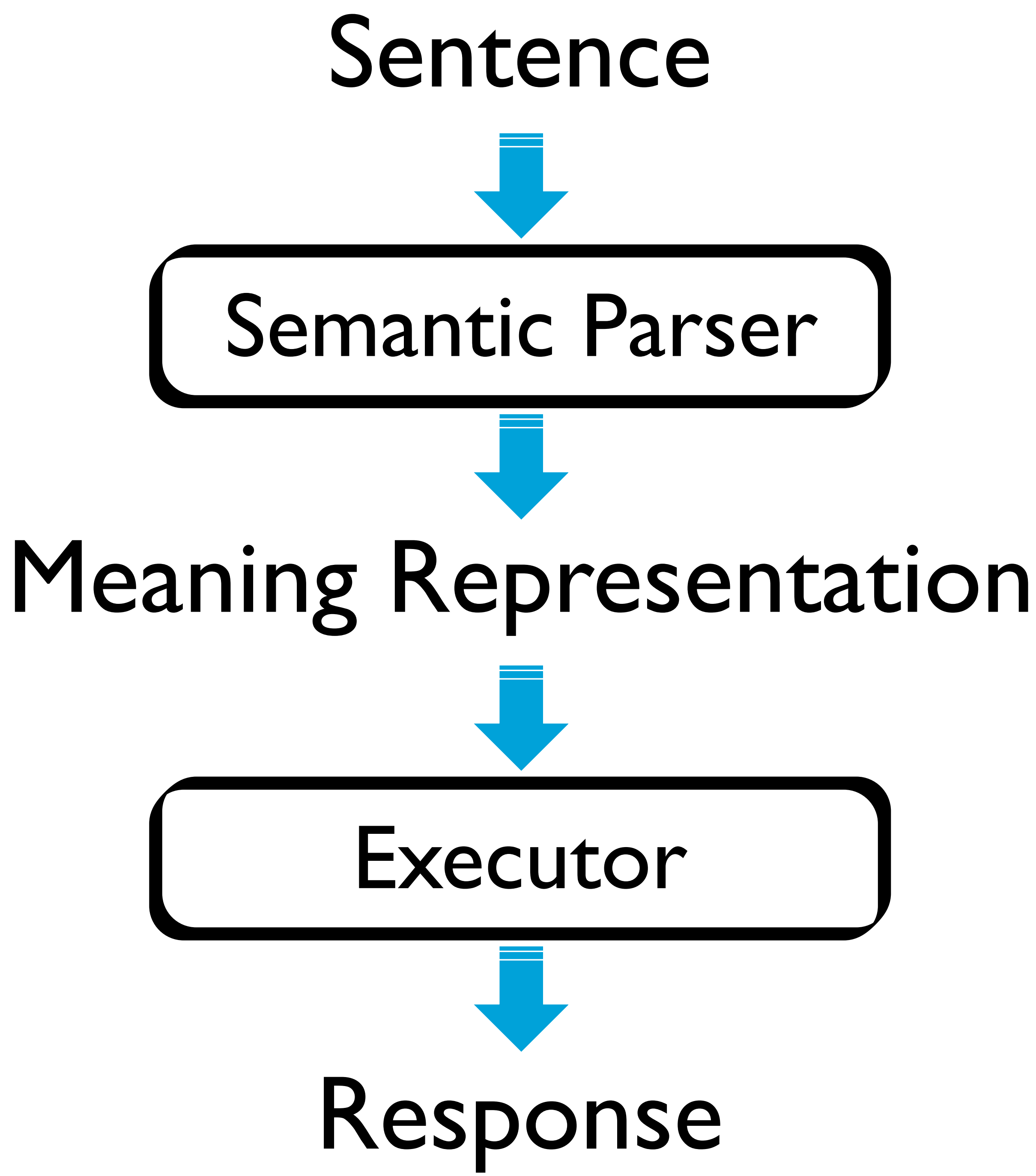
Interactive Learning of Parsers from Weak Supervision

Luke Zettlemoyer

with Luheng He, Kenton Lee, Mike Lewis, Julian Michael

Interpreting Language



Semantic Parsing: QA

How many people live in Seattle?

Semantic Parser

```
SELECT Population FROM CityData
where City=="Seattle”;
```

Executor

620,778

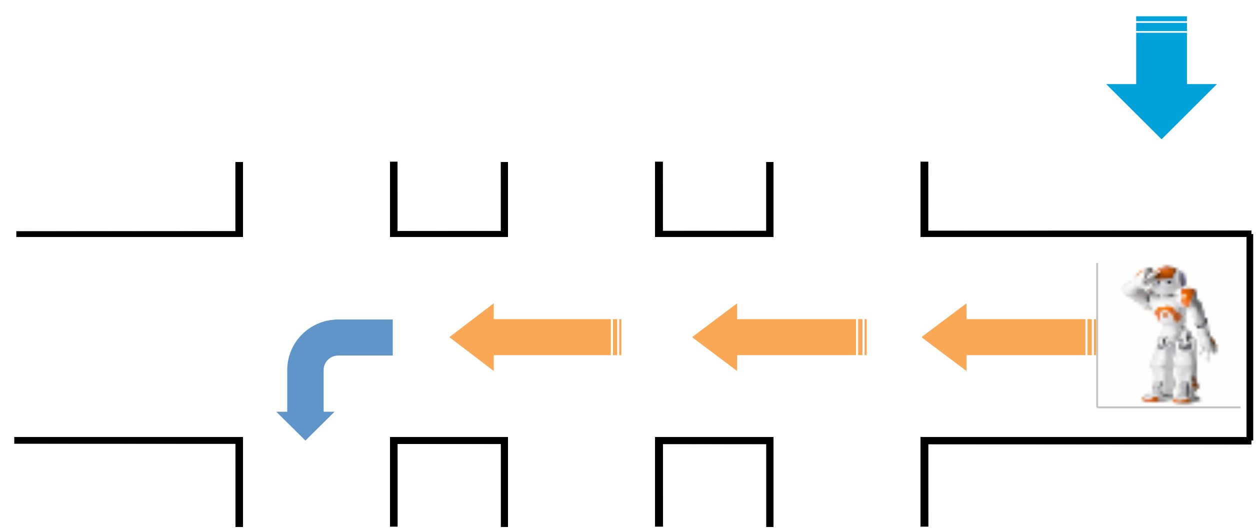
[Wong & Mooney 2007],
[Zettlemoyer & Collins 2005, 2007],
[Kwiatkowski et.al 2010, 2011],
[Liang et.al. 2011], [Cai & Yates 2013],
[Berant et.al. 2013,2014,2015],
[Kwiatkowski et.al. 2013],
[Reddy et.al, 2014,2016]

Semantic Parsing: Instructions

Go to the third junction and take a left

Semantic Parser

```
(do-seq(do-n-times 3
  (move-to forward-loc
    (do-until
      (junction current-loc
        (move-to forward-loc))))
  (turn-right))
```



[Chen & Mooney 2011]
[Matuszek et.al. 2012]
[Artzi & Zettlemoyer 2013]
[Mei et.al. 2015]

Semantic Parsing: IE

Somerset Maugham was a British playwright, novelist and short story writer.

Semantic Parser

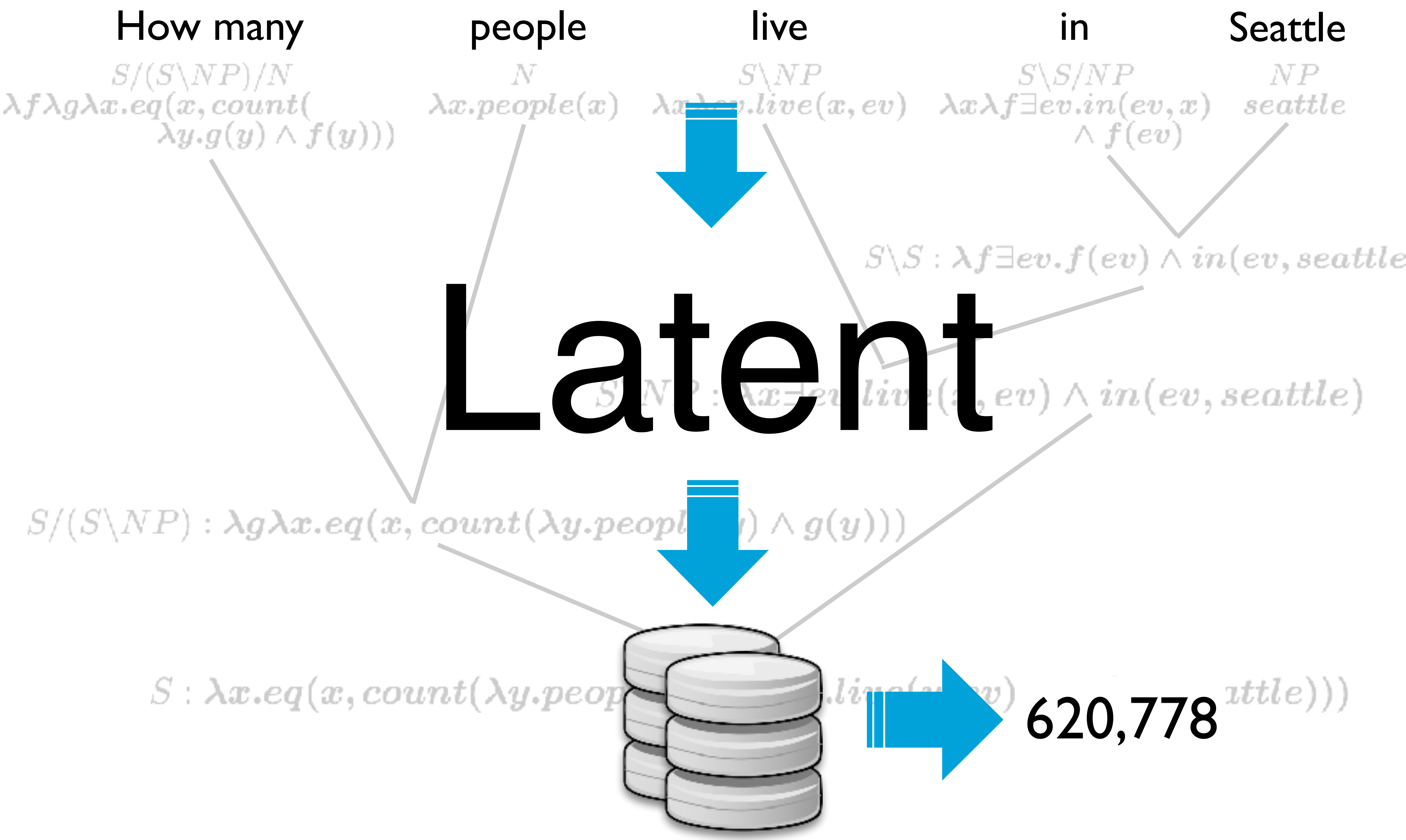
Knowledge Base (KB)

Freebase

S. Maugham	Nationality	United Kingdom
S. Maugham	Profession	Novelist

[Krishnamurthy and Mitchell; 2012, 2014][Choi et al., 2015]

Semantic Parsing: Complex Structure



Lots of Different Applications

We are doing semantic analysis for:

- Visual Semantic Role Labeling [Yatskar et al, 2016]
- Visual Question Answering [FitzGerald et al, in prep]
- Language to Code [Lin et al, in prep]
- Entity-entity sentiment [Choi et al, 2016]
- Understanding Cooking Recipes [Kiddon et al, 2016]
- Zero-shot Relation Extraction [Levy et al, in review]
- Interactive Learning for NLIDBs [Iyer, et al, in review]

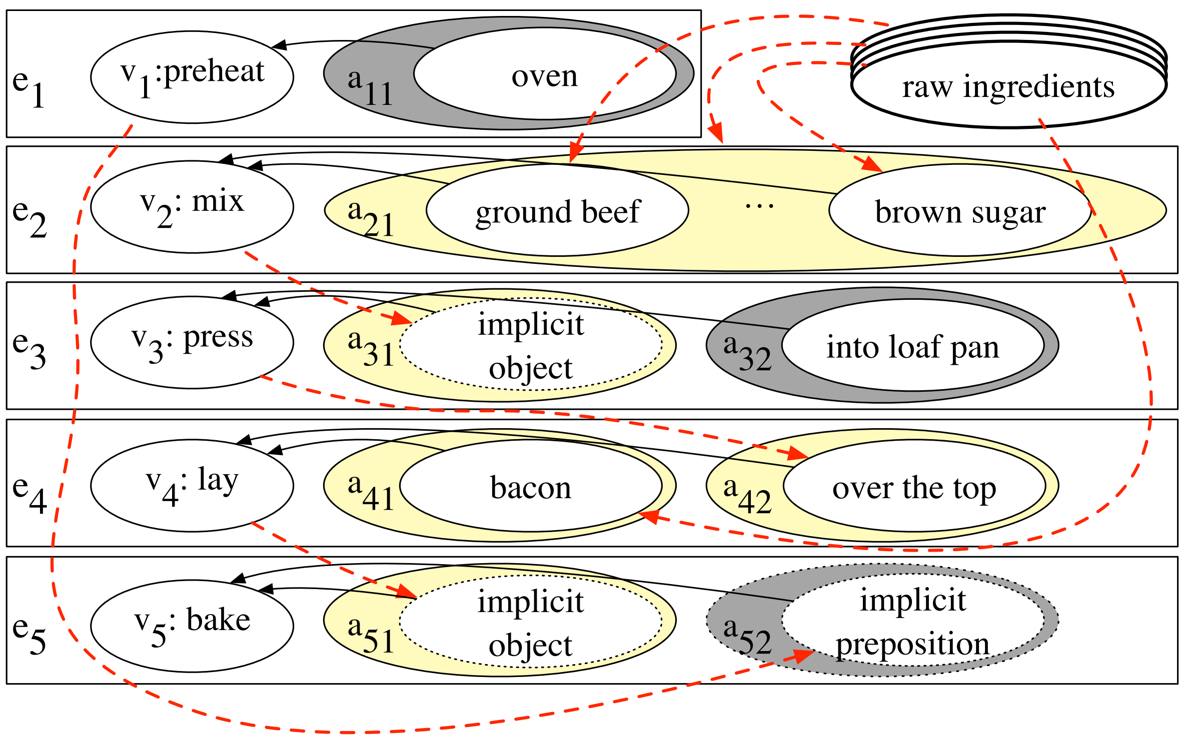
Challenge: typically gather data and learn model from scratch in each case…

Understanding Cooking Recipes

Amish Meatloaf (http://allrecipes.com/recipe/amish-meatloaf/, recipe condensed)

Ingredients
2 pounds ground beef
2 1/2 cups crushed butter-flavored crackers
1 small onion, chopped
2 eggs
3/4 cup ketchup
1/4 cup brown sugar
2 slices bacon

Preheat the oven to 350 degrees F (175 degrees C).
In a medium bowl, mix together ground beef, crushed crackers, onion, eggs, ketchup, and brown sugar until well blended.
Press into a 9x5 inch loaf pan.
Lay the two slices of bacon over the top.
Bake for 1 hour, or until cooked through.



Approach: unsupervised learning for actions and object flow

Open Question:

- Can we build an off-the-shelf parser that would help here?

[Kiddon et al 2015, 2016]

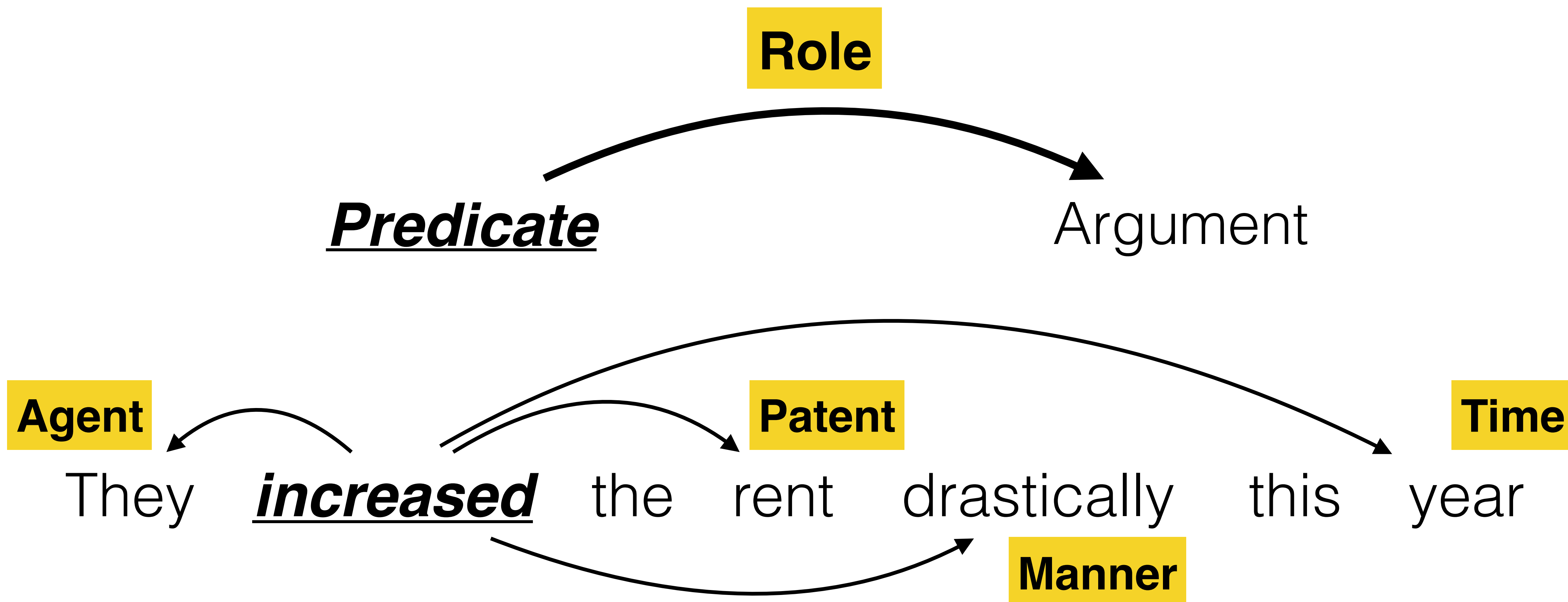
Intel Science and Technology Center for Pervasive Computing

Towards Broad Coverage Semantic Parsing

- Can we crowdsource semantics?
- Train with latent syntax?
- Build fast and accurate parsers?
- Actively select which data to label?

Semantic Role Labeling (SRL)

who did **what** to **whom**, **when** and **where**?



- Defining a set of roles can be difficult
- Existing formulations have used different sets

Existing SRL Formulations and Their Frame Inventories

FrameNet
1000+ semantic frames,
roles (frame elements)
shared across frames

Frame: Change_position_on_a_scale
This frame consists of words that indicate the change of an Item's position on a scale (the Attribute) from a starting point (Initial_value) to an end point (Final_value).
The direction (Path) …
Lexical Units:
…, reach.v, rise.n, ***rise.v,*** *rocket.v, shift.n, …*

PropBank
10,000+ frame files
with predicate-specific roles

Roleset Id: rise.01 ***, go up***

Arg1-: *Logical subject, patient, thing rising*
Arg2-EXT: *EXT, amount risen*
Arg3-DIR: *start point*
Arg4-LOC: *end point*
Argm-LOC: *medium*

Unified Verb Index, University of Colorado http://verbs.colorado.edu/verb-index/
PropBank Annotation Guidelines, Bonial et al., 2010
FrameNet II: Extended theory and practice, Ruppenhofer et al., 2006

Our Annotation Scheme

Given sentence and a verb:

They ***increased*** the rent this year .

Step 1: Ask a question about the verb:

Step 2: Answer with words in the sentence:

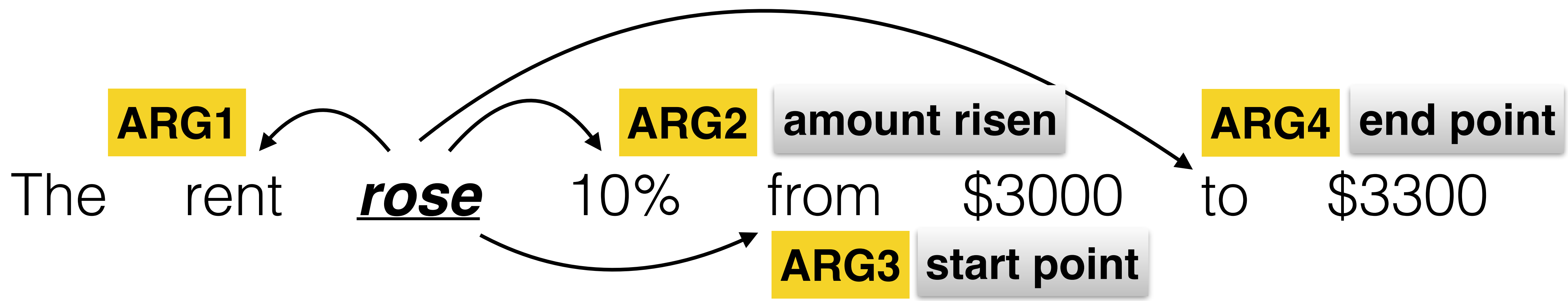
Who increased something ?	They

Step 3: Repeat, write as many QA pairs as possible …

What is increased ?	the rent
When is something increased ?	this year

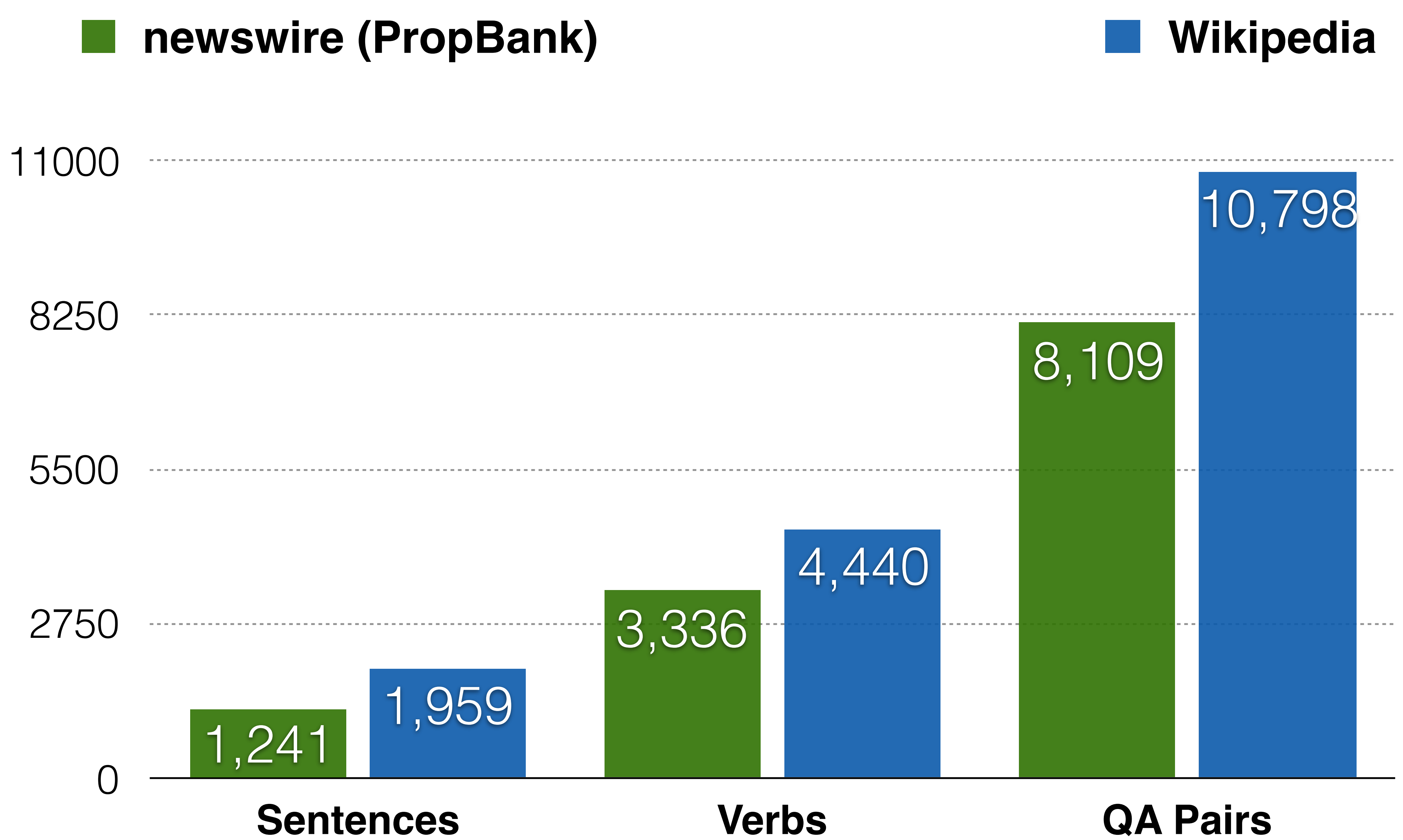
[He et al 2015]

Our Method: Q/A Pairs for Semantic Relations

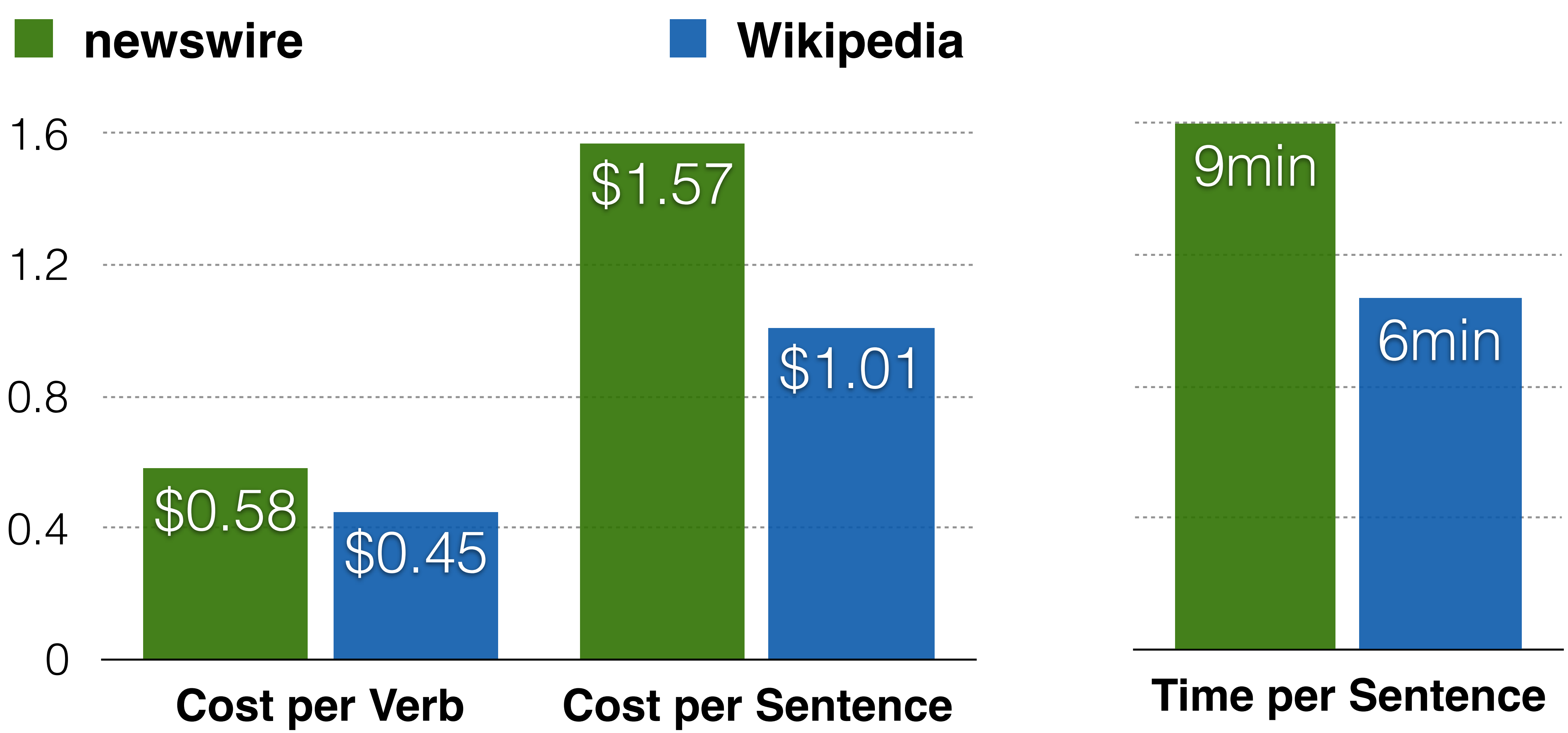


Wh-Question	**Answer**
What ***rose*** ?	the rent
How much did something ***rise*** ?	10%
What did something ***rise*** from ?	$3000
What did something ***rise*** to ?	$3300

