

# Learnability-based Syntactic Annotation Design

Roy Schwartz, Omri Abend  
and Ari Rappoport

The Hebrew University  
In proceedings of COLING 2012



# Overview

- In many cases, there is more than one plausible way to annotate syntactic structures
  - A single annotation must be selected
- We propose **learnability** as a selection criterion
  - A principled learnability-based methodology
  - Use parsers for annotation design
- Selecting the more learnable annotation may result in up to **35.3%** error reduction in parsing performance



# Different Syntactic Formalisms

## Example

- I want to eat the apple



# Different Syntactic Formalisms

## Example

- I want to eat the apple

PRP	VBP	TO	VB	DT	NN
I	want	to	eat	the	apple



# Different Syntactic Formalisms

## Example

- I want to eat the apple

PRP	VBP	TO	VB	DT	NN
I	want	to	eat	the	apple

(S (NP I) (VP (VP want) (VP to (VP eat (NP the apple))))))



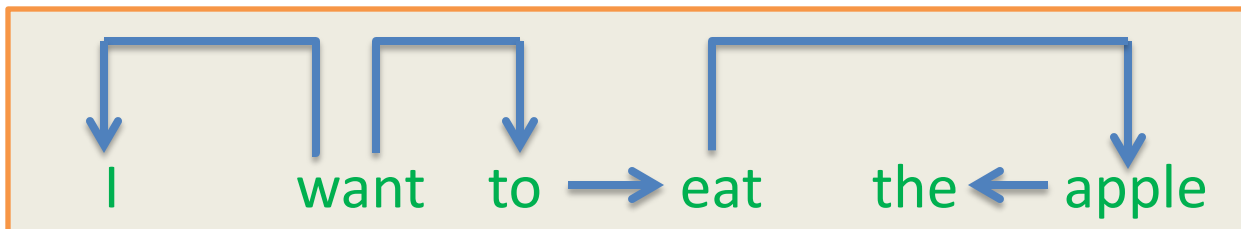
# Different Syntactic Formalisms

## Example

- I want to eat the apple

PRP	VBP	TO	VB	DT	NN
I	want	to	eat	the	apple

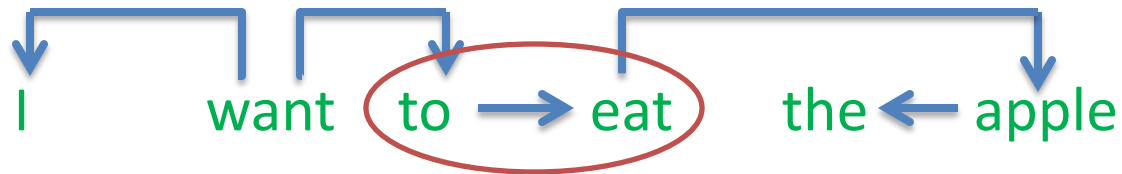
(S (NP I) (VP (VP want) (VP to (VP eat (NP the apple))))))



# Disagreement within the same Formalism

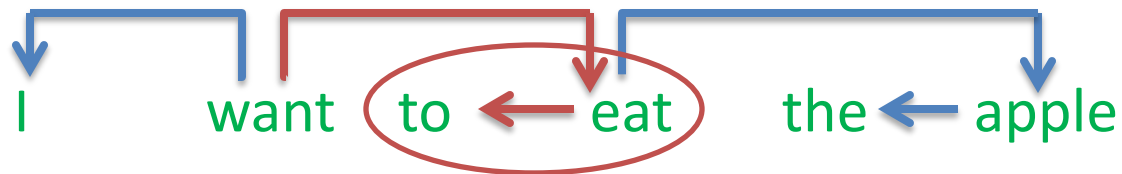
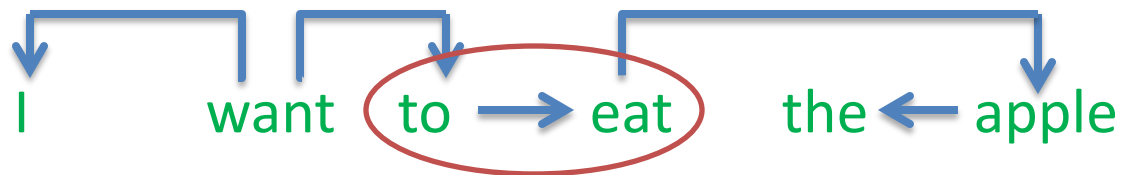


# Disagreement within the same Formalism

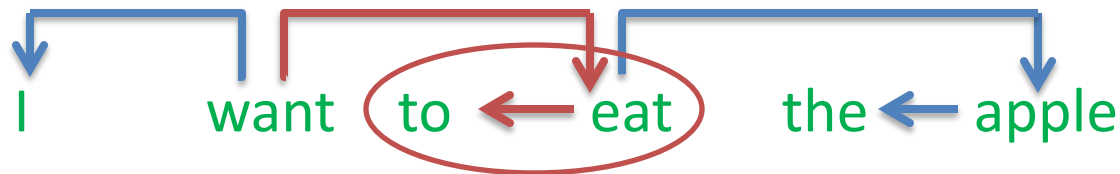
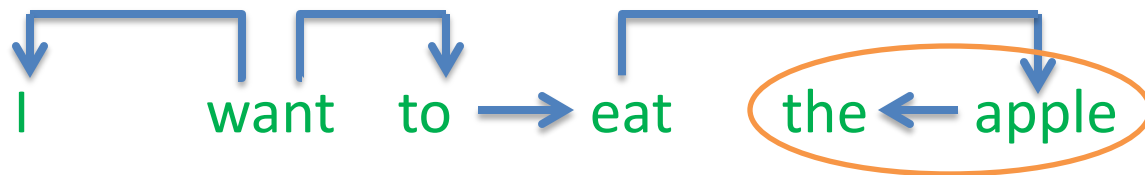




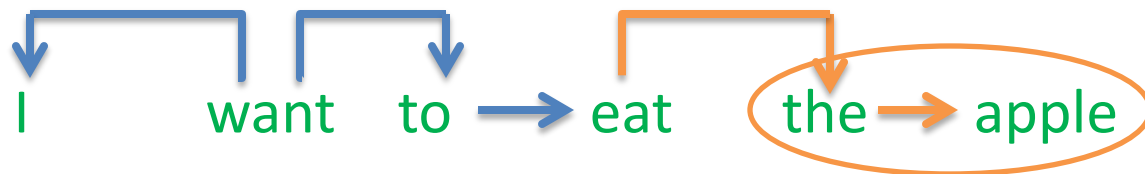
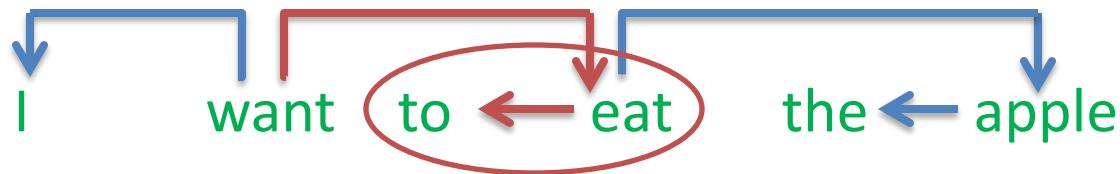
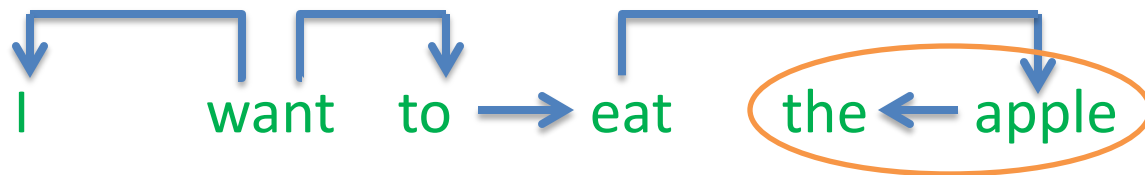
# Disagreement within the same Formalism



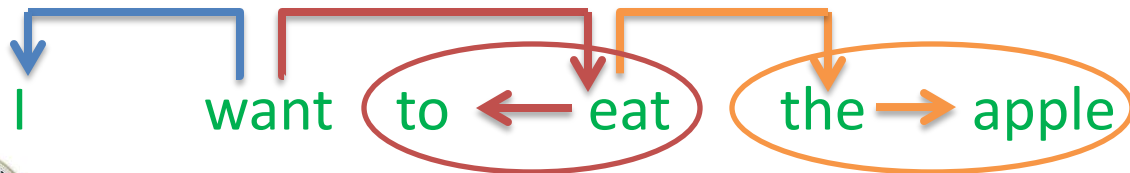
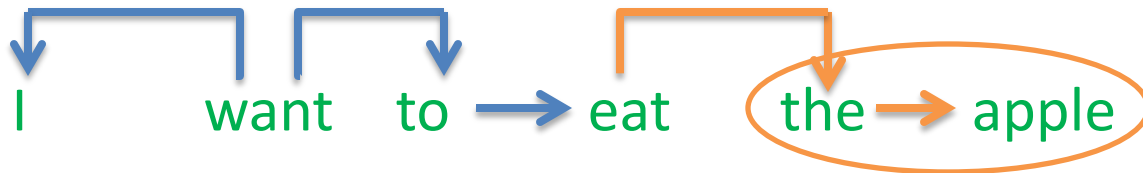
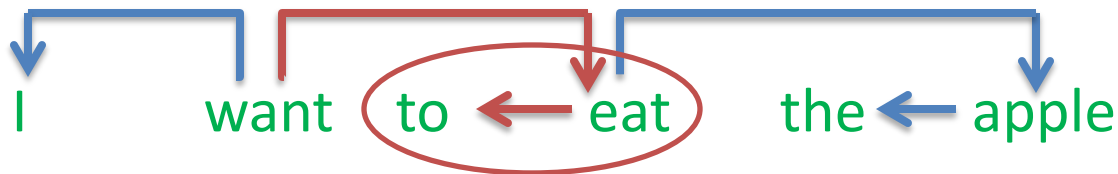
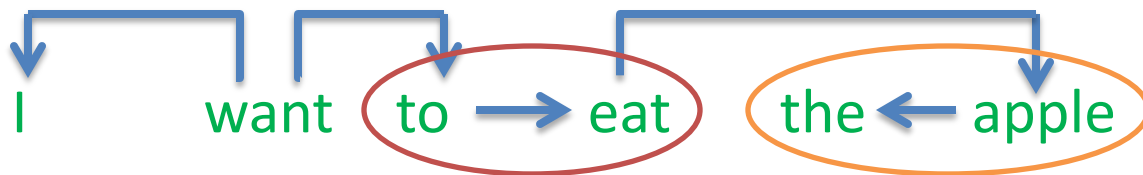
# Disagreement within the same Formalism



