Global Neural CCG Parsing with Optimality Guarantees

Kenton Lee Mike Lewis† Luke Zettlemoyer

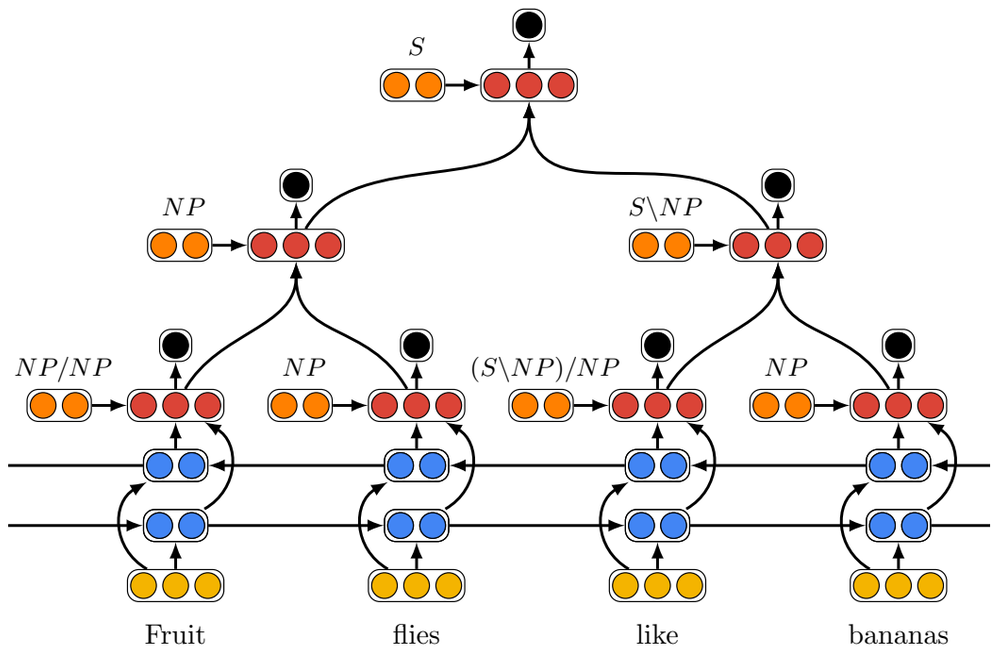
University of Washington

† Now at Facebook AI Research

This Talk

Challenge:

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



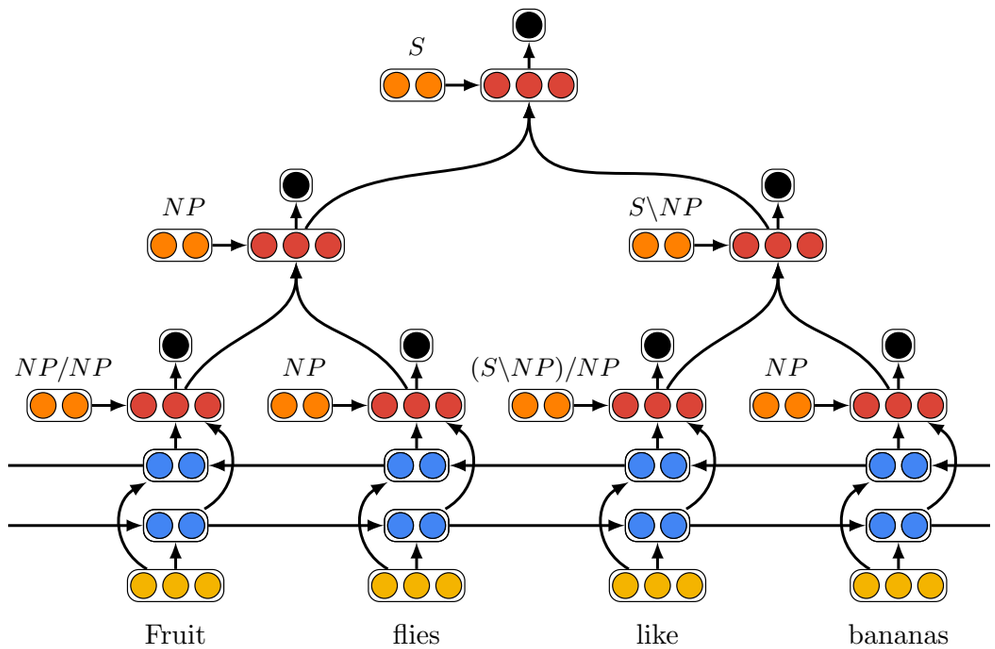
This Talk

Challenge:

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

Our approach:

Combine local and global models in A* parser



Result:

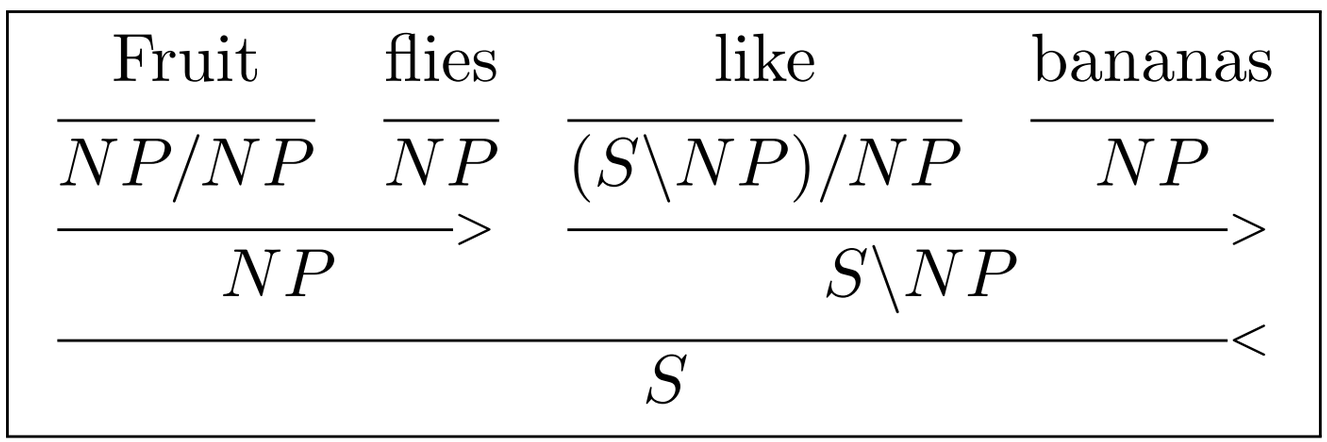
Global model with exact inference

Parsing with Hypergraphs

Input

Fruit flies like bananas

Output



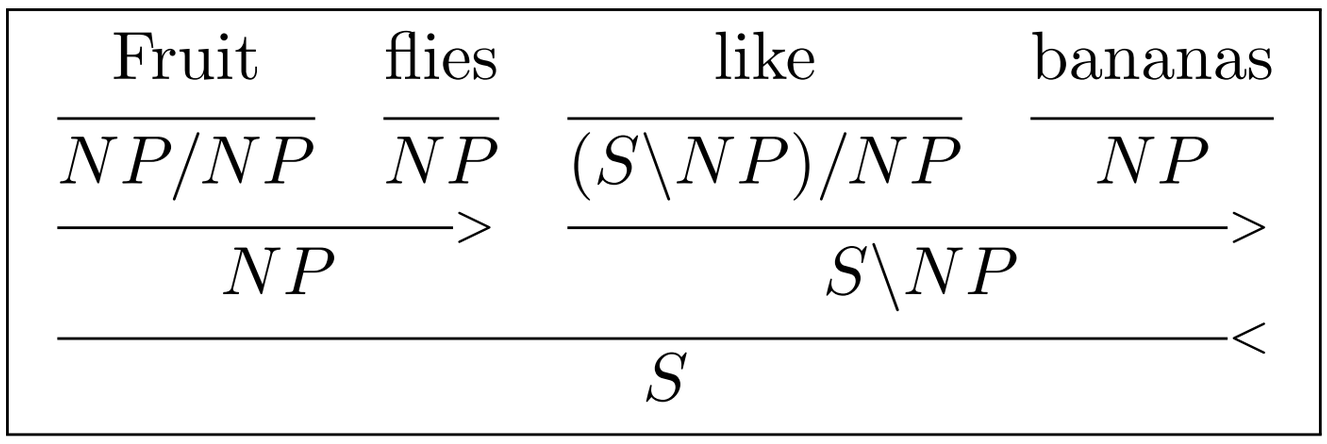
Klein and Manning, 2001

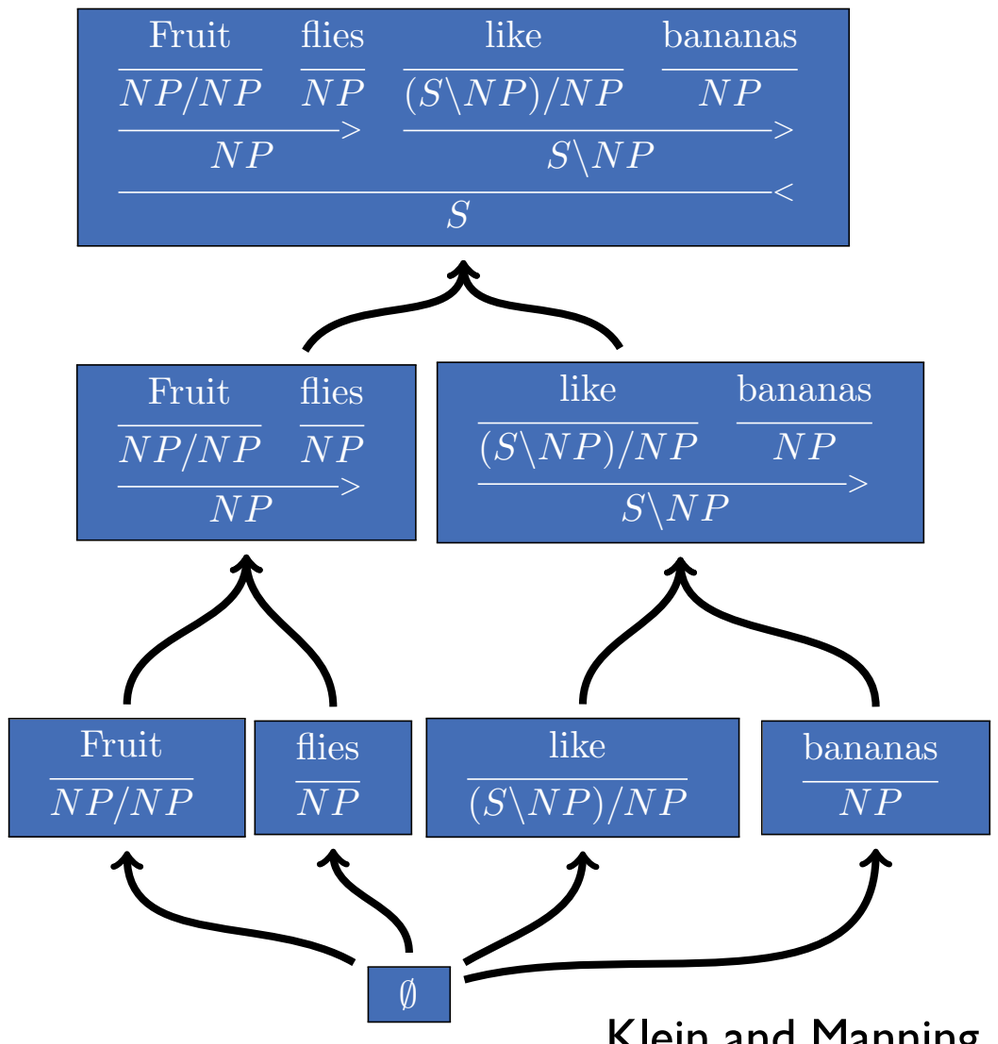
Parsing with Hypergraphs

Input

Fruit flies like bananas

Output





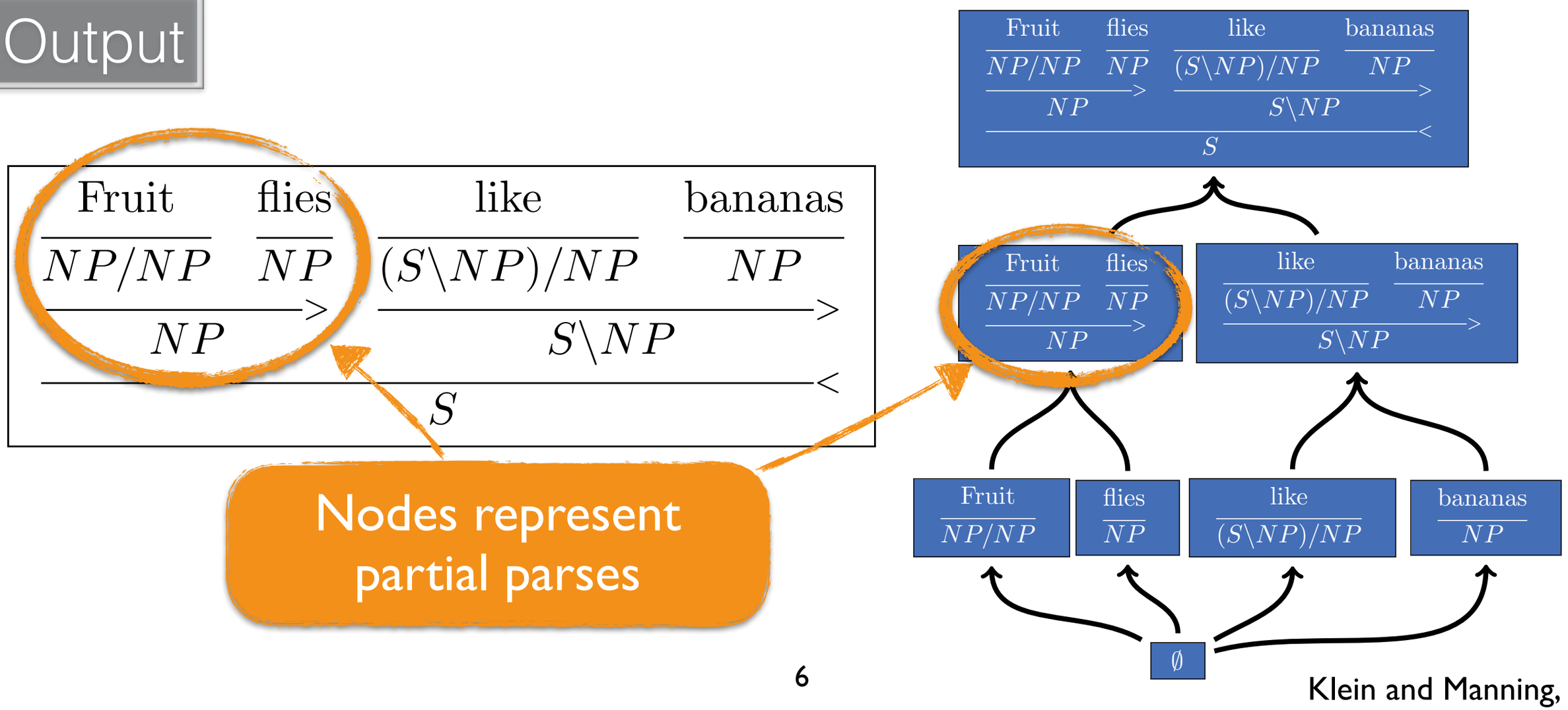
Klein and Manning, 2001

Parsing with Hypergraphs

Input

Fruit flies like bananas

Output