Dataset Statistics



Cost and Speed



- Part-time freelancers from upwork.com (hourly rate: $10)
- ~2h screening process for native English proficiency

Wh-words vs. PropBank Roles

	Who	What	When	Where	Why	How	HowMuch
ARG0	1575	414	3	5	17	28	2
ARG1	285	2481	4	25	20	23	95
ARG2	85	364	2	49	17	51	74
ARG3	11	62	7	8	4	16	31
ARG4	2	30	5	11	2	4	30
ARG5	0	0	0	1	0	2	0
AM-ADV	5	44	9	2	25	27	6
AM-CAU	0	3	1	0	23	1	0
AM-DIR	0	6	1	13	0	4	0
AM-EXT	0	4	0	0	0	5	5
AM-LOC	1	35	10	89	0	13	11
AM-MNR	5	47	2	8	4	108	14
AM-PNC	2	21	0	1	39	7	2
AM-PRD	1	1	0	0	0	1	0
AM-TMP	2	51	341	2	11	20	10

Advantages

- Easily explained
- No pre-defined roles, few syntactic assumptions
- Can capture implicit arguments
- Generalizable across domains

Limitations

- Only modeling verbs (for now)
- Not annotating verb senses directly
- Can have multiple equivalent questions

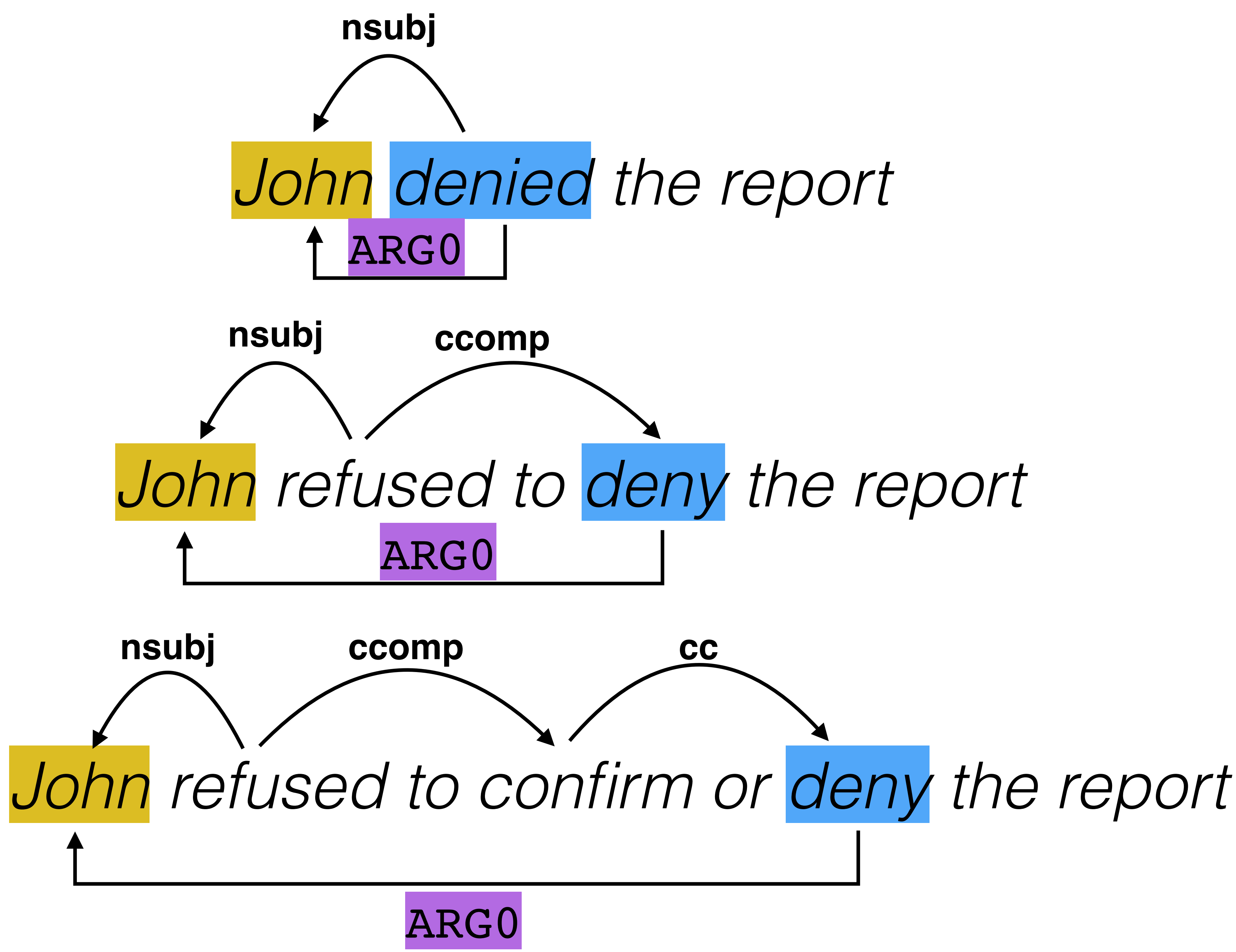
Challenges

- What questions to ask?
- Quality - Can we get good Q/A pairs?
- Coverage - Can we get all the Q/A pairs?

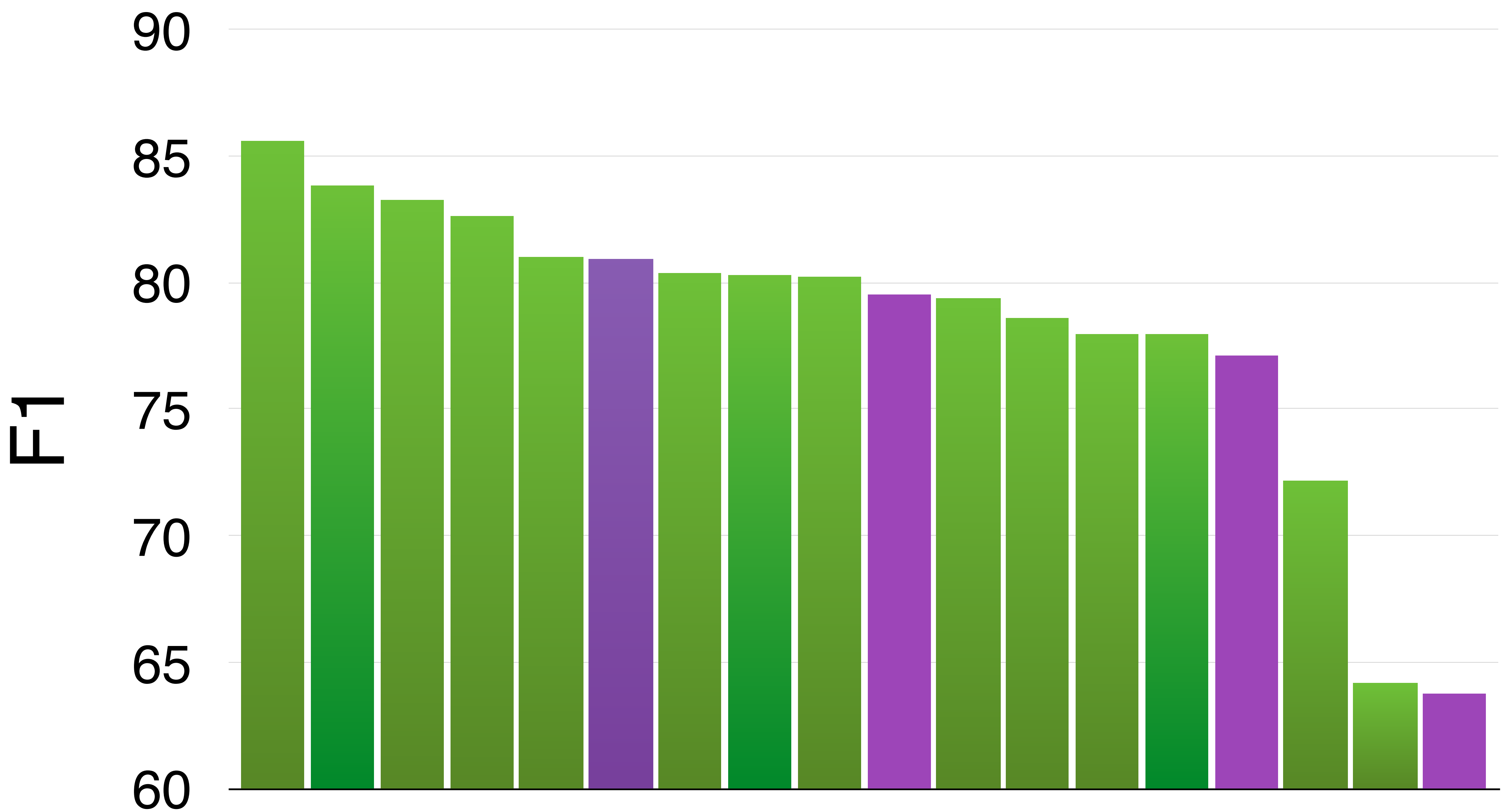
Towards Broad Coverage Semantic Parsing

- Can we crowdsource semantics?
- Train with latent syntax?
- Build fast and accurate parsers?
- Actively select which data to label?

SRL Challenge: Sparsity



Joint vs. Pipelines



CCG Dependencies

Include nearly all SRL dependencies:

John	wanted	to confirm	the report
NP_{John}	$(S\backslash NP_x)/(S\backslash NP_x)$	$(S\backslash NP_x)/NP_y$	NP_{report}
	wanted—>x	confirm—>x , confirm—>y	

$S\backslash NP_x$

confirm—>report , confirm—>x

$S\backslash NP_x$

confirm—>report , confirm—>x , wanted—>x

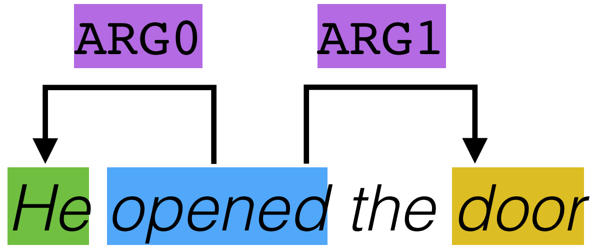
S

confirm—>report , confirm—>john , wanted—>john

[Lewis et al, 2015]

Training

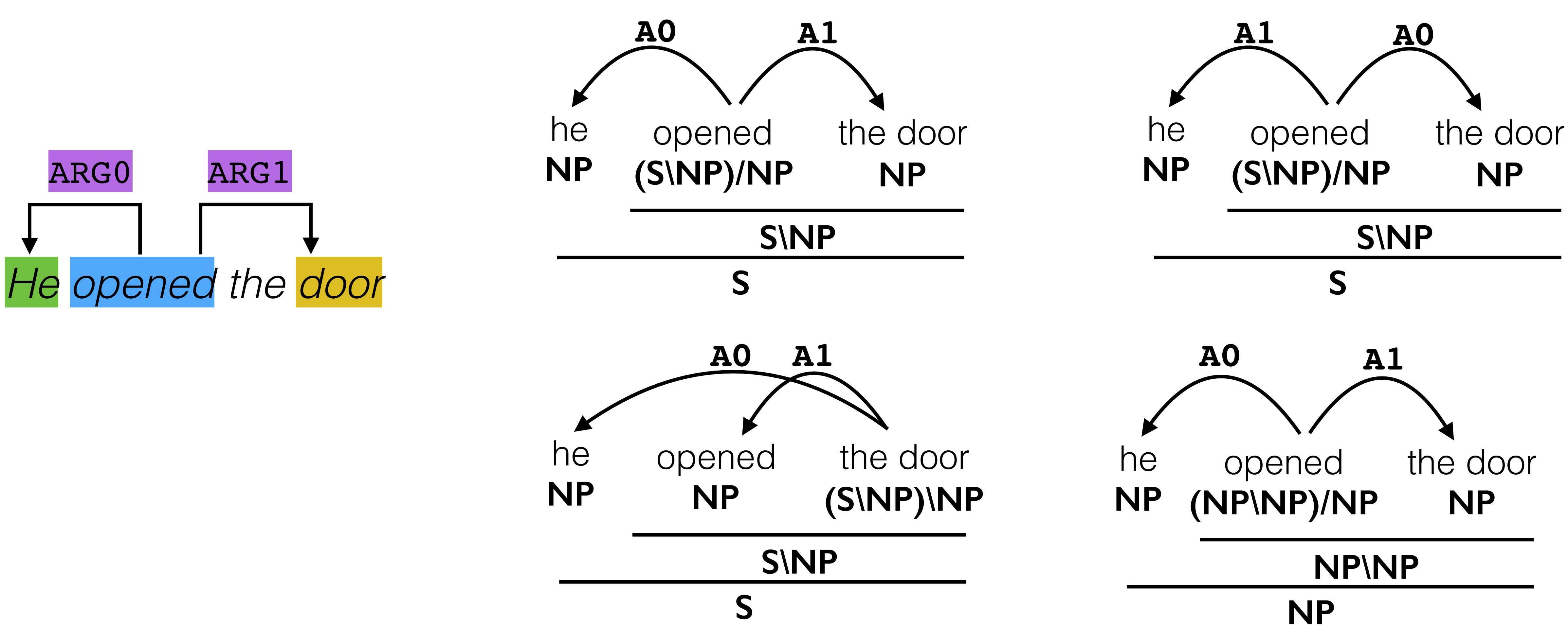
Learn latent CCG that recovers SRL



Training

Learn latent CCG that recovers SRL

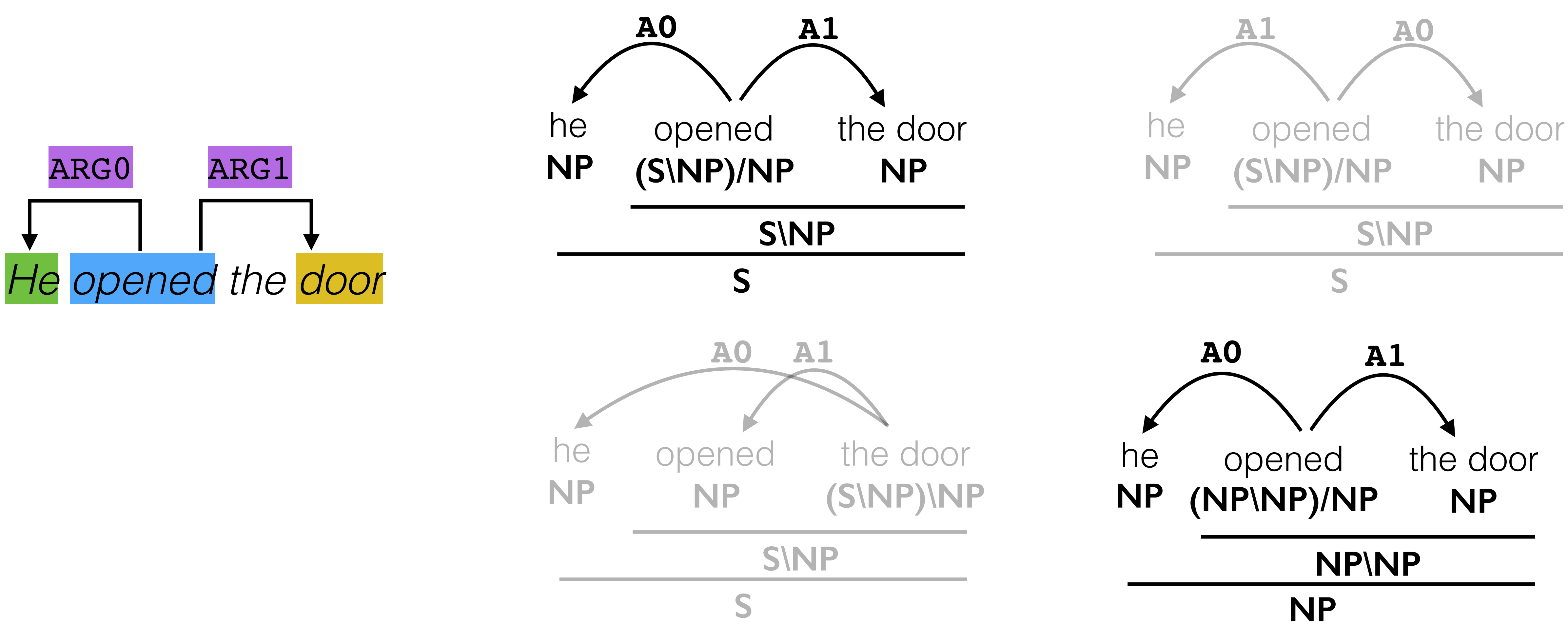
- Generate *consistent* CCG/SRL parses for training sentences



Training

Learn latent CCG that recovers SRL

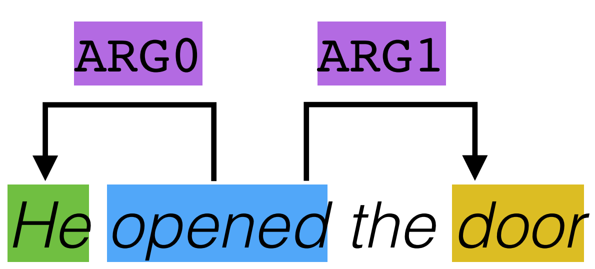
- Mark subset as correct, based on semantic dependencies

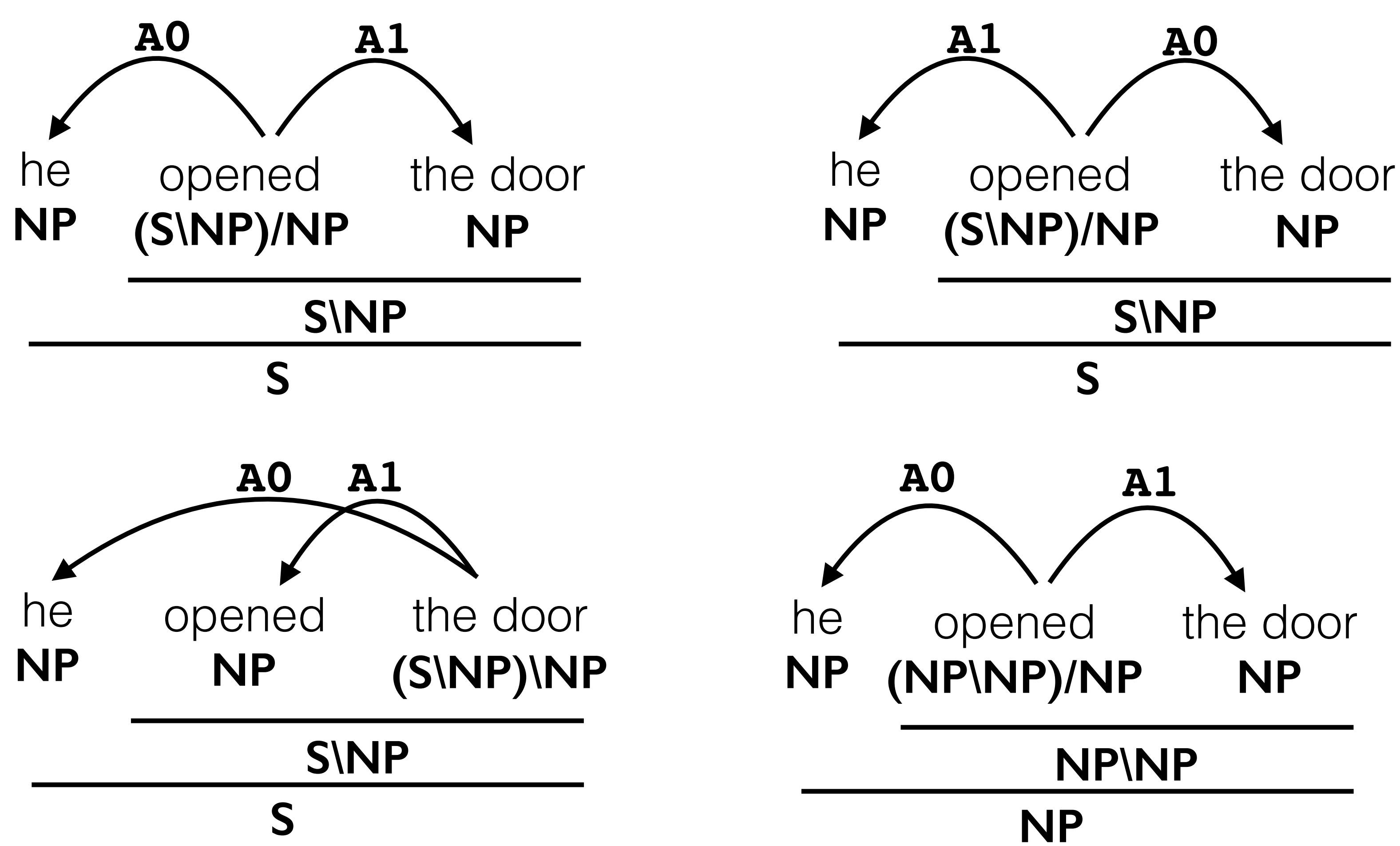


Training

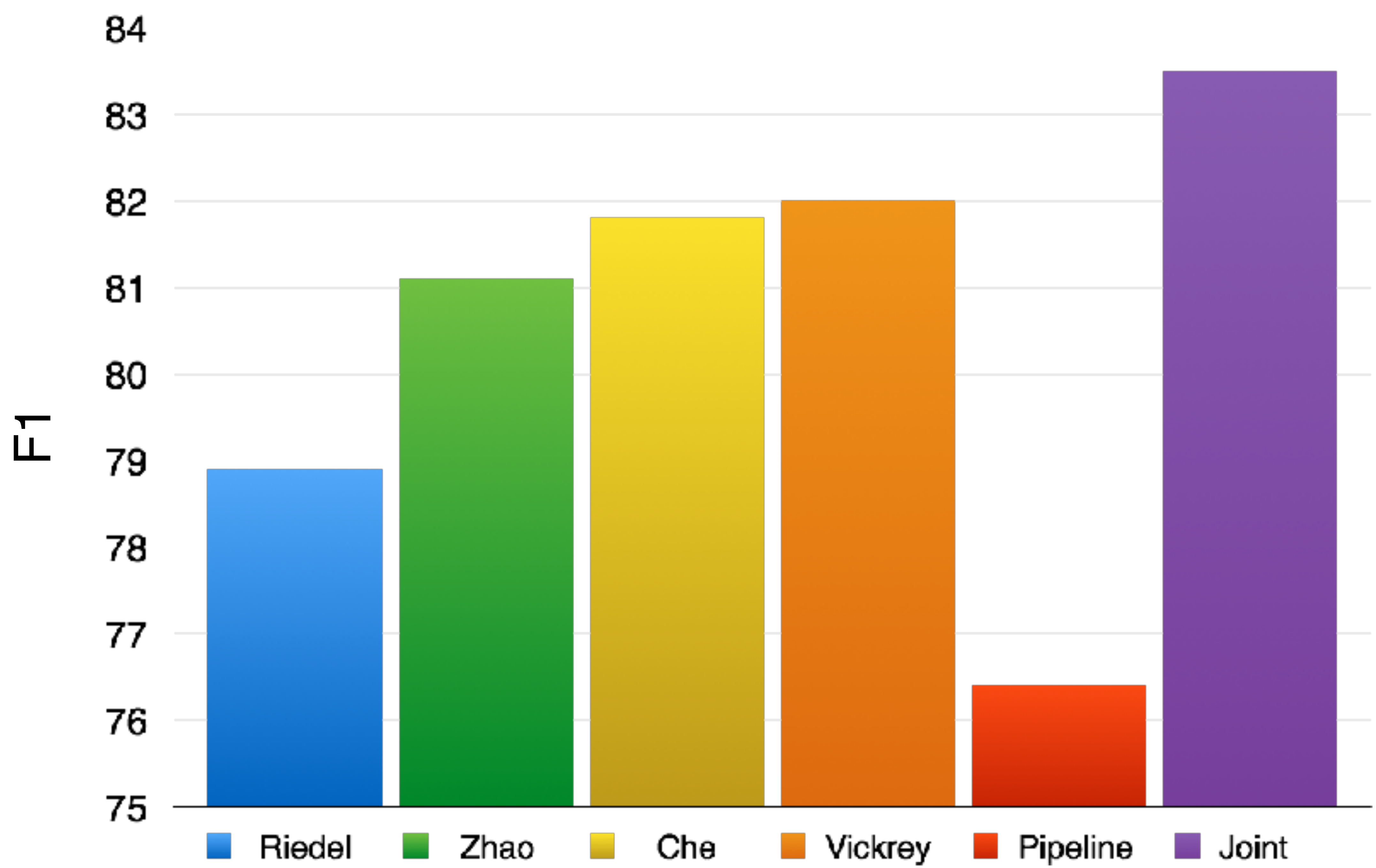
Learn latent CCG that recovers SRL

- Optimize marginal likelihood



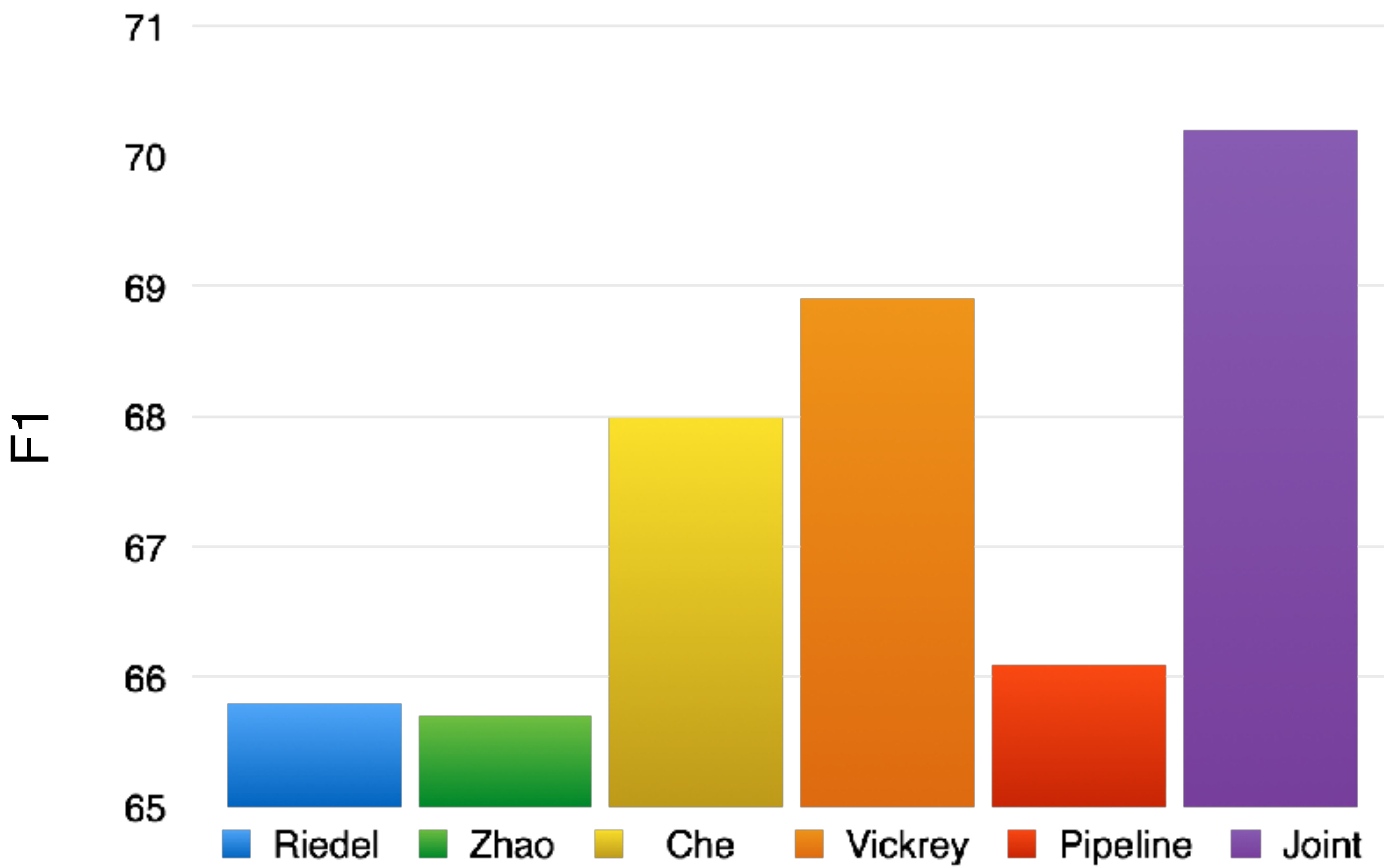


SRL Results



[Lewis et al 2015]

Out-of-domain SRL Results



Towards Broad Coverage Semantic Parsing

- Can we crowdsource semantics?
- Train with latent syntax?
- Build fast and accurate parsers?
- Actively select which data to label?

