

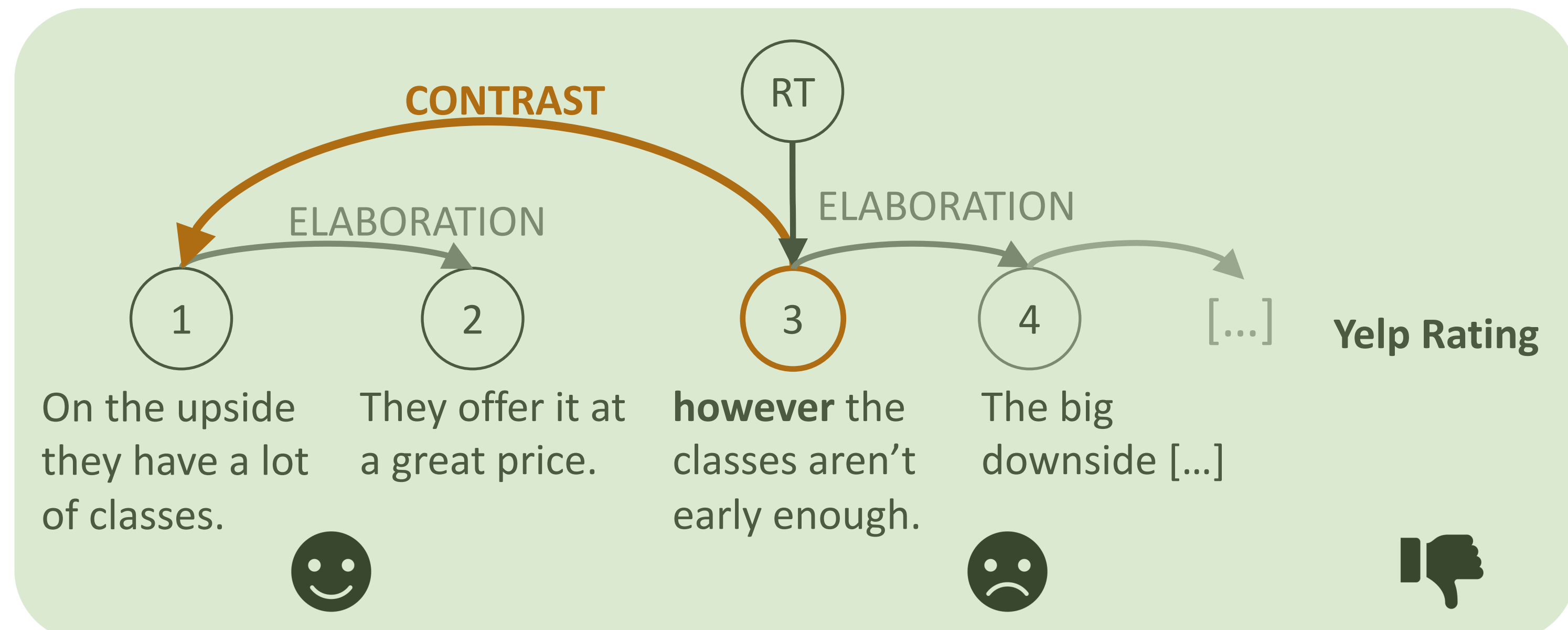
Elisa Ferracane, Greg Durrett, Jessy Li, Katrin Erk

elisa@ferracane.com, gdurrett@cs.utexas.edu, jessy@utexas.edu, katrin.erk@mail.utexas.edu

## Structured Text Representations

**Structured text representations** such as trees generated by Rhetorical Structure Theory (RST) are helpful for NLP end tasks including sentiment analysis.