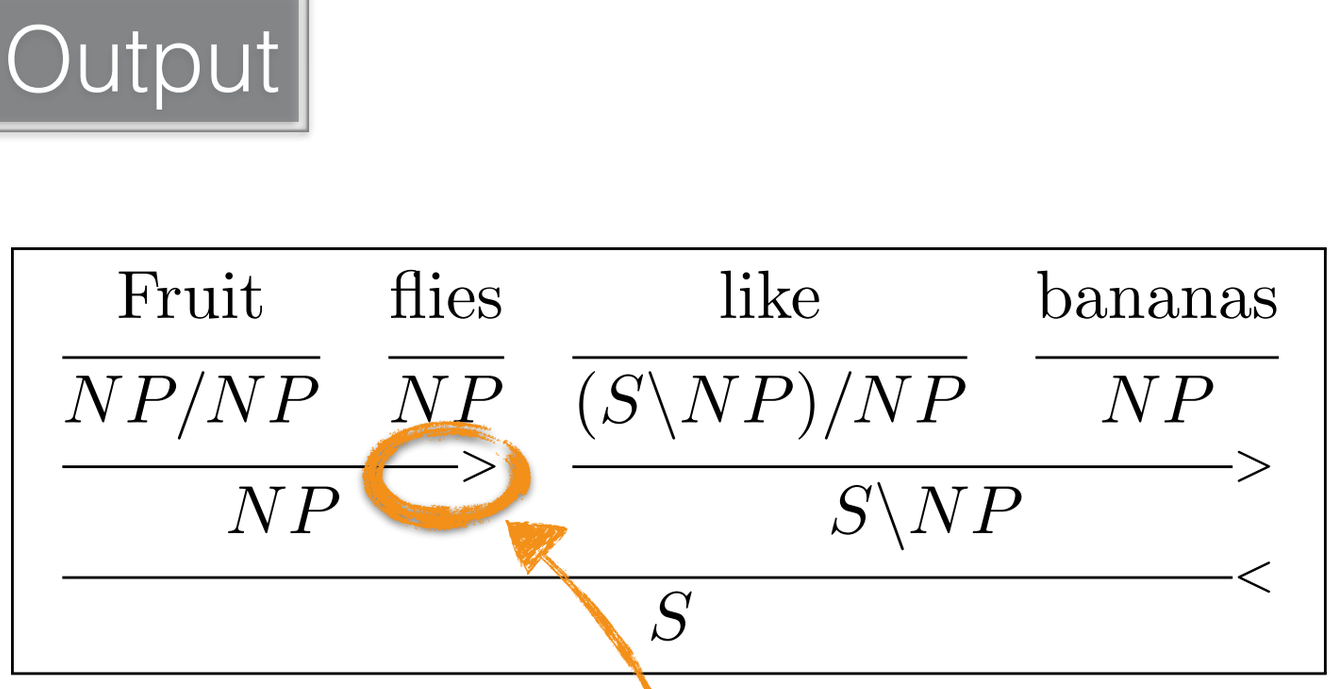
Klein and Manning, 2001

Parsing with Hypergraphs

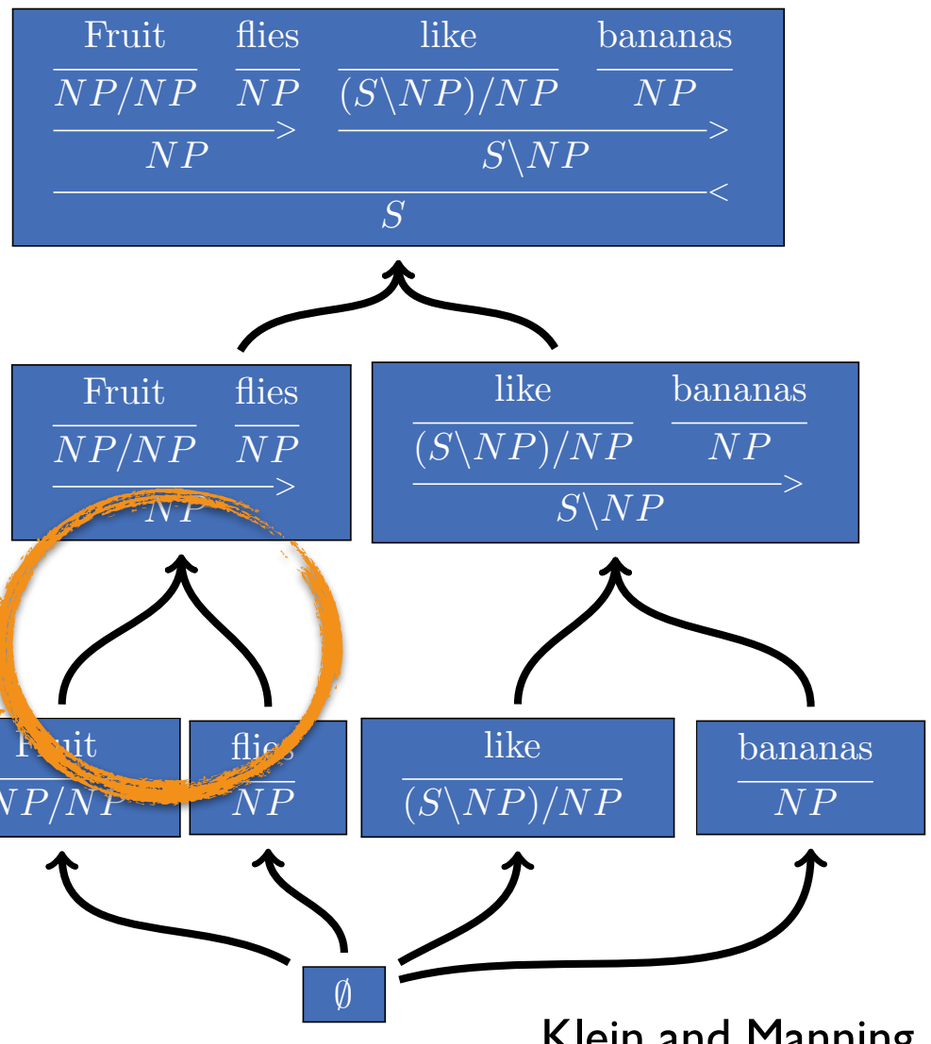
Input

Fruit flies like bananas

Output



Hyperedges represent rule productions



Klein and Manning, 2001

Parsing with Hypergraphs

Input

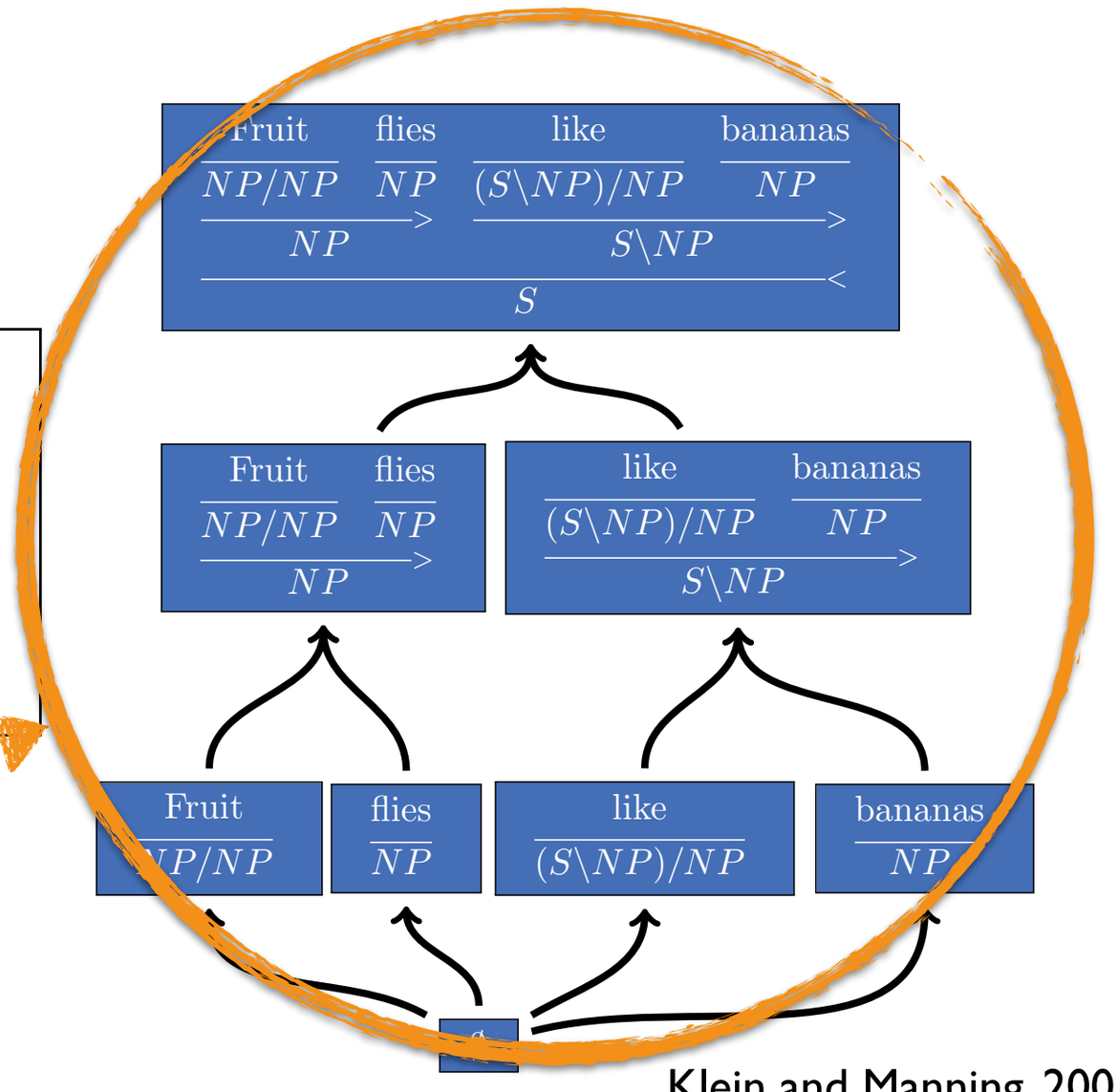
Fruit flies like bananas

Output

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

$$\dfrac{\text{Fruit flies like bananas}}{S}<$$

Path $y = \{e_1, \ldots, e_m\}$