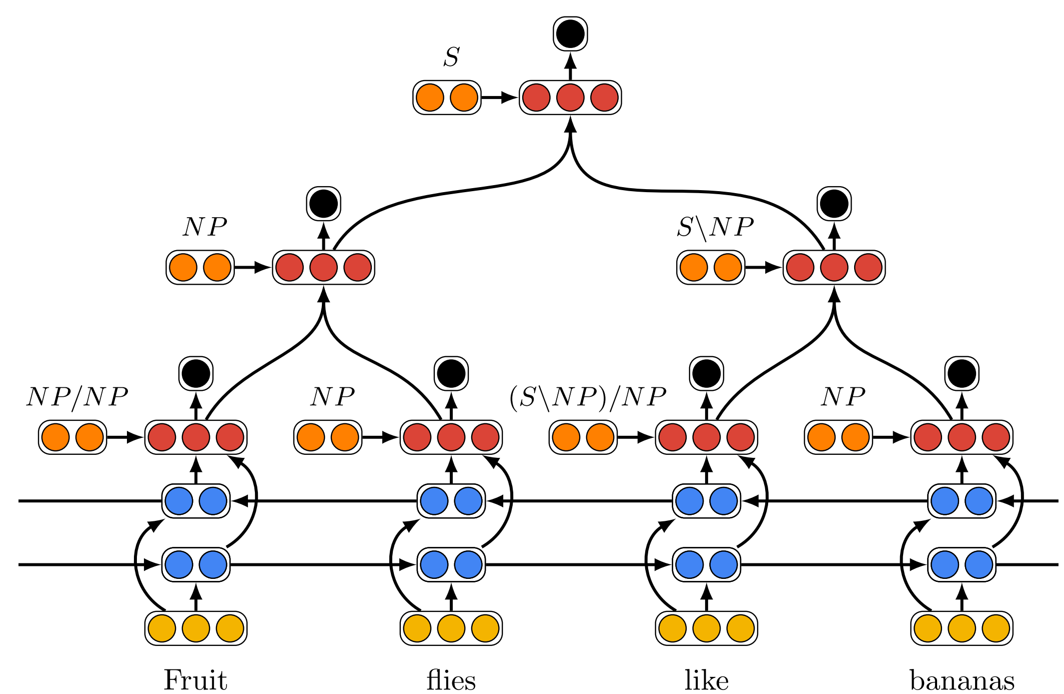
Global A* Parsing

Challenge:

Global models (e.g. Recursive NNs) break dynamic programs

Our approach:

Combine local and global models in A^* parser



Result:

Accurate models with formal guarantees

[Lee et al, 2016, EMNLP best paper]

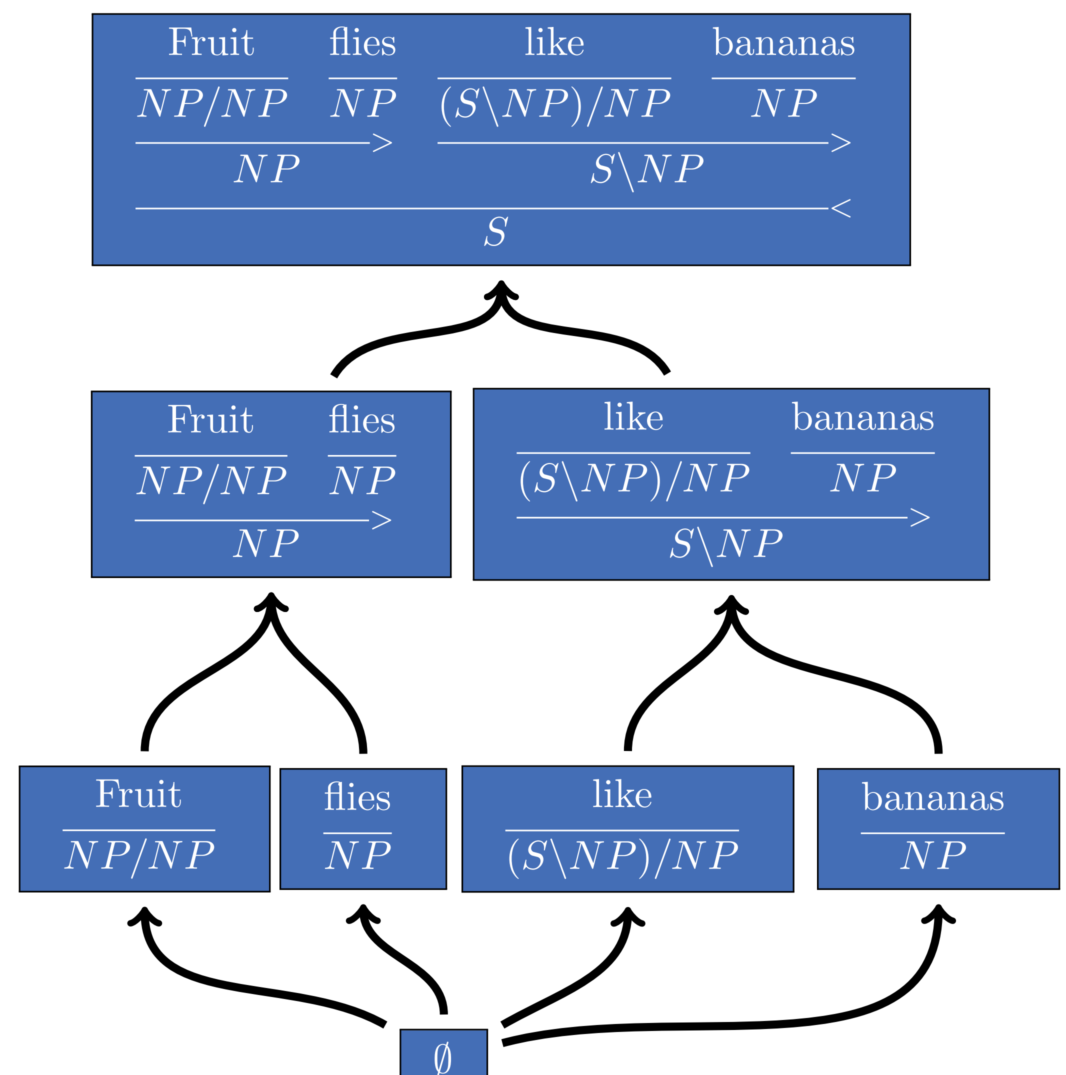
Parsing with Hypergraphs

Input

Fruit flies like bananas

Output

$$
\begin{array}{cccc}
\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\
\overline{NP/NP} & \overline{NP} & \overline{(S\backslash NP)/NP} & \overline{NP} \\
\multicolumn{2}{c}{\overline{\qquad\qquad NP \qquad\qquad}{>}} & \multicolumn{2}{c}{\overline{\qquad\qquad S\backslash NP \qquad\qquad}{>}} \\
\multicolumn{4}{c}{\overline{\qquad\qquad\qquad\qquad S \qquad\qquad\qquad\qquad}{<}}
\end{array}
$$



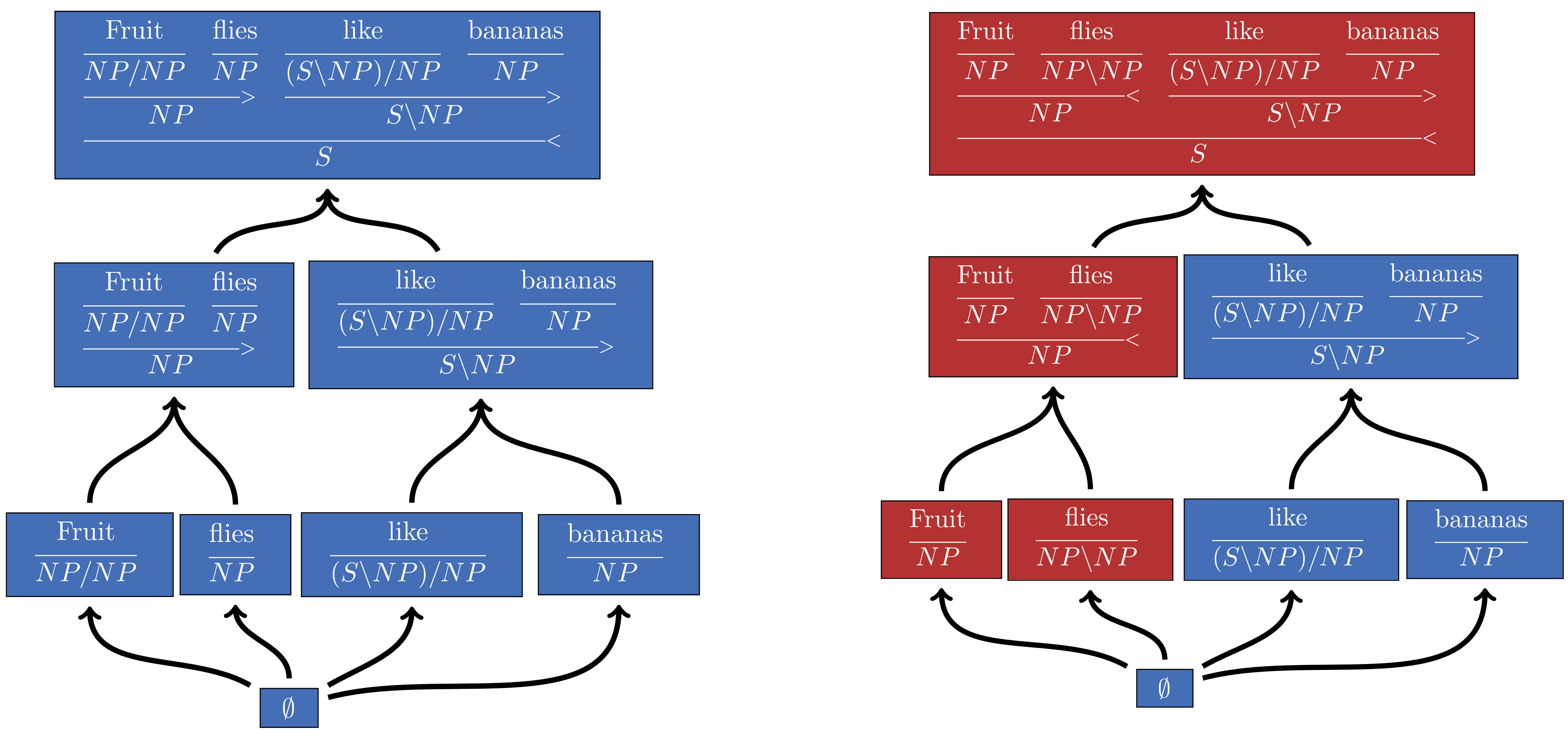
Klein and Manning, 2001

Parsing with Hypergraphs

Input

Fruit flies like bananas

Output

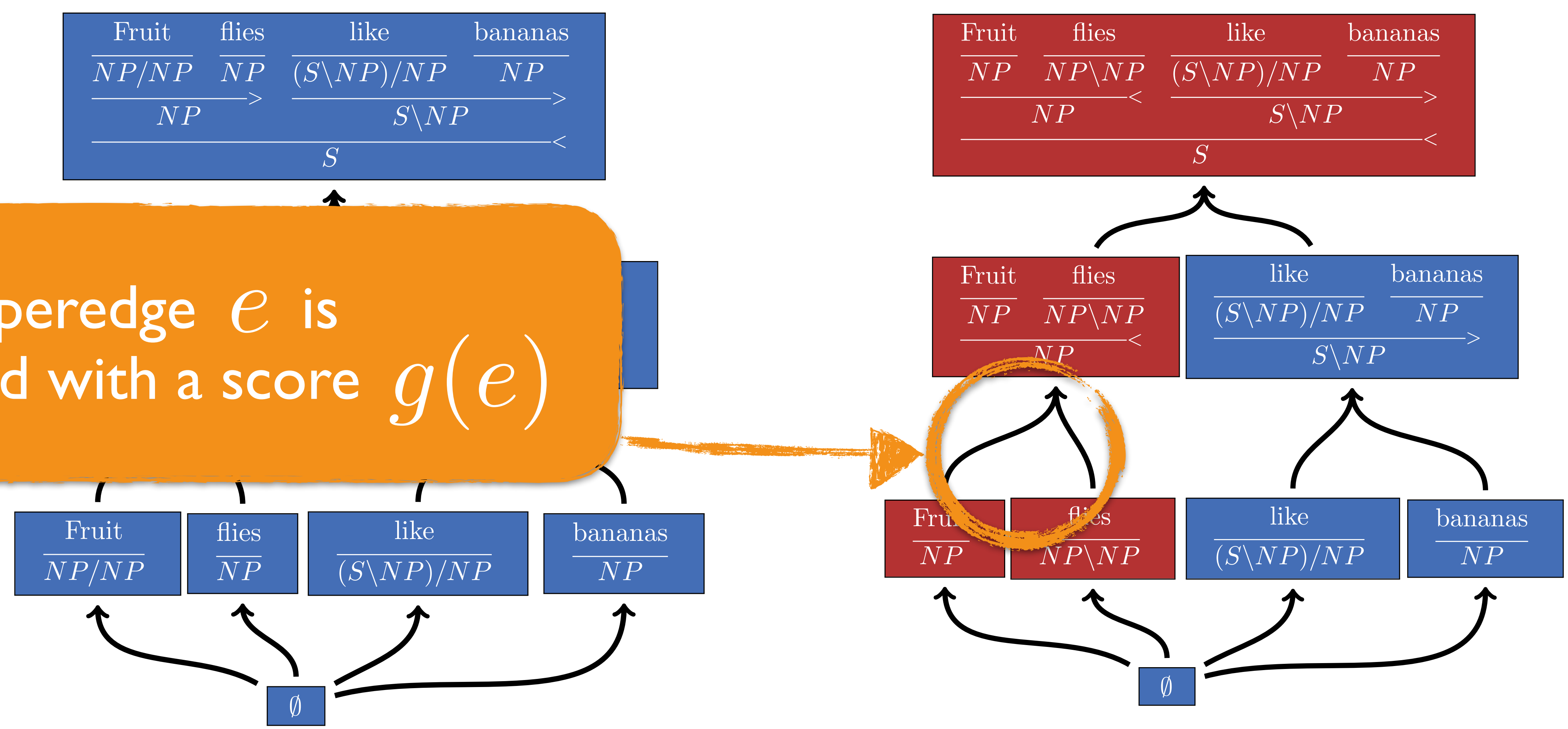


Parsing with Hypergraphs

Input

Fruit flies like bananas

Output

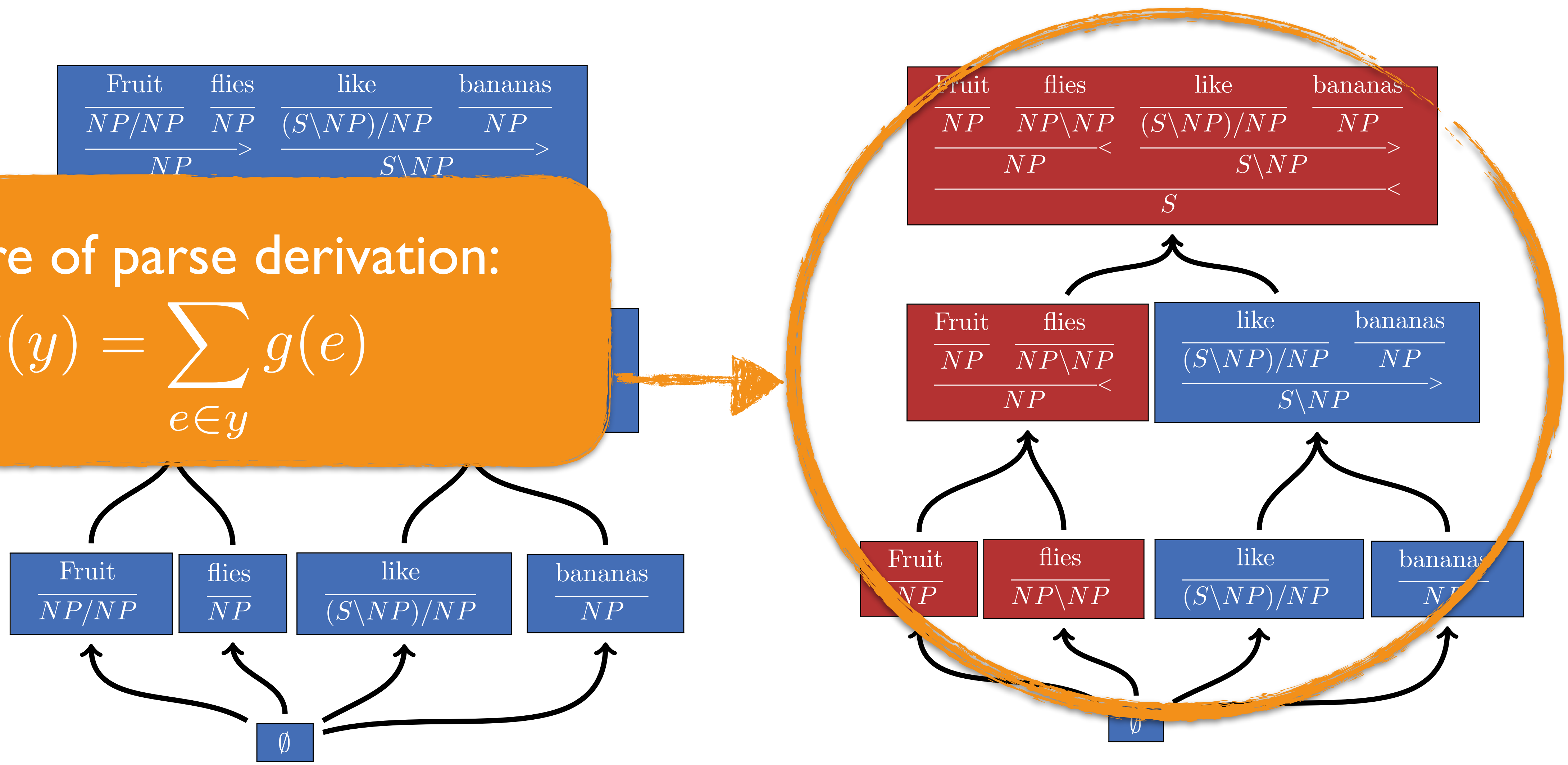


Each hyperedge e is weighted with a score $g(e)$

Parsing with Hypergraphs

Input Fruit flies like bananas

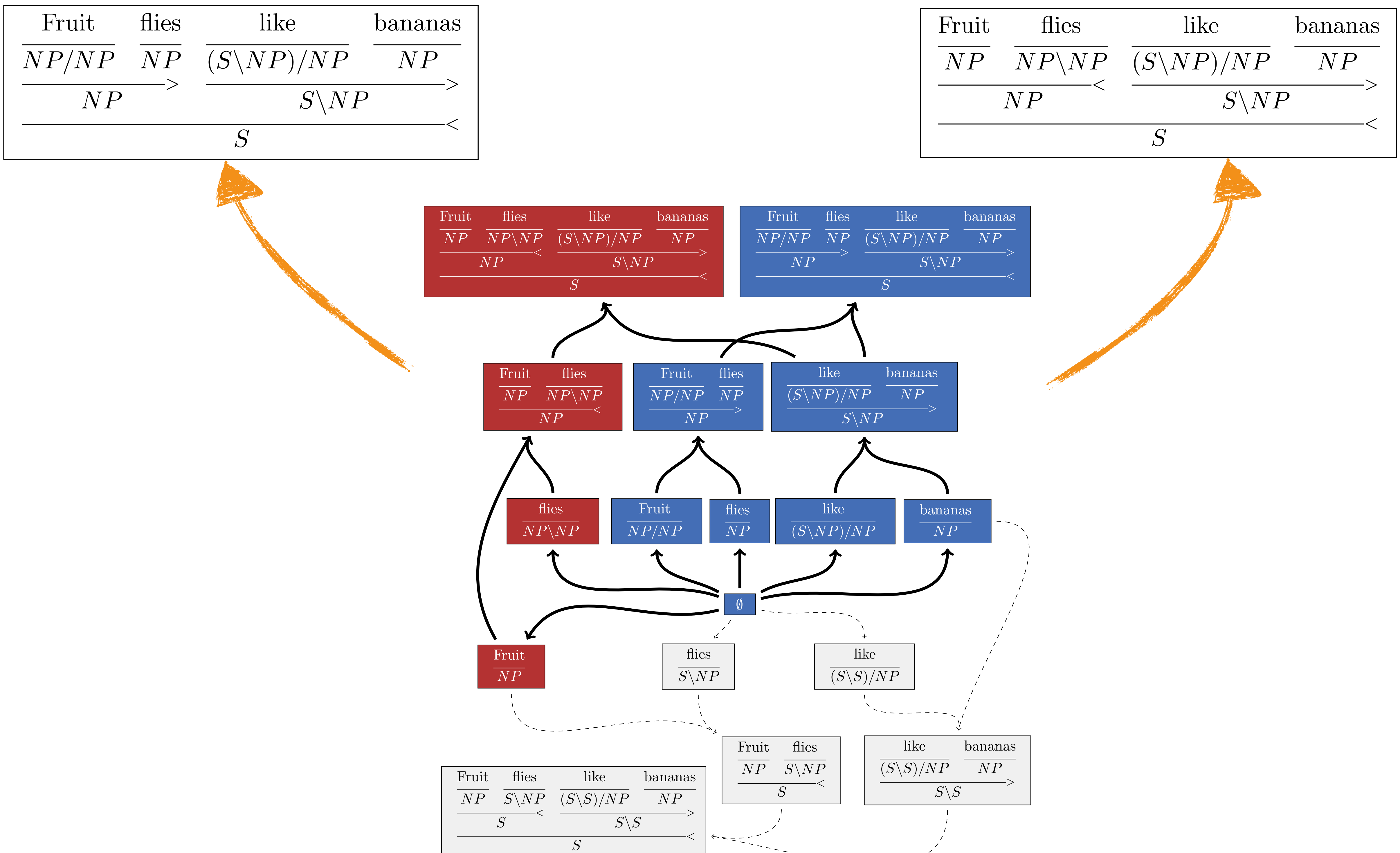
Output



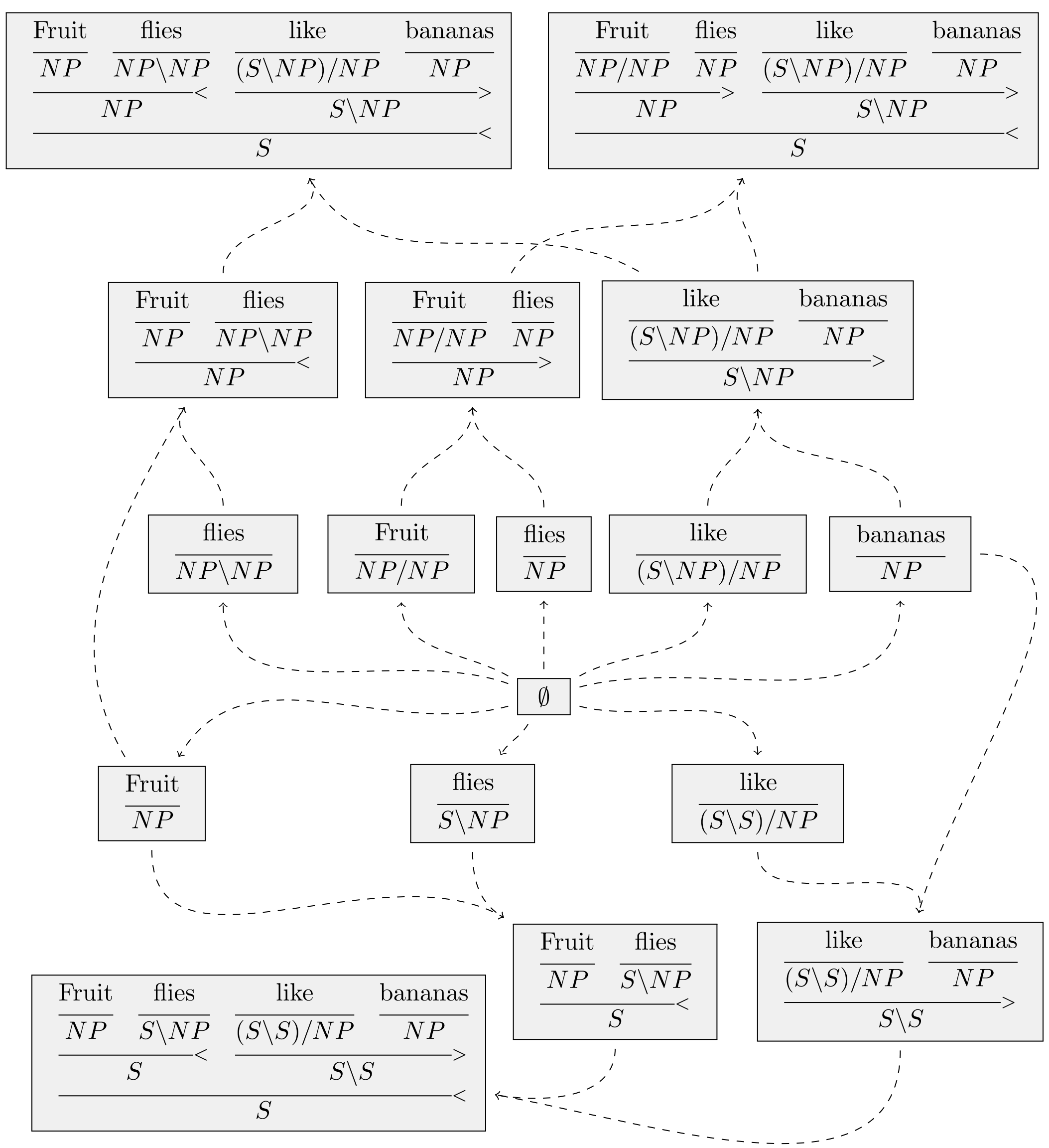
Score of parse derivation:

$$g(y) = \sum_{e \in y} g(e)$$

Parsing with Hypergraphs



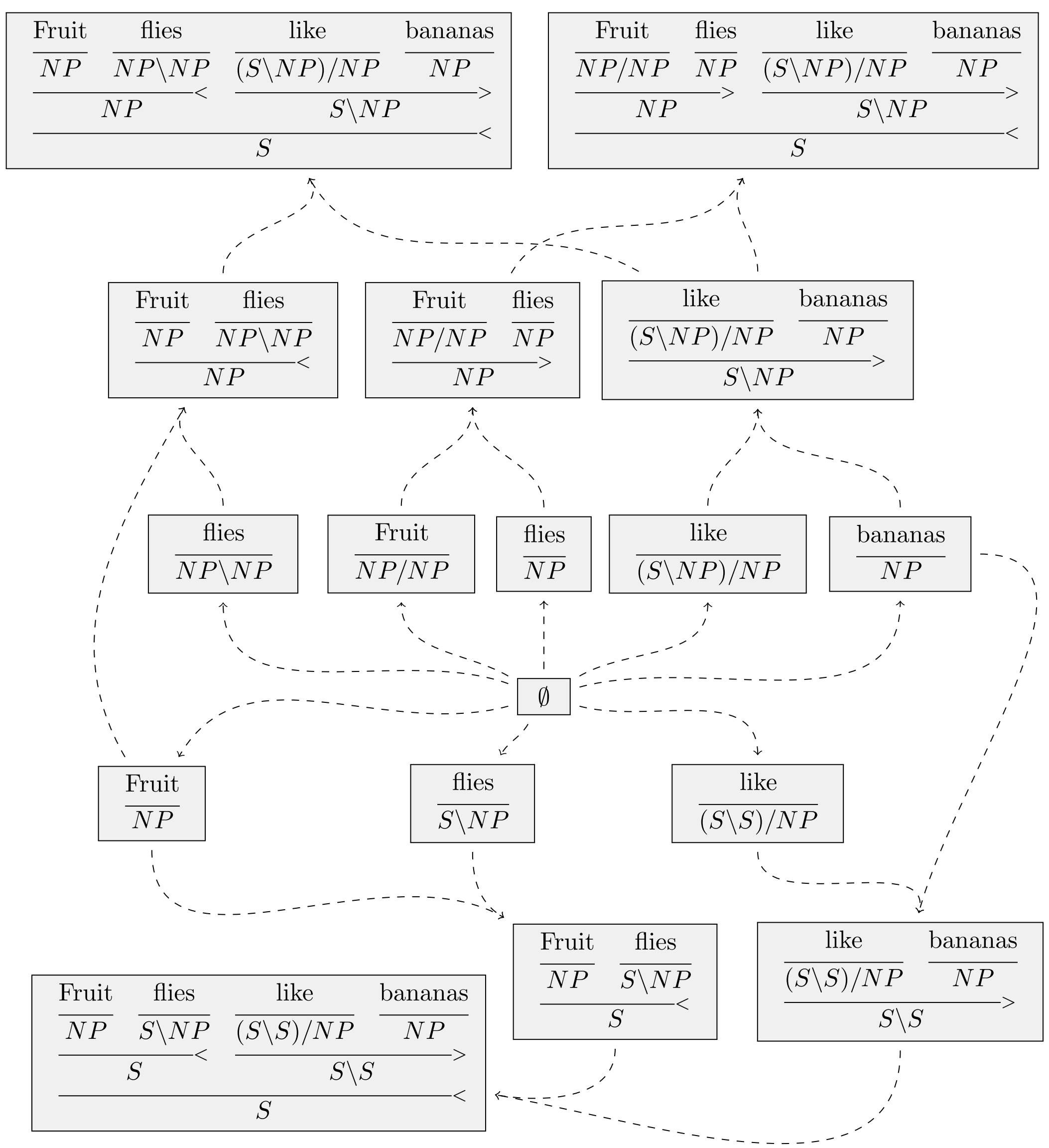
Parsing with Hypergraphs



- Predicted parse: $$y^* = \operatorname*{argmax}_{y \in Y} \; g(y)$$
- Exponential number of nodes

