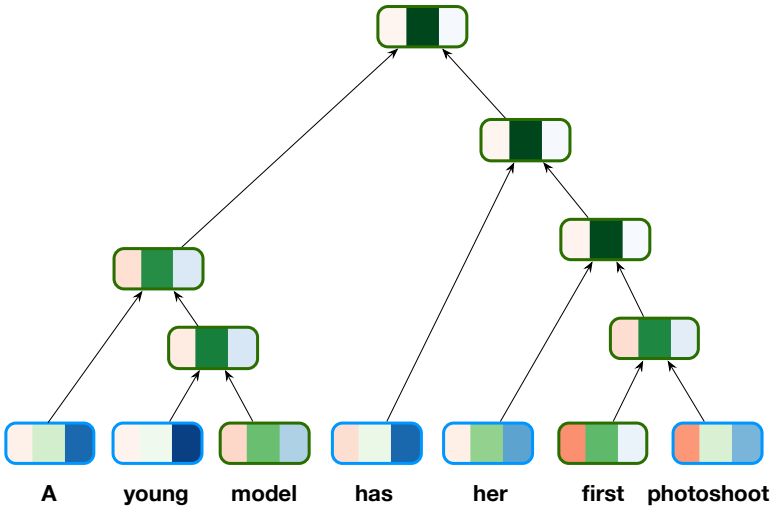
Hypothesis: A man is waiting for green light.

Entailment ?

Premise: A boy is splashing through the ocean.

Hypothesis: A boy is in Kansas.

Contradiction ?



FIN

Thank you!

Any questions?