represents a parse derivation



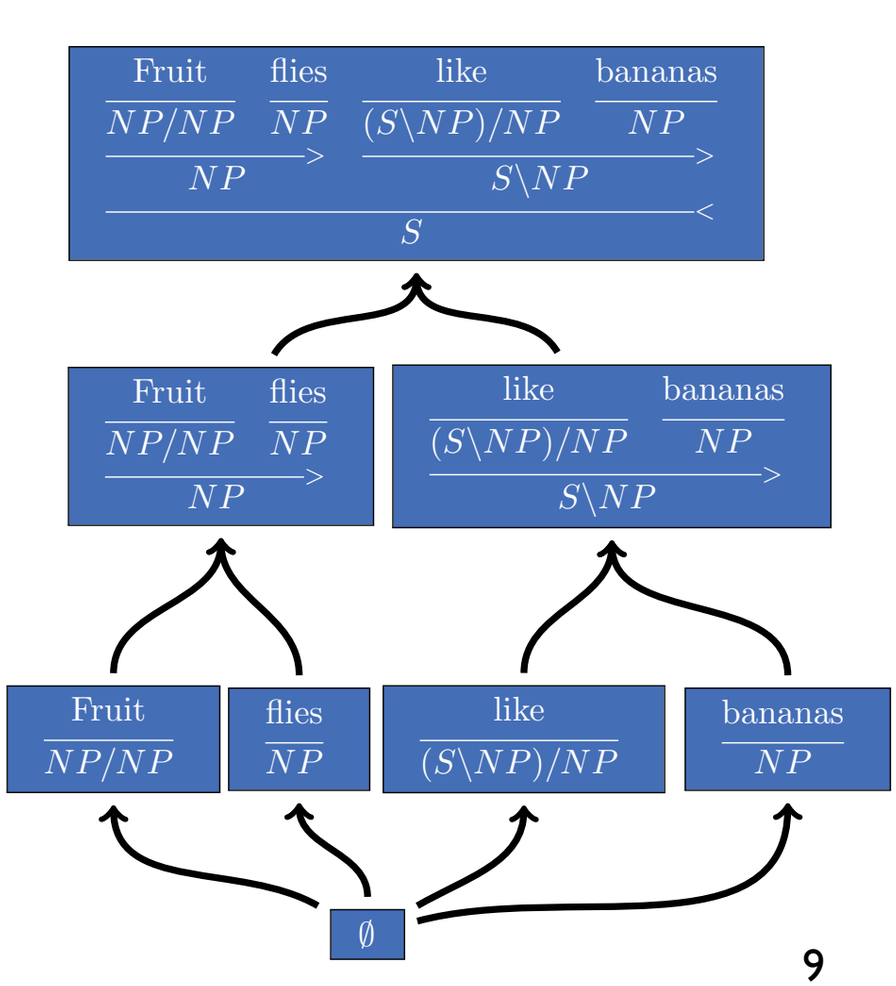
Klein and Manning, 2001

Parsing with Hypergraphs

Input

Fruit flies like bananas

Output

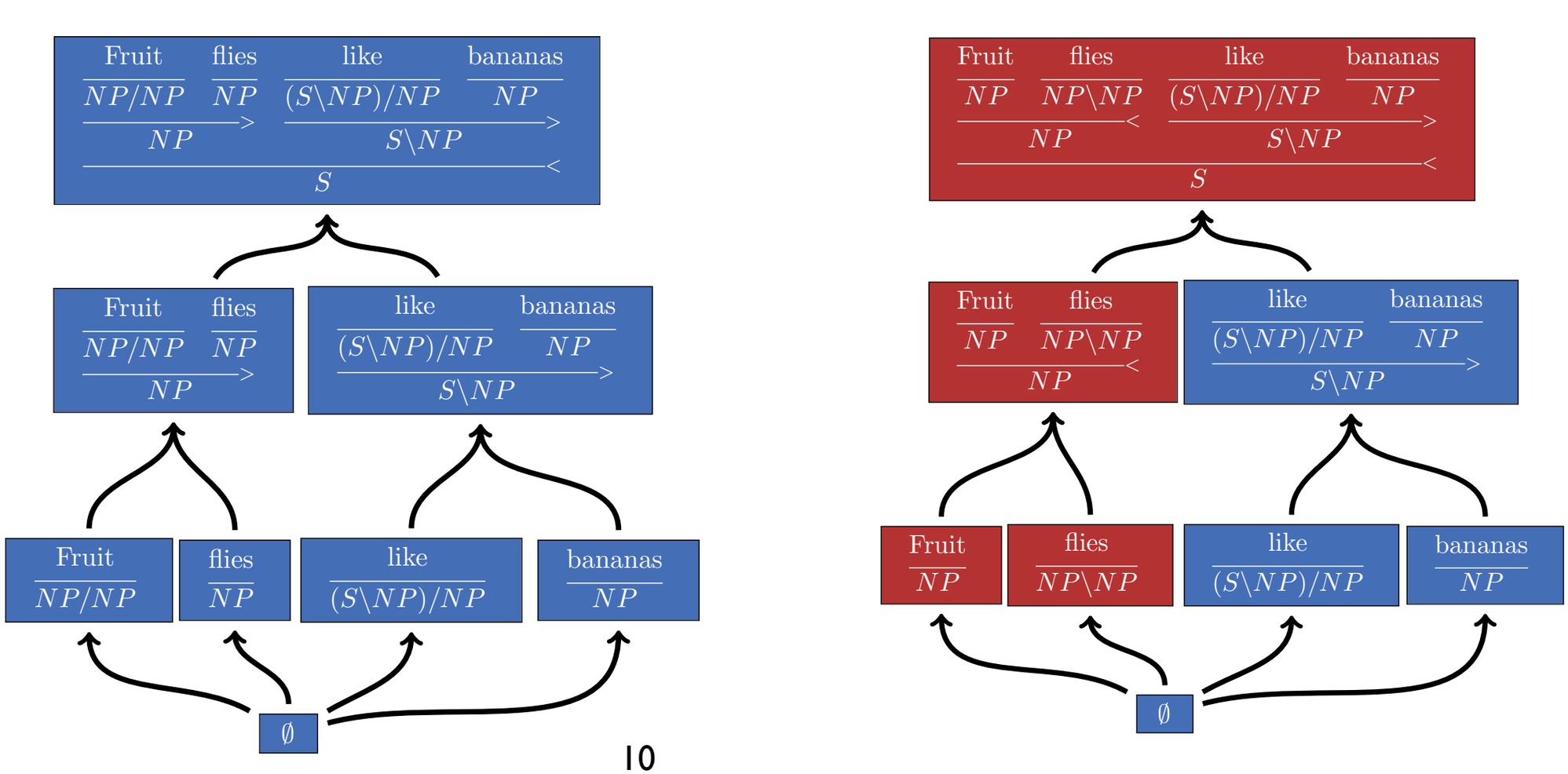


Parsing with Hypergraphs

Input

Fruit flies like bananas

Output

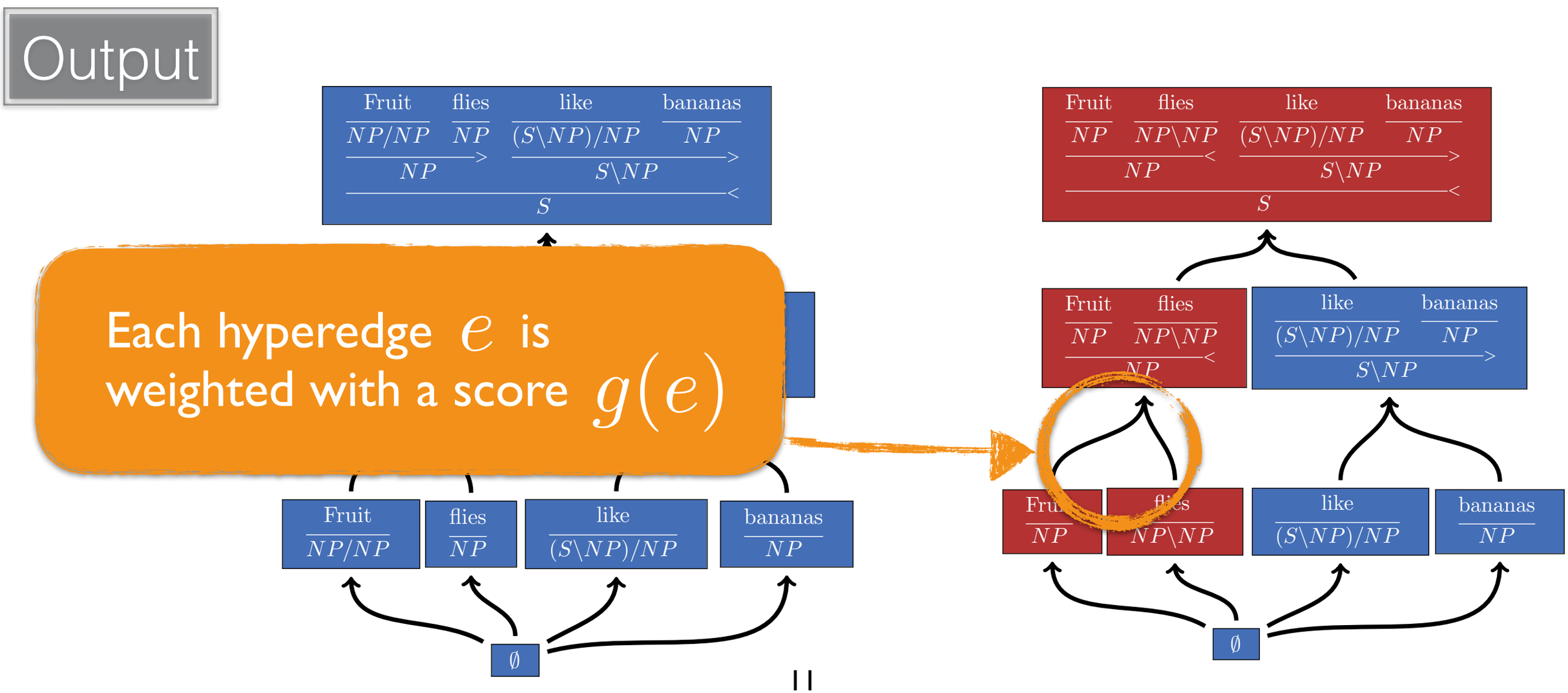


Parsing with Hypergraphs

Input

Fruit flies like bananas

Output

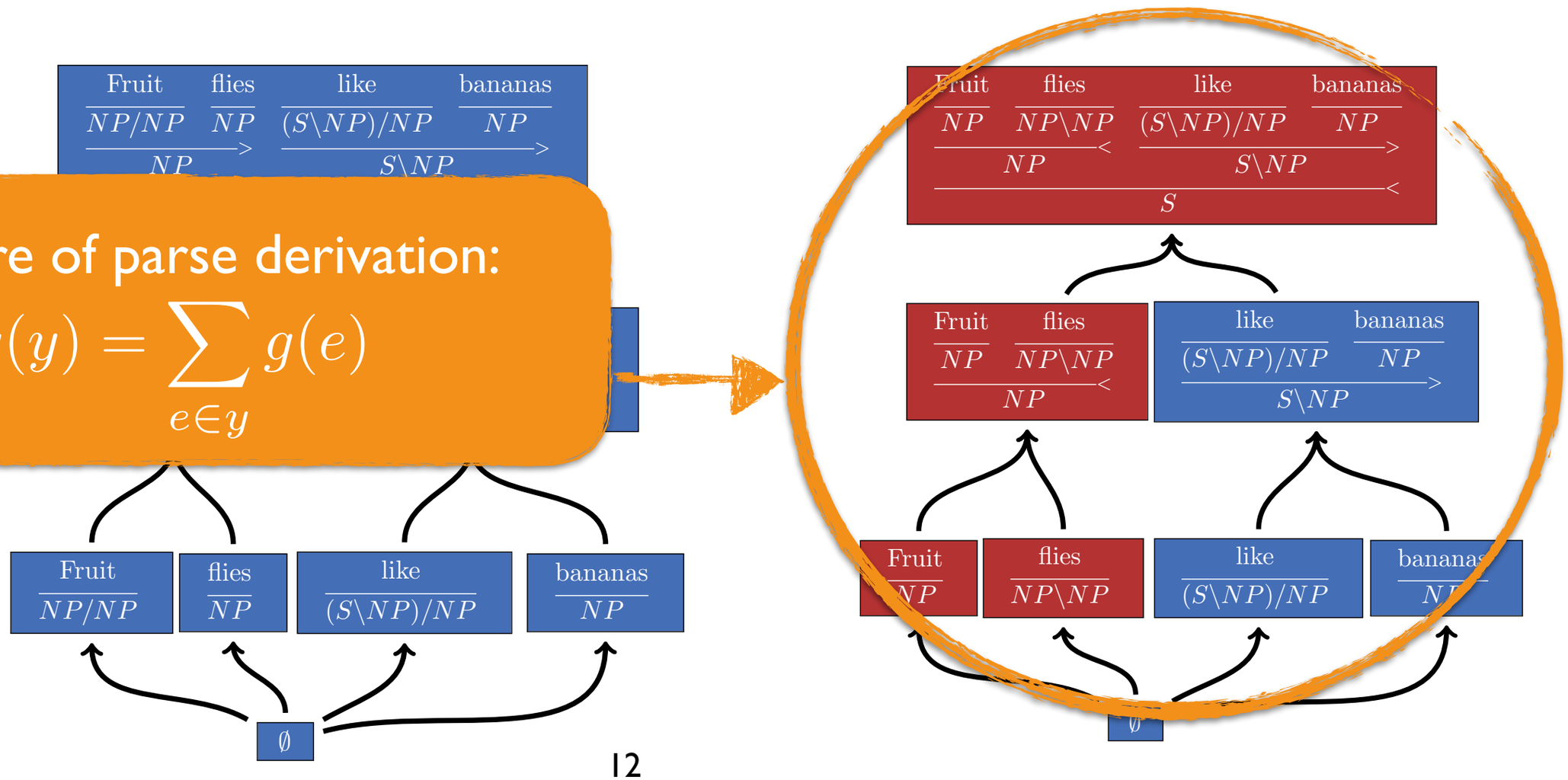


Parsing with Hypergraphs

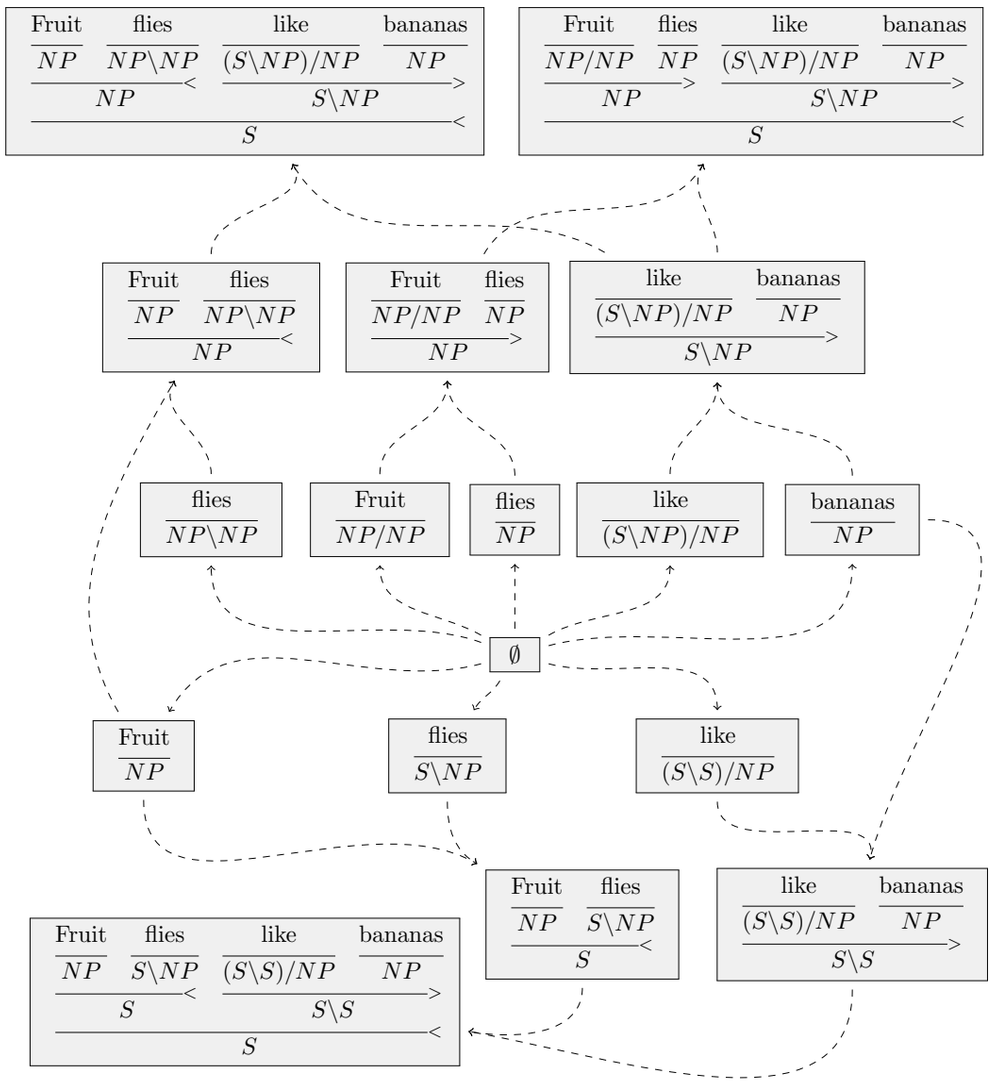