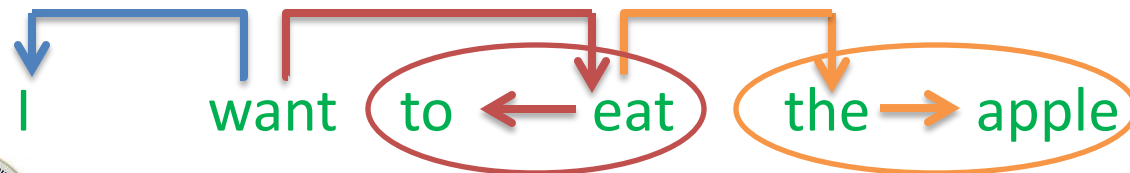
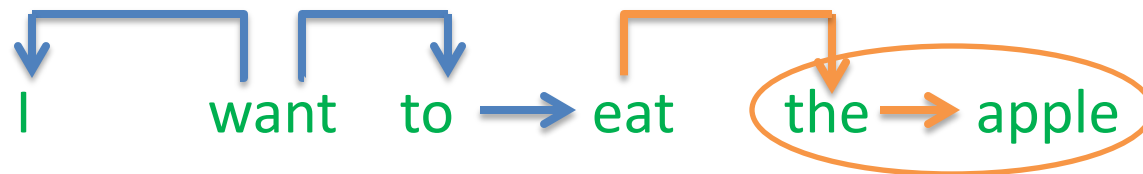
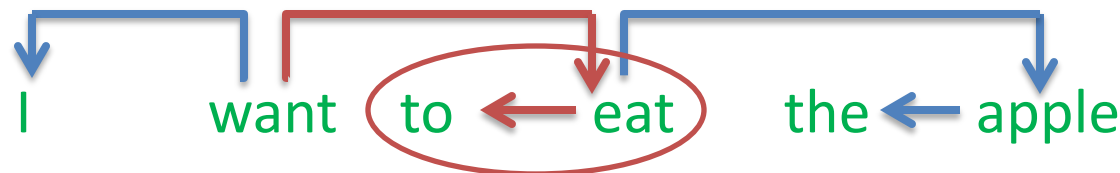
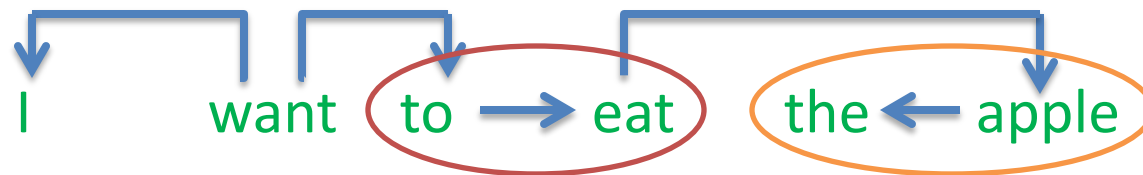
# Disagreement within the same Formalism



# Disagreement within the same Formalism



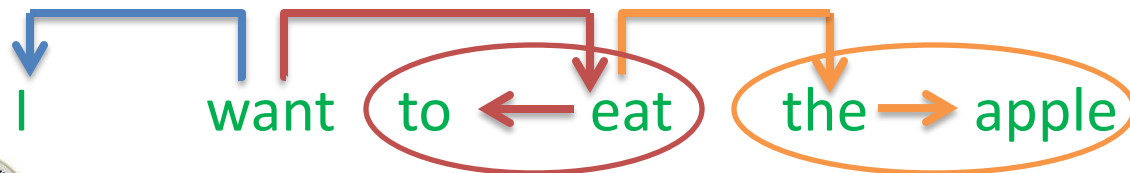
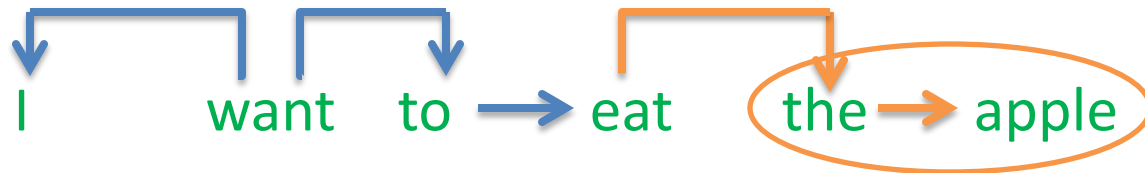
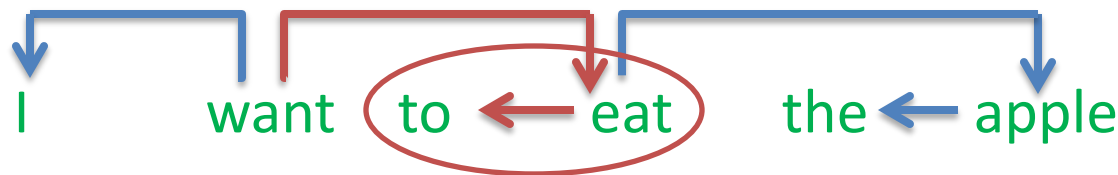
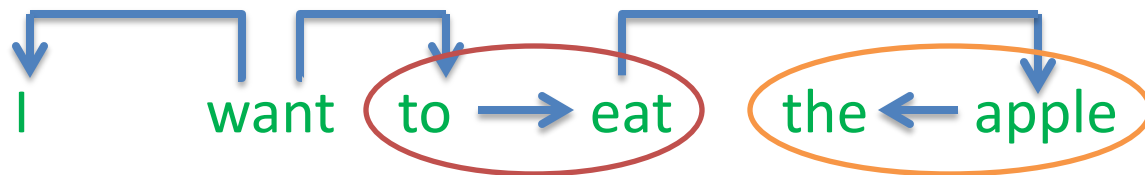
# Disagreement within the same Formalism



Varying Syntactic Structures (VSS)



# Disagreement within the same Formalism



Varying Syntactic Structures (VSS)



Annotation scheme



# Varying Syntactic Structures (VSS)

- VSSs are very frequent
  - More than 40% of the tokens in PTB participate in at least one VSS\*

\* Schwartz et al., ACL 2011



# Varying Syntactic Structures (VSS)

- VSSs are very frequent
  - More than 40% of the tokens in PTB participate in at least one VSS\*
- Evaluation Problems
  - Different parsers train and evaluate against different annotation schemes
- Selecting one alternative over the other in a VSS can affect the performance of a specific parser\*\*

\* Schwartz et al., ACL 2011

\*\* Nilsson et al., ACL 2006





# Performance Differences

- Learning the correct annotation for a VSS is easy
  - Usually the direction of a single edge

the<sub>DT</sub> ↔ apple<sub>NN</sub>



# Performance Differences

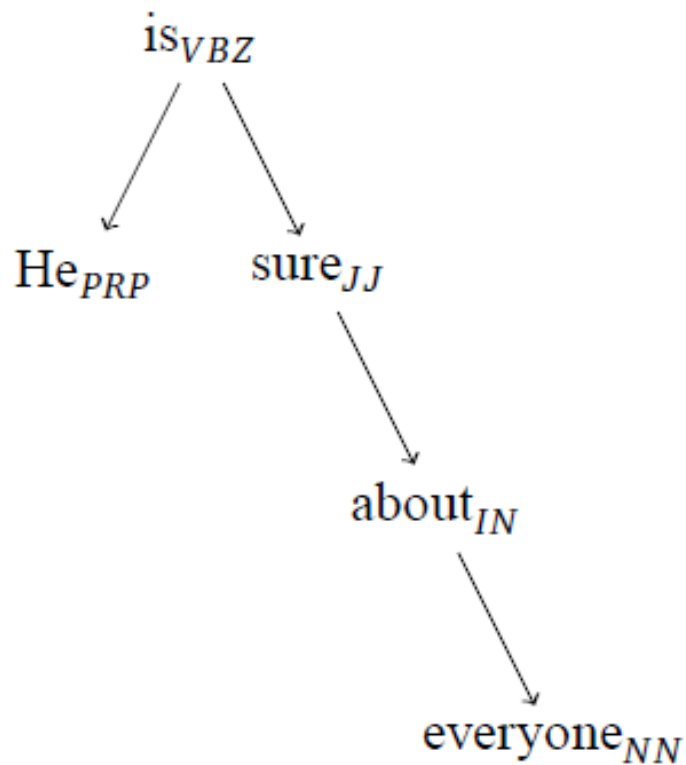
- Learning the correct annotation for a VSS is easy
  - Usually the direction of a single edge
- An annotation scheme is learned **as a whole**
  - Not each VSS alone
- There are **second order** effects
  - The way in which the VSS attaches to the rest of the sentence
  - These can lead to **performance differences**