### RST Dependency Tree:

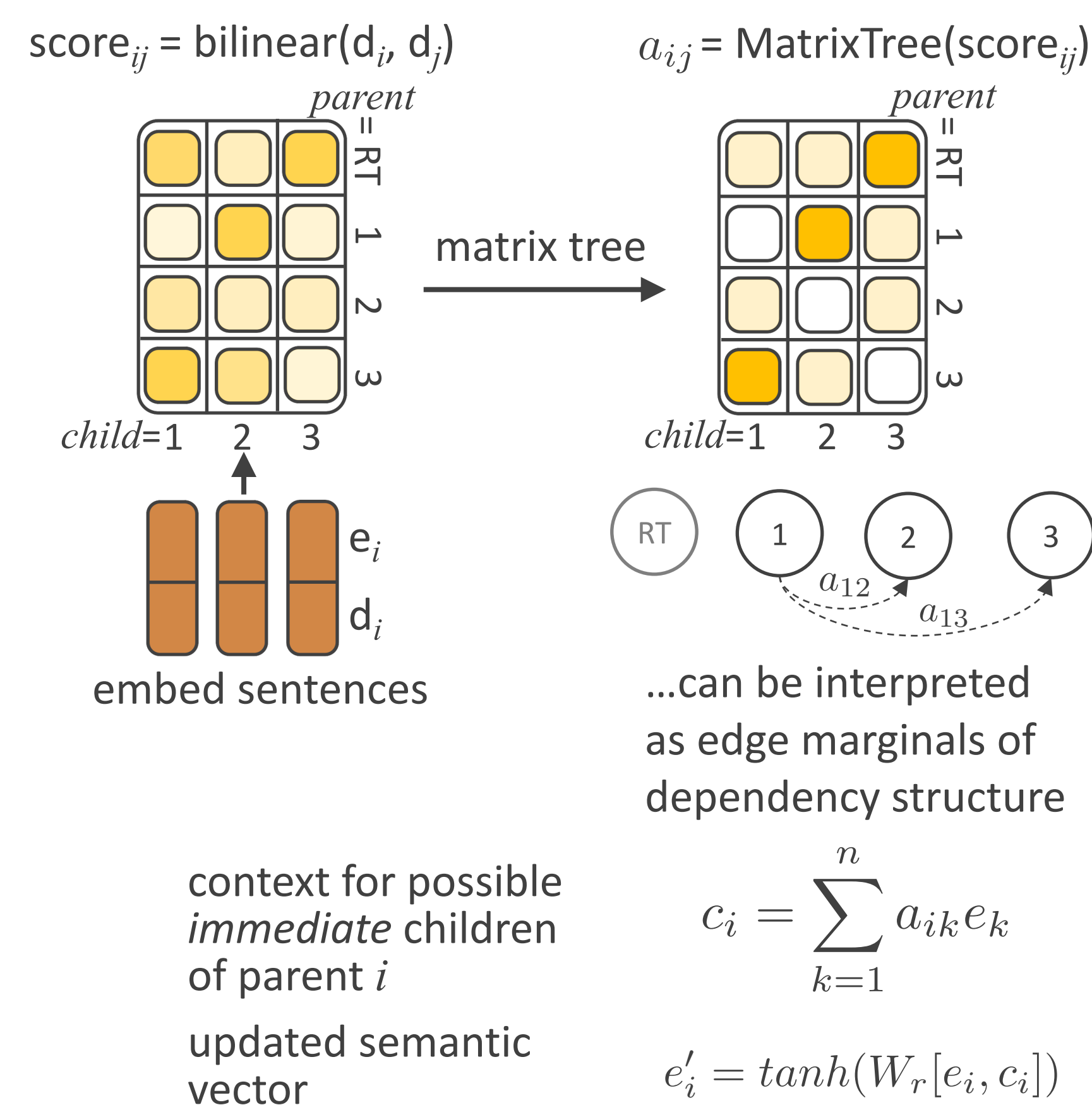


**Problem:** Although RST trees are linguistically defined and motivated, they are hard to exploit because annotations are scarce and limited by genre.