Input

Fruit flies like bananas

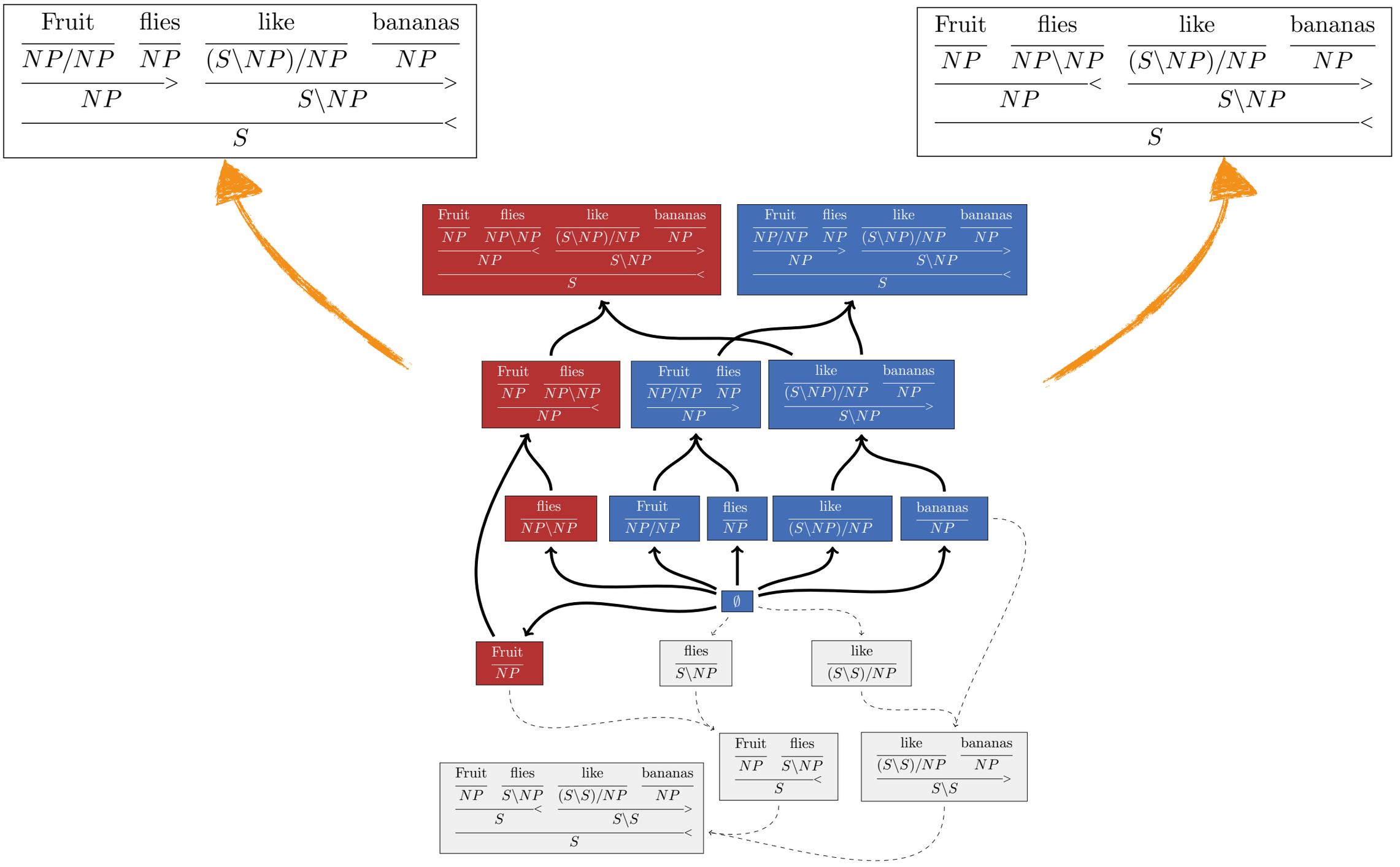
Output



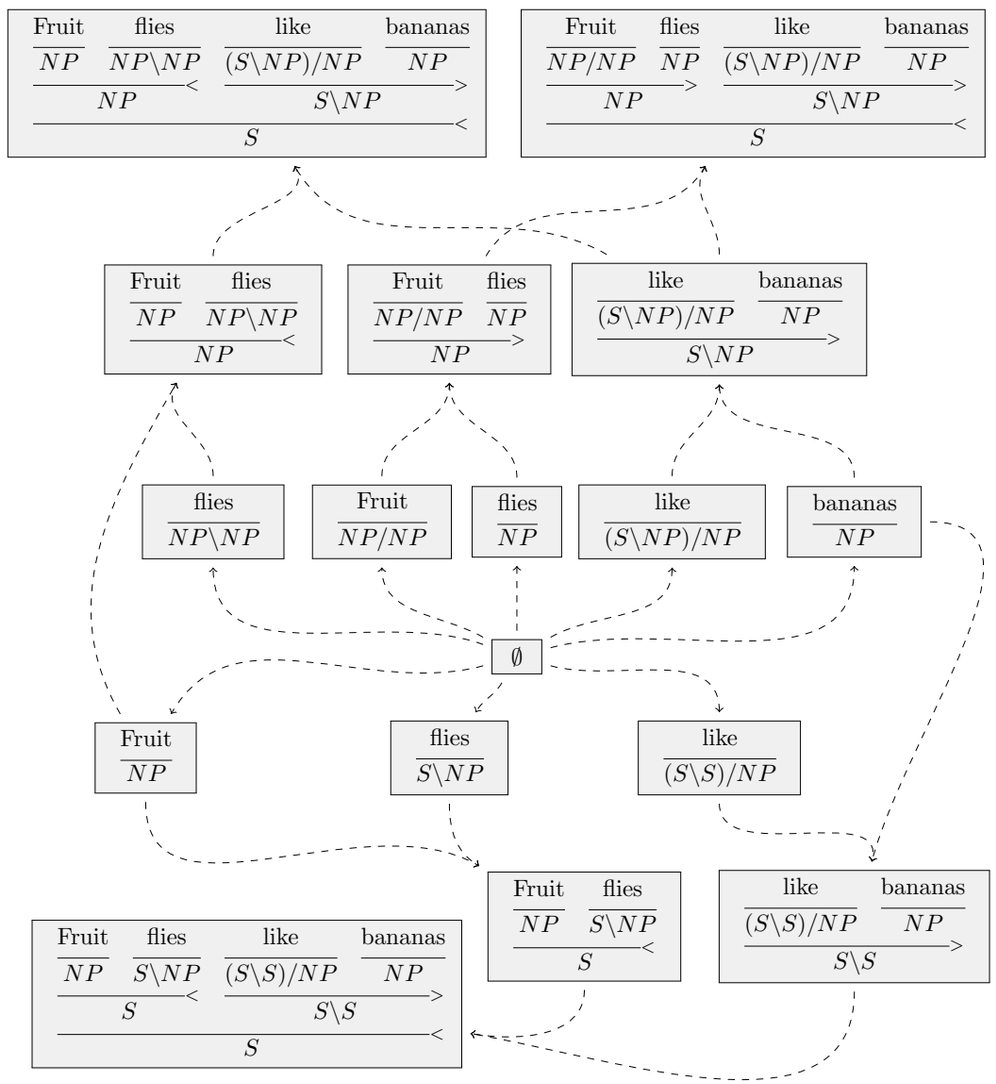
Parsing with Hypergraphs



Parsing with Hypergraphs



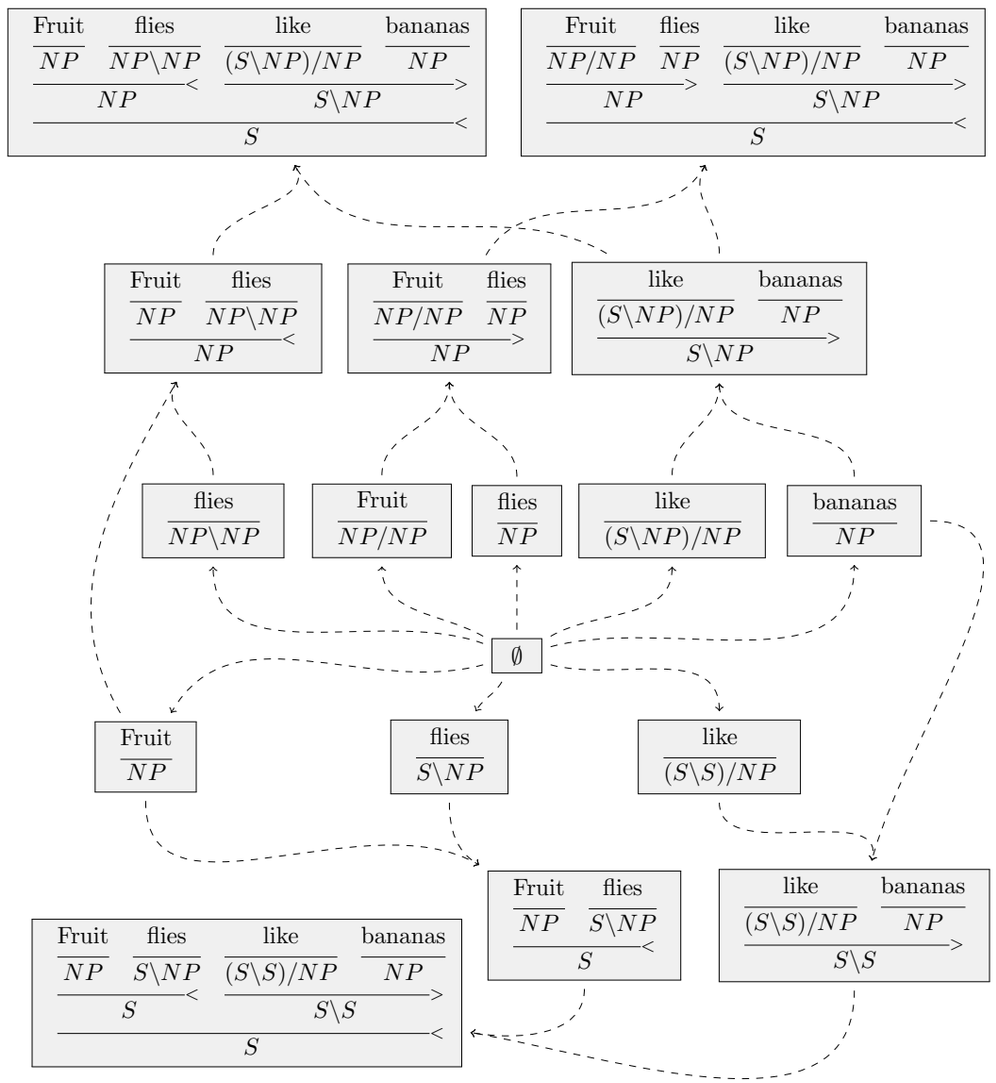
Parsing with Hypergraphs



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

→ Intractable inference

Managing Intractable Search Spaces

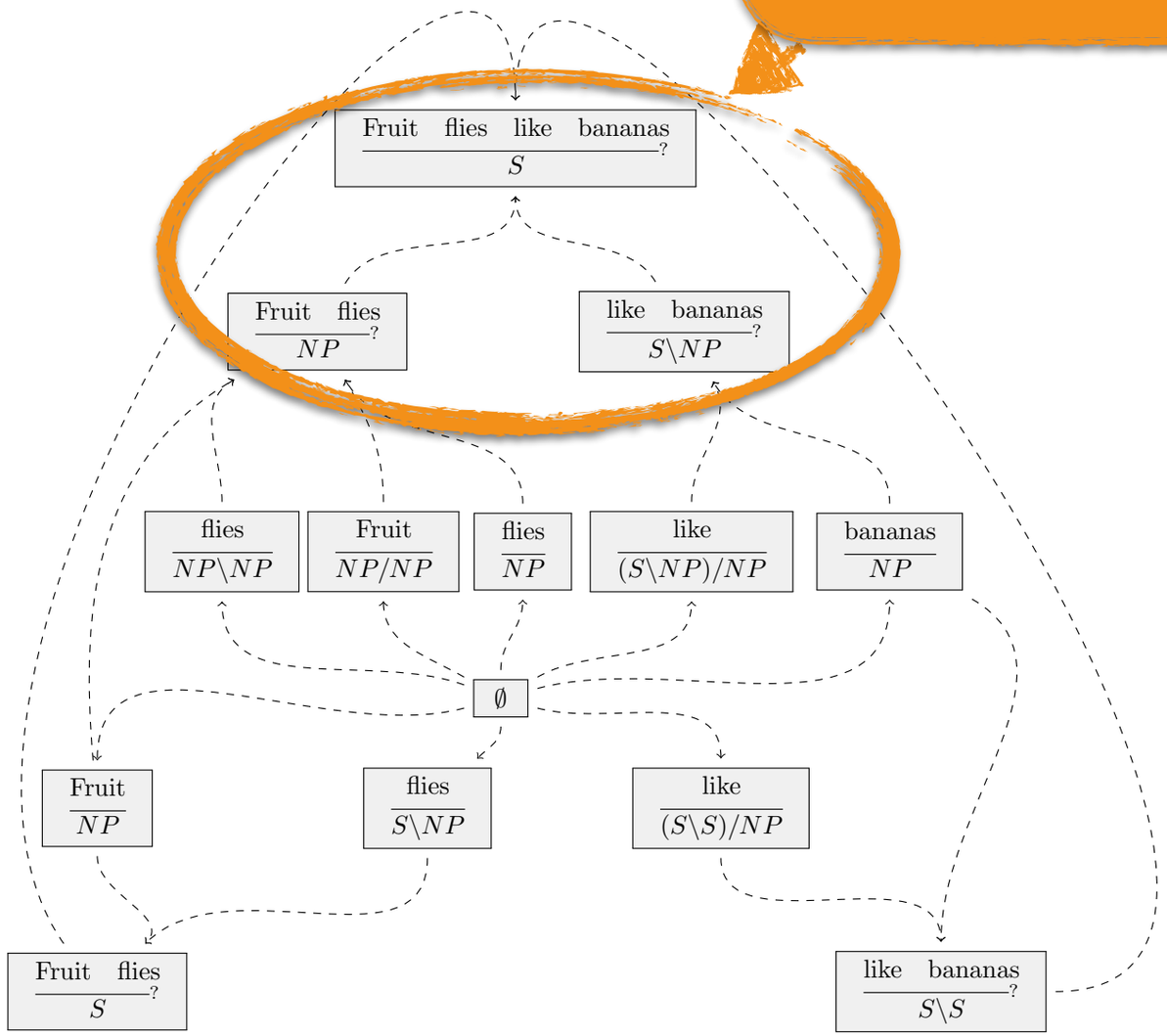


Approximate inference with **global expressivity**, e.g.