$\longrightarrow$ Intractable inference

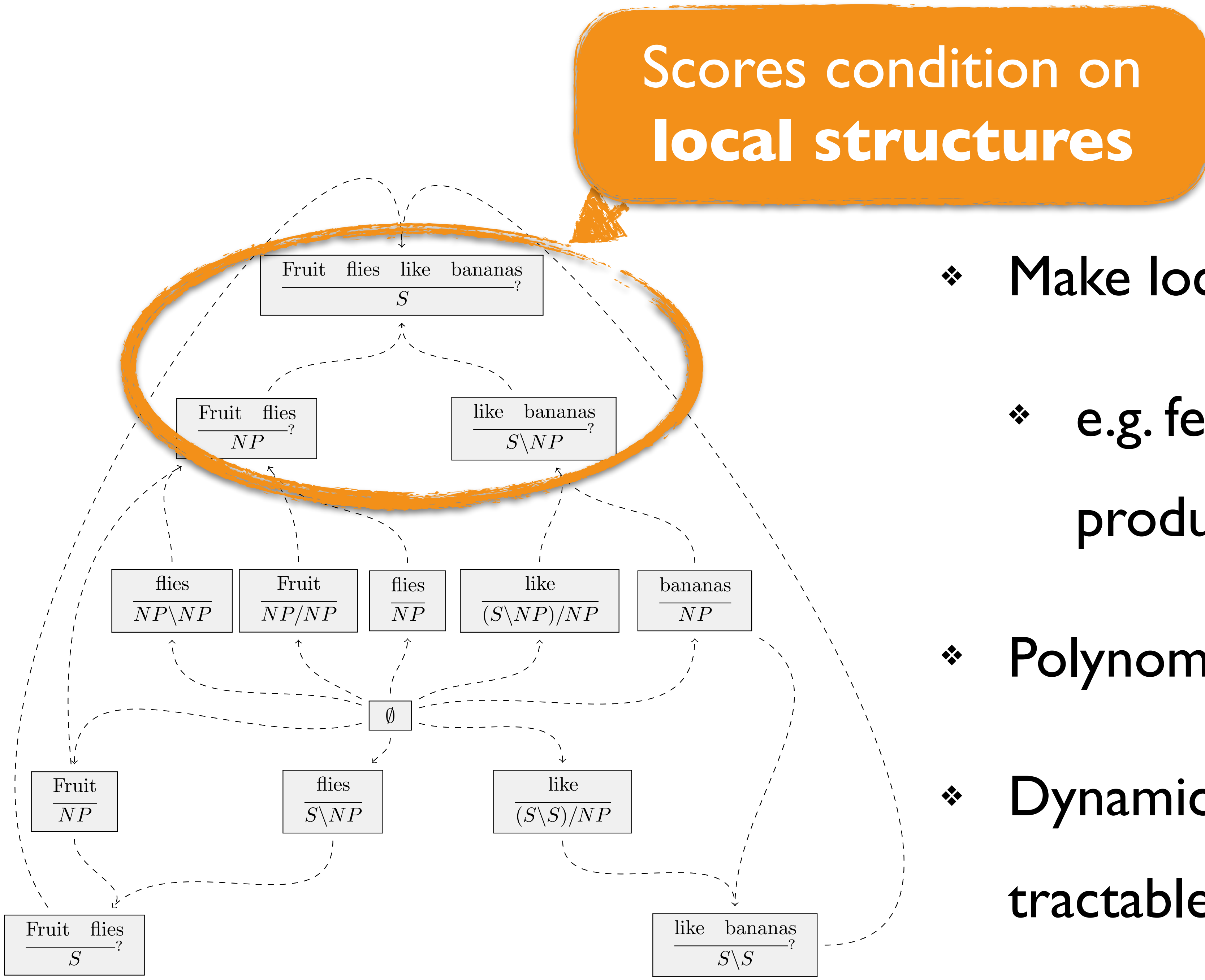
Managing Intractable Search Spaces



Approximate inference with **global expressivity**, e.g.

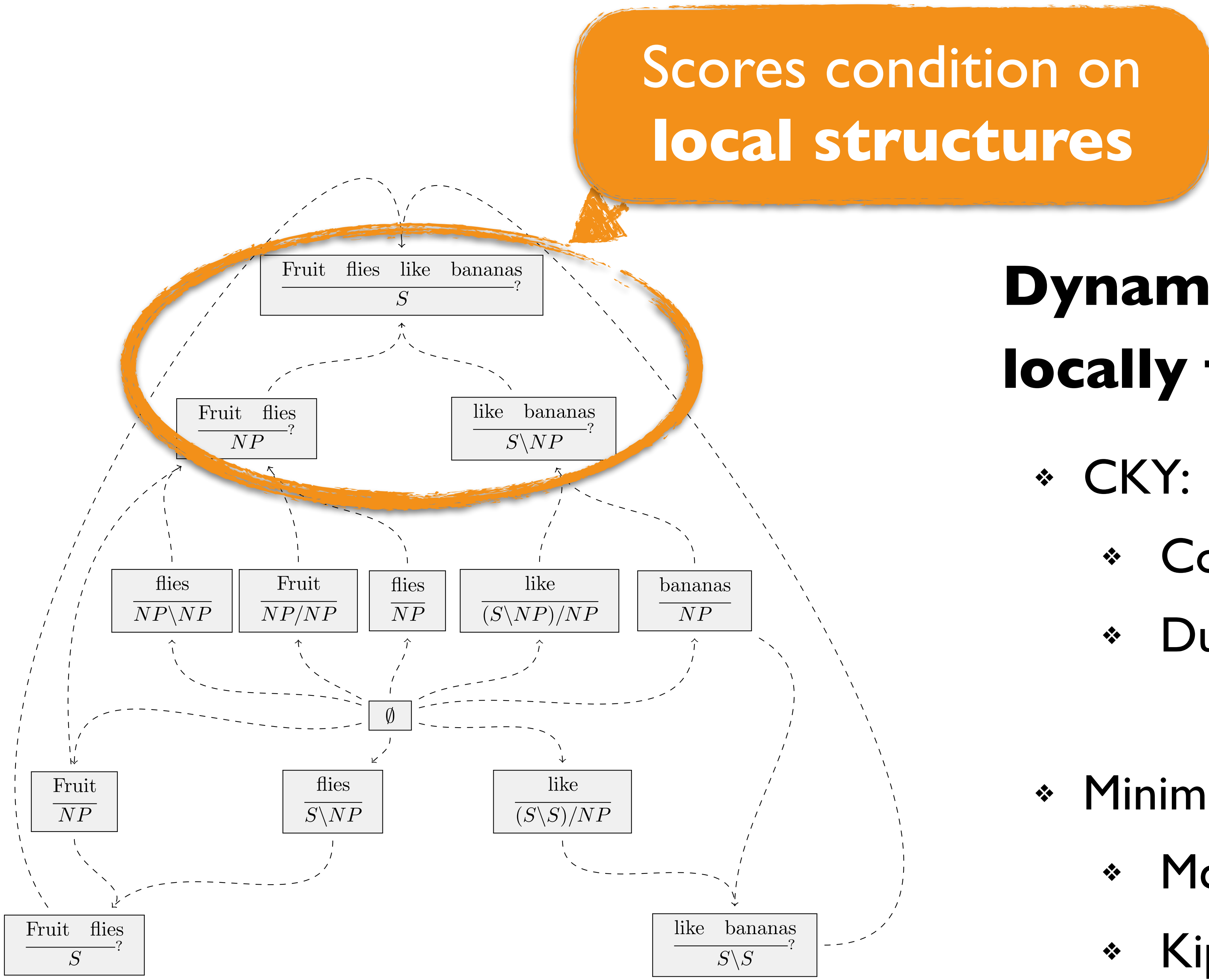
- Greedy / beam search:
  - Nivre, 2008
  - Chen and Manning, 2014
  - Andor et al., 2016
- Reranking:
  - Charniak and Johnson, 2005
  - Huang, 2008
  - Socher et al., 2013

Locally Factored Parsing



- Make locality assumptions:
  - e.g. features are local to CFG productions
- Polynomial number of nodes
- Dynamic programs enable tractable inference

Locally Factored Parsing

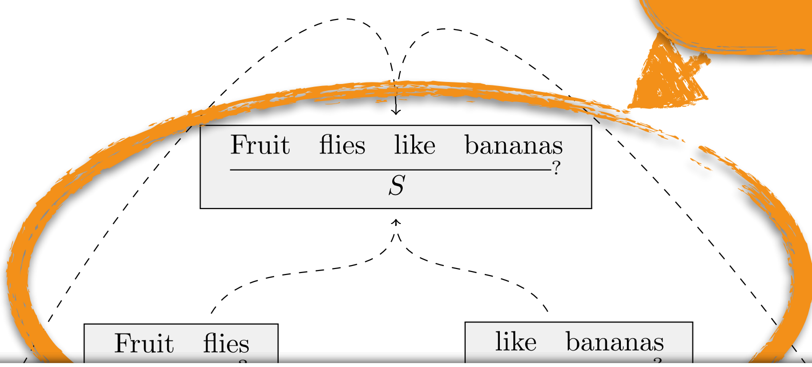


Dynamic programs with **locally factored models**, e.g.

- CKY:
  - Collins, 1997
  - Durrett and Klein, 2015
- Minimum spanning tree:
  - McDonald et al., 2005
  - Kiperwasser and Goldberg, 2016

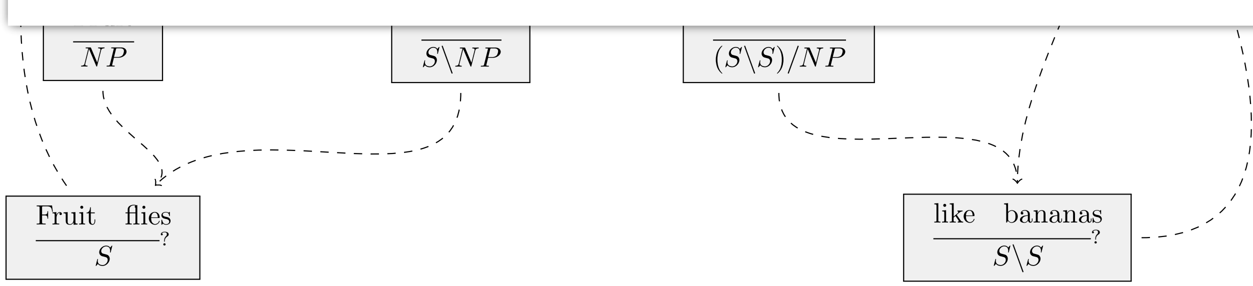
Locally Factored Parsing

Scores condition on **local structures**



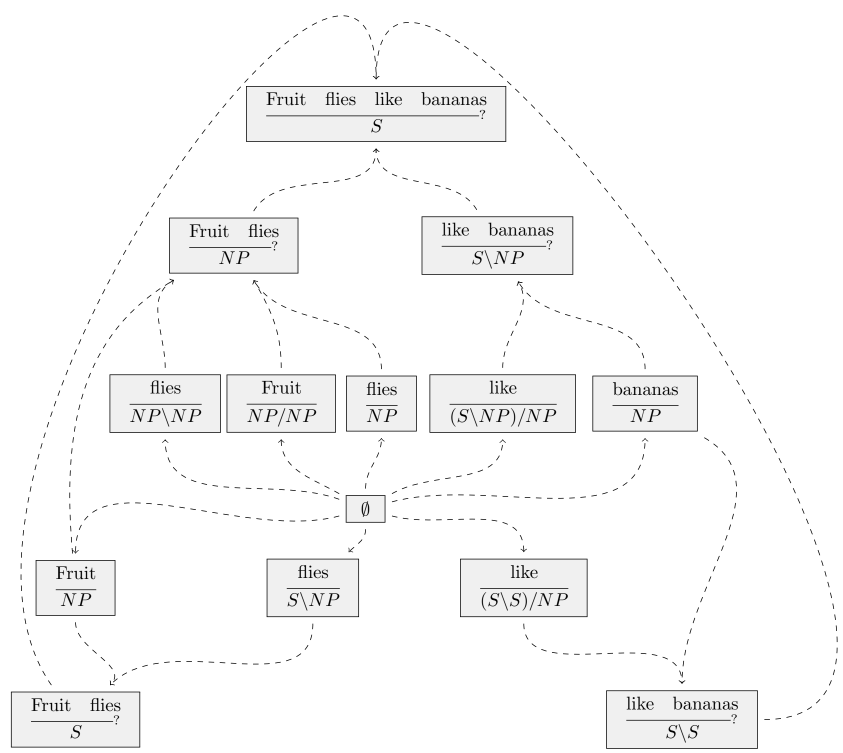
Dynamic programs with **locally factored models**, e.g.

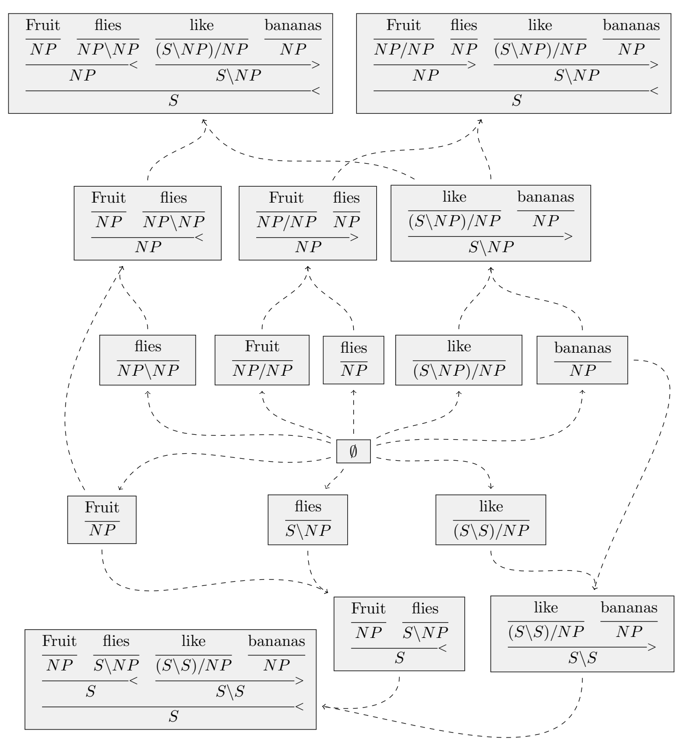
Recursive neural networks break dynamic programs!



- Minimum spanning tree:
  - McDonald et al., 2005
  - Kiperwasser and Goldberg, 2016

Local vs. Global Models





Local model:

$$y^* = \operatorname*{argmax}_{y \in Y} \left(g_{local}(y) \right)$$

Efficient

Inexpressive

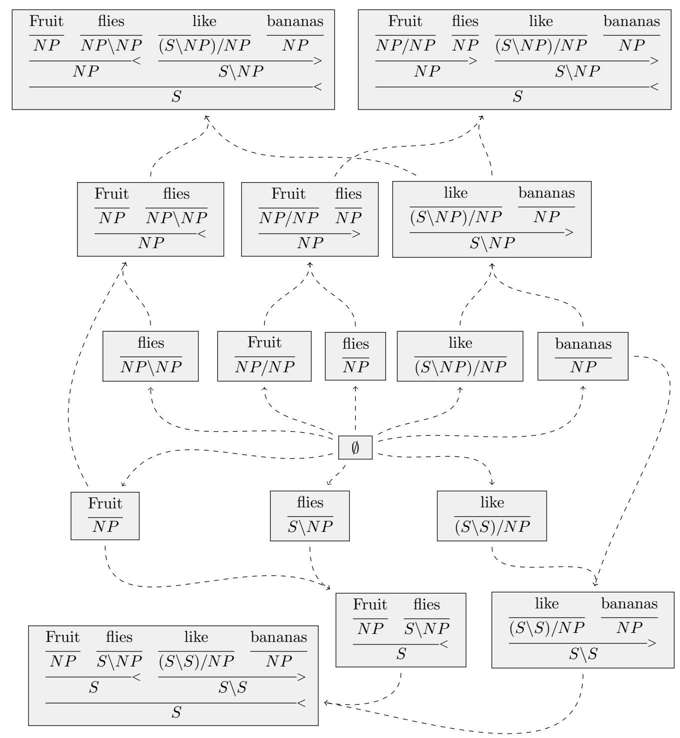
Global model:

$$y^* = \operatorname*{argmax}_{y \in Y} \left(g_{global}(y) \right)$$

Intractable

Expressive

This Work



Combined model:

$$y^* = \operatorname*{argmax}_{y \in Y} \left(g_{local}(y) + g_{global}(y) \right)$$

Efficient

Expressive

A^* Parsing

$$y^* = \underset{y \in Y}{\operatorname{argmax}} \; g(y)$$

- Search in the space of partial parses
- First explored full parse **guaranteed to be optimal**

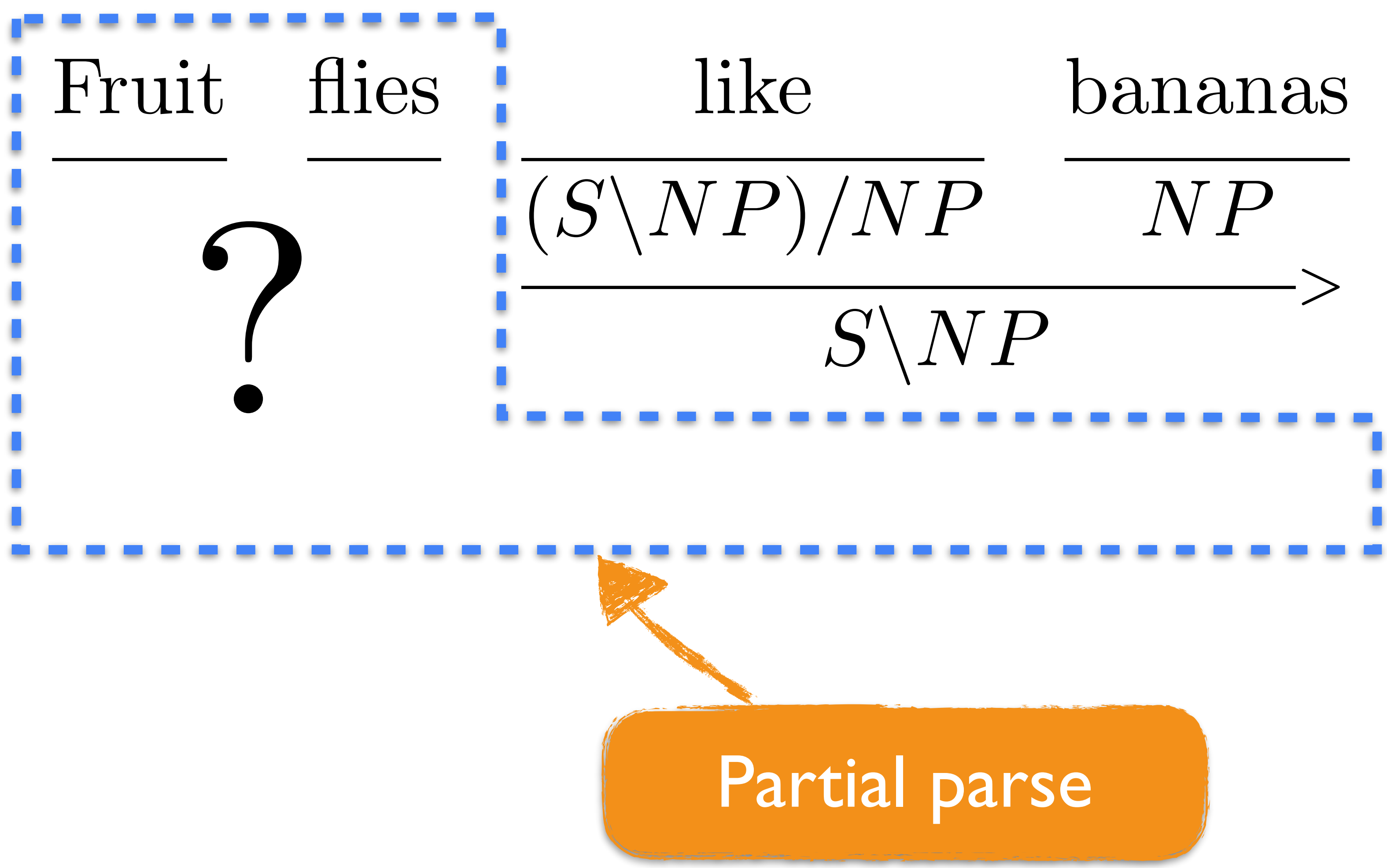
Klein and Manning, 2003

A* Parsing

$$
\begin{array}{cccc}
\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\
\cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}
\multicolumn{2}{c}{?} & (S\backslash NP)/NP & NP \\
\cline{3-4}
 & & \multicolumn{2}{c}{S\backslash NP \quad >}
\end{array}
$$

Partial parse

A* Parsing



A^* Parsing

$$f\left(\begin{array}{cccc}\text{Fruit} & \text{flies} & \text{like} & \text{bananas}\\ \cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4} \multicolumn{2}{c}{?} & (S\backslash NP)/NP & NP \\ \cline{3-4} & & \multicolumn{2}{c}{S\backslash NP \quad >} \end{array}\right)$$

Exploration priority

Partial parse

A* Parsing

Exploration priority

$$f\left(\begin{array}{cccc} \text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\ \cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4} \multicolumn{2}{c}{?} & (S\backslash NP)/NP & NP \\ \cline{3-4} & & \multicolumn{2}{c}{S\backslash NP \quad >} \end{array}\right) = g\left(\begin{array}{cccc} \text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\ \cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4} \multicolumn{2}{c}{?} & (S\backslash NP)/NP & NP \\ \cline{3-4} & & \multicolumn{2}{c}{S\backslash NP \quad >} \end{array}\right) + h\left(\begin{array}{cccc} \text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\ \cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4} \multicolumn{2}{c}{?} & (S\backslash NP)/NP & NP \\ \cline{3-4} & & \multicolumn{2}{c}{S\backslash NP \quad >} \end{array}\right)$$

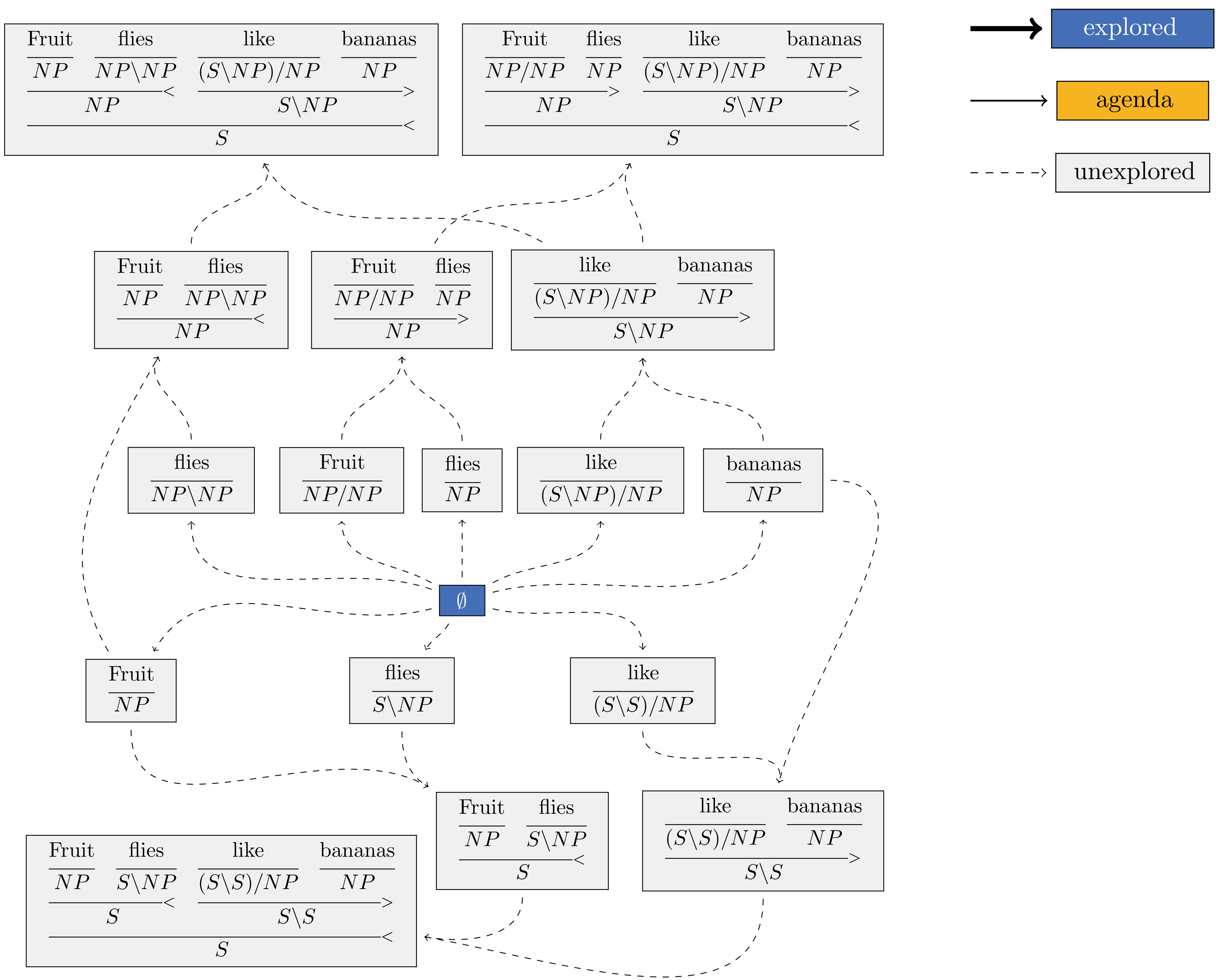
Inside score

$$\begin{array}{cccc} \text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\ \cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4} \multicolumn{2}{c}{?} & (S\backslash NP)/NP & NP \\ \cline{3-4} & & \multicolumn{2}{c}{S\backslash NP \quad >} \end{array}$$