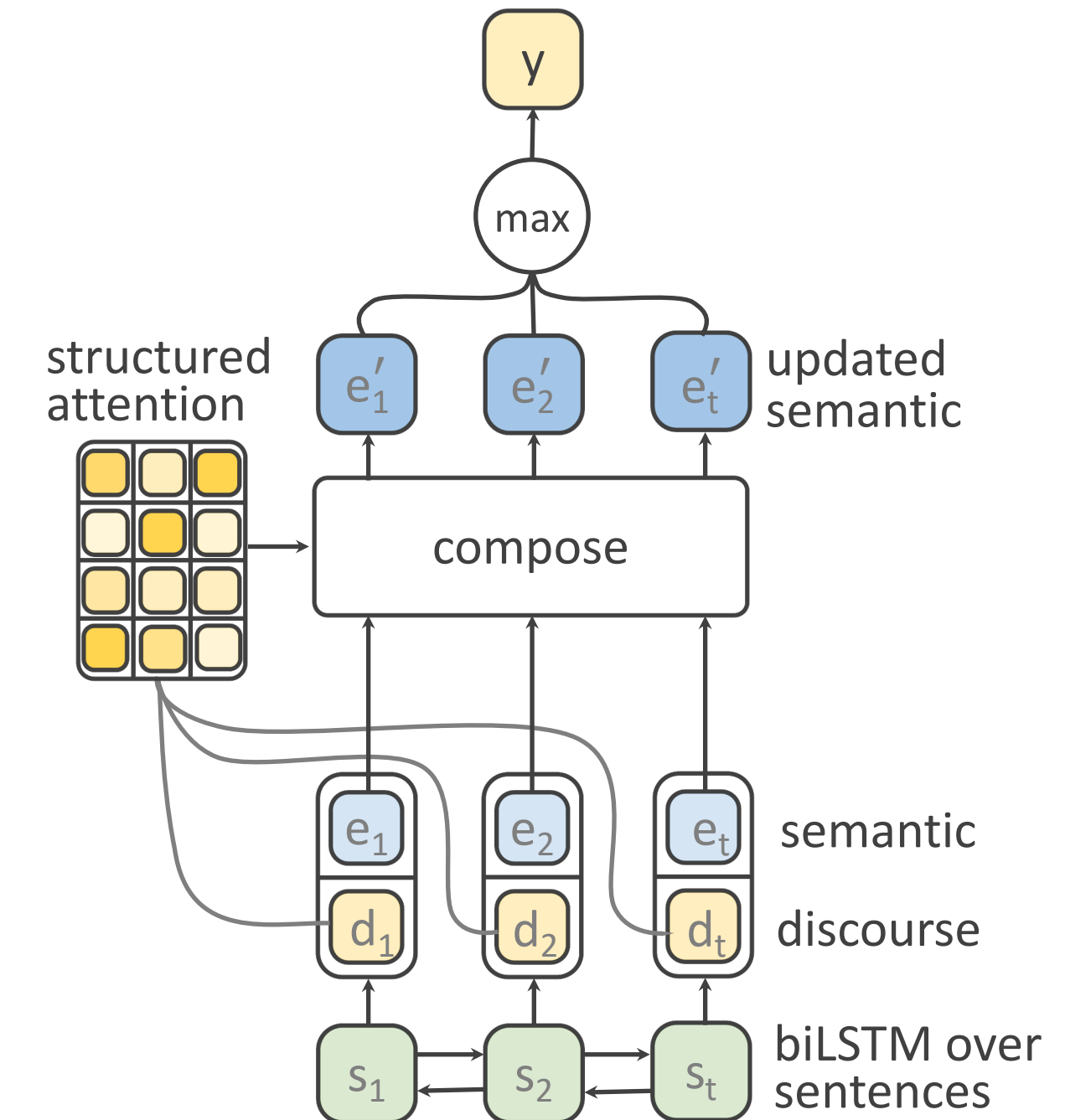
## Induced Trees

**Solution...?** Liu & Lapata (2018) train on text classification tasks and use structured attention to **induce** dependency trees over the text, akin to RST discourse dependency trees.