the<sub>DT</sub> ↔ apple<sub>NN</sub>



# Example

## Performance Differences

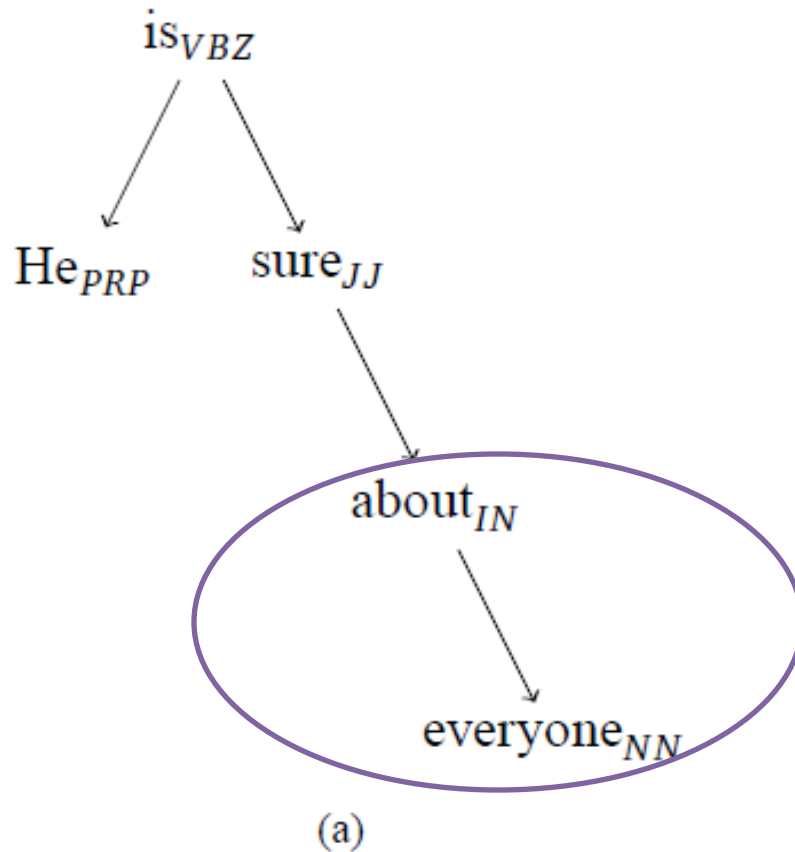


(a)