- 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

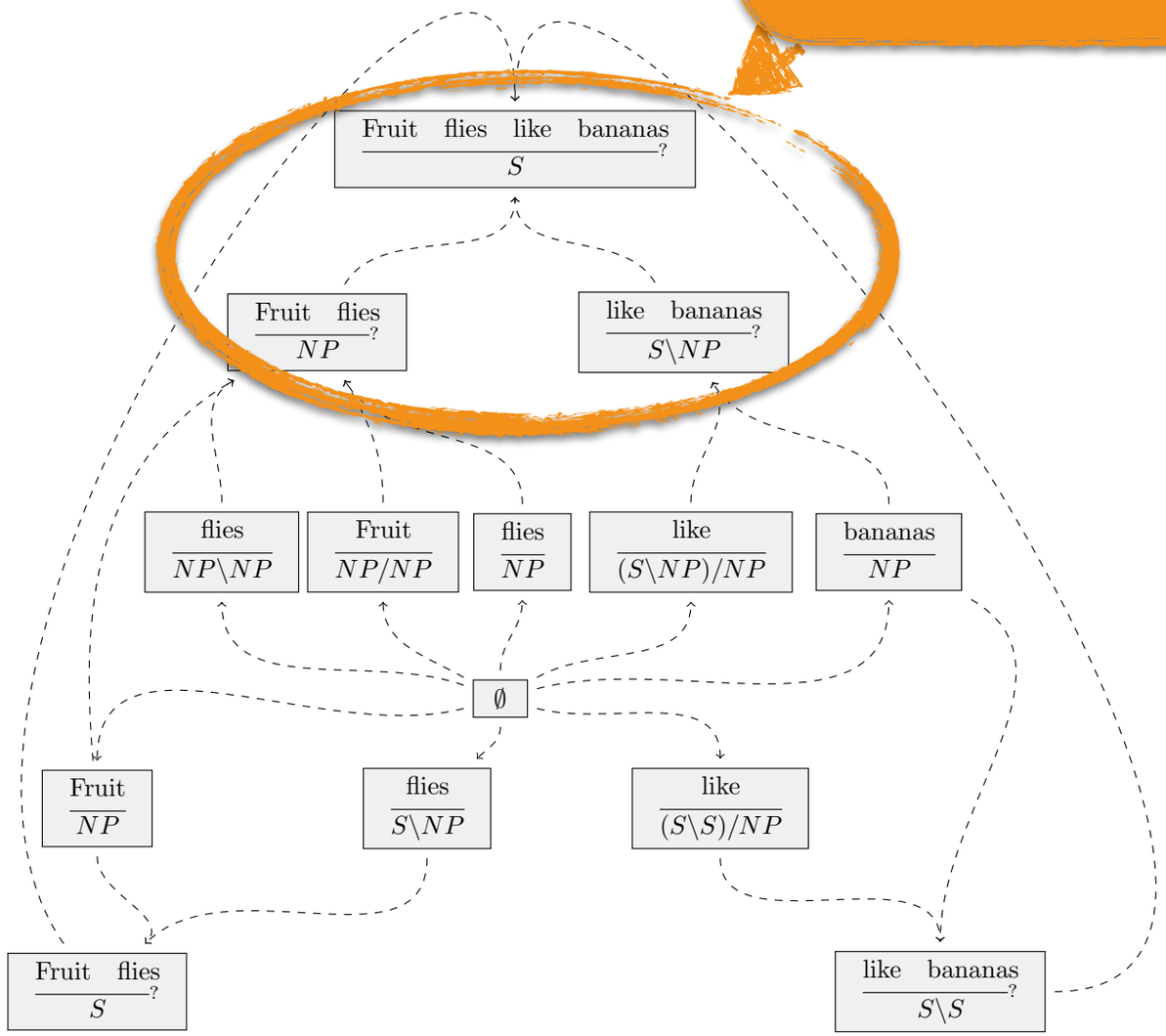
Scores condition on **local structures**



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

Locally Factored Parsing

Scores condition on **local structures**

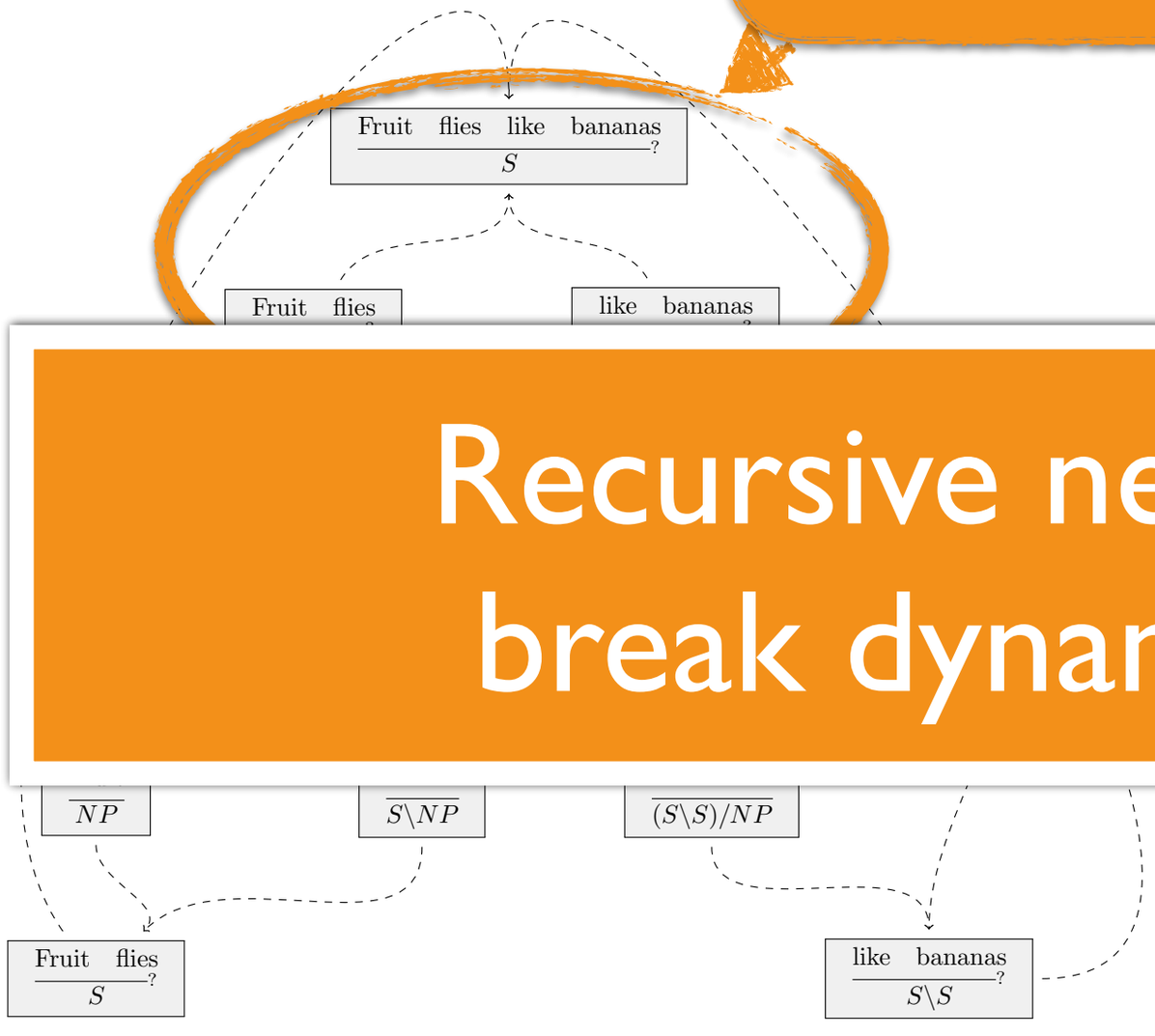


Dynamic programs with **locally factored models**, e.g.

- 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**

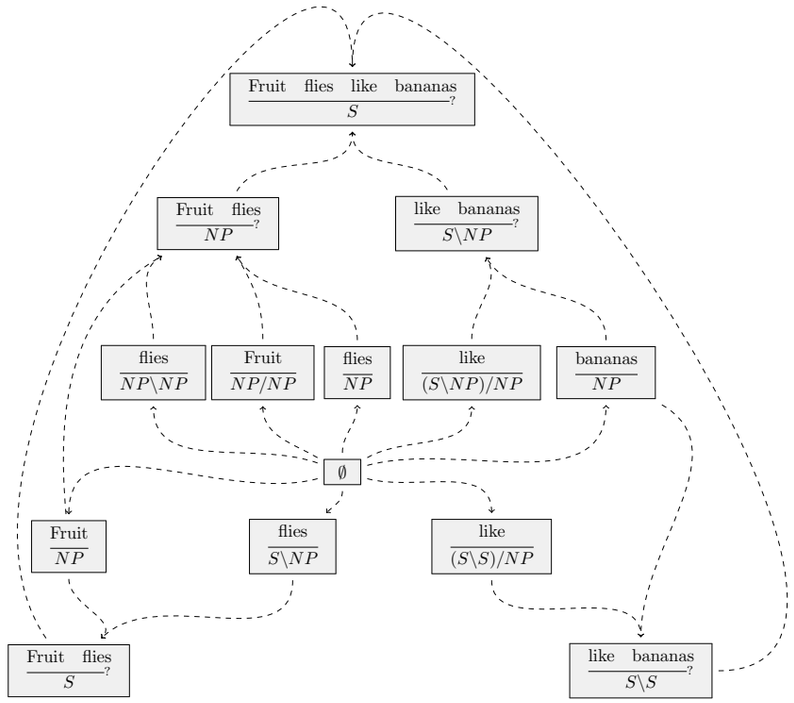


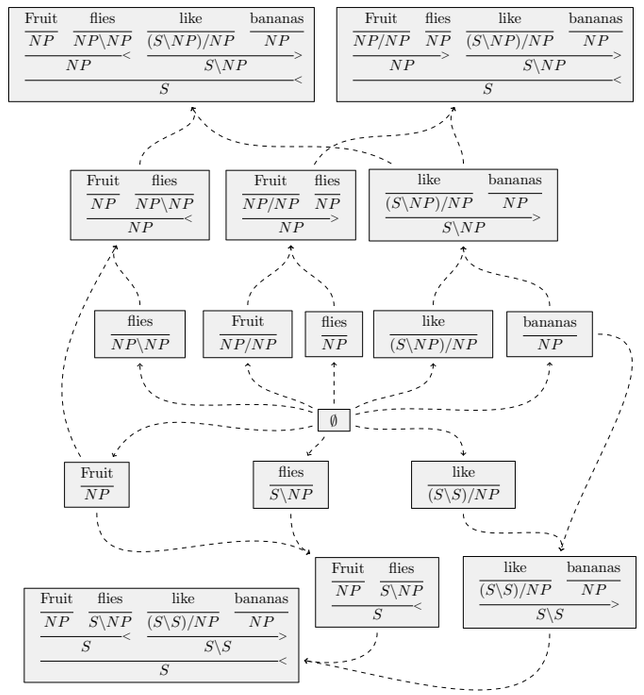
Dynamic programs with **locally factored models**, e.g.