### Structured Attention



### Model



## Experiments

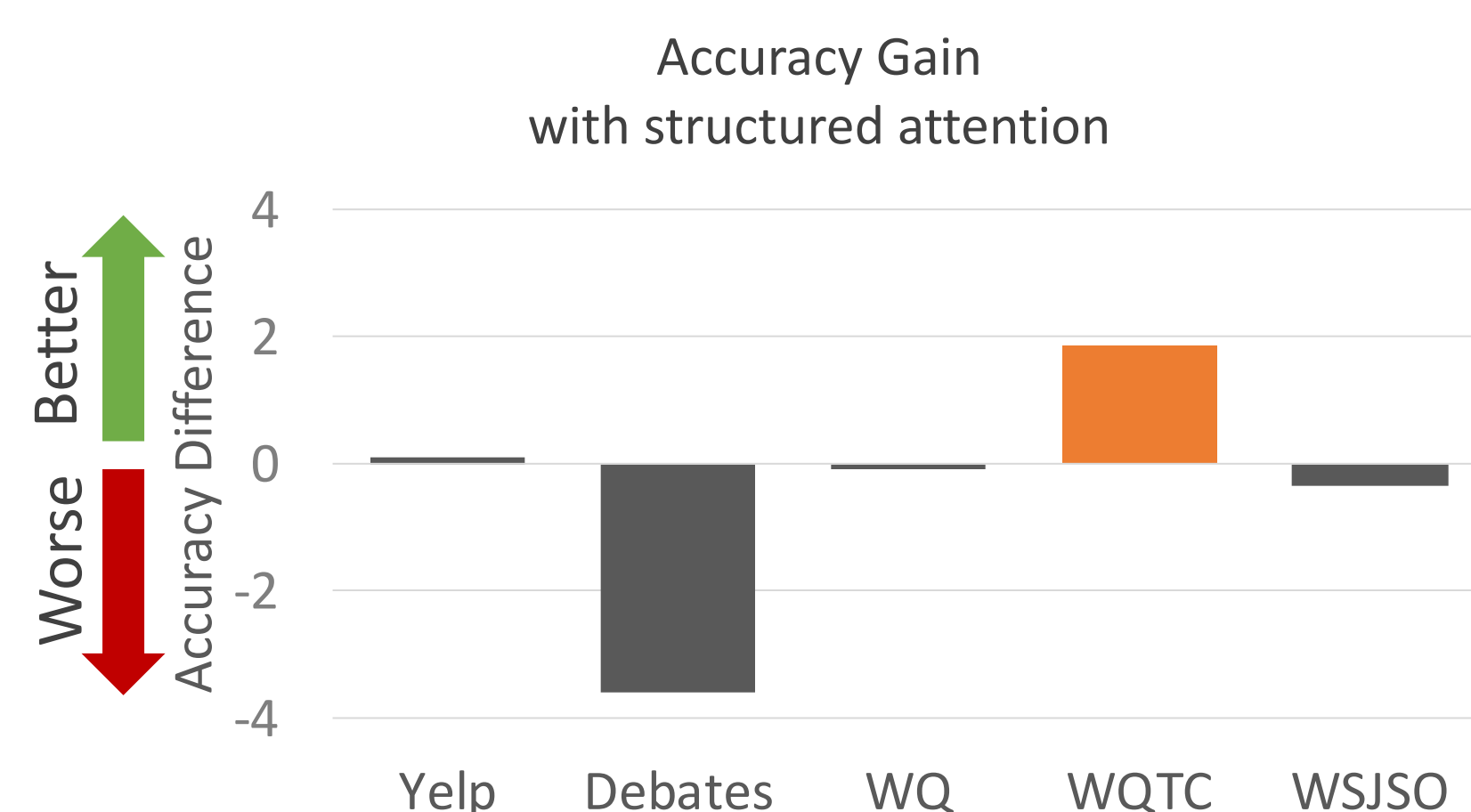
**Q** Does structured attention help?

**A** Generally, no. It only significantly helps one task.

Dataset	Label	Task
Yelp	1-5	Review sentiment
Debates	0/1	Vote prediction
Writing Quality (WQ)	0/1	Good writing
WQ Topic-Controlled (WQTC)	0/1	Good writing (topic-controlled)
WSJ Sentence Ordering (WSJSO)	-	Sentence order discrimination

**Setup:** Train 4x with different random seeds and report mean (see paper for std dev, max)

■ Performance gain with structured attention:



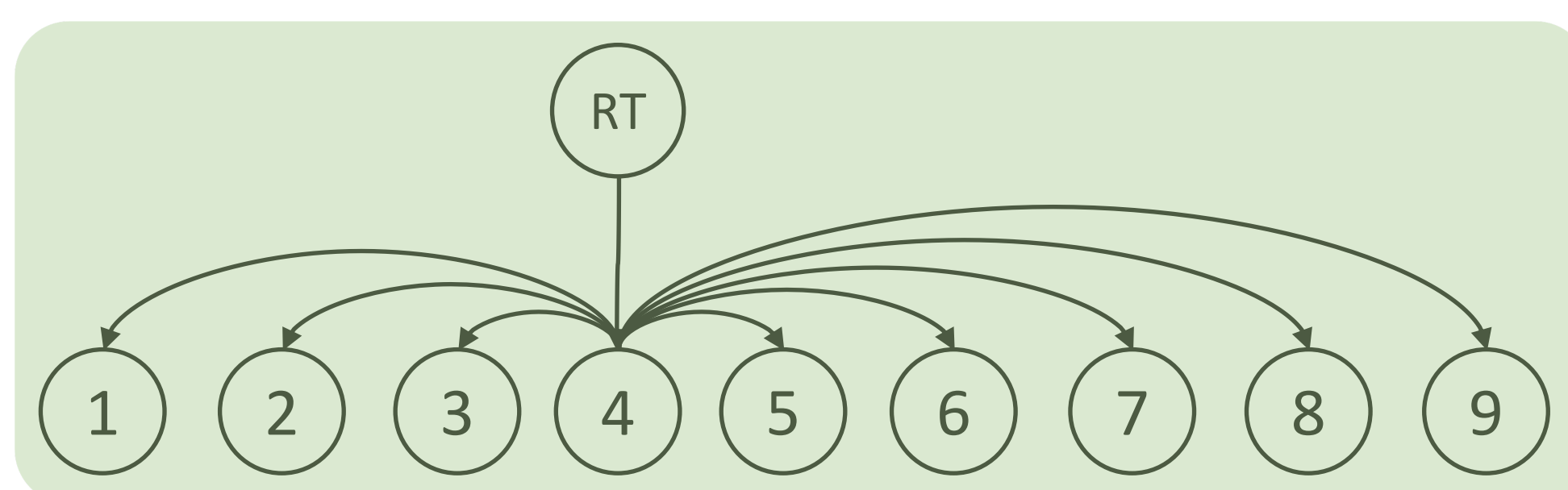
Adding structured attention **helps only WQTC**. On Yelp, WQ, WSJSO, there is **no difference**. On Debates, the attention **hurts**.