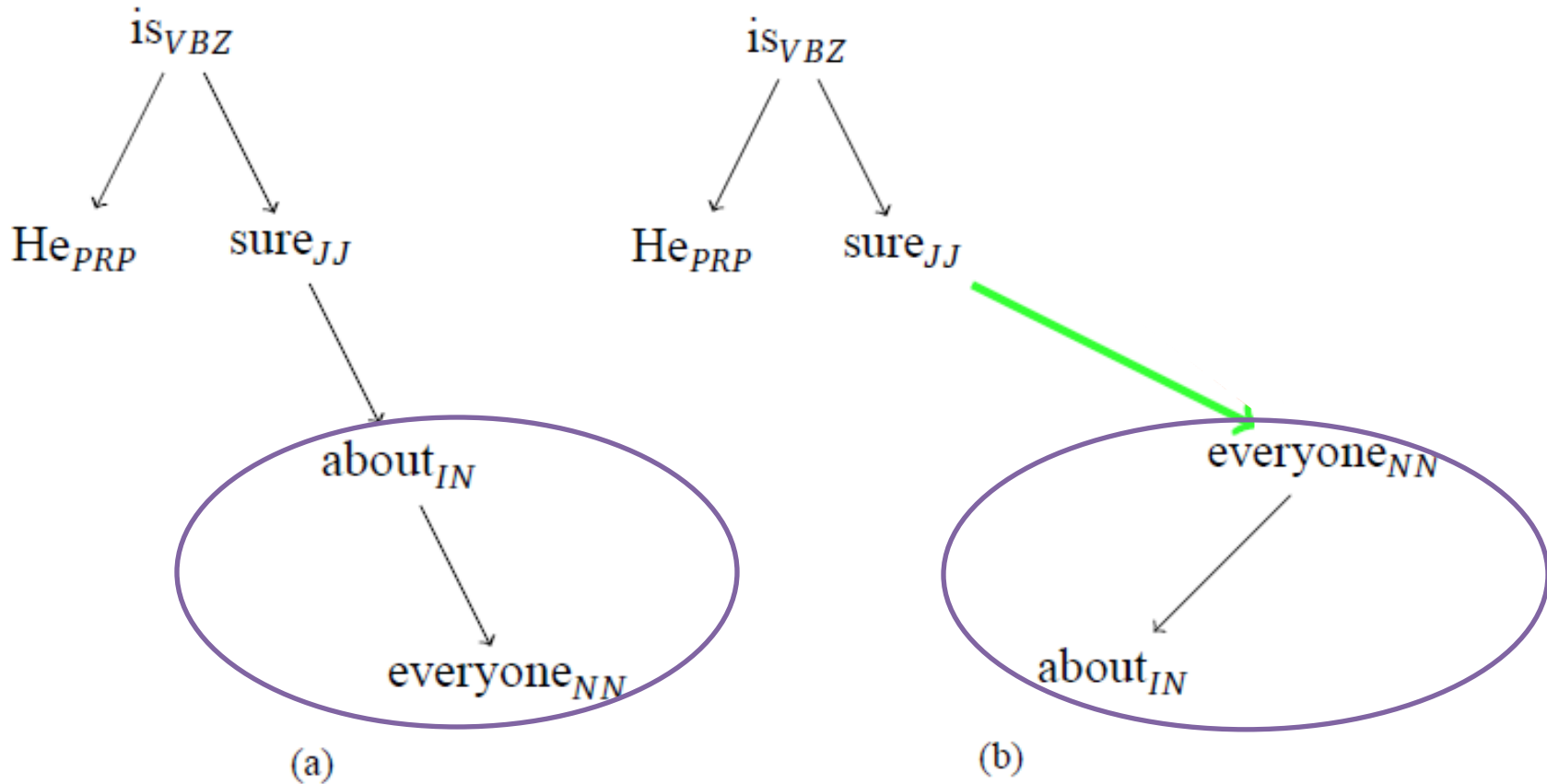
# Example

## Performance Differences



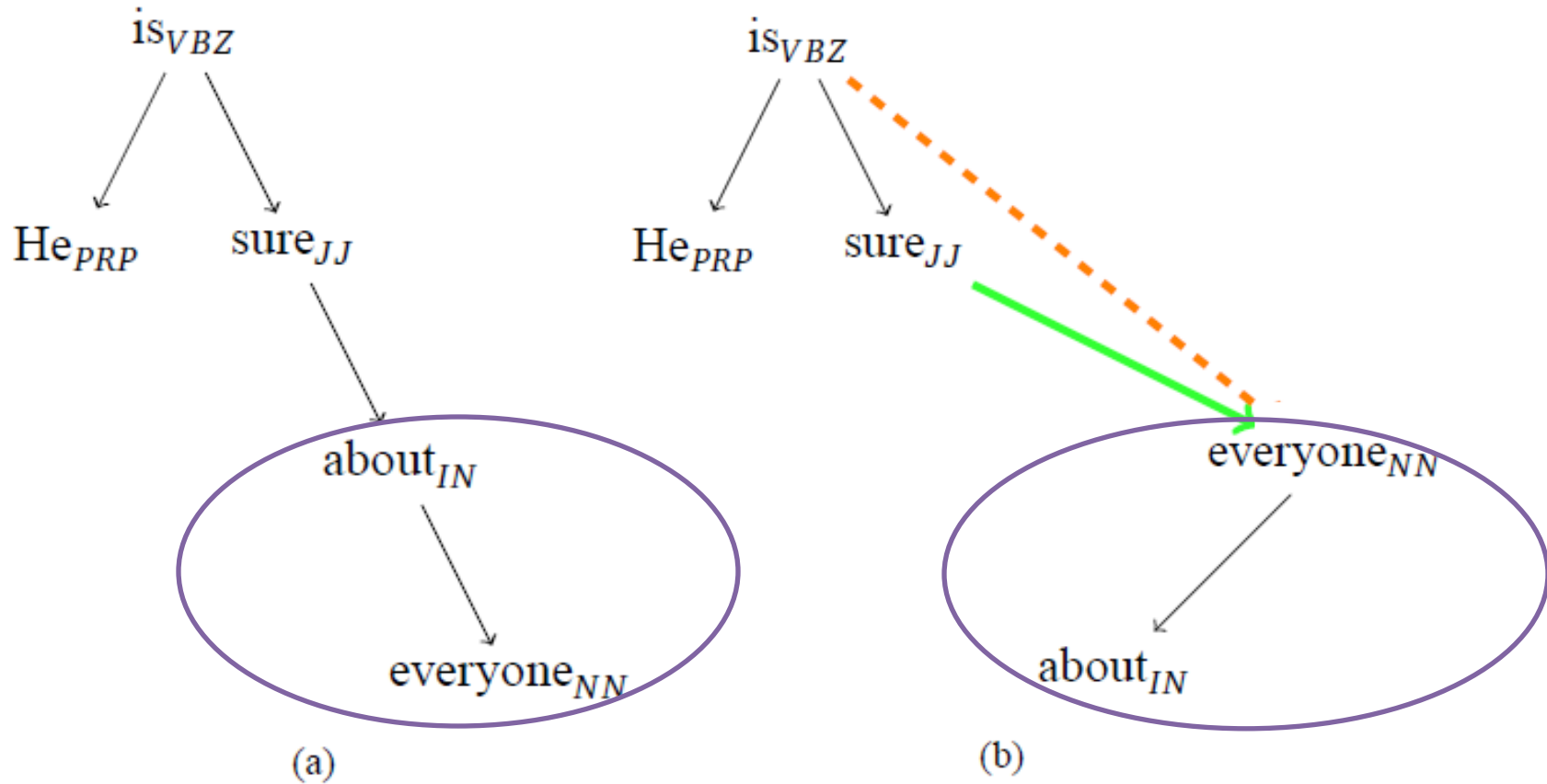
# Example

## Performance Differences



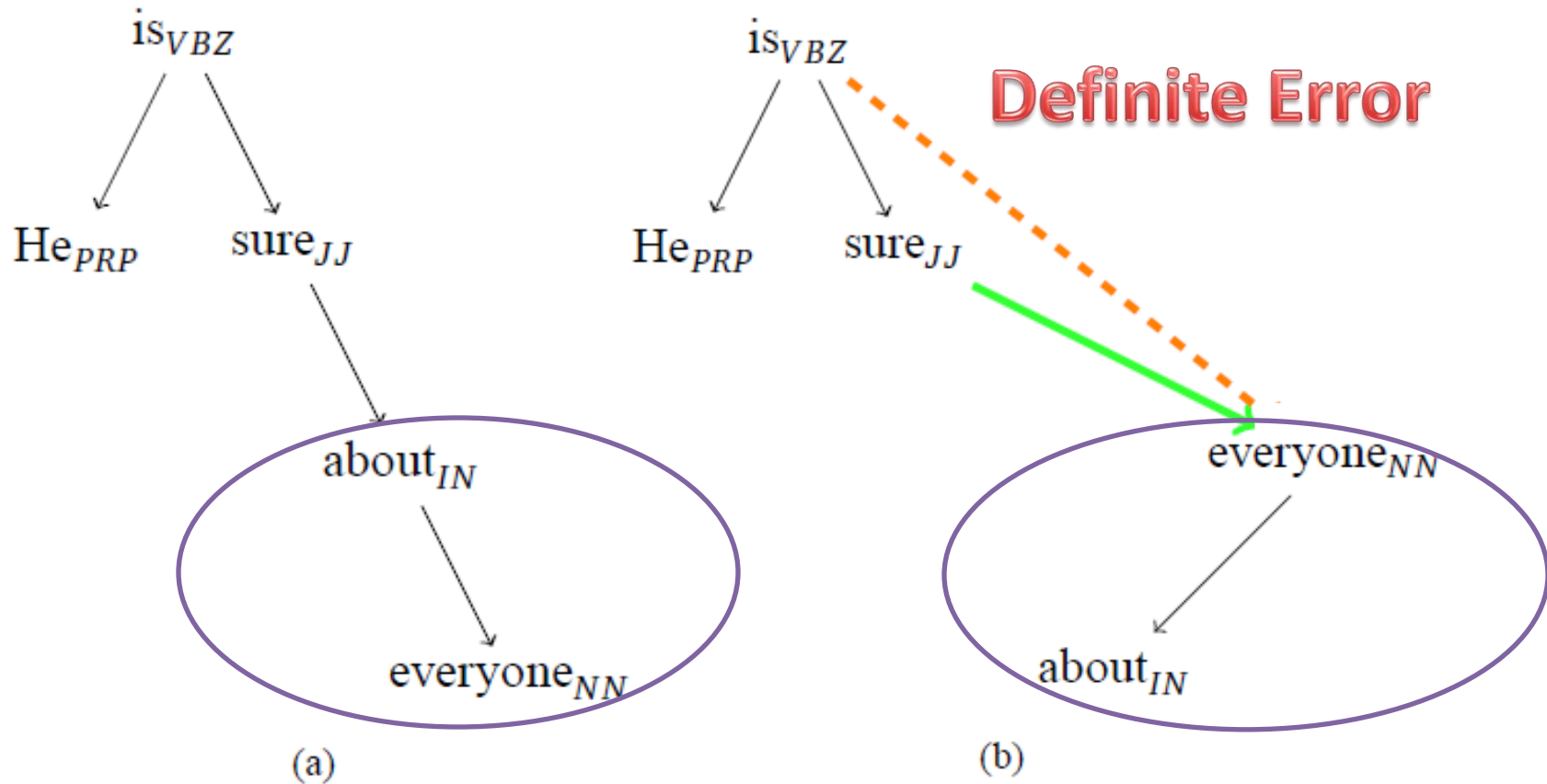
# Example

## Performance Differences



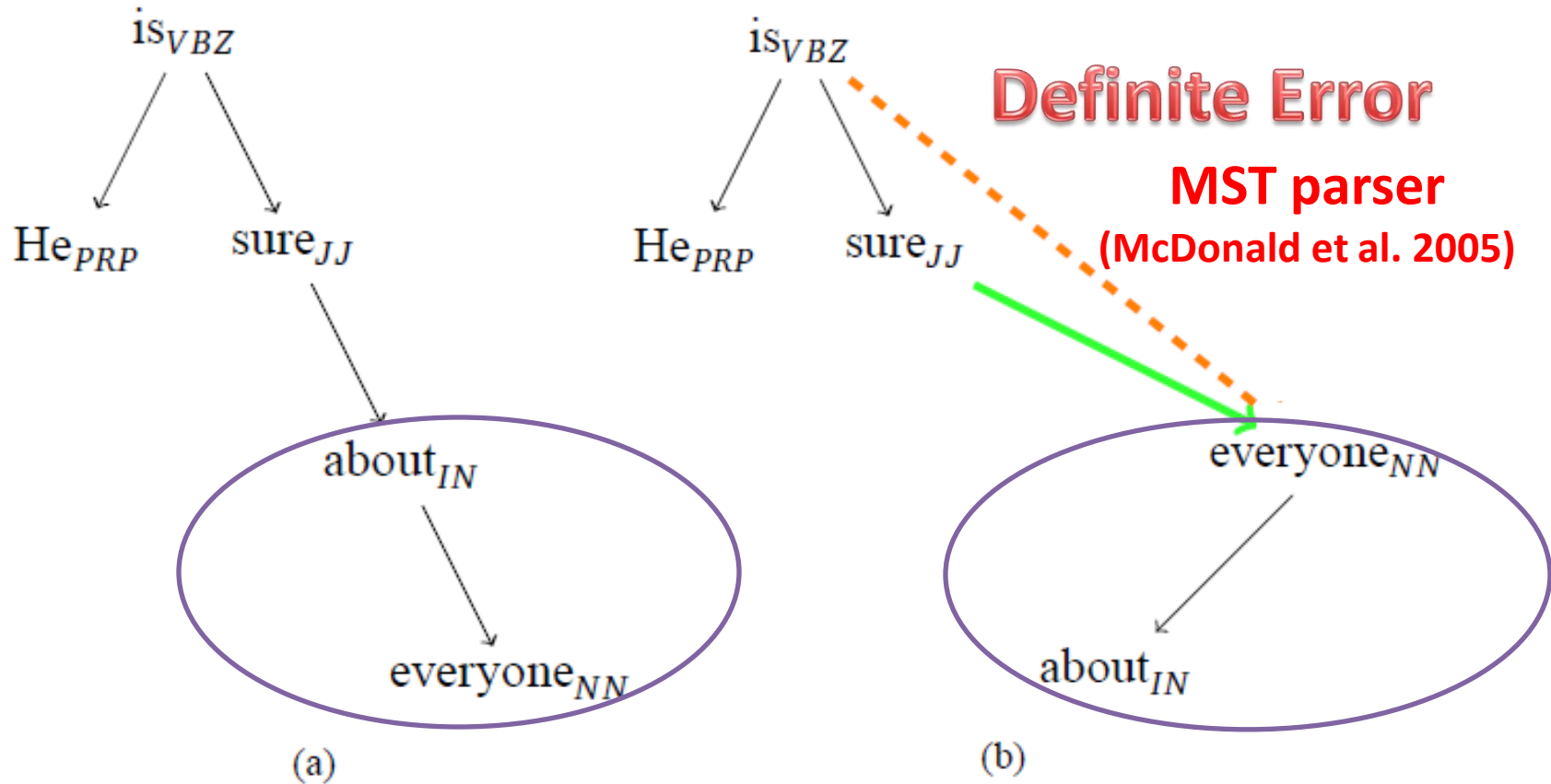
# Example

## Performance Differences



# Example

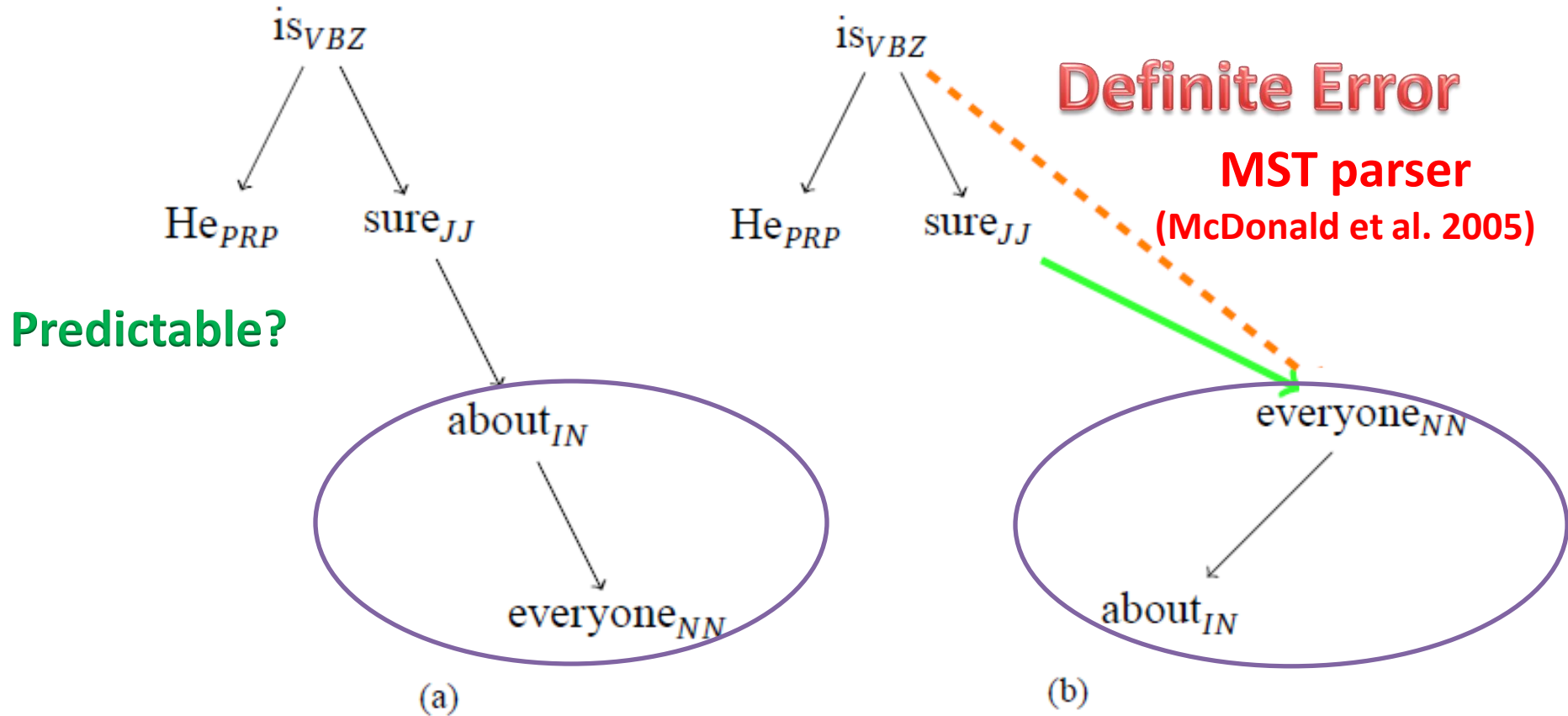
## Performance Differences





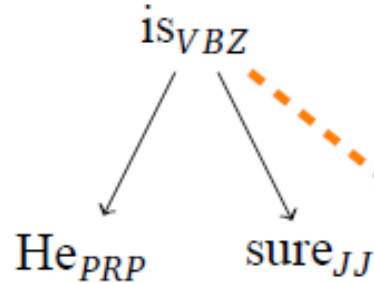
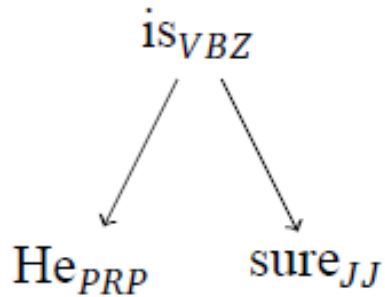
# Example

## Performance Differences



# Example

## Performance Differences



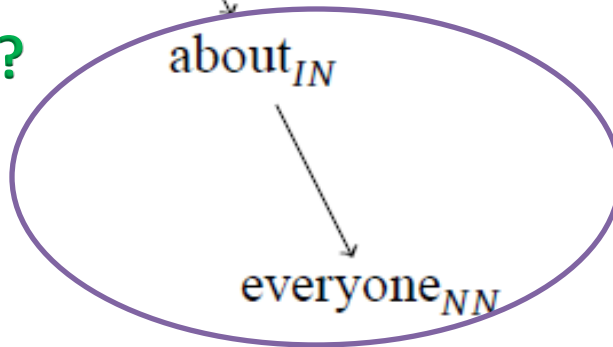
**Definite Error**

**MST parser**

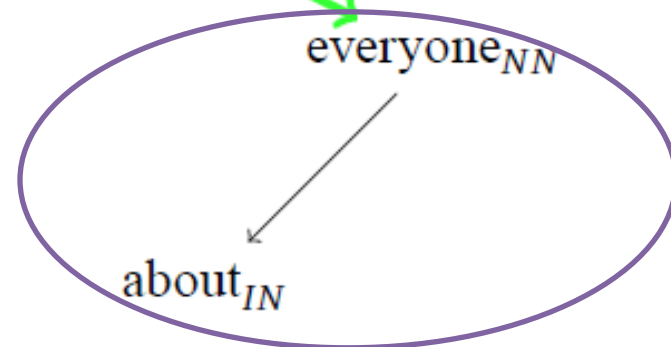
(McDonald et al. 2005)

Predictable?

Parser specific?



(a)

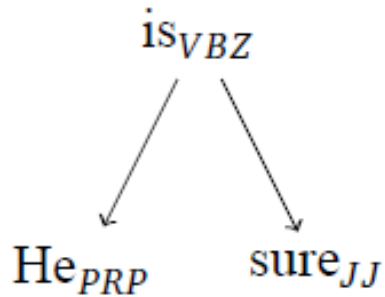


(b)

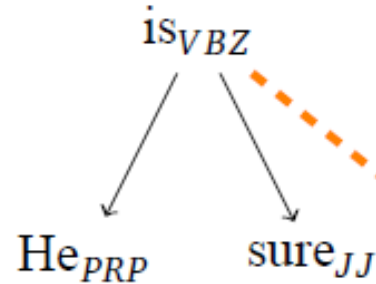


# Example

## Performance Differences



(a)



**Definite Error**

**MST parser**

(McDonald et al. 2005)

(b)

Predictable?