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

Recursive neural networks break dynamic programs!

Local vs. Global Models





Local model:

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

Efficient

Inexpressive

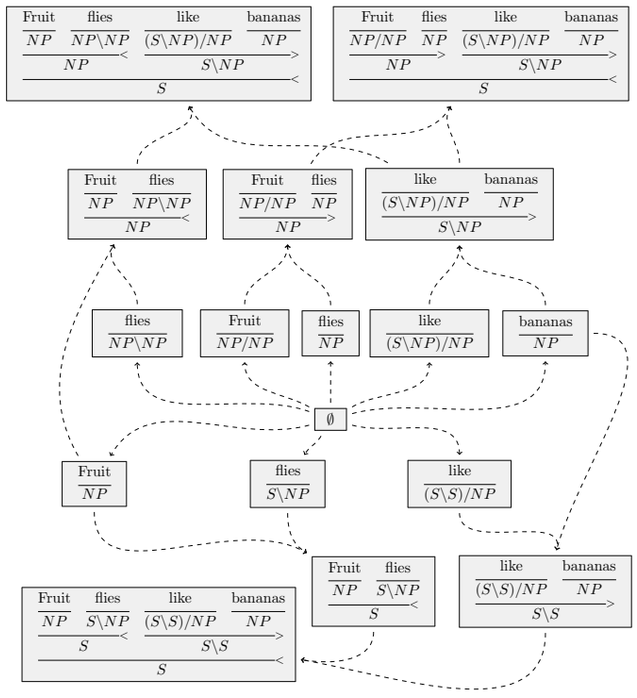
Global model:

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

Intractable

Expressive

This Work



Combined model:

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

Efficient

Expressive

Outline

- **Background: A* parsing**
- Combined global and local parsing model
- Learning to search accurately and efficiently
- Experiments on CCGBank

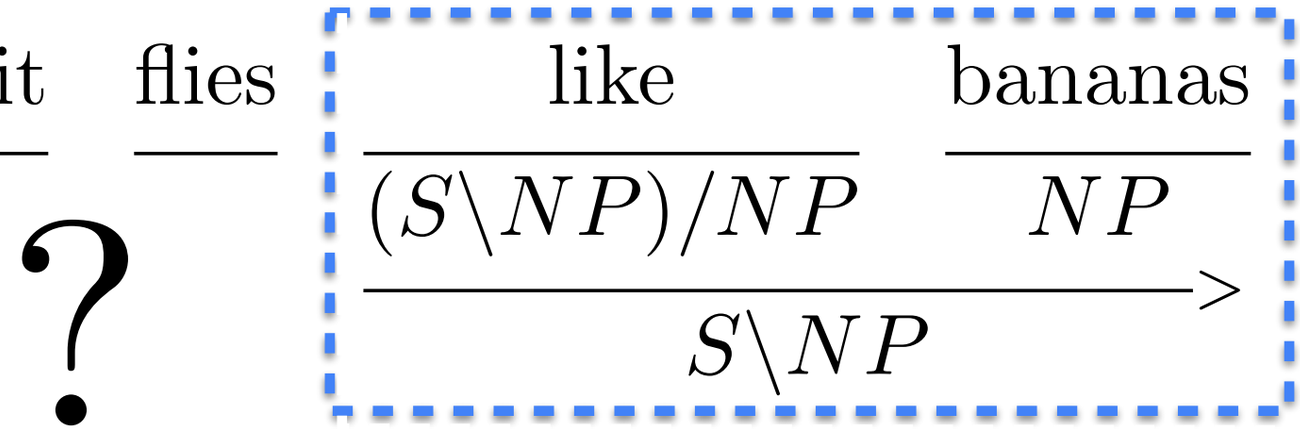
A* Parsing

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

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

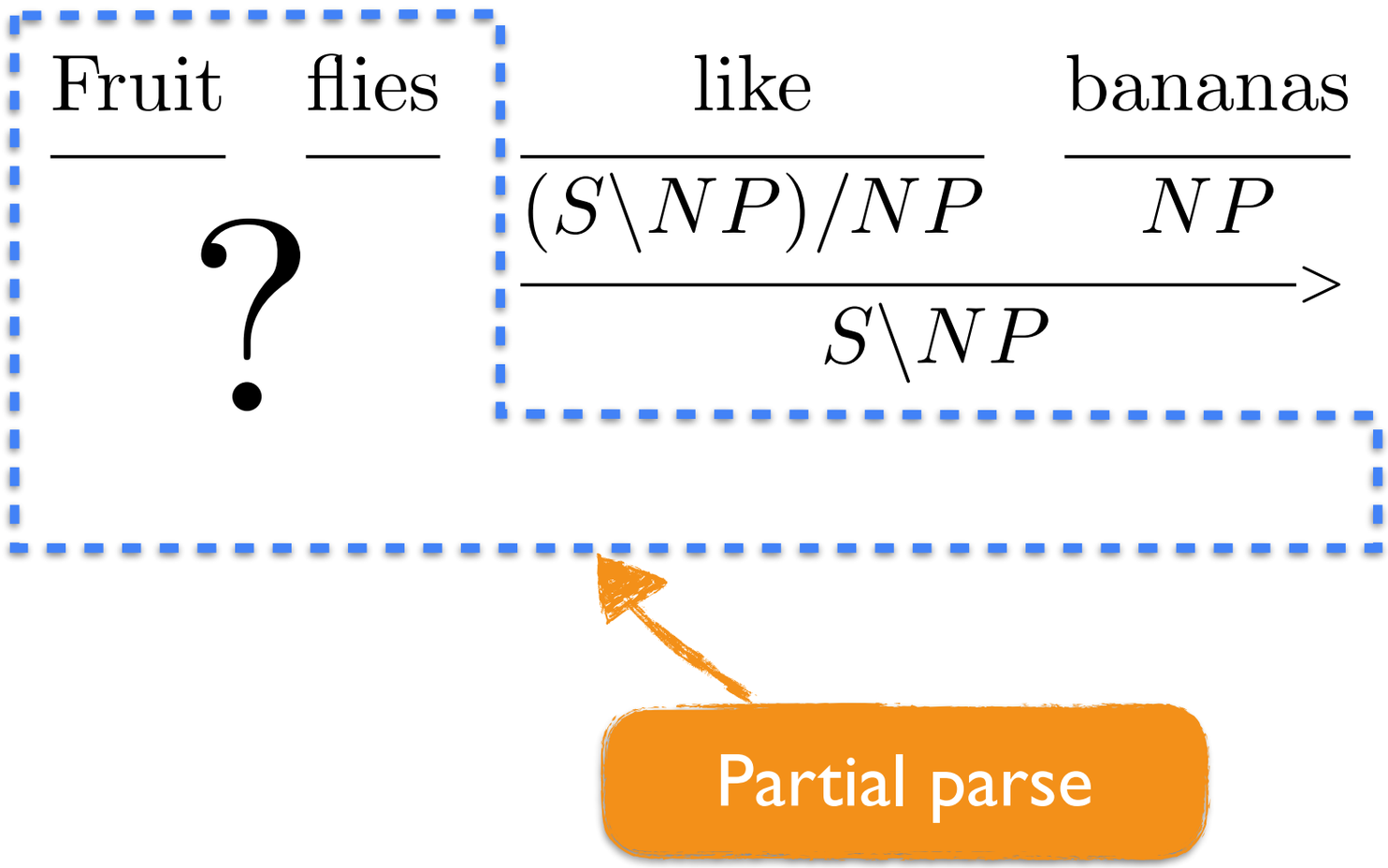
Klein and Manning, 2003

A* Parsing



Partial parse

A* Parsing



A* Parsing

$$f\Big(\underline{\text{Fruit}}\quad\underline{\text{flies}}\quad\underline{\underline{\text{like}}}\quad\underline{\text{bananas}},\ ?\Big)$$

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

Exploration priority

Partial parse

A* Parsing

Exploration priority

$$f\left(\frac{\text{Fruit}}{?}\ \frac{\text{flies}}{}\ \frac{\text{like}}{\frac{(S\backslash NP)/NP}{S\backslash NP}>}\ \frac{\text{bananas}}{NP}\right) = g\left(\frac{\text{Fruit}}{?}\ \frac{\text{flies}}{}\ \frac{\text{like}}{\frac{(S\backslash NP)/NP}{S\backslash NP}>}\ \frac{\text{bananas}}{NP}\right) + h\left(\frac{\text{Fruit}}{?}\ \frac{\text{flies}}{}\ \frac{\text{like}}{\frac{(S\backslash NP)/NP}{S\backslash NP}>}\ \frac{\text{bananas}}{NP}\right)$$