**Q** Do the induced trees learn discourse?

**A** No. The model focuses on lexical cues.

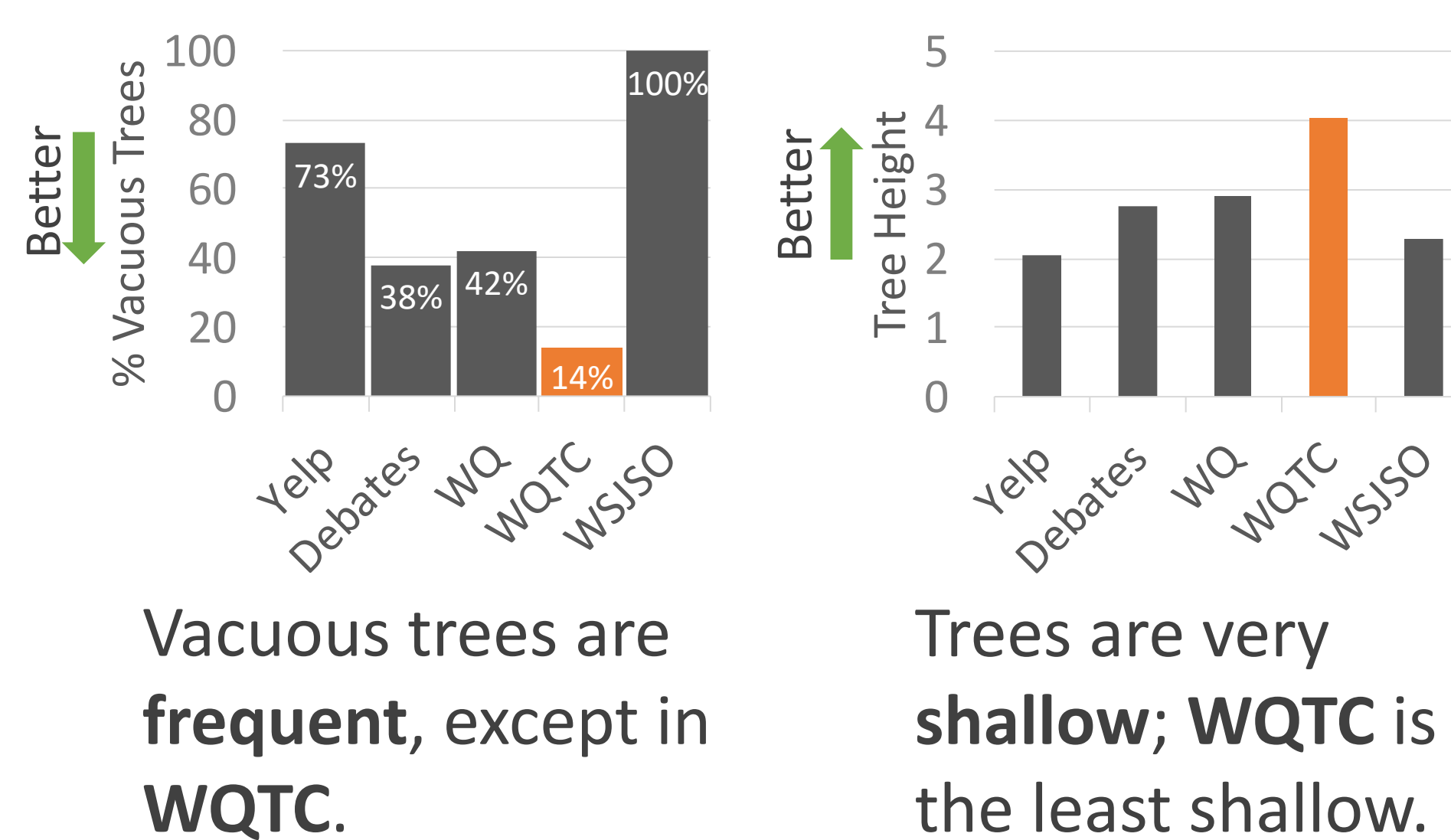
■ Tree analysis:

**Induced Tree:** (same Yelp review as above)



**Vacuous tree:** Flat, uninformative discourse structure. Root is one sentence at beginning (or end) of text, and all other sentences are children.

■ Tree statistics:



Vacuous trees are **frequent**, except in WQTC.

Trees are very **shallow**; WQTC is the least shallow.

■ Root sentence analysis:

Yelp	uuu, Sterne, star, rating, deduct, 0, edit
Debates	oppose, republican, majority, thank
WQ	valley, mp3, firm, capital, universal

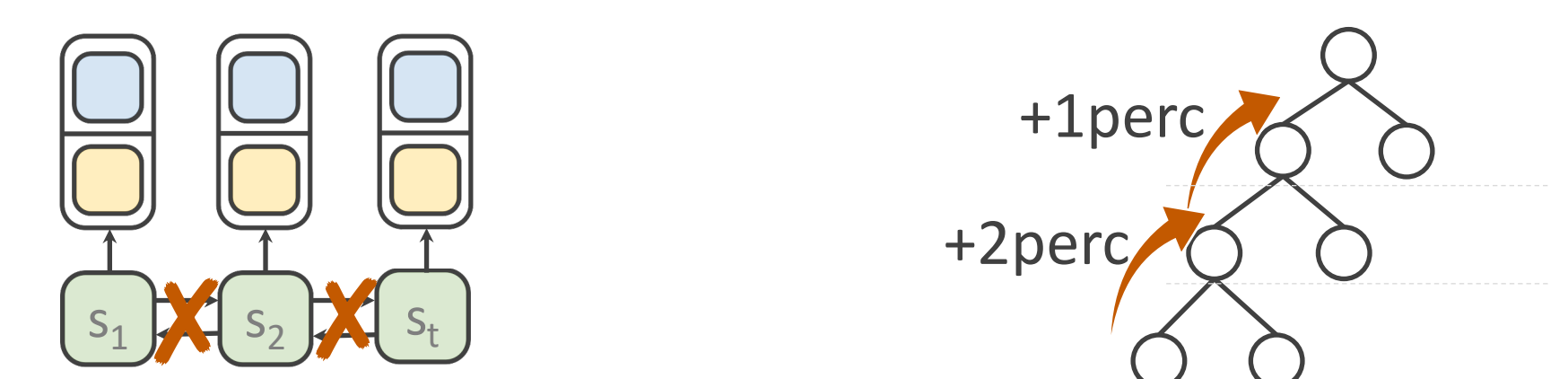
Top PPMI words in root sentence are **indicative of label**: rating or sentiment (Yelp), stance or politeness (Debates), topic (WQ).