Admissible A* heuristic

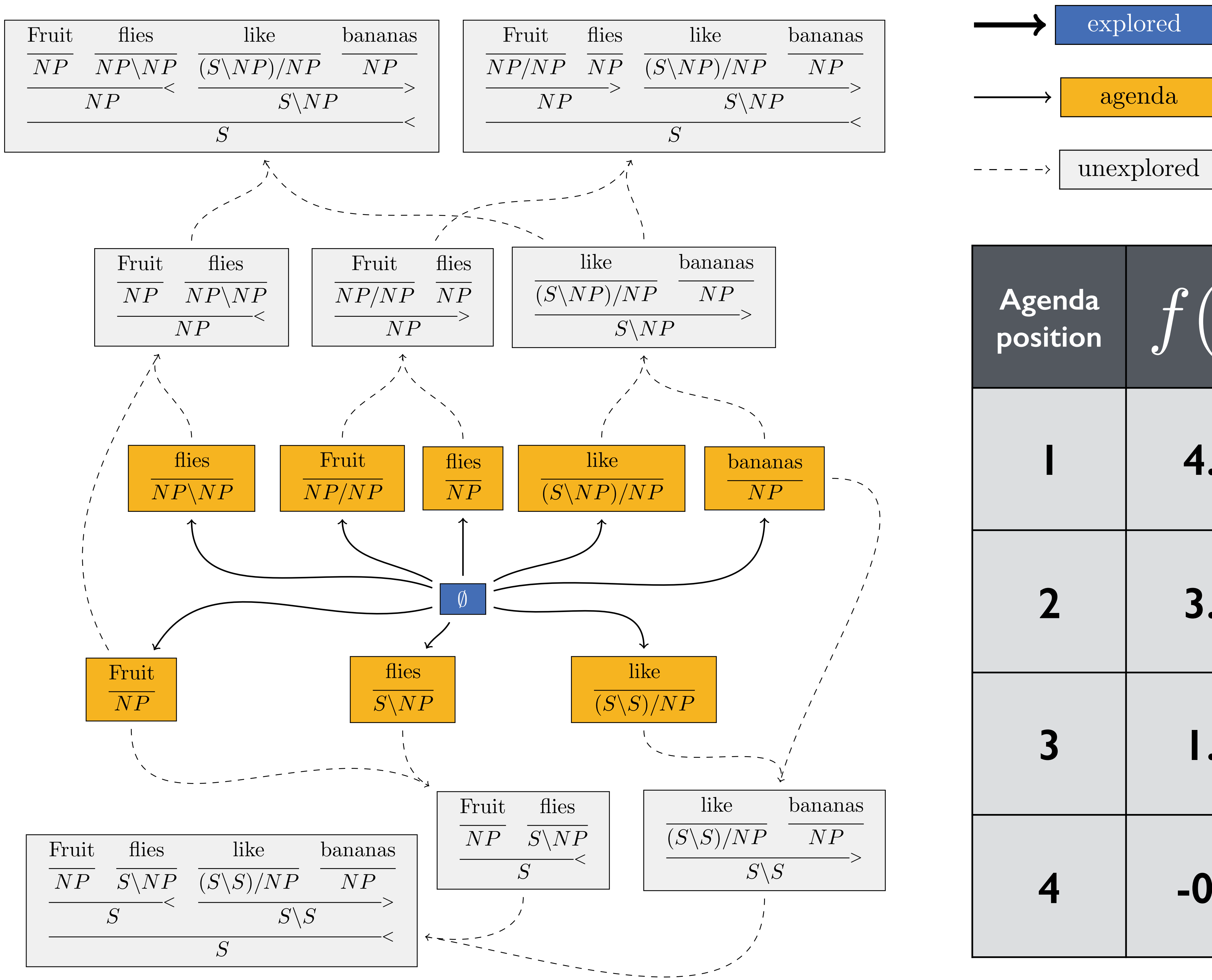
$$\begin{array}{cccc} \text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\ \cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4} \multicolumn{2}{c}{?} & (S\backslash NP)/NP & NP \\ \cline{3-4} & & \multicolumn{2}{c}{S\backslash NP \quad >} \end{array}$$

A* Parsing



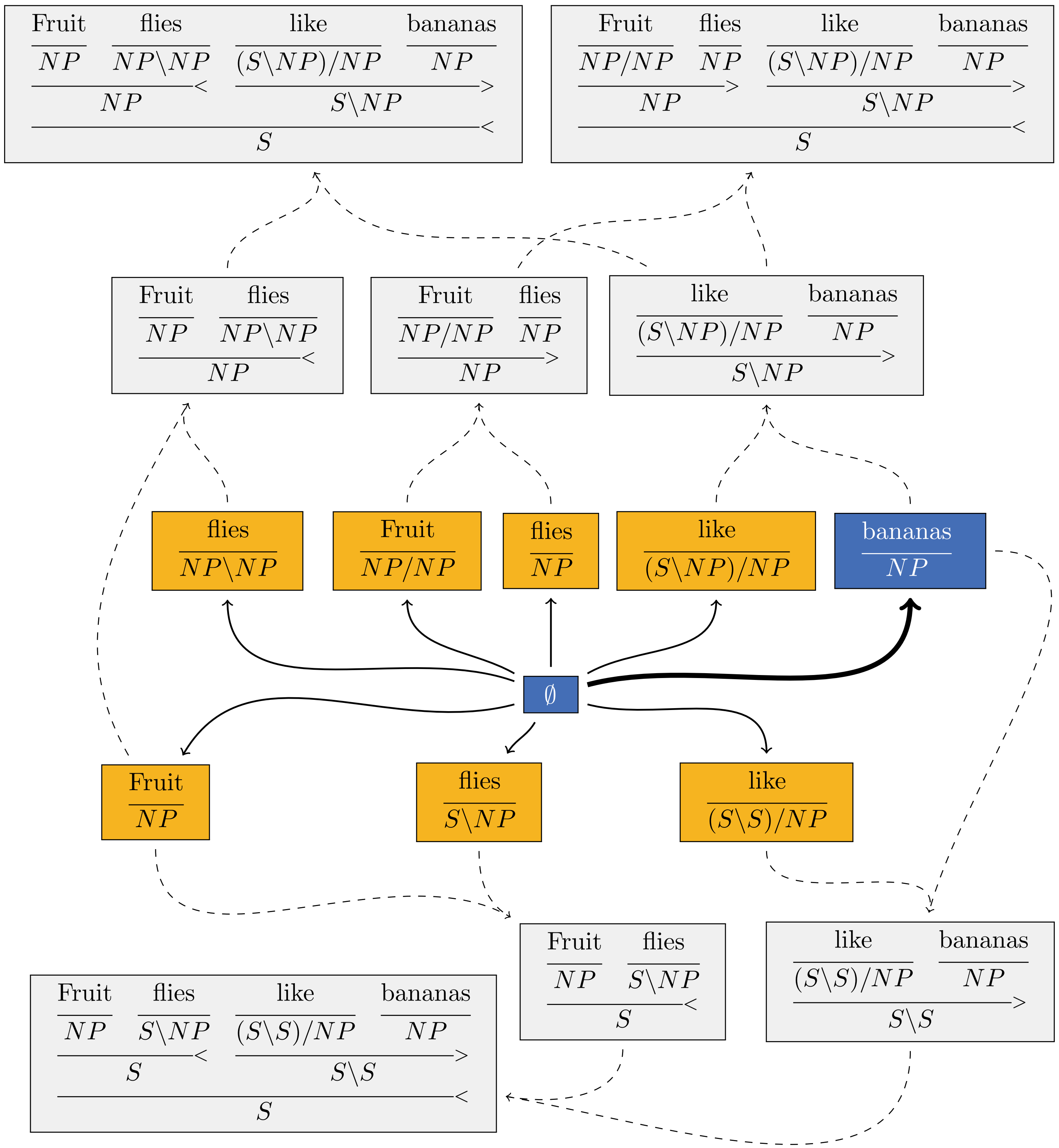
explored
agenda
unexplored
Fruit flies like bananas
NP $NP\backslash NP$ $(S\backslash NP)/NP$ NP
NP $S\backslash NP$
S
Fruit flies like bananas
NP/NP NP $(S\backslash NP)/NP$ NP
NP $S\backslash NP$
S
Fruit flies
NP $NP\backslash NP$
NP
Fruit flies
NP/NP NP
NP
like bananas
$(S\backslash NP)/NP$ NP
$S\backslash NP$
flies
$NP\backslash NP$
Fruit
NP/NP
flies
NP
like
$(S\backslash NP)/NP$
bananas
NP
$\emptyset$
Fruit
NP
flies
$S\backslash NP$
like
$(S\backslash S)/NP$
Fruit flies
NP $S\backslash NP$
S
like bananas
$(S\backslash S)/NP$ NP
$S\backslash S$
Fruit flies like bananas
NP $S\backslash NP$ $(S\backslash S)/NP$ NP
S $S\backslash S$
S

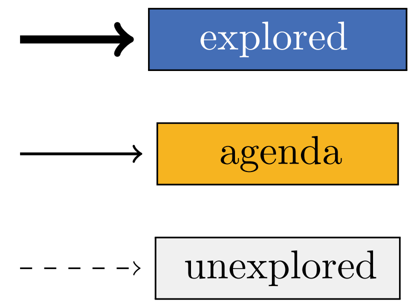
A* Parsing



Agenda position	$f(y)$	y
1	4.5	bananas $\frac{}{NP}$
2	3.1	like $\frac{}{(S\backslash NP)/NP}$
3	1.9	Fruit $\frac{}{NP}$
4	-0.5	Fruit $\frac{}{NP/NP}$

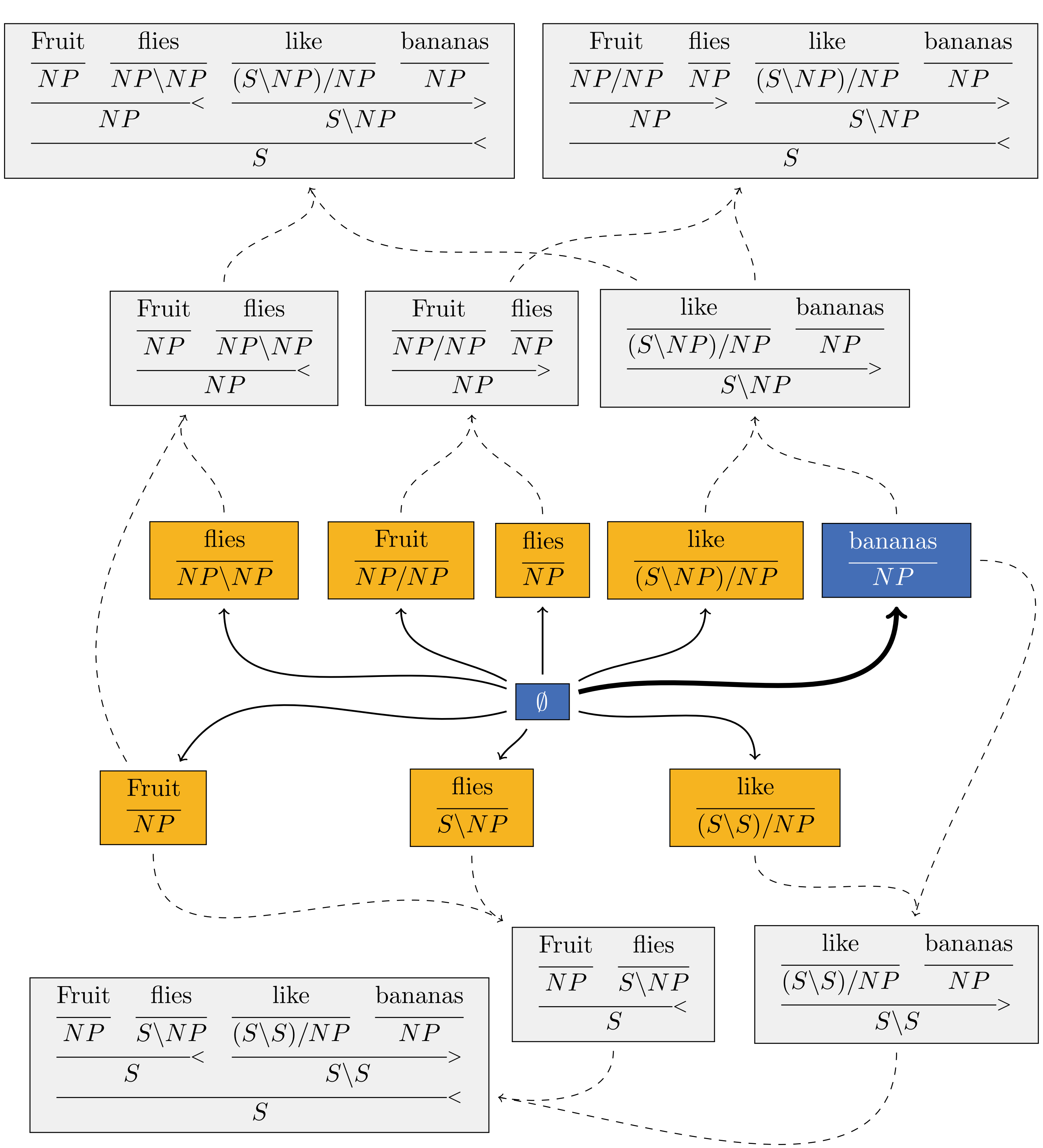
A* Parsing

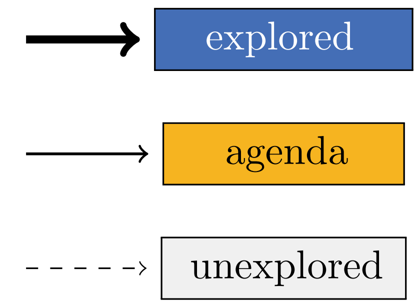




Agenda position	$f(y)$	y
1	4.5	$\frac{\text{bananas}}{NP}$
2	3.1	$\frac{\text{like}}{(S\backslash NP)/NP}$
3	1.9	$\frac{\text{Fruit}}{NP}$
4	-0.5	$\frac{\text{Fruit}}{NP/NP}$

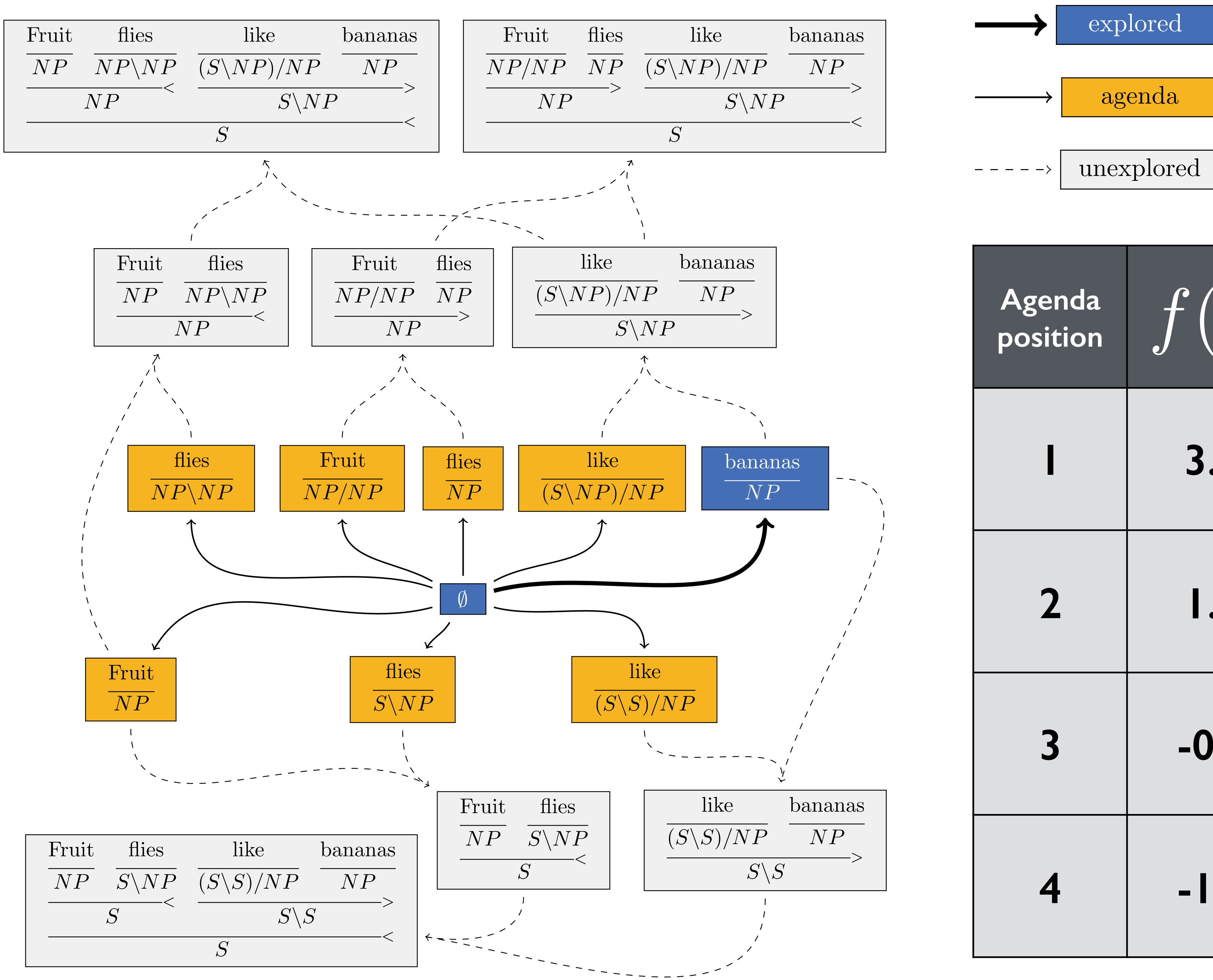
A* Parsing





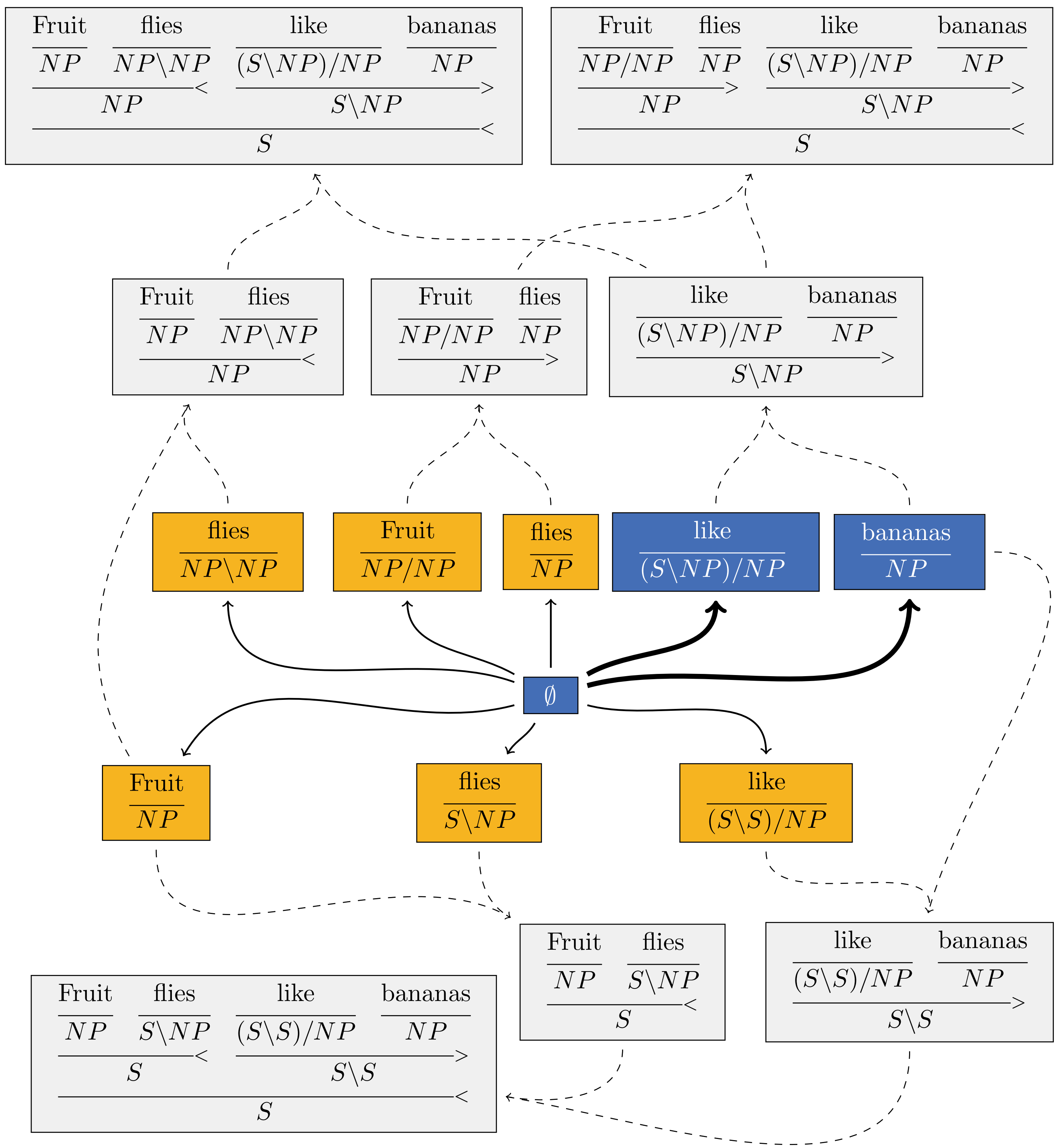
Agenda position	$f(y)$	y
2	3.1	like $\frac{}{(S\backslash NP)/NP}$
3	1.9	Fruit $\frac{}{NP}$
4	-0.5	Fruit $\frac{}{NP/NP}$

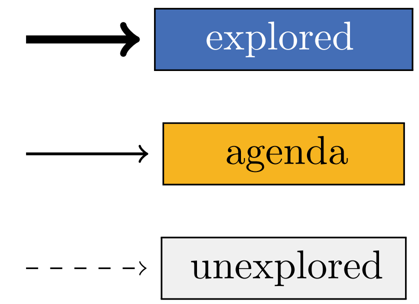
A* Parsing



Agenda position	$f(y)$	y
1	3.1	like $\frac{}{(S\backslash NP)/NP}$
2	1.9	Fruit $\frac{}{NP}$
3	-0.5	Fruit $\frac{}{NP/NP}$
4	-1.3	flies $\frac{}{NP}$

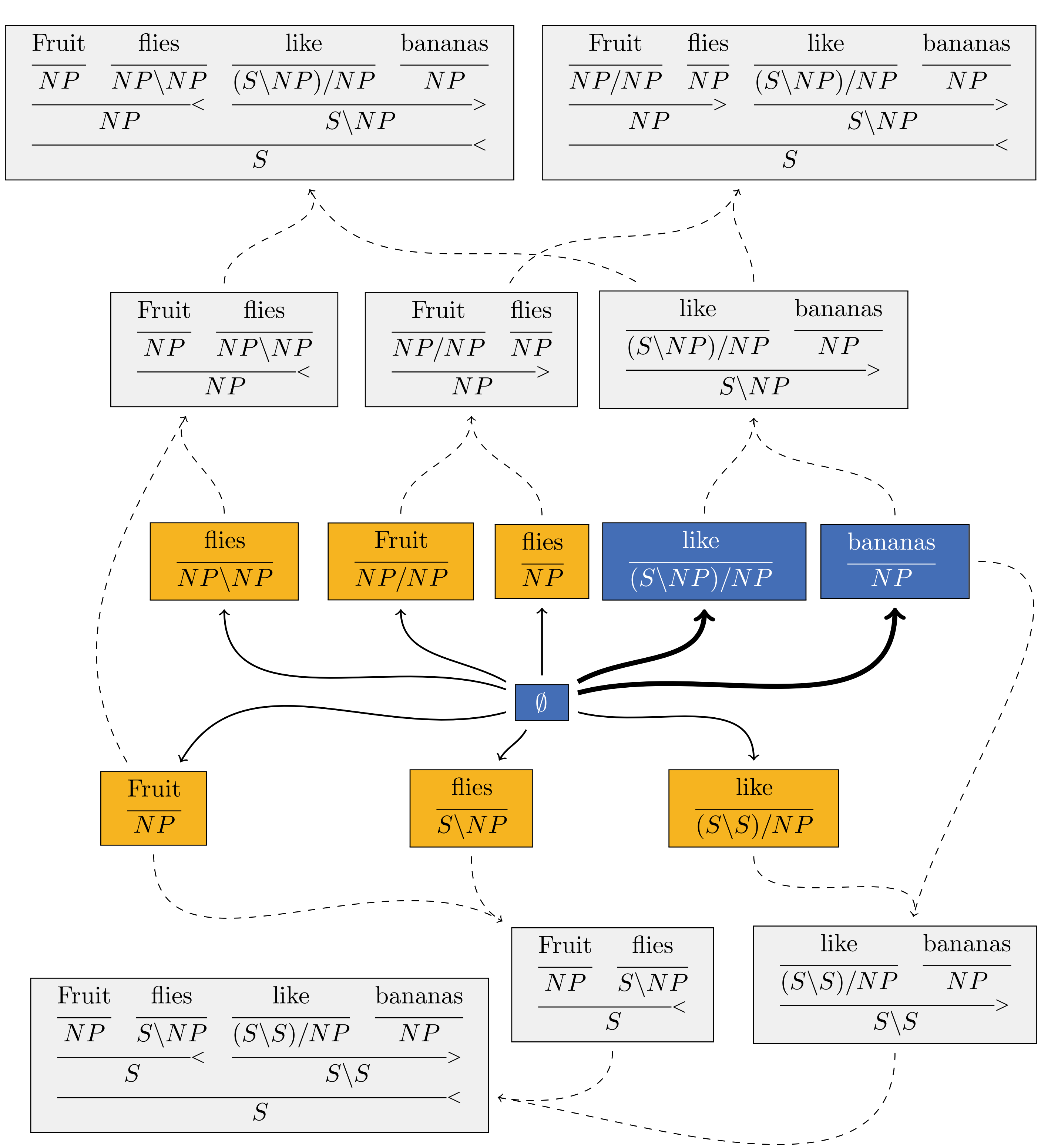
A* Parsing

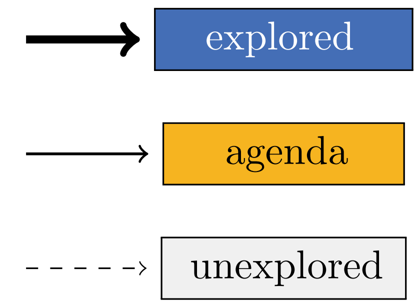




Agenda position	$f(y)$	y
1	3.1	like / $(S\backslash NP)/NP$
2	1.9	Fruit / NP
3	-0.5	Fruit / NP/NP
4	-1.3	flies / NP

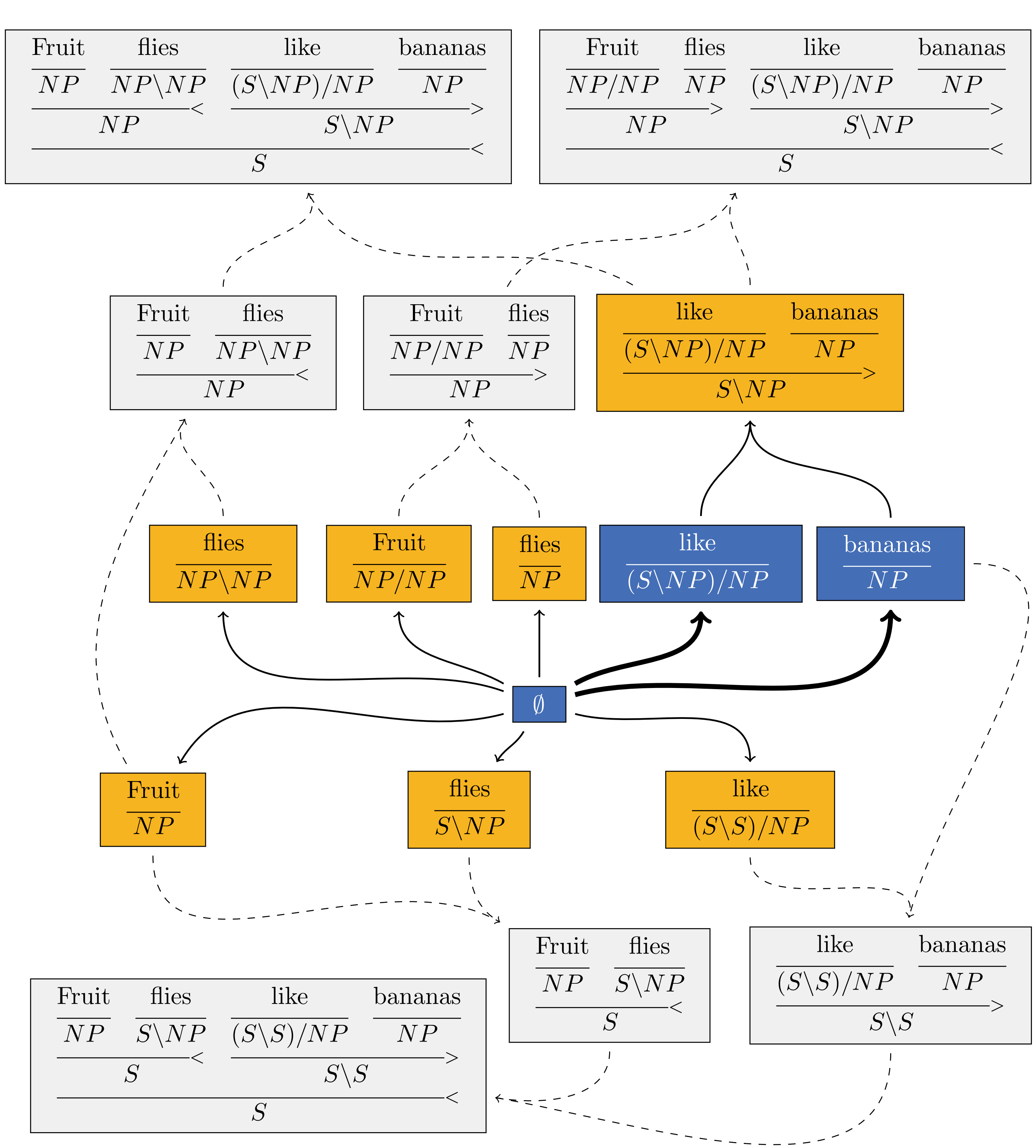
A* Parsing

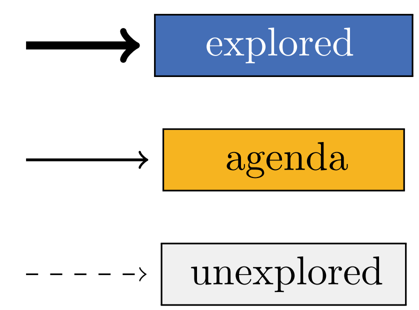




Agenda position	$f(y)$	y
2	1.9	Fruit $\frac{}{NP}$
3	-0.5	Fruit $\frac{}{NP/NP}$
4	-1.3	flies $\frac{}{NP}$