Parser specific?

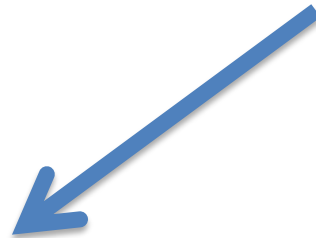
Magnitude?



# Varying Syntactic Structures (VSS)



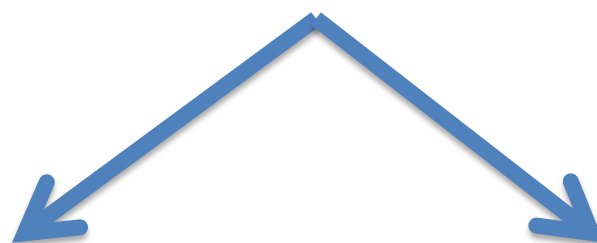
# Varying Syntactic Structures (VSS)



**Selecting one  
Annotation is Required**



# Varying Syntactic Structures (VSS)



**Selecting one  
Annotation is Required**

**Selection Can Affect  
Parsing Performance**



# Our Solution: Learnability

- A straightforward selection criterion
  - Namely, **how easy it is to learn** a given annotation scheme using statistical tools
- Learnability is justified practically
  - Training on more learnable schemes results in more accurate parsers
- This criterion is only applied on linguistically **plausible** annotations



# Learnability

- Learnability is widely used theoretically
  - Learnability using distributional methods has been used as an important consideration in designing the **phrase structure** formalism\*
  - It is also used recurrently in cognitive science\*\*