**Q** Can we learn better structure?

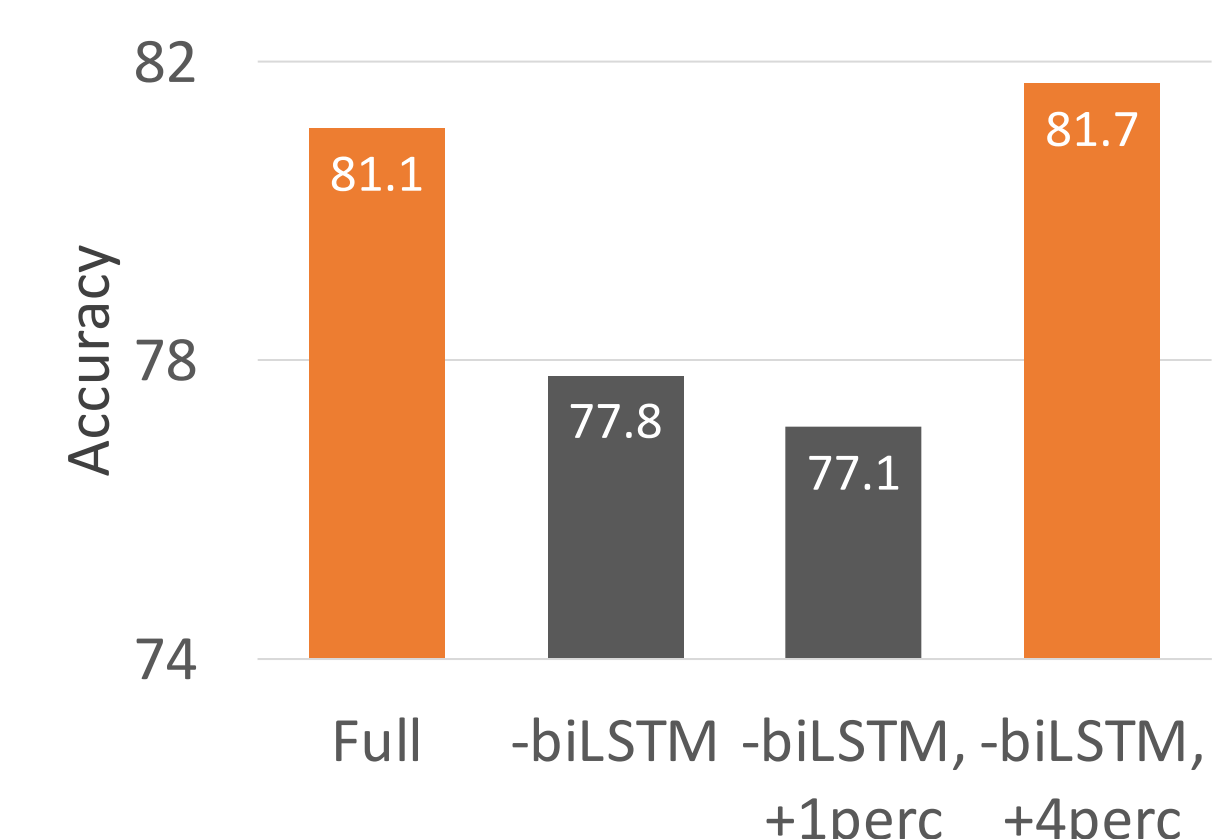
**A** No. Model changes induce better trees, but are still far from discourse.