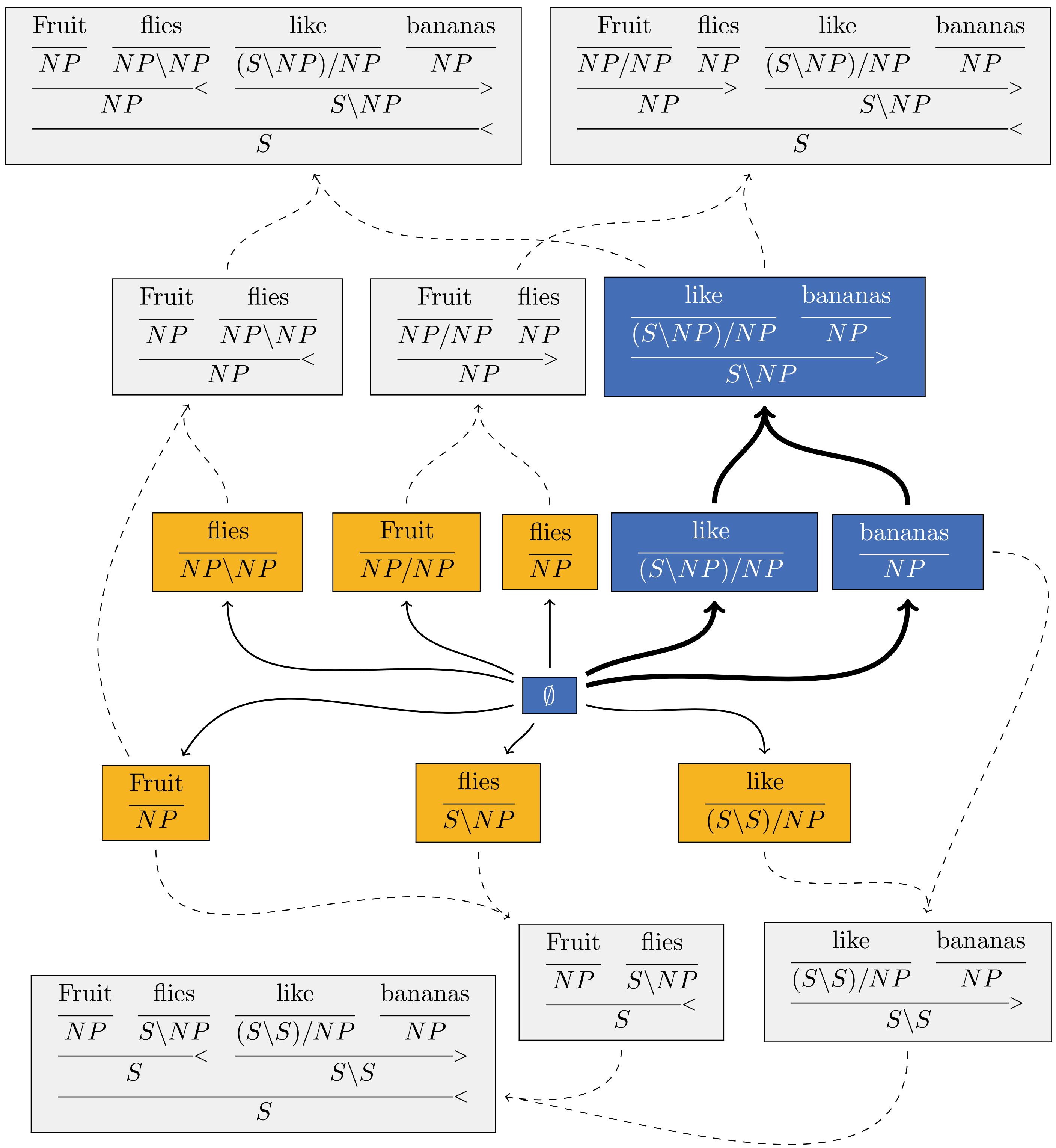
A* Parsing

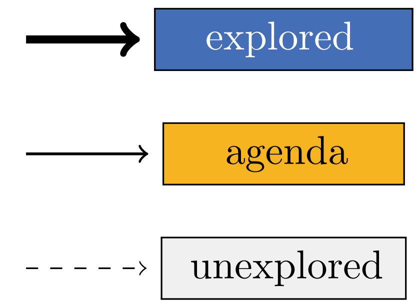




Agenda position	$f(y)$	y
1	2.1	like bananas: $\dfrac{(S\backslash NP)/NP \quad NP}{S\backslash NP}>$
2	1.9	Fruit: $\dfrac{}{NP}$
3	-0.5	Fruit: $\dfrac{}{NP/NP}$
4	-1.3	flies: $\dfrac{}{NP}$

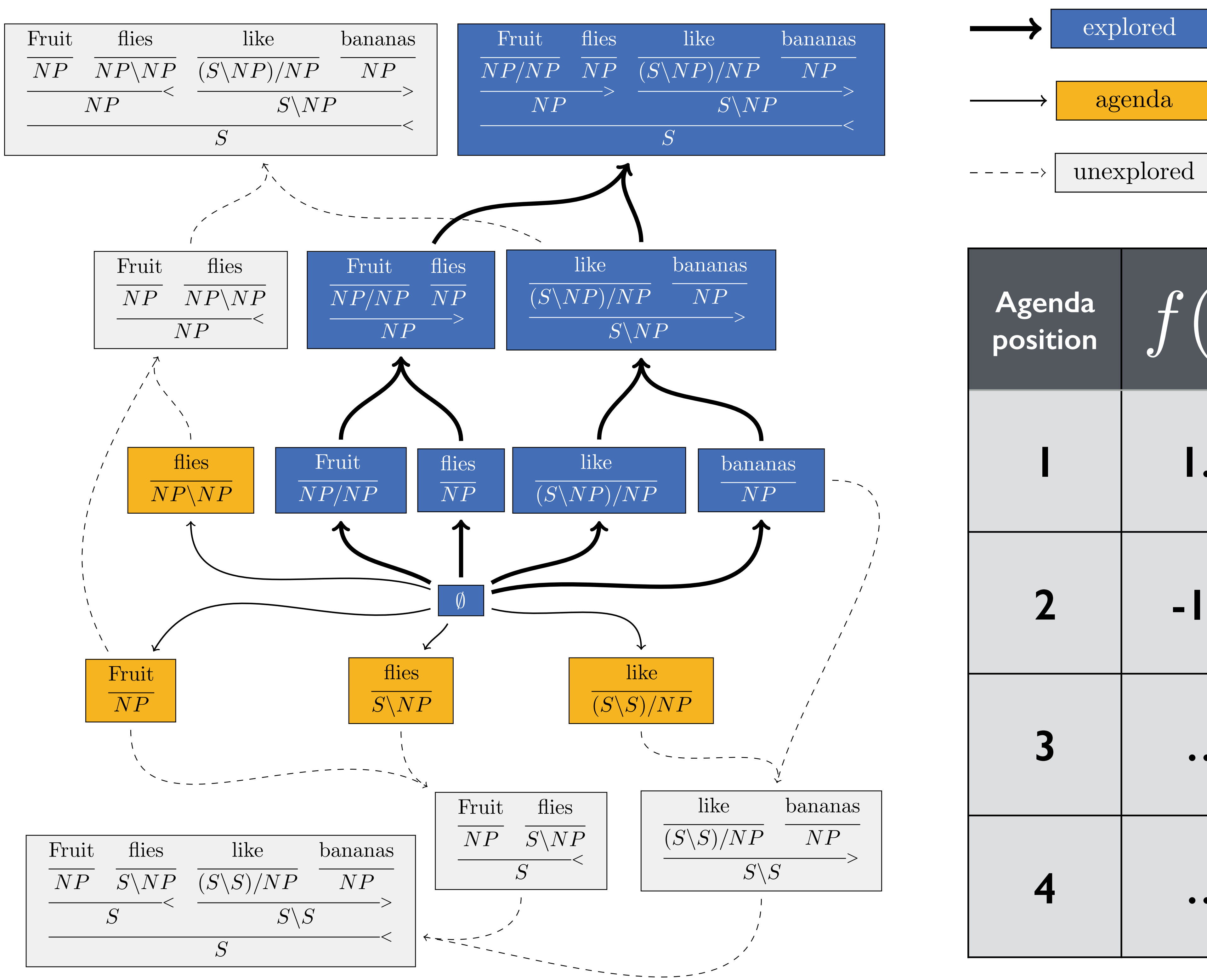
A* Parsing





Agenda position	$f(y)$	y
1	2.1	like: $(S\backslash NP)/NP$, bananas: NP $\Rightarrow$ $S\backslash NP$
2	1.9	Fruit: NP
3	-0.5	Fruit: NP/NP
4	-1.3	flies: NP

A* Parsing



Agenda position	$f(y)$	y
1	1.9	Fruit $\frac{}{NP}$
2	-1.5	like $\frac{}{(S\backslash S)/NP}$
3	...	...
4	...	...

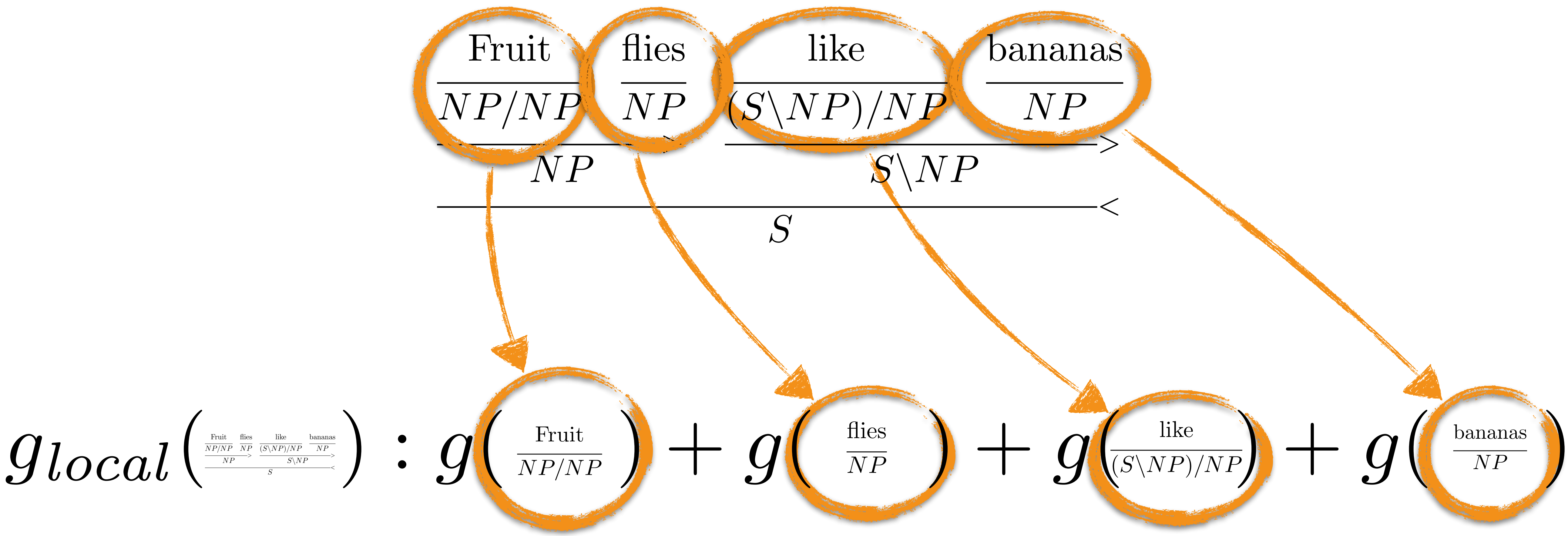
Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

$$
\begin{array}{cccc}
\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\
\hline
NP/NP & NP & (S\backslash NP)/NP & NP \\
\hline
\multicolumn{2}{c}{NP \quad >} & \multicolumn{2}{c}{S\backslash NP \quad >} \\
\hline
\multicolumn{4}{c}{S \quad <}
\end{array}
$$

Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):



Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

$$
\dfrac{\text{Fruit}}{} \quad \dfrac{\text{flies}}{} \quad \dfrac{\dfrac{\text{like}}{(S\backslash NP)/NP} \quad \dfrac{\text{bananas}}{NP}}{S\backslash NP}>
$$

$$?$$

Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

$$
\begin{array}{cccc}
\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\
\hline & \hline & \hline (S\backslash NP)/NP & \hline NP \\
\multicolumn{2}{c}{?} & \multicolumn{2}{c}{\overline{\quad\quad S\backslash NP \quad\quad}>}
\end{array}
$$

$$
g_{local}\left(
\begin{array}{cccc}
\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\
\hline & \hline & \hline (S\backslash NP)/NP & \hline NP \\
\multicolumn{2}{c}{?} & \multicolumn{2}{c}{\overline{\quad S\backslash NP \quad}>}
\end{array}
\right) = g\left(\frac{\text{like}}{(S\backslash NP)/NP}\right) + g\left(\frac{\text{bananas}}{NP}\right)
$$

Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

$$\begin{array}{cccc}
\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\
\hline
 & & (S\backslash NP)/NP & NP \\
\hline
 & & \multicolumn{2}{c}{S\backslash NP \;>}
\end{array}$$

?

$$g_{local}\left(\begin{array}{cccc}\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\ \hline ? & & (S\backslash NP)/NP & NP \\ \hline & & \multicolumn{2}{c}{S\backslash NP \;>}\end{array}\right) : g\left(\frac{\text{like}}{(S\backslash NP)/NP}\right) + g\left(\frac{\text{bananas}}{NP}\right)$$

$$h_{local}\left(\begin{array}{cccc}\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\ \hline ? & & (S\backslash NP)/NP & NP \\ \hline & & \multicolumn{2}{c}{S\backslash NP \;>}\end{array}\right) : \max_{\textbf{tag}} g\left(\frac{\text{Fruit}}{\textbf{tag}}\right) + \max_{\textbf{tag}} g\left(\frac{\text{flies}}{\textbf{tag}}\right)$$

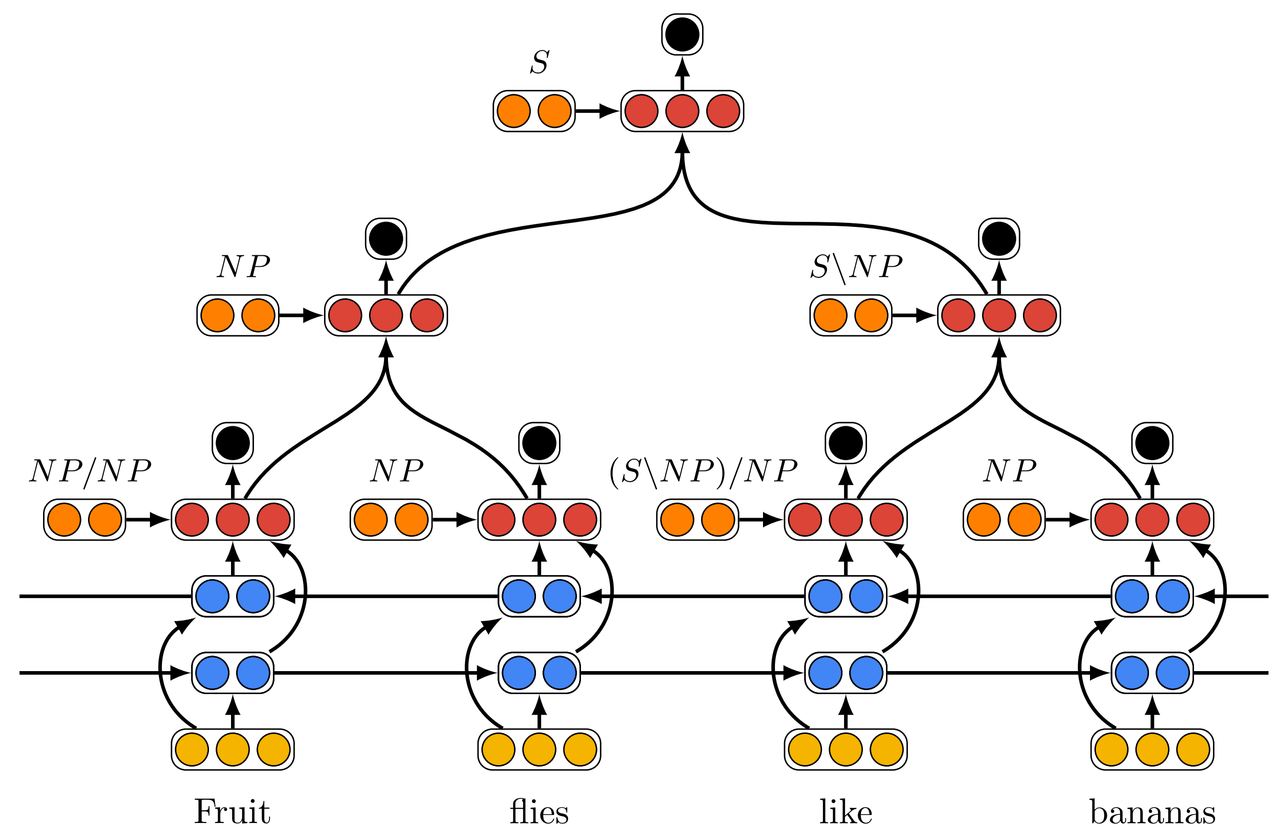
Global A^* Parsing

$$y^* = \operatorname*{argmax}_{y \in Y} \; g(y)$$

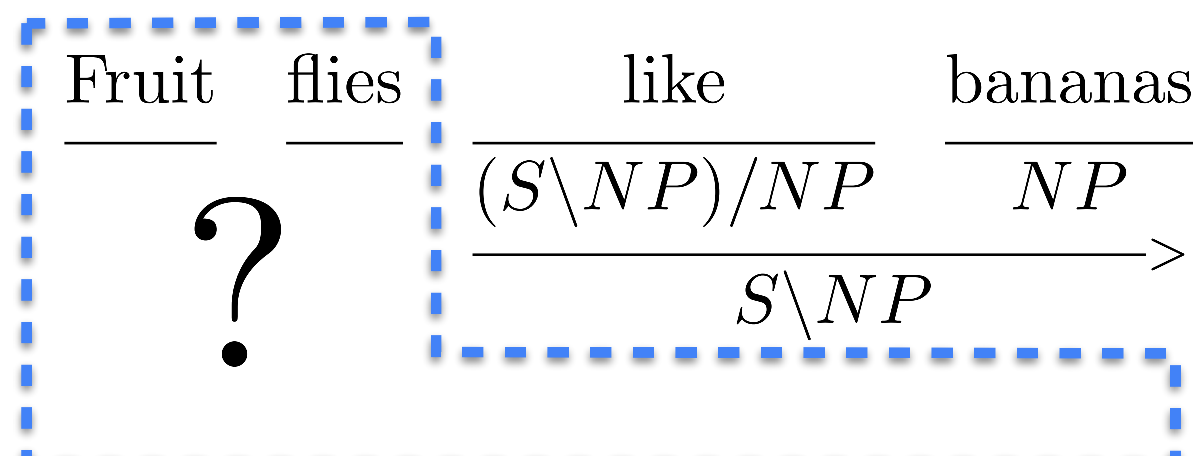
- First explored full parse **guaranteed to be optimal**
- Global search graph is **exponential** in sentence length
- Open question: Can we still **learn to search** efficiently?

Modeling Global Structure

$g_{global}(y)$:



$h_{global}(y)$:



Modeling Global Structure

$$g(y) = g_{global}(y)$$

Non-positive
global model

$$h(y) = 0$$

Modeling Global Structure

$$g(y) = g_{local}(y) + g_{global}(y)$$

Any locally factored model with an admissible A* heuristic

Non-positive global model

$$h(y) = h_{local}(y) + 0$$

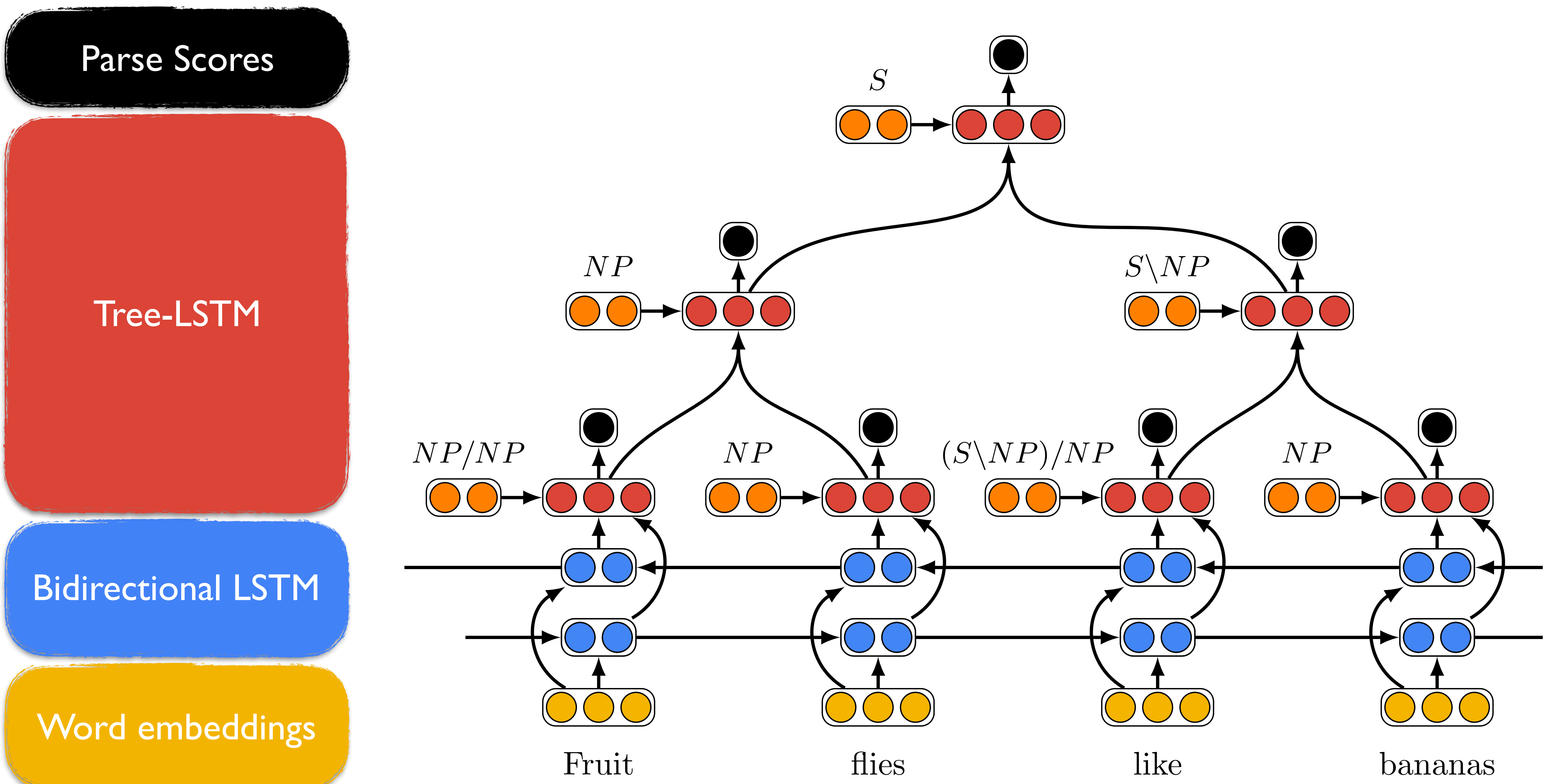
Division of Labor

$$g(y) = g_{local}(y) + g_{global}(y)$$

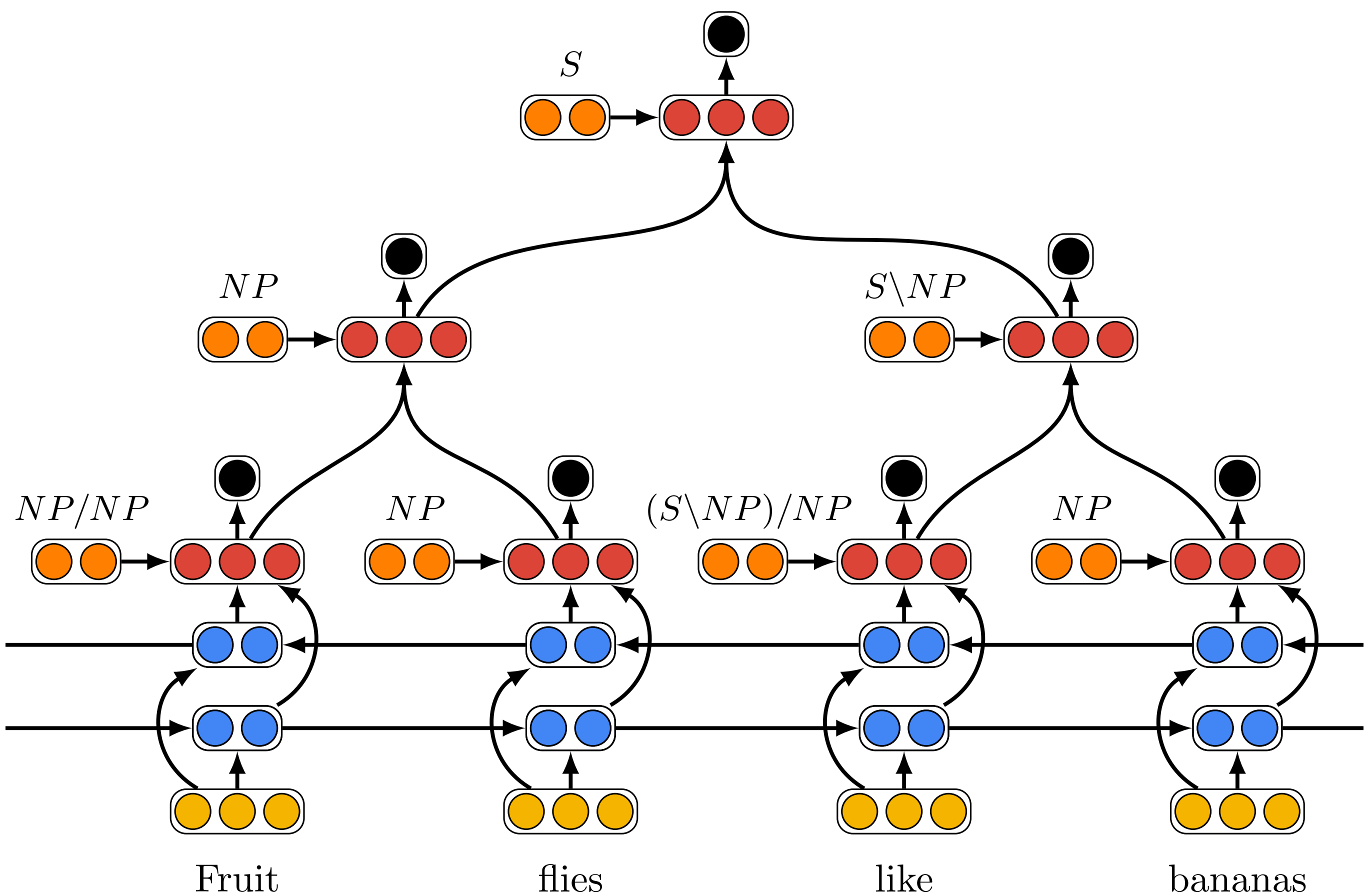
- Limited expressivity
- Provides guidance with an A^* heuristic