\* Chomsky 2006

\*\* Chater and Vitányi 2003, Perfors et al. 2011



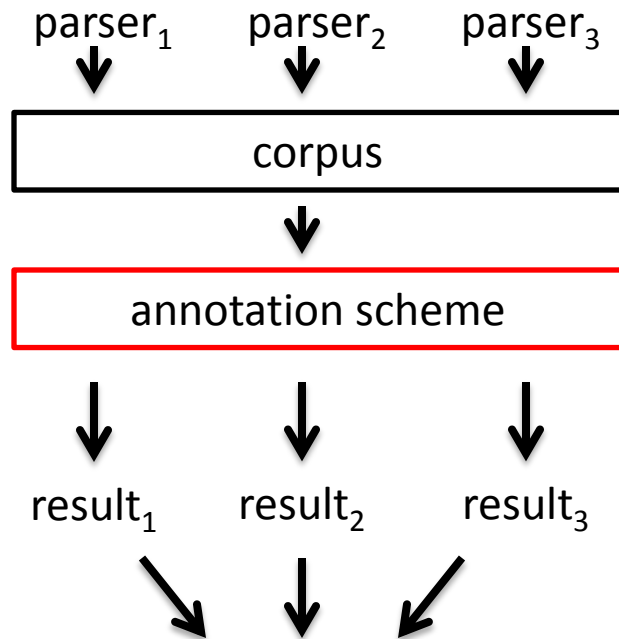


# Learnability-based Methodology 1



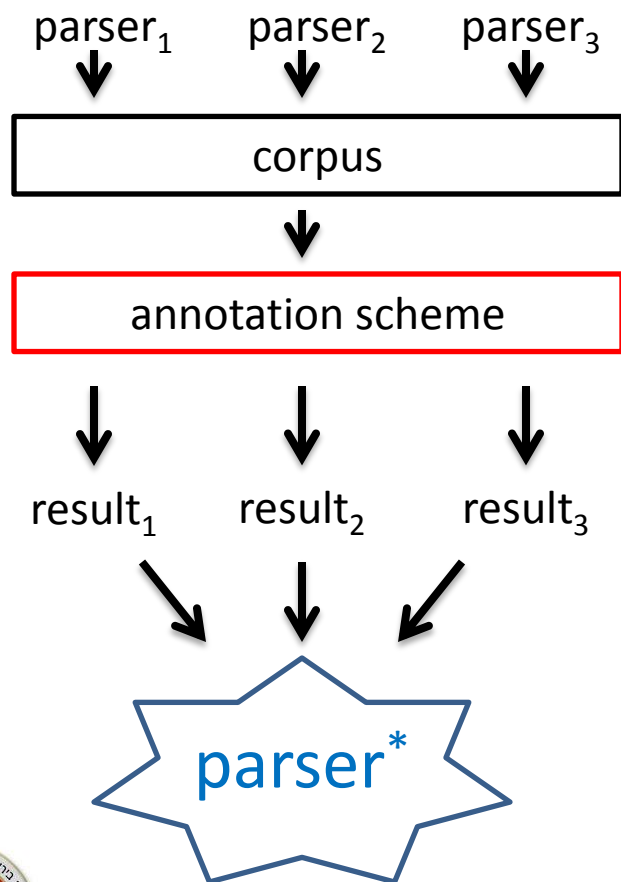
# Learnability-based Methodology 1

## Standard Parsing Evaluation



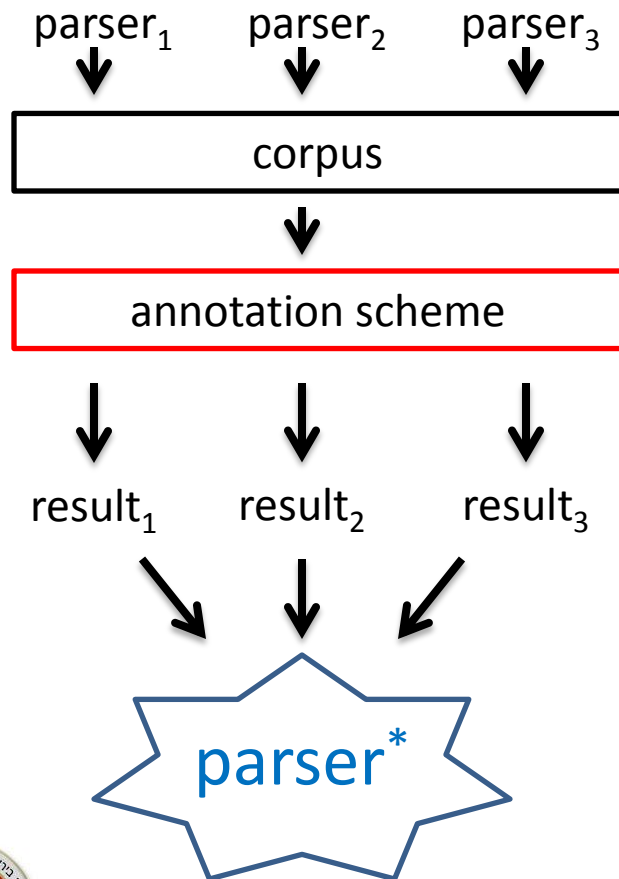
# Learnability-based Methodology 1

## Standard Parsing Evaluation

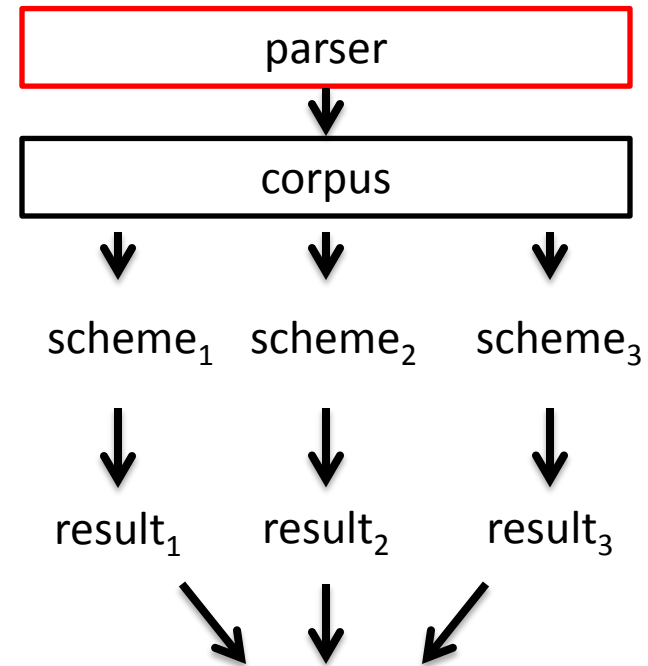


# Learnability-based Methodology 1

## Standard Parsing Evaluation

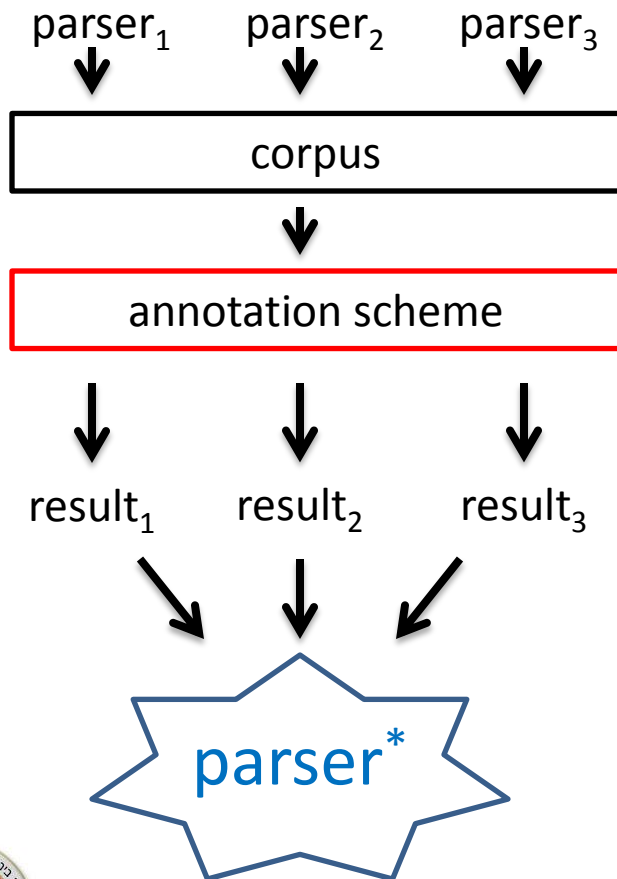


## Our Approach

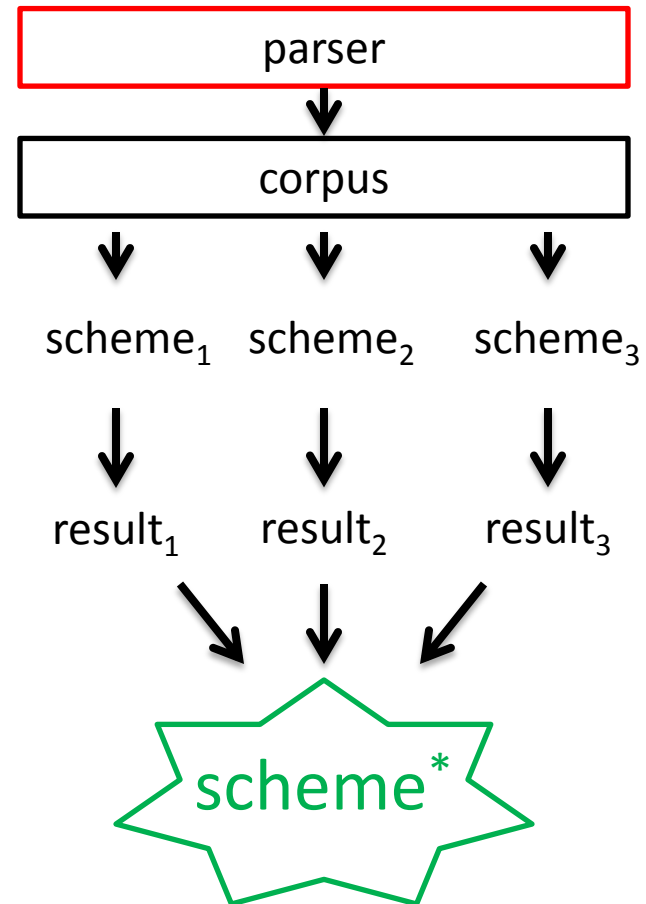


# Learnability-based Methodology 1

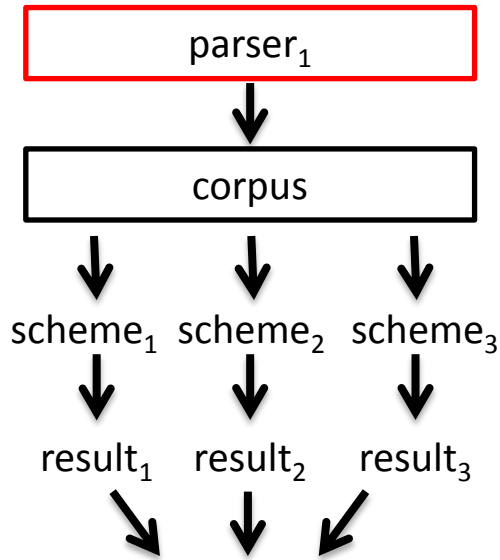
## Standard Parsing Evaluation



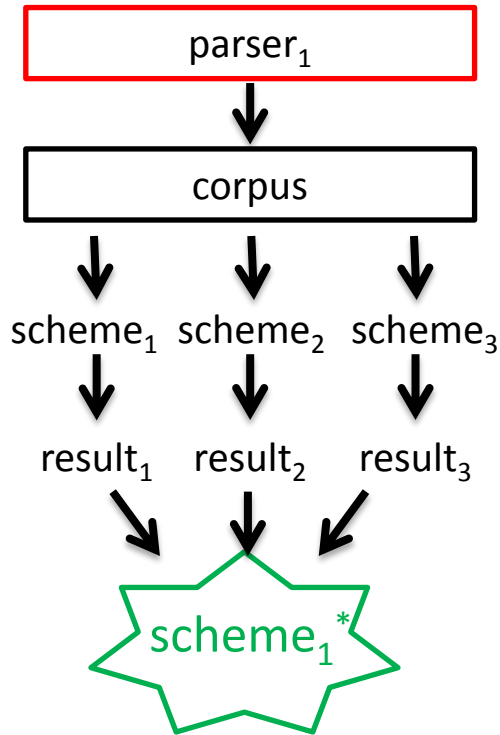
## Our Approach



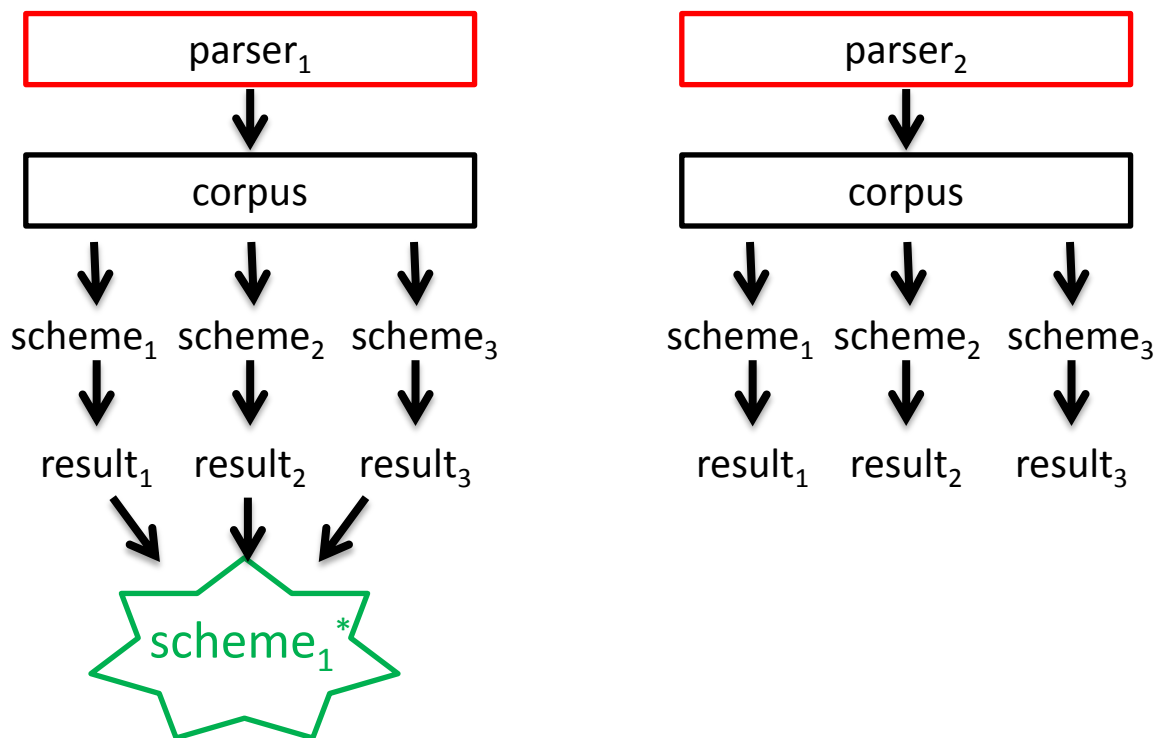
# Learnability-based Methodology 2



# Learnability-based Methodology 2

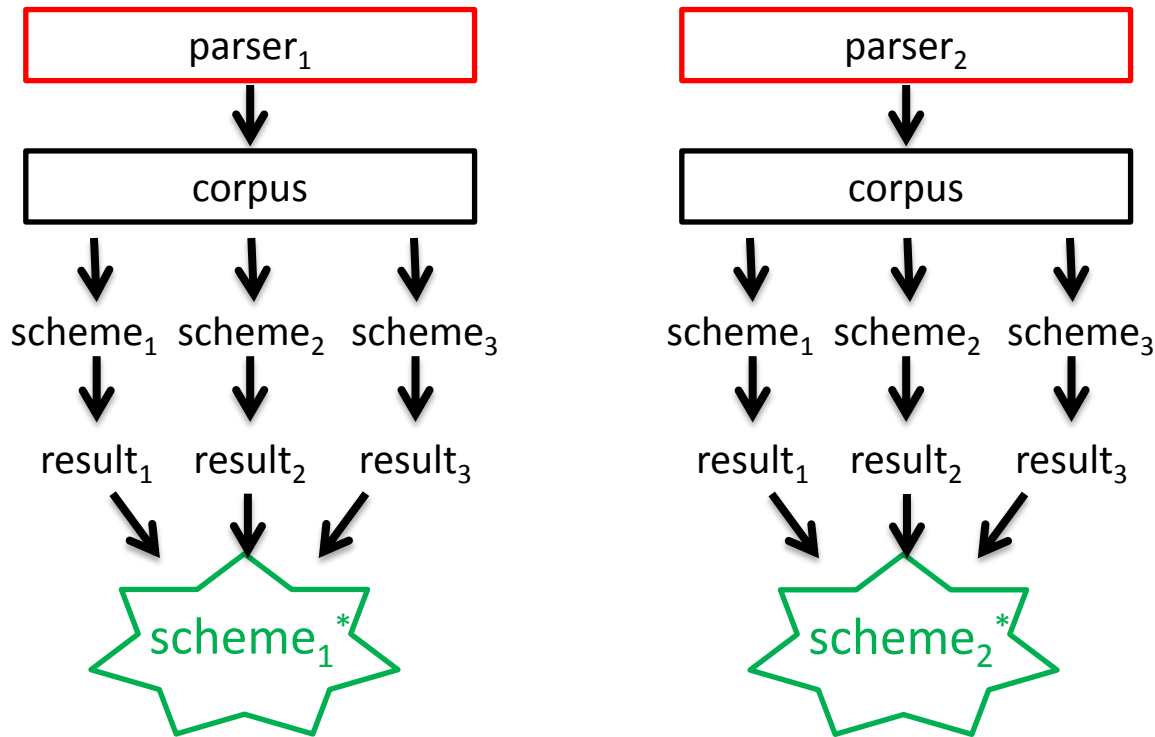


# Learnability-based Methodology 2

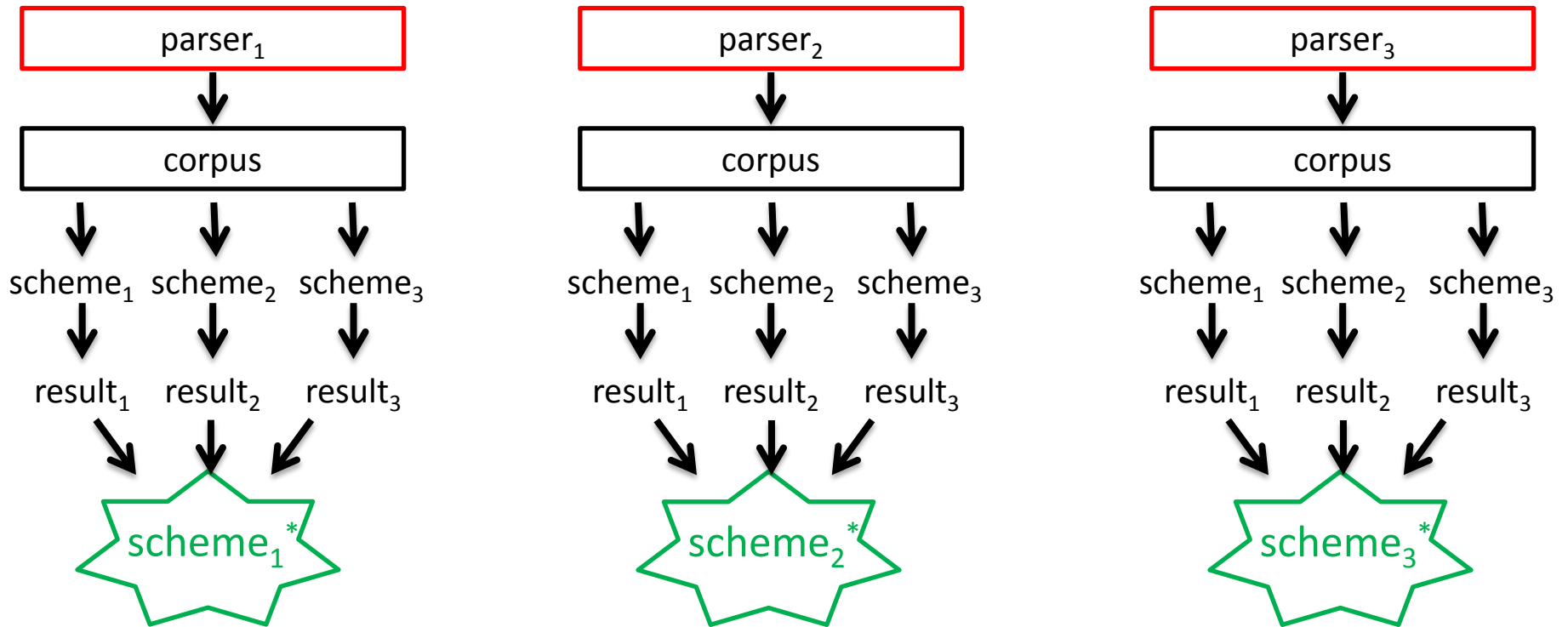




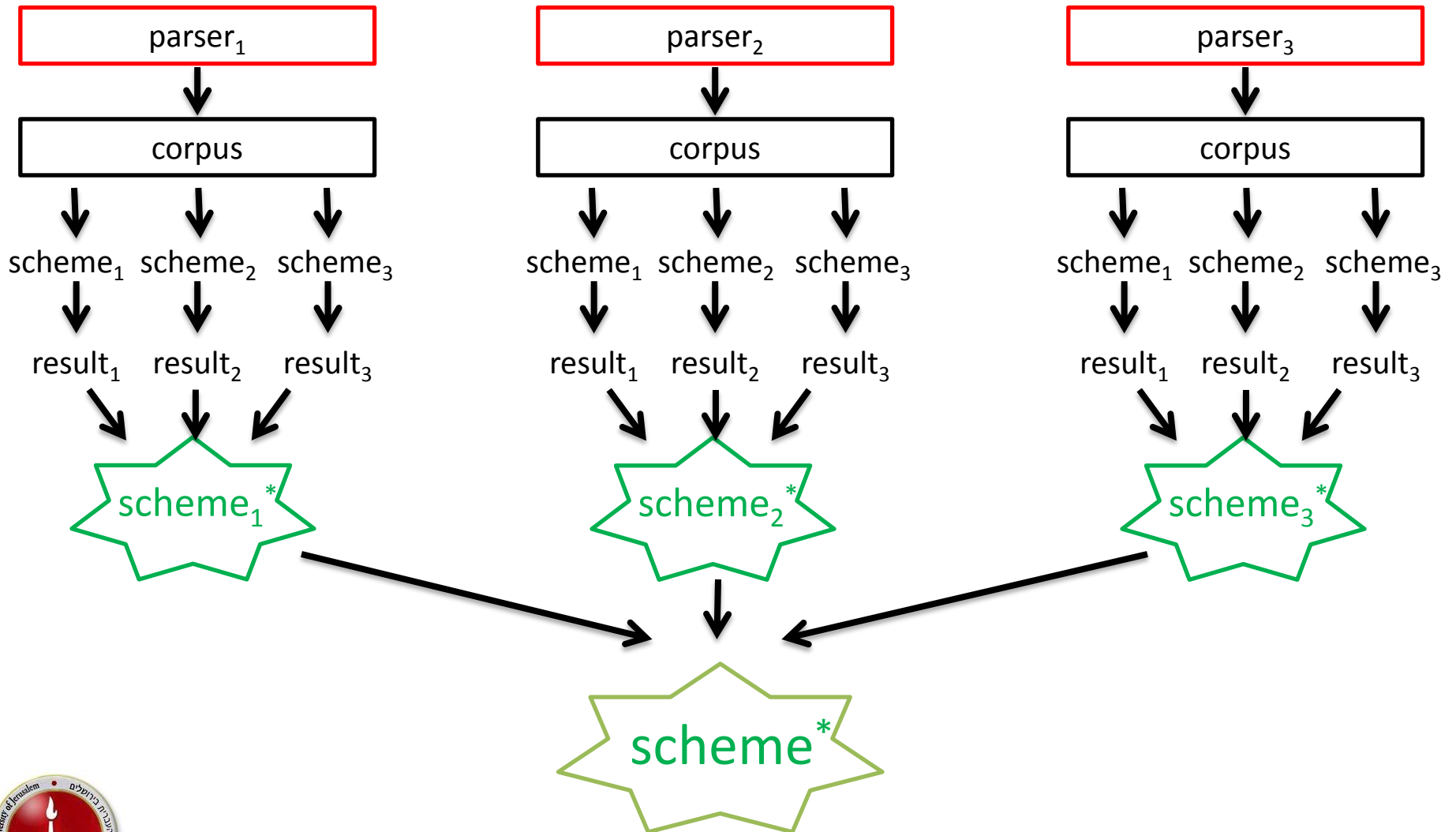
# Learnability-based Methodology 2



# Learnability-based Methodology 2



# Learnability-based Methodology 2



# Experimental Setup 1

- 6 VSSs,  $2^6 = 64$  annotation schemes

about<sub>IN</sub> ⇔ everyone<sub>NN</sub>

(a) Prepositional Phrases

can<sub>MD</sub> ⇔ come<sub>VB</sub>

(d) Verb Groups


the<sub>DT</sub> ⇔ apple<sub>NN</sub>

(b) Noun Phrases

John<sub>NNP</sub> ⇔ Doe<sub>NNP</sub>

(e) Noun Sequences

John<sub>NNP</sub> → and<sub>CC</sub> → Mary<sub>NNP</sub>



(c) Coordinations

to<sub>TO</sub> ⇔ eat<sub>VB</sub>

(f) Infinitive Verbs



# Experimental Setup 2

- 5 parsers of different types
  - Graph based parsers
    - MST parser (McDonald et al. 2005)
    - DMV parser (Klein and Manning 2004)
  - Transition based parsers
    - $S_u$  parser (Nivre 2009)
    - Clear parser (Choi and Nicolov 2009)
  - Other
    - NonDir Parser (Goldberg and Elhadad 2010)