■ Model modifications to increase reliance on structure:

a. remove biLSTM over sentences (-biLSTM)  
b. percolate context from children of *subtrees*  $n$  levels down (+ $n$ perc)

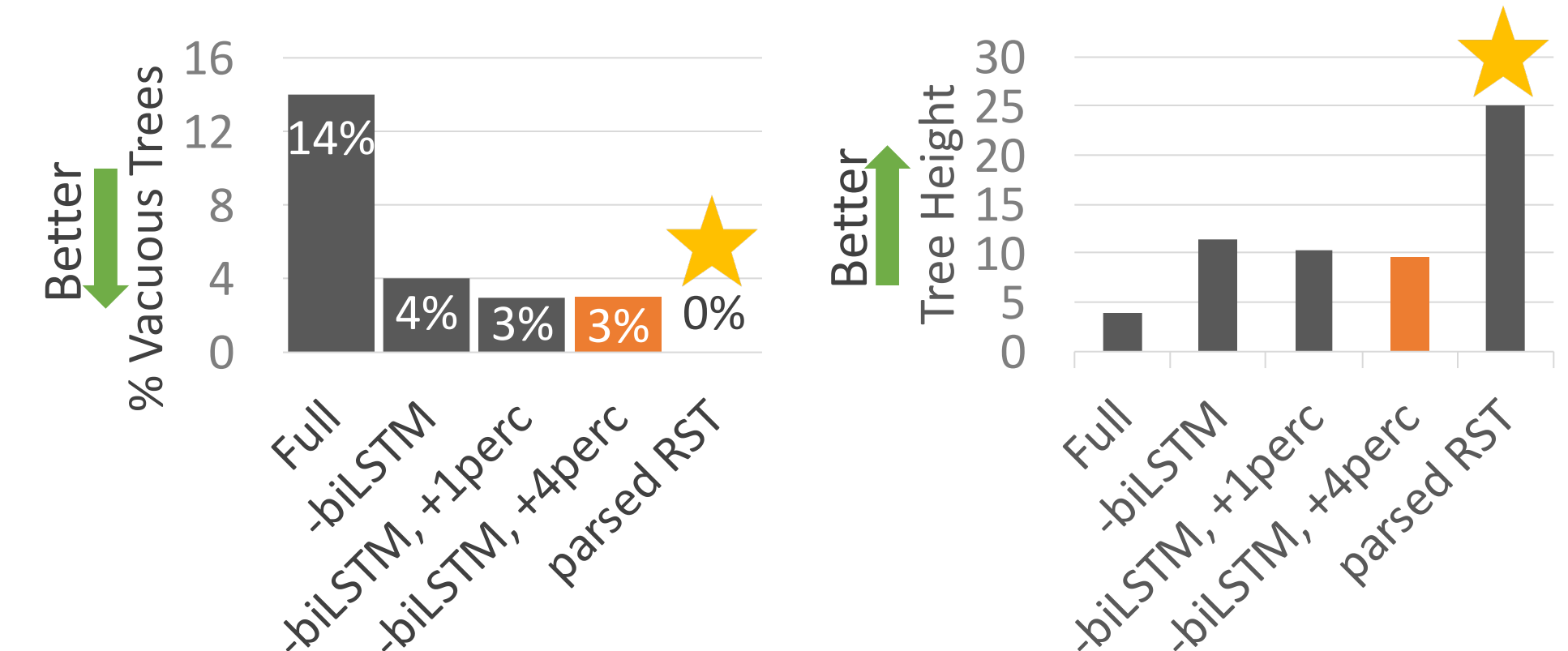


■ Performance:



Removing biLSTM + adding 4 levels of children percolation yields **similar performance** as Full model.

■ Tree statistics:



Best model produces **less vacuous** and **deeper** trees, but these are **still far from 'gold'** parsed RST discourse dependency trees.