- Global expressivity
- Discriminative only when necessary

Global Model: $g_{global}(y)$



Non-positive Global Model



Log-probability of a logistic regression layer

$$g_{global}(\quad) = \log(\sigma(w\cdot \quad))$$

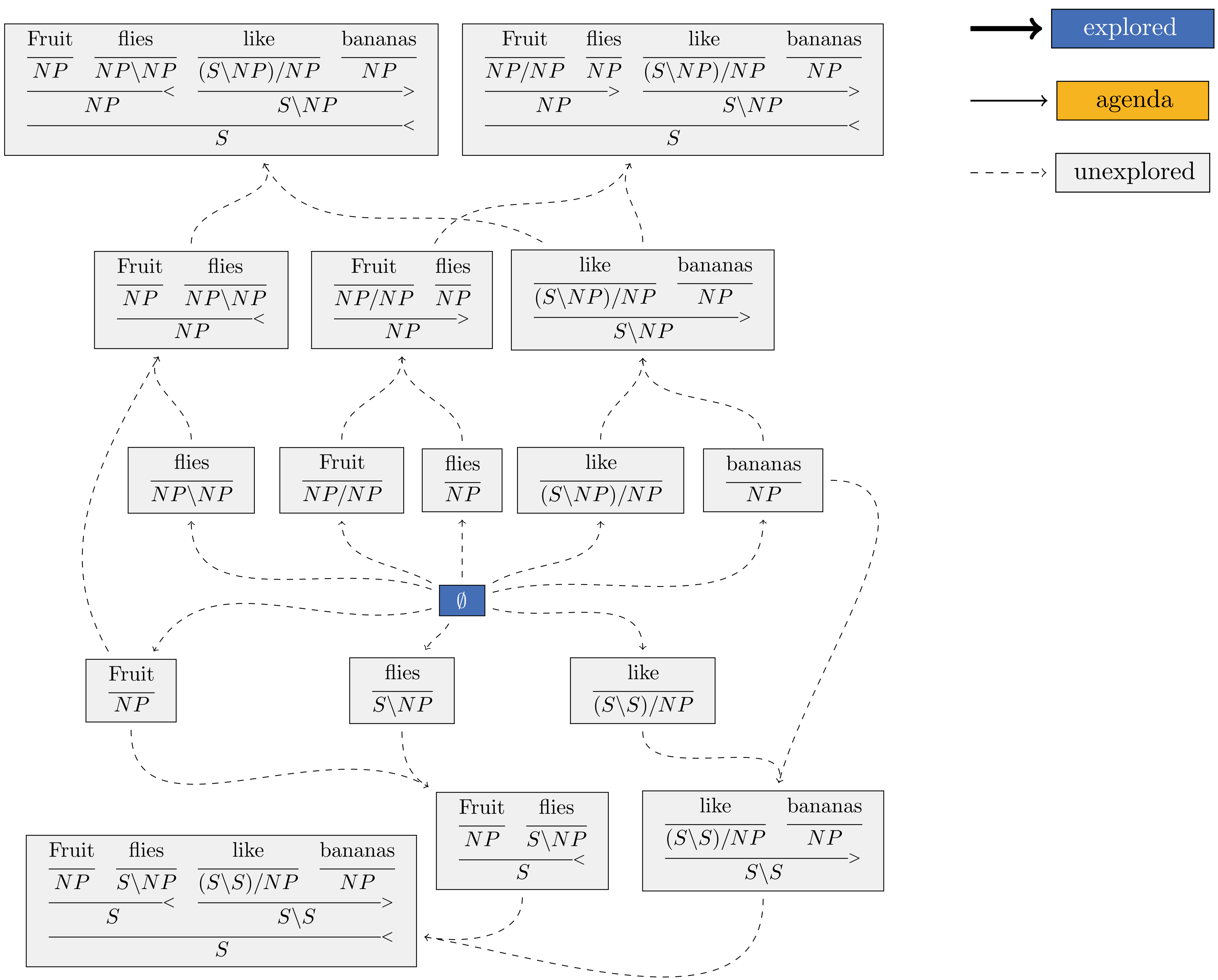
Division of Labor

$$g(y) = g_{local}(y) + g_{global}(y)$$

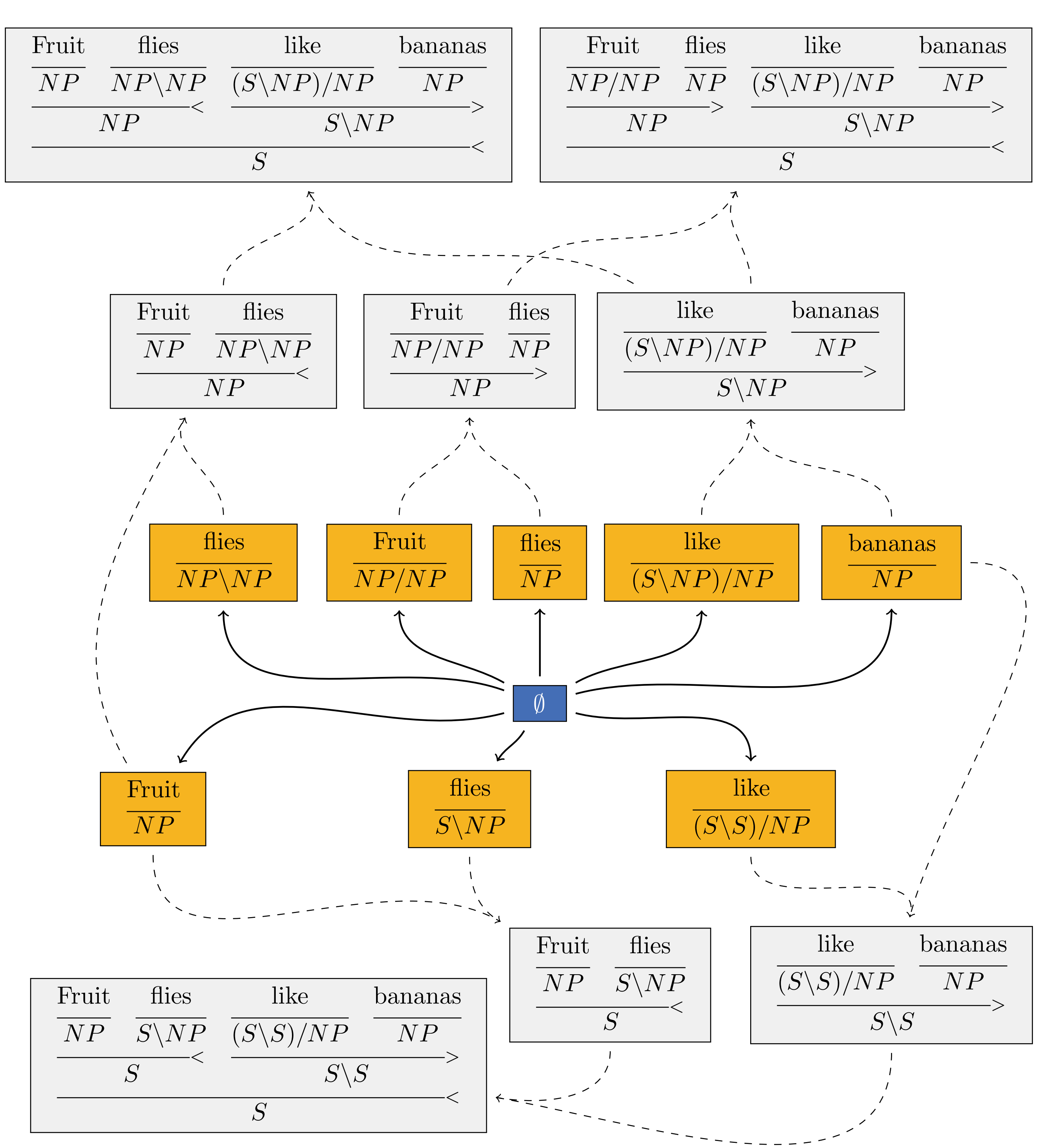
- Limited expressivity
- Provides guidance with an A* heuristic

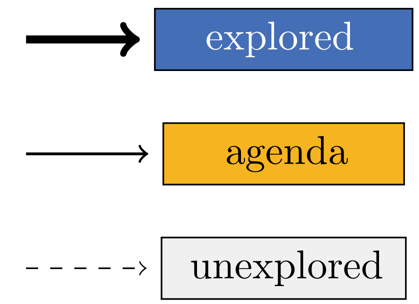
- Global expressivity
- **Discriminative only when necessary**

Learning with A*



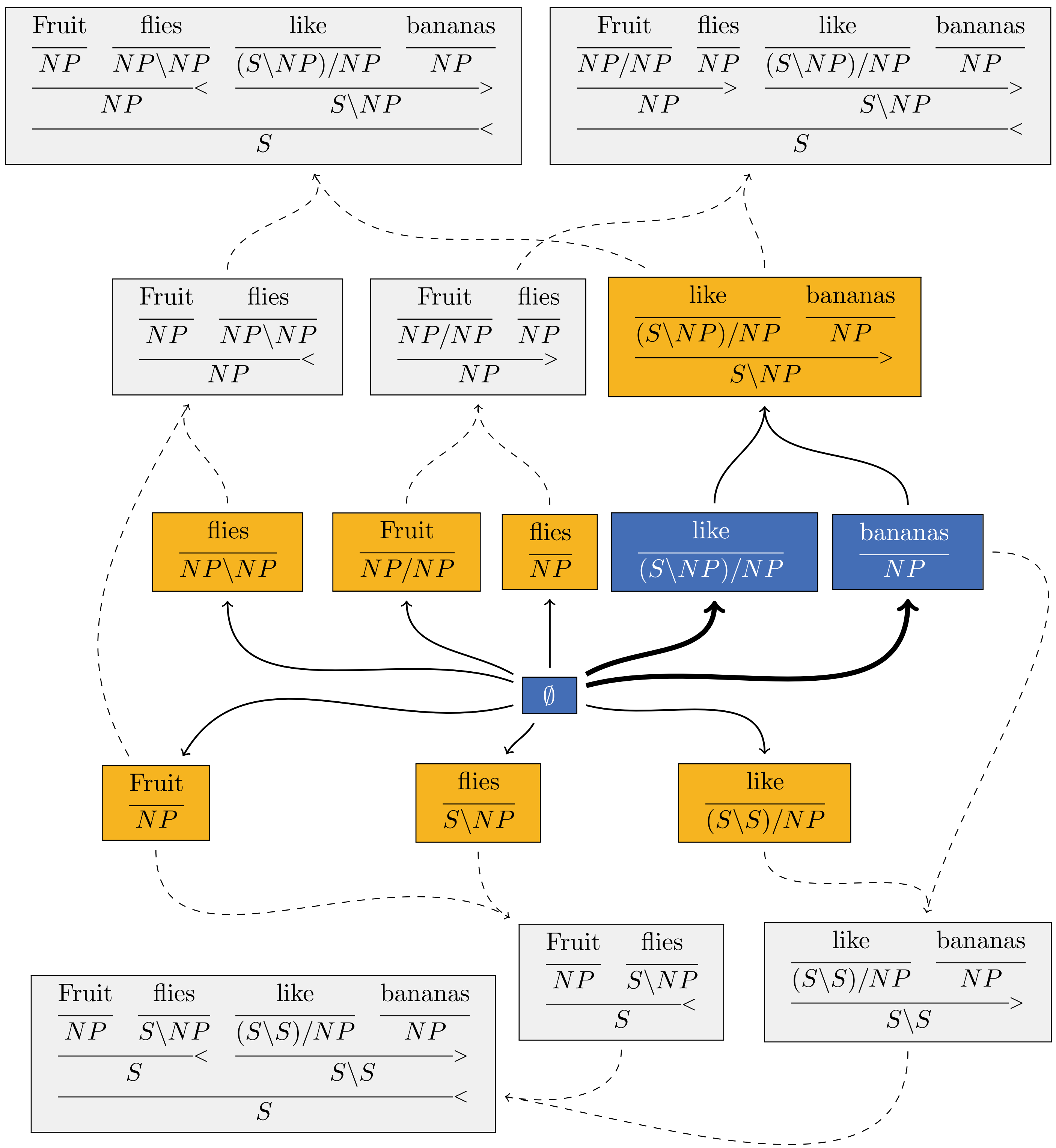
Learning with A*

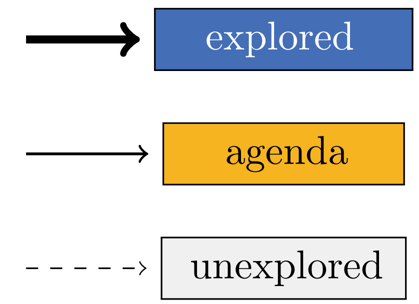




Agenda position	$f(y)$	y	Is correct?
1	4.5	bananas NP	✓
2	3.1	like $(S\backslash NP)/NP$	✓
3	1.9	Fruit NP	✗
4	-0.5	Fruit NP/NP	✓

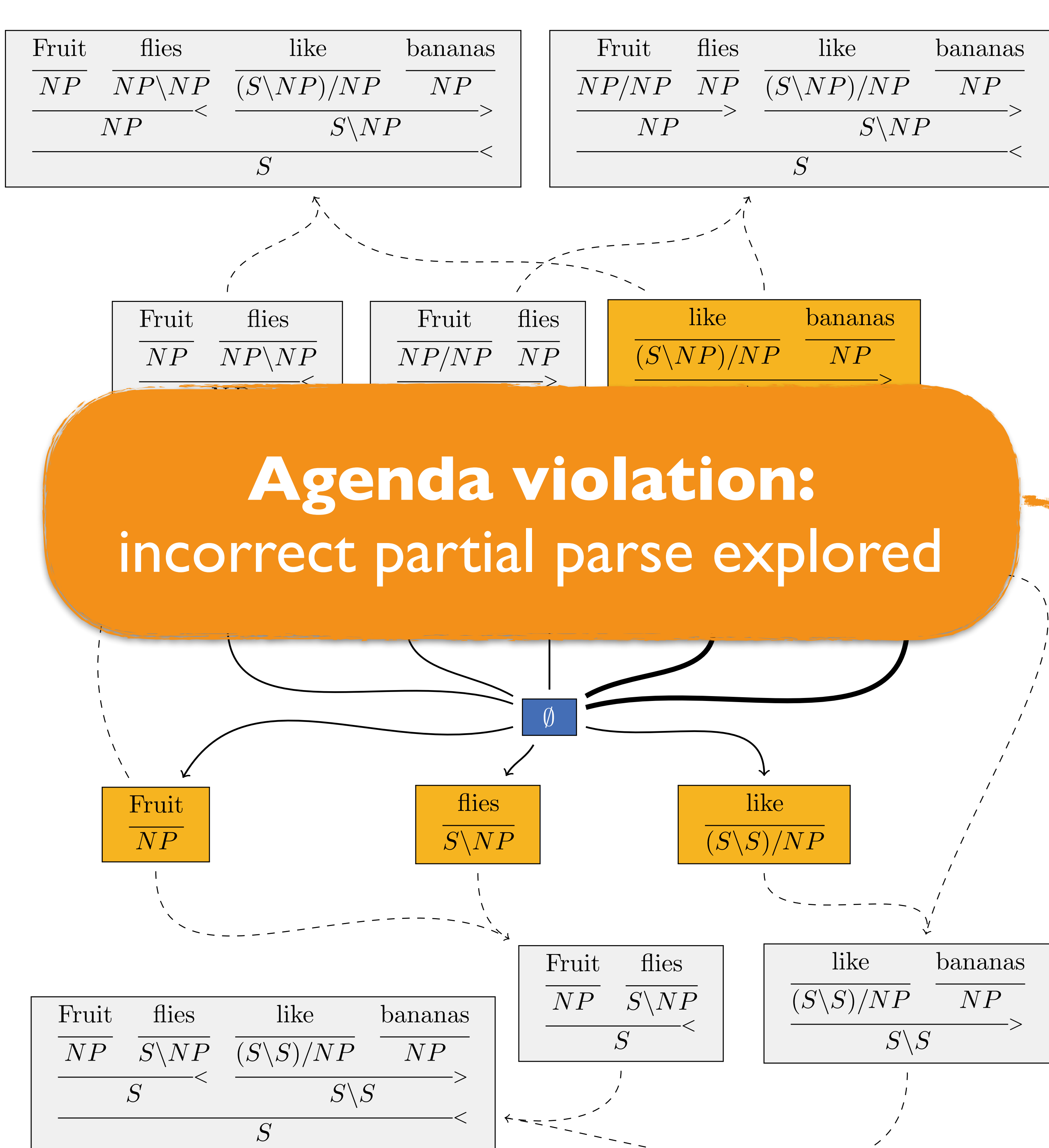
Learning with A*

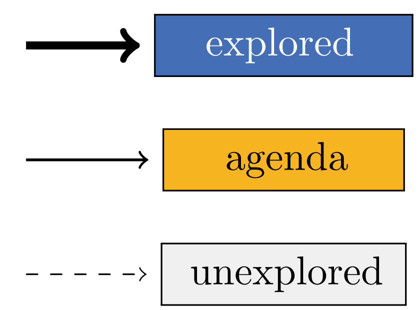




Agenda position	$f(y)$	y	Is correct?
1	1.9	Fruit $\frac{}{NP}$	✗
2	-0.5	Fruit $\frac{}{NP/NP}$	✓
3	...	...	...
4	...	...	...

Learning with A*





Agenda position	$f(y)$	y	Is correct?
1	1.9	Fruit NP	✗
2	-0.5	Fruit NP/NP	✓
3	...	...	...
4	...	...	...

Violation-based Loss

$\mathcal{A} : [$

Agenda position	$f(y)$	y	Is correct?
1	4.5	$\frac{\text{bananas}}{NP}$	✓
2	3.1	$\frac{\text{like}}{(S\backslash NP)/NP}$	✓
3	1.9	$\frac{\text{Fruit}}{NP}$	✗
4	-0.5	$\frac{\text{Fruit}}{NP/NP}$	✓

$\ldots$

Agenda position	$f(y)$	y	Is correct?
1	1.9	$\frac{\text{Fruit}}{NP}$	✗
2	-0.5	$\frac{\text{Fruit}}{NP/NP}$	✓
3	...	...	...
4	...	...	...

$\ldots]$

Violation-based Loss

$\mathcal{A}$: [

Agenda position	$f(y)$	y	Is correct?
1	4.5	$\frac{\text{bananas}}{NP}$	✓
2	3.1	$\frac{\text{like}}{(S\backslash NP)/NP}$	✓
3	1.9	$\frac{\text{Fruit}}{NP}$	✗
4	-0.5	$\frac{\text{Fruit}}{NP/NP}$	✓

. . .

Agenda position	$f(y)$	y	Is correct?
1	1.9	$\frac{\text{Fruit}}{NP}$	✗
2	-0.5	$\frac{\text{Fruit}}{NP/NP}$	✓
3	...	...	...
4	...	...	...

. . .]

$$L(\mathcal{A}) = \sum_{t=1}^{T} \underbrace{\max_{y \in \mathcal{A}_t} f(y)}_{\text{Top of agenda}} - \underbrace{\max_{y \in \text{GOLD}(\mathcal{A}_t)} f(y)}_{\text{Best gold partial parse}}$$

Jointly Optimizing Accuracy and Efficiency

Correct partial parse can still be predicted via backtracking

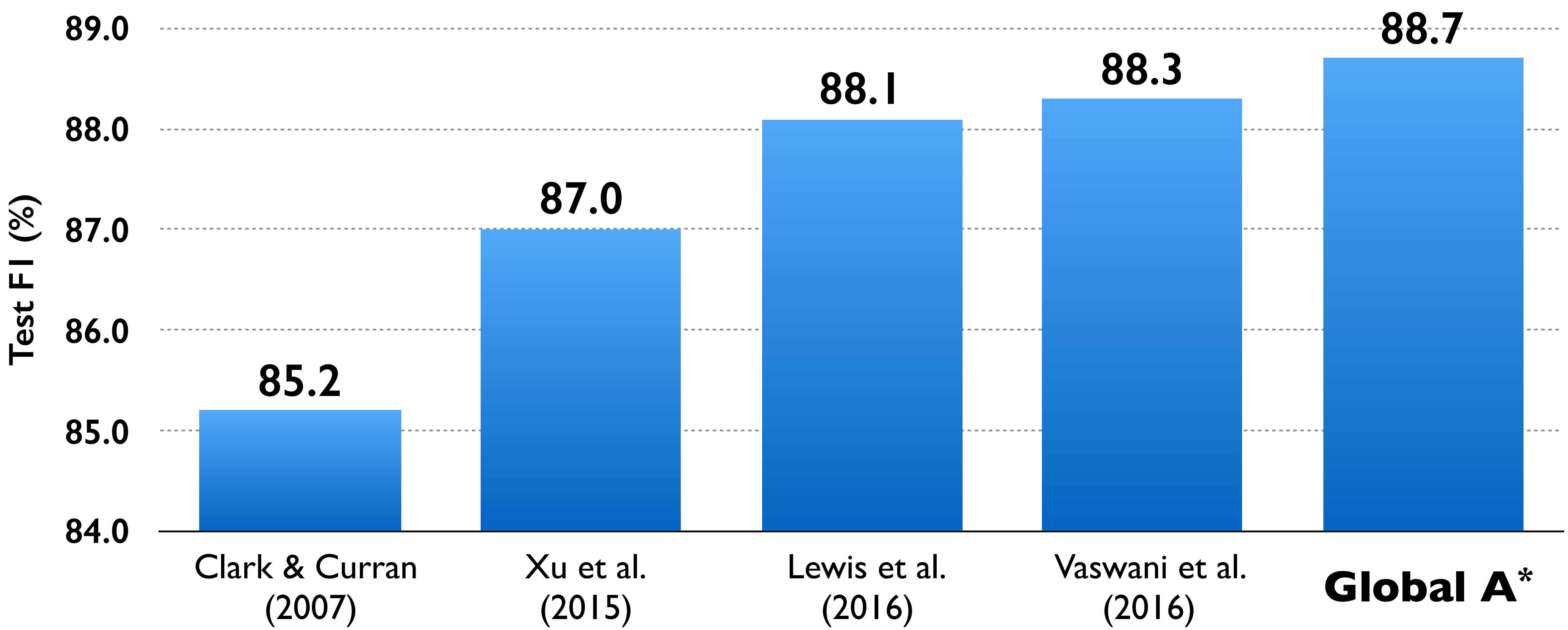
Agenda position	$f(y)$	y	Is correct?
1	1.9	$\frac{\text{Fruit}}{NP}$	✗
2	-0.5	$\frac{\text{Fruit}}{NP/NP}$	✓
3	…	…	…
4	…	…	…

Jointly Optimizing Accuracy and Efficiency

Agenda position	$f(y)$	y	Is correct?
3	$\dots$	$\dots$	$\dots$
4	$\dots$	$\dots$	$\dots$

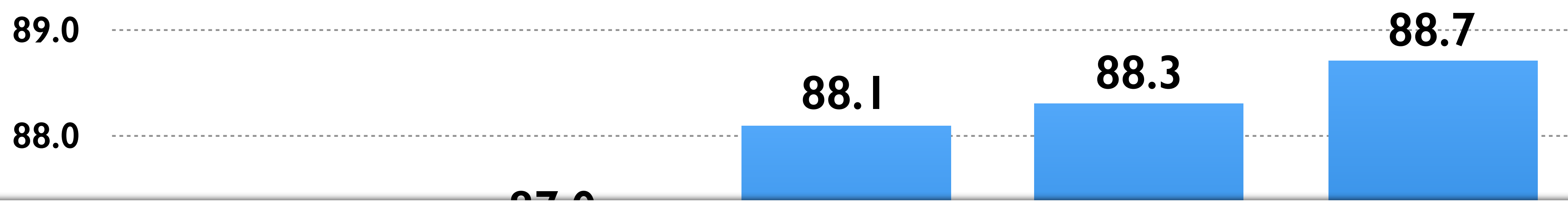
Explicitly optimize for search efficiency!

CCG Parsing Results



	Clark & Curran (2007)	Xu et al. (2015)	Lewis et al. (2016)	Vaswani et al. (2016)	Global A*
Is global?		✓		✓	✓
Is exact?			✓		✓

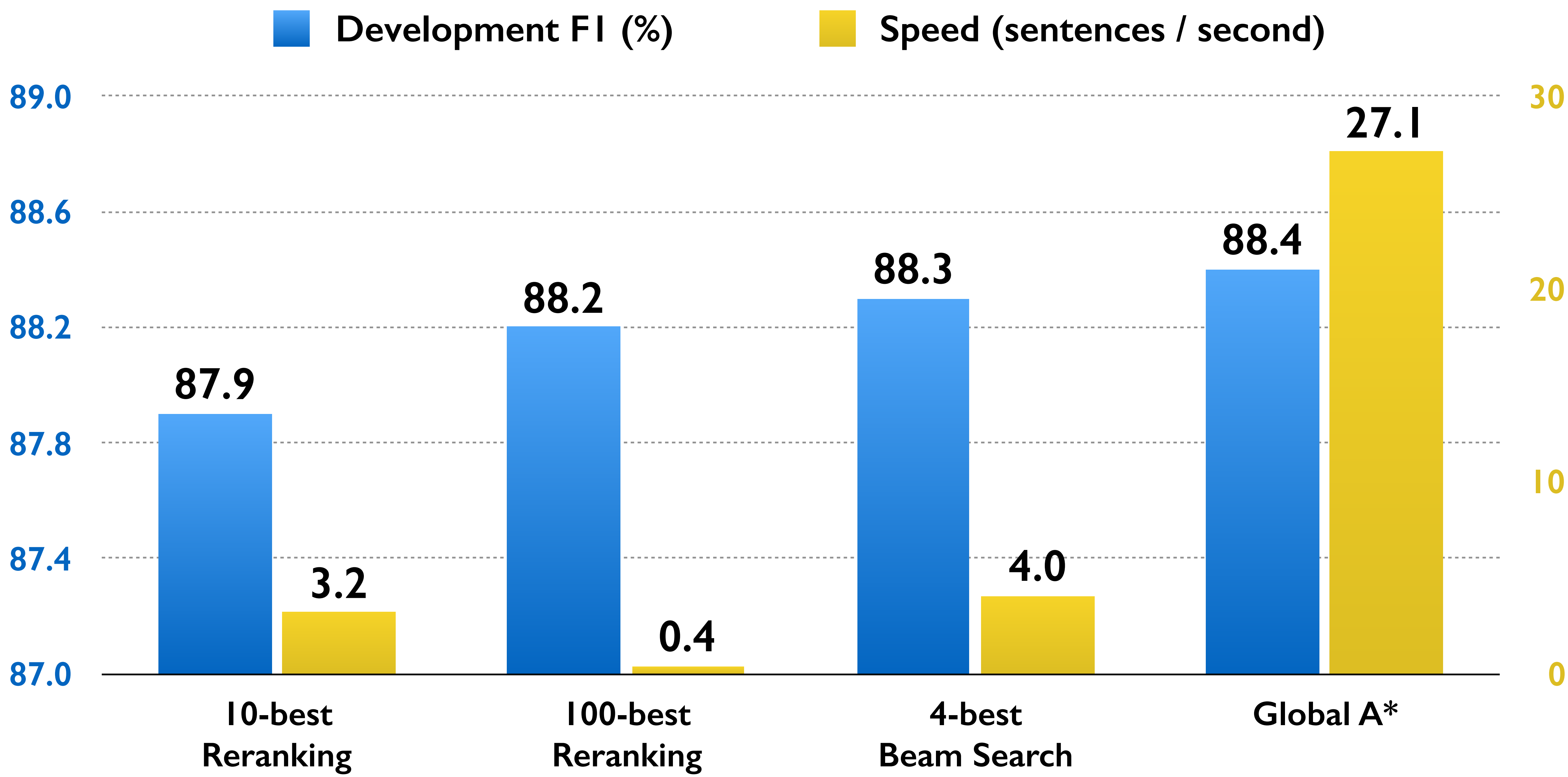
CCG Parsing Results



- Optimal parse found for 99.9% of sentences
- Explores only 190 partial parses on average

Is global?		√		√	√
Is exact?			√		√

Decoder Comparisons



Garden Paths

Incorrect partial parse (syntactically plausible in isolation):

$$
\begin{array}{ccccc}
\text{U.S.} & \text{small} & \text{business} & \text{is} & \text{one} \\
\hline
N/N & (N/N)\backslash(N/N) & N & (S\backslash NP)/NP & N \\
\cline{1-2} \cline{5-5}
\multicolumn{2}{r}{<} & & & \\
\cline{1-3} \cline{4-5}
\multicolumn{3}{r}{>} & \multicolumn{2}{r}{>} \\
\cline{1-3}
\multicolumn{5}{c}{} \\
\hline
\multicolumn{5}{r}{<} \\
\multicolumn{5}{c}{S}
\end{array}
$$

Heavily penalized by the global model

Input sentence:

The favorite **U.S. small business is one** whose research and development can be milked for future Japanese use.