# Results

- In 3/6 structures, a **unanimously** more learnable annotation was found:

- **Prepositions** (and not NPs) as heads of **PPs**

about<sub>IN</sub> → everyone<sub>NN</sub>

- **Nouns** (and not their determiners) as heads of **NPs**

the<sub>DT</sub> ← apple<sub>NN</sub>

- **Conjuncts** as heads of **coordination** structures

John<sub>NNP</sub> → and<sub>CC</sub> → Mary<sub>NNP</sub>



# Results

- In 3/6 structures, a **unanimously** more learnable annotation was found:

- **Prepositions** (and not NPs) as heads of **PPs**

about<sub>IN</sub> → everyone<sub>NN</sub>

- **Nouns** (and not their determiners) as heads of **NPs**

the<sub>DT</sub> ← apple<sub>NN</sub>

- **Conjuncts** as heads of **coordination** structures

John<sub>NNP</sub> → and<sub>CC</sub> Mary<sub>NNP</sub>

**Consistently** more learnable,  
independently of the **parser**  
and the **other VSSs**



# Magnitude

- Gains are substantial
  - Up to **19.8%** error reduction for a single structure



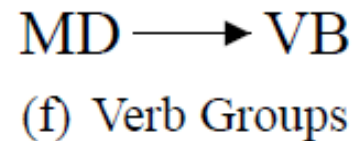
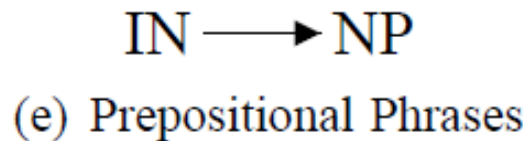
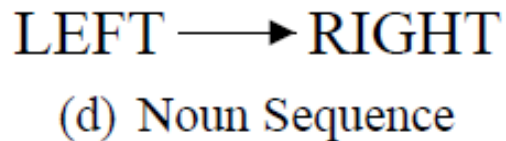
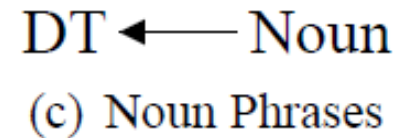
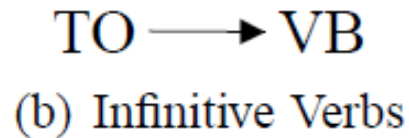
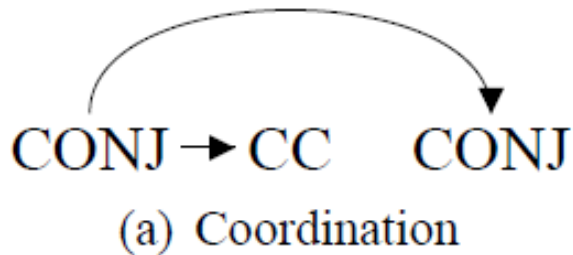


# Magnitude

- Gains are substantial
  - Up to **19.8%** error reduction for a single structure
- Gains are additive
  - Selecting the more learnable annotation in all 3 VSSs results in an **even more learnable** scheme
  - Up to **35.3%** error reduction by selecting the most vs. least learnable annotation



# And the Winner is...



Available @ <http://www.cs.huji.ac.il/~roys02/>



# Additional Experiments

- Two learnability measures
- High agreement between different parsers
- *Predictability* – a simple information-theoretic measure that yields similar results to learnability



# Additional Experiments

- Two learnability measures
- High agreement between different parsers
- *Predictability* – a simple information-theoretic measure that yields similar results to learnability

See paper!



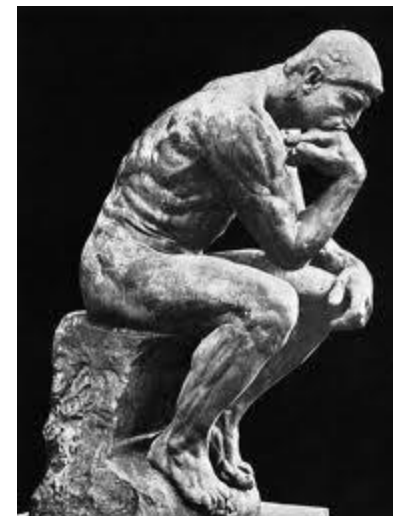
# Some more Ideas

- Apply our methodology to different tasks
  - POS tagging , Phrase Structure parsing, etc.
- Apply our methodology to different languages
  - Ballesteros and Nivre, CL 2013



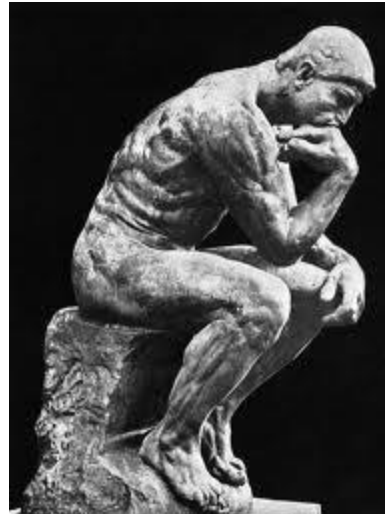
# What does it all Mean?

- Powerful results
  - Some annotations are clearly more learnable than others



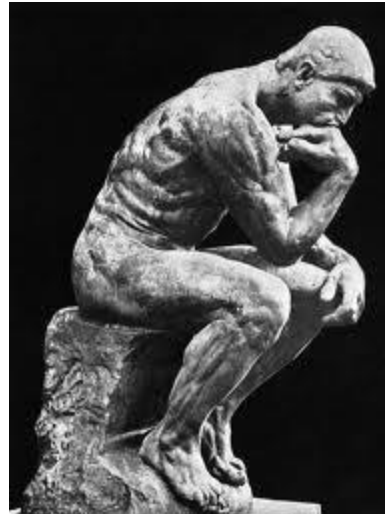
# What does it all Mean?

- Powerful results
  - Some annotations are clearly more learnable than others
- Linguistic implications?



# What does it all Mean?

- Powerful results
  - Some annotations are clearly more learnable than others
- Linguistic implications?
- Cognitive implications?





# Summary

- Varying Syntactic Structures (VSS)
  - Sometimes you have to choose
- **Learnability** as a selection criterion
  - A principled learnability-based methodology
  - Use parsers as research tools
- Selecting one alternative has a **substantial** and **predictable** effect on **parsing** performance
  - Parser independent
  - up to **35.3%** error reduction



# Summary

- Varying Syntactic Structures (VSS)
  - Sometimes you have to choose
- **Learnability** as a selection criterion
  - A principled learnability-based methodology
  - Use parsers as research tools
- Selecting one alternative has a **substantial** and **predictable** effect on **parsing** performance
  - Parser independent
  - up to **35.3%** error reduction

**Direct** implications for  
designing more **learnable**  
annotation schemes



- Richard Johansson
- Roi Reichart
- Shai Shalev-Shwartz
- Valentin I. Spitzkovsky





*[roys02@cs.huji.ac.il](mailto:roys02@cs.huji.ac.il)*

*<http://www.cs.huji.ac.il/~roys02/>*