Towards Broad Coverage Semantic Parsing

- Can we crowdsource semantics?
- Train with latent syntax?
- Build fast and accurate parsers?
- Actively select which data to label?

Our key hypothesis:
Anyone who **understands the meaning of a sentence** should be able to correct **parser mistakes**.

✓ ✗

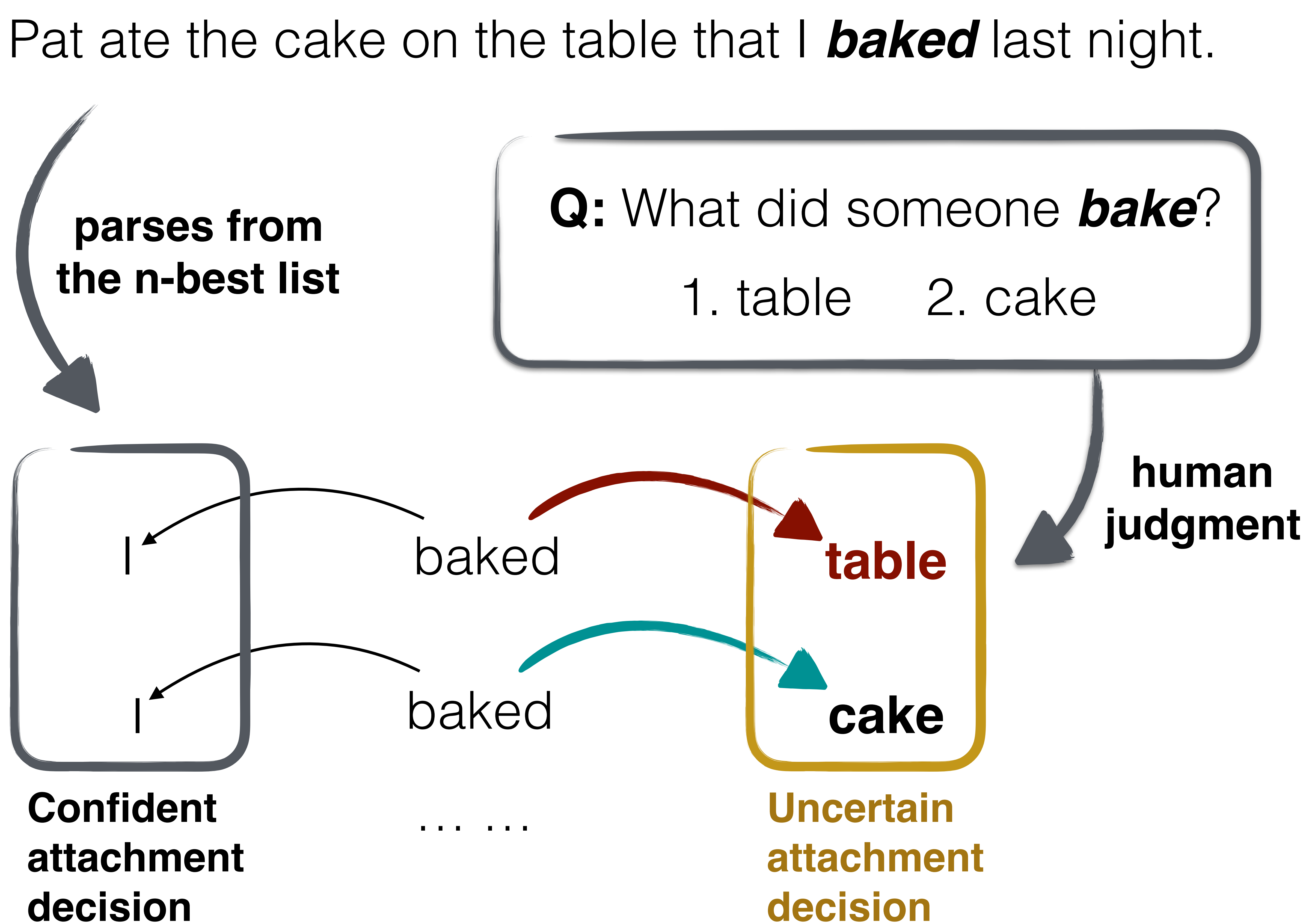
Pat ate the cake on the table that I ***baked*** last night.

Parser: I baked **table**

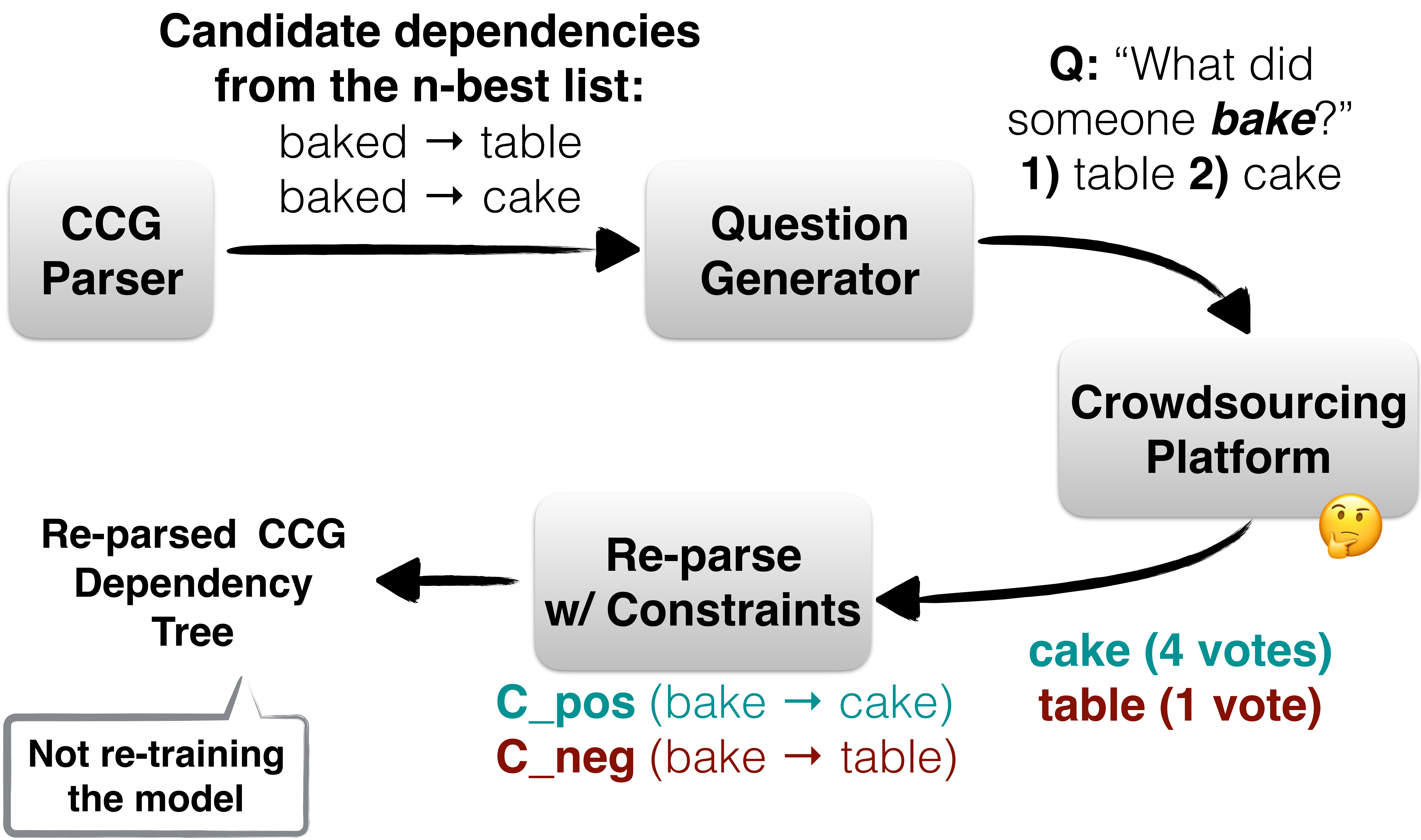
Human understanding: I baked **cake**

Can we use human judgements to improve parse?

[He et al, 2016]


Pat ate the cake on the table that I ***baked*** last night.
parses from the n-best list
Q: What did someone ***bake***?
1. table 2. cake
human judgment
I
baked
table
I
baked
cake
Confident attachment decision
… …
Uncertain attachment decision

Workflow



Candidate dependencies from the n-best list:
baked → table
baked → cake
CCG Parser
Question Generator
Q: "What did someone *bake*?"
1) table 2) cake
Crowdsourcing Platform
cake (4 votes)
table (1 vote)
Re-parse w/ Constraints
C_pos (bake → cake)
C_neg (bake → table)
Re-parsed CCG Dependency Tree
Not re-training the model

Generate Q/A Pairs from CCG Dependencies

Predicted CCG category of ***baked***: $(S\backslash \mathbf{NP_1})/\mathbf{NP_2}$

Convert to template: $\boxed{\mathbf{NP_1}}$ bake $\boxed{\mathbf{NP_2}}$

I **baked** **table**

I **baked** **cake**

Filling-in the Slots:

$\boxed{\textbf{what}}$ bake $\boxed{\text{sth.}}$	**What** baked something? — **I**	**What** baked something? — **I**
$\boxed{\text{sth.}}$ bake $\boxed{\textbf{what}}$	**What** did someone bake? — **the table**	**What** did someone bake? — **the cake**

Infer **someone/something** and the **answer spans** based on the n-best parses

Used "**what**" for all questions

Group Q/A Pairs into Queries

Questions	Answers	Scores	Question Confidence	Answer Uncertainty (Entropy)
What baked something?	**I**	1.0	**1.0**	**0.0**
What did someone bake?	**the table**	0.7	**1.0**	**0.88**
	the cake	0.3		
What was baked something something?	**the table**	0.1	**0.1**	**0.0**

Non-sensical question

No uncertainty

Our Annotation Task

Sentence:

Pat ate the cake on the table that I **baked** last night.

Question:

What did someone bake?

Check one or more

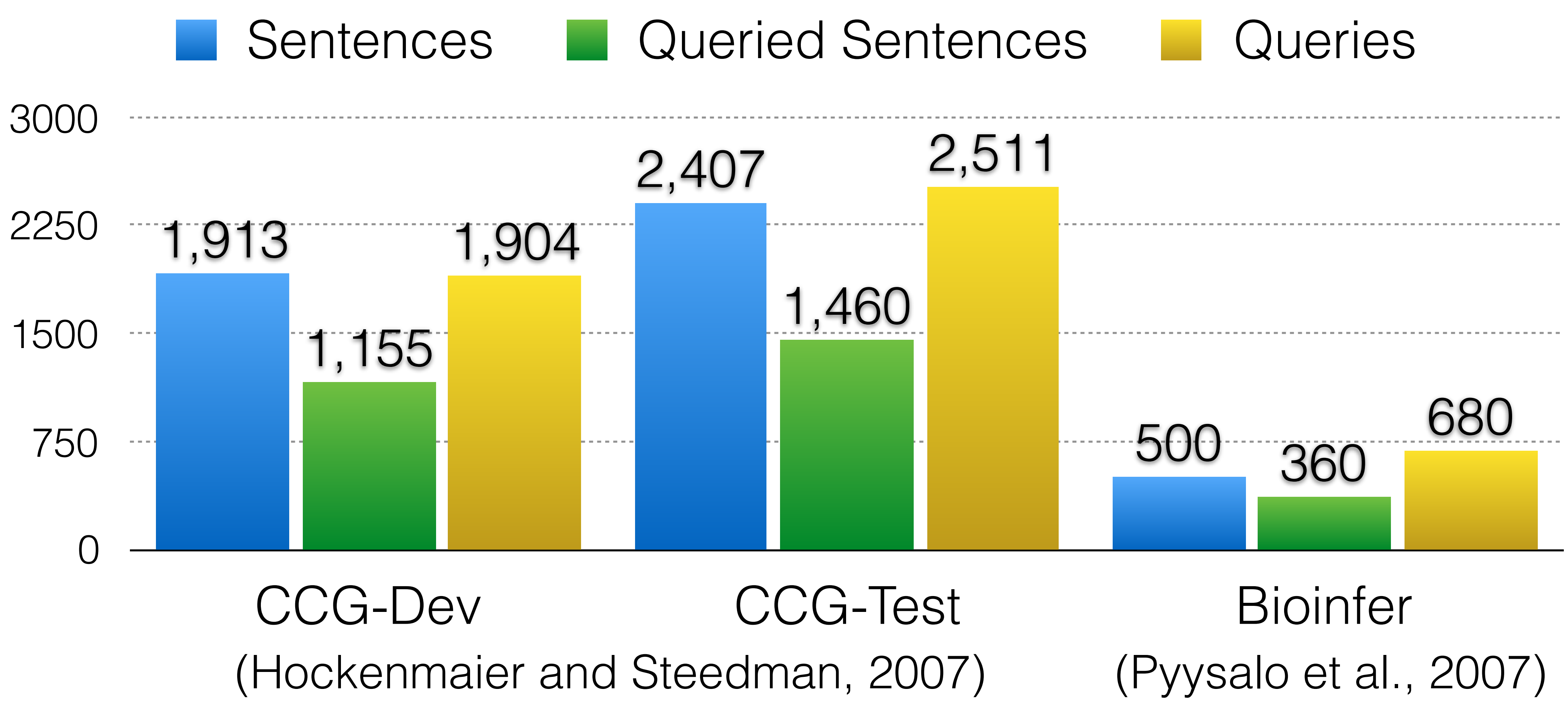
- [ ] the cake
- [ ] the table
- [ ] None of the above.

Comment:

- Annotators are instructed to choose options that "***explicitly and directly***" answer the question.
- Multiple answers are allowed.
- 5 judgements per query.

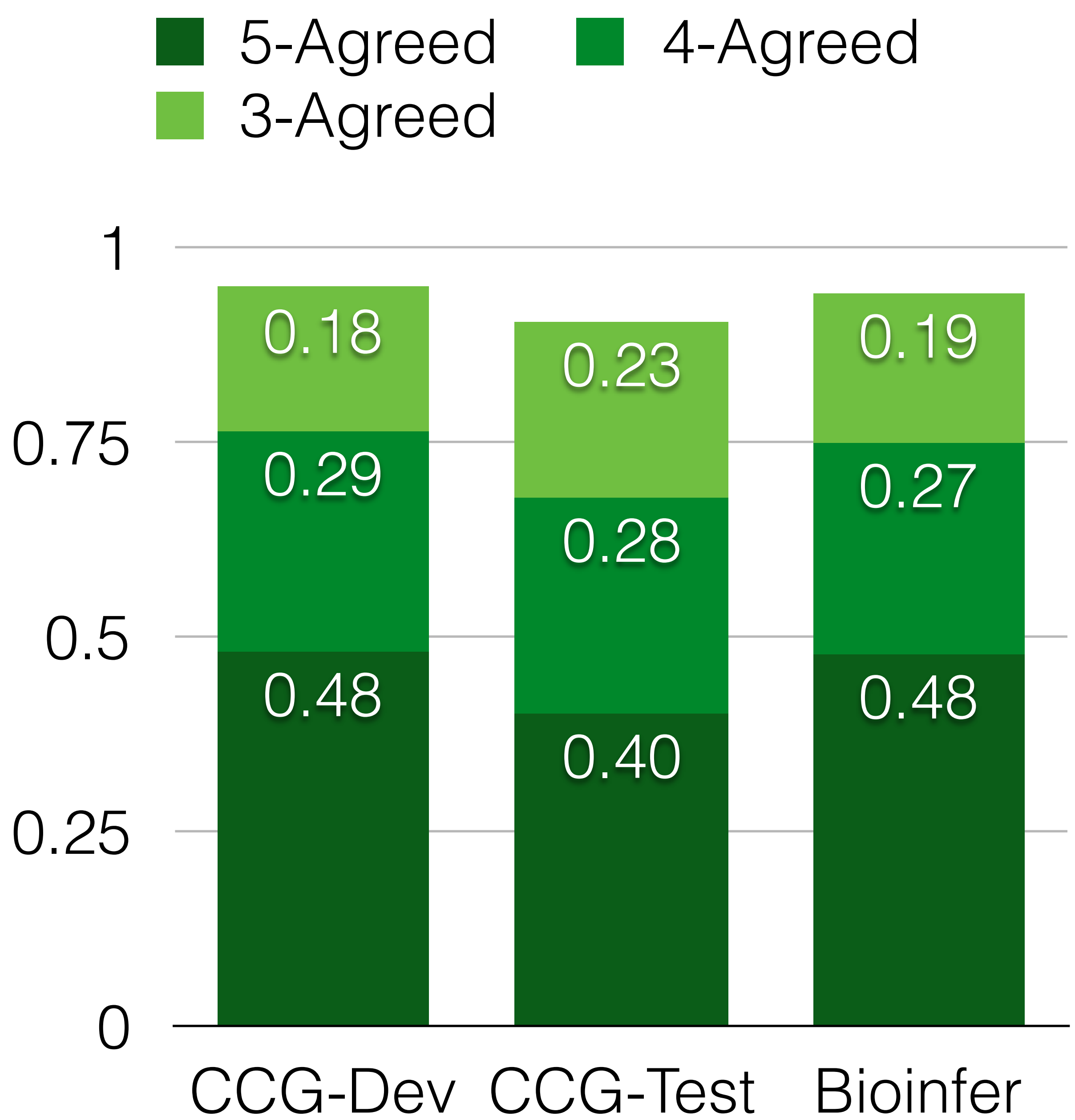
* Crowdsourcing platform: https://www.crowdflower.com/.

Data Collection with Crowdsourcing



- All developments are done on CCG-Dev only.
- Less than 2 queries per sentence, for about 60% of the sentences.
- **Cost:** 46 cents per query.
- **Speed:** 200 queries per hour.

Inter-Annotator Agreement



- Agreement is computed only for matching the exact set of answers. i.e. (A, B) and (B) are considered disagreement.
- Unanimous agreement for over 40% of the queries.
- Over 90% absolute majority.

Putting our hypothesis to the test:
How well does annotators' **human understanding** align with the **gold syntax?**

- Successes: Long-range attachment decisions
- Challenges: **Syntax-semantics mismatch**
- Use heuristics to fix the mismatch problems at re-parsing time.

Success - Long-range Dependency

Temple also said Sea Containers' plan raises numerous legal, regulatory, financial and fairness issues, but didn't ***elaborate***.

What *didn't* ***elaborate*** something?

4 Temple

1 Sea Containers' plan

0 None of the above.

Success - Coordination

To ***avoid*** these costs, and a possible default, immediate action is imperative.

What would something ***avoid***?

4	these costs
3	a possible default
0	None of the above.

Challenge - Coreference

Kalipharma is a New Jersey-based pharmaceuticals concern that ***sells*** products under the Purepac label.

What ***sells*** something?

5 Kalipharma

0 a New Jersey-based pharmaceuticals concern

0 None of the above.

- Syntax-semantics mismatch
- Also happens with pronouns and appositives.
- Some cases are heuristically fixed during reparsing.

Challenge - Headedness

Timex had requested duty-free treatment for many types of watches, **covered** by 58 different U.S. tariff classifications.

What would be ***covered*** ?

0	Timex	**2**	many types of watches
0	duty-free treatment	**3**	watches
0	None of the above.		

- Annotators tend to struggle with headedness.
- We add "disjunctive constraint", forcing the re-parser to produce either of the two dependencies.

Re-Parsing with Crowdsourced Constraints

Q1: What did someone ***bake***?

votes(*cake*) = 4

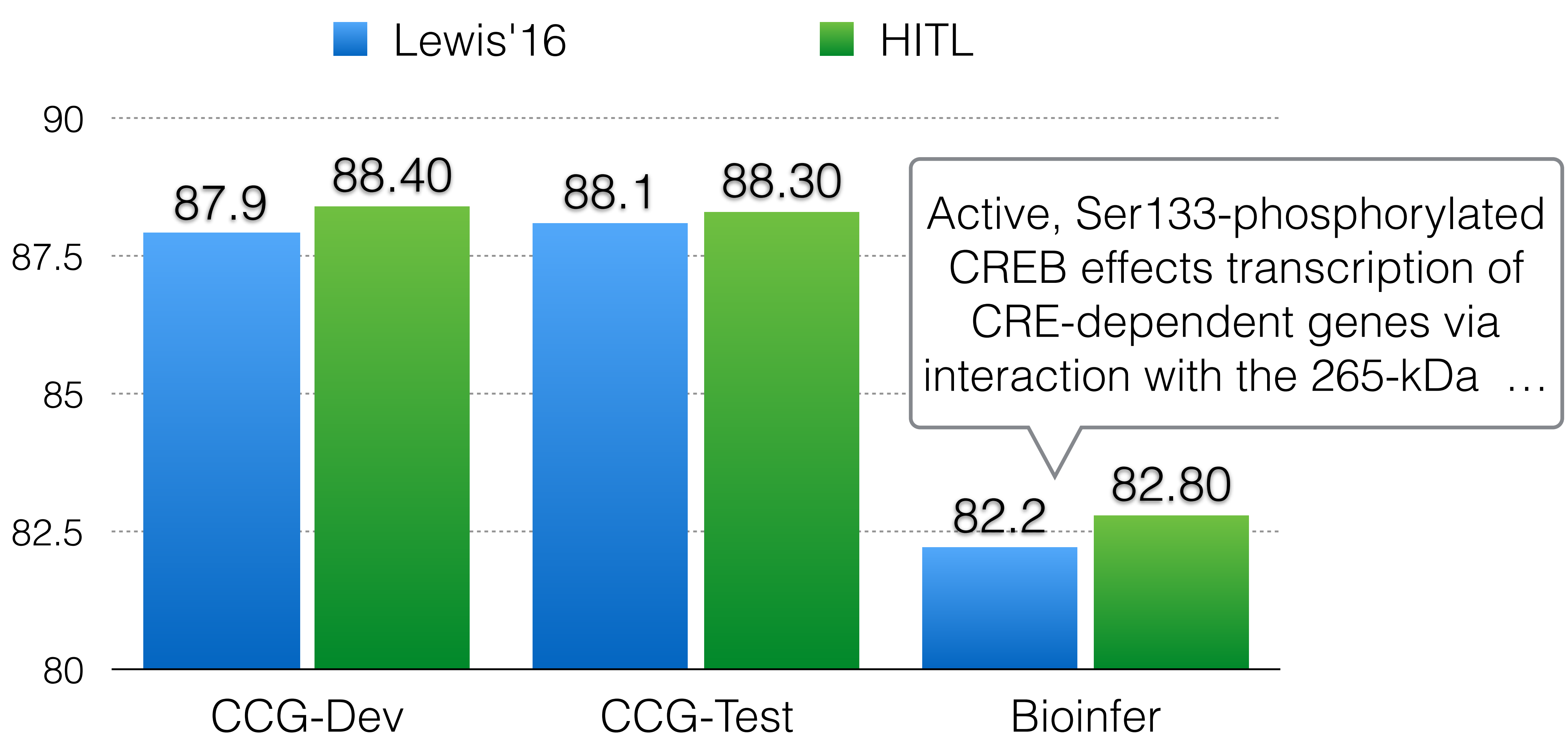
votes(*table*) = 1

votes(*None of the above*) = 0

$$
\begin{aligned}
y^{\text{new}} = \arg\max_y \quad & \text{base\_parser\_score}(y) \\
& - T^{+} \times \mathbb{1}(\text{baked} \rightarrow \text{cake} \in y) \\
& - T^{-} \times \mathbb{1}(\text{baked} \rightarrow \text{table} \in y)
\end{aligned}
$$

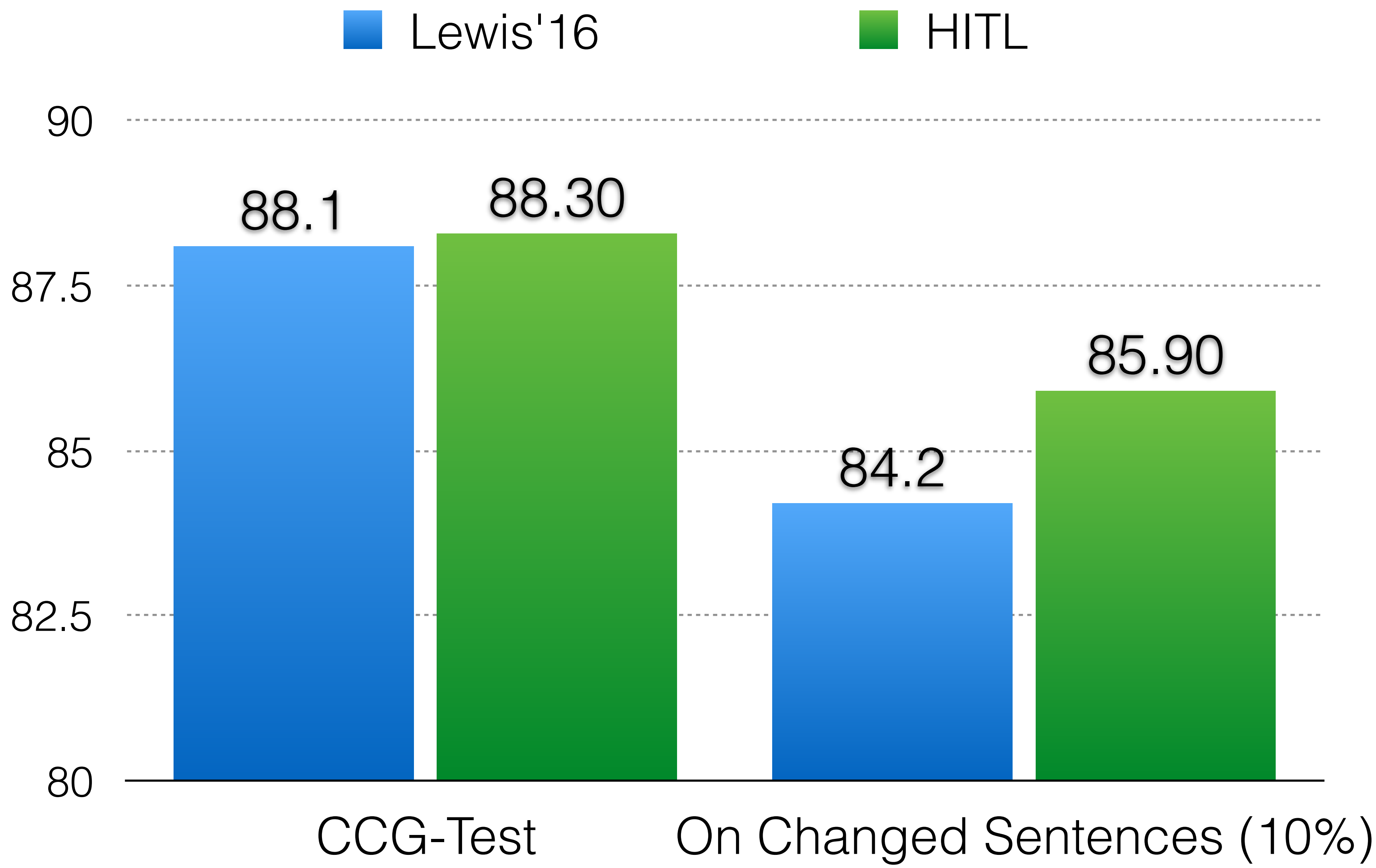
- Penalizes parses that disagree with crowdsourced judgments.
- Constraints are decomposed by dependencies.
- Thresholds and penalties are tuned on CCG-Dev.

Re-parsing Results (Labeled F1)



- Modest improvement due to syntax-semantics mismatch.
- Larger improvement on out-of-domain data.

Re-parsing Results



- Modified parse trees for about 10% of the sentences after incorporating human judgments.
- Larger gain on changed sentences.
- Changed sentences are “more difficult” on average.

Towards Broad Coverage Semantic Parsing

- Can we crowdsource semantics?
  - Yes, but need more than verbs….
- Train with latent syntax?
  - Yes, but must extend to QA supervision…
- Build fast and accurate parsers?
  - Yes, but need to extend to latent-variable case…
- Actively select which data to label?
  - Yes, but need to scale up…

Questions

?