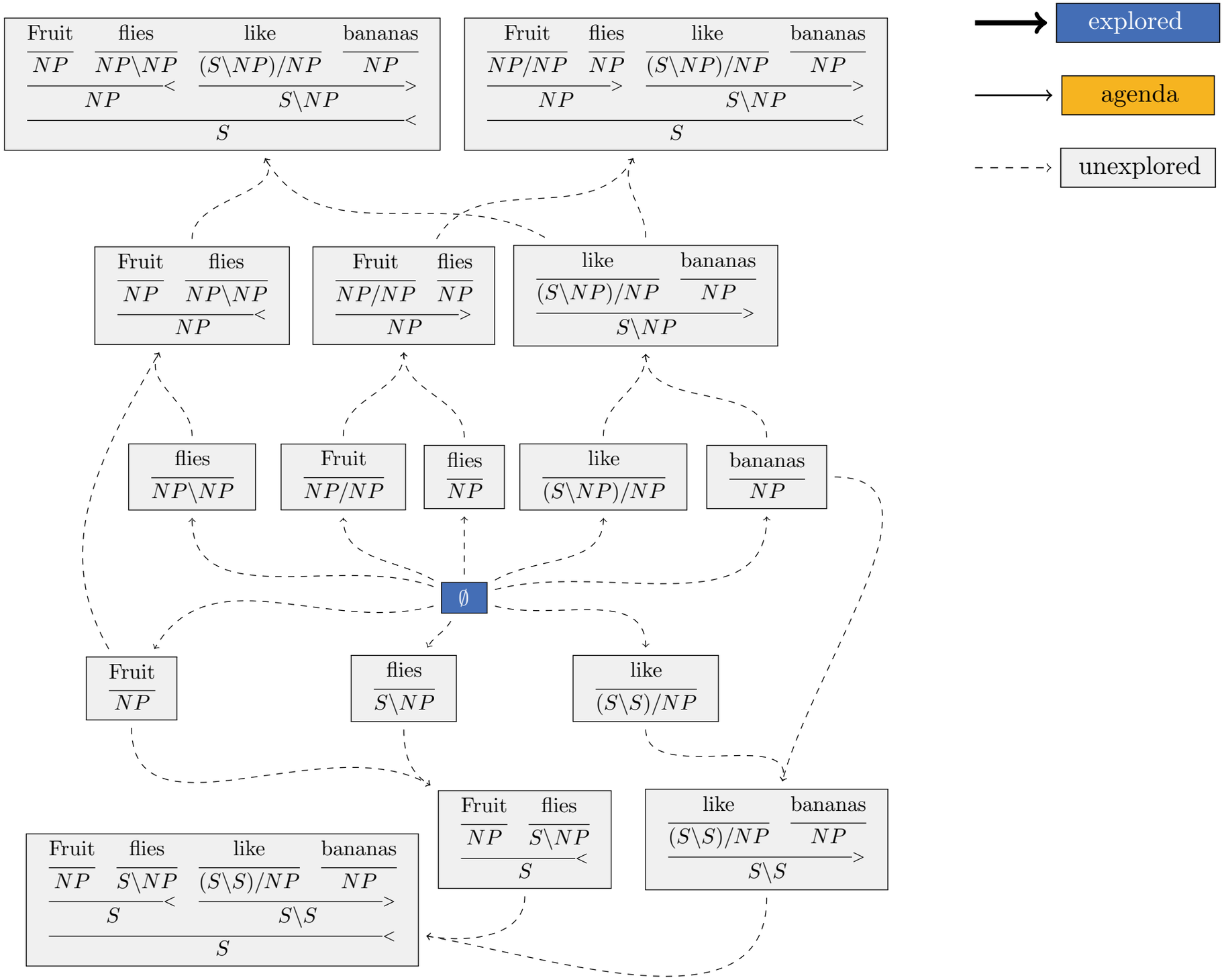
Inside score

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

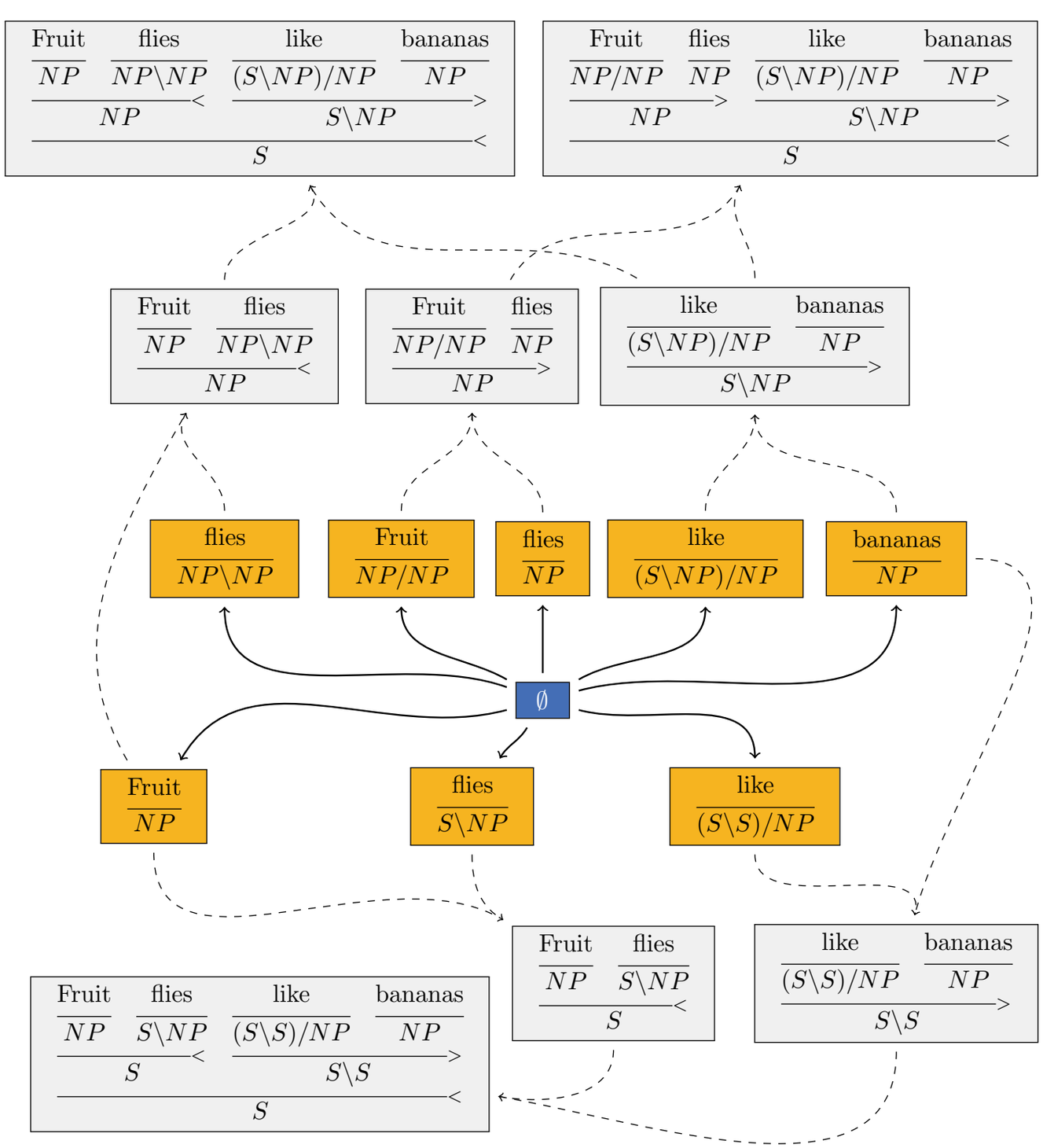
Admissible A* heuristic

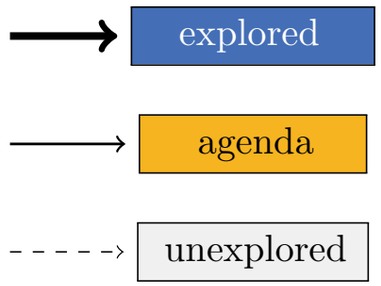
$$\frac{\text{Fruit}}{?}\ \frac{\text{flies}}{}\ \frac{\text{like}}{\frac{(S\backslash NP)/NP}{S\backslash NP}>}\ \frac{\text{bananas}}{NP}$$

A* Parsing


Fruit flies like bananas
NP NP\NP (S\NP)/NP NP
NP
S\NP
S
Fruit flies like bananas
NP/NP NP (S\NP)/NP NP
NP
S\NP
S
Fruit flies
NP NP\NP
NP
Fruit flies
NP/NP NP
NP
like bananas
(S\NP)/NP NP
S\NP
flies
NP\NP
Fruit
NP/NP
flies
NP
like
(S\NP)/NP
bananas
NP
∅
Fruit
NP
flies
S\NP
like
(S\S)/NP
Fruit flies
NP S\NP
S
like bananas
(S\S)/NP NP
S\S
Fruit flies like bananas
NP S\NP (S\S)/NP NP
S
S\S
S
explored
agenda
unexplored

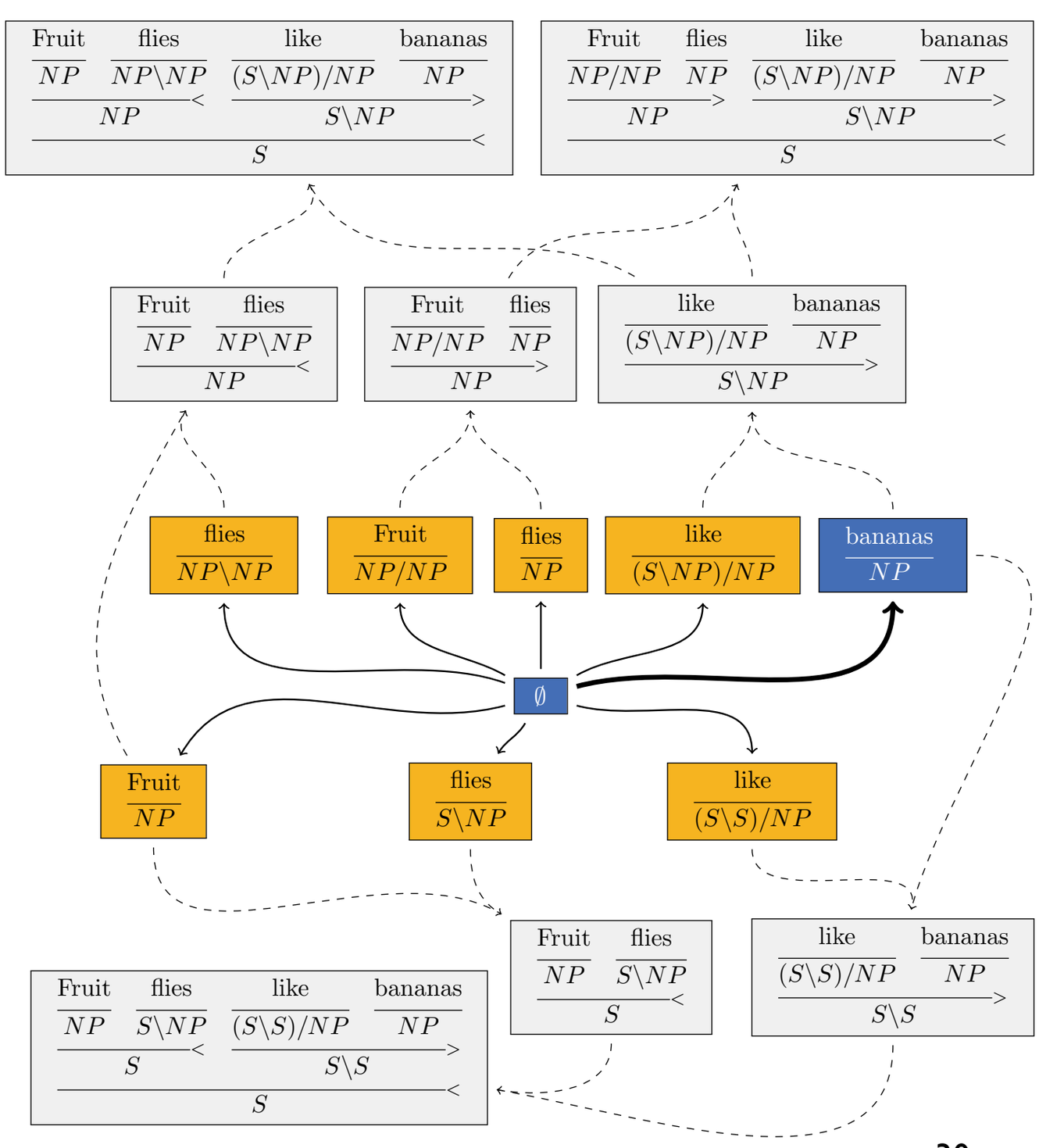
A* Parsing

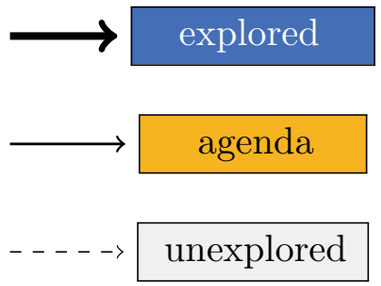




Agenda position	$f(y)$	y
1	4.5	bananas $\overline{NP}$
2	3.1	like $\overline{(S\backslash NP)/NP}$
3	1.9	Fruit $\overline{NP}$
4	-0.5	Fruit $\overline{NP/NP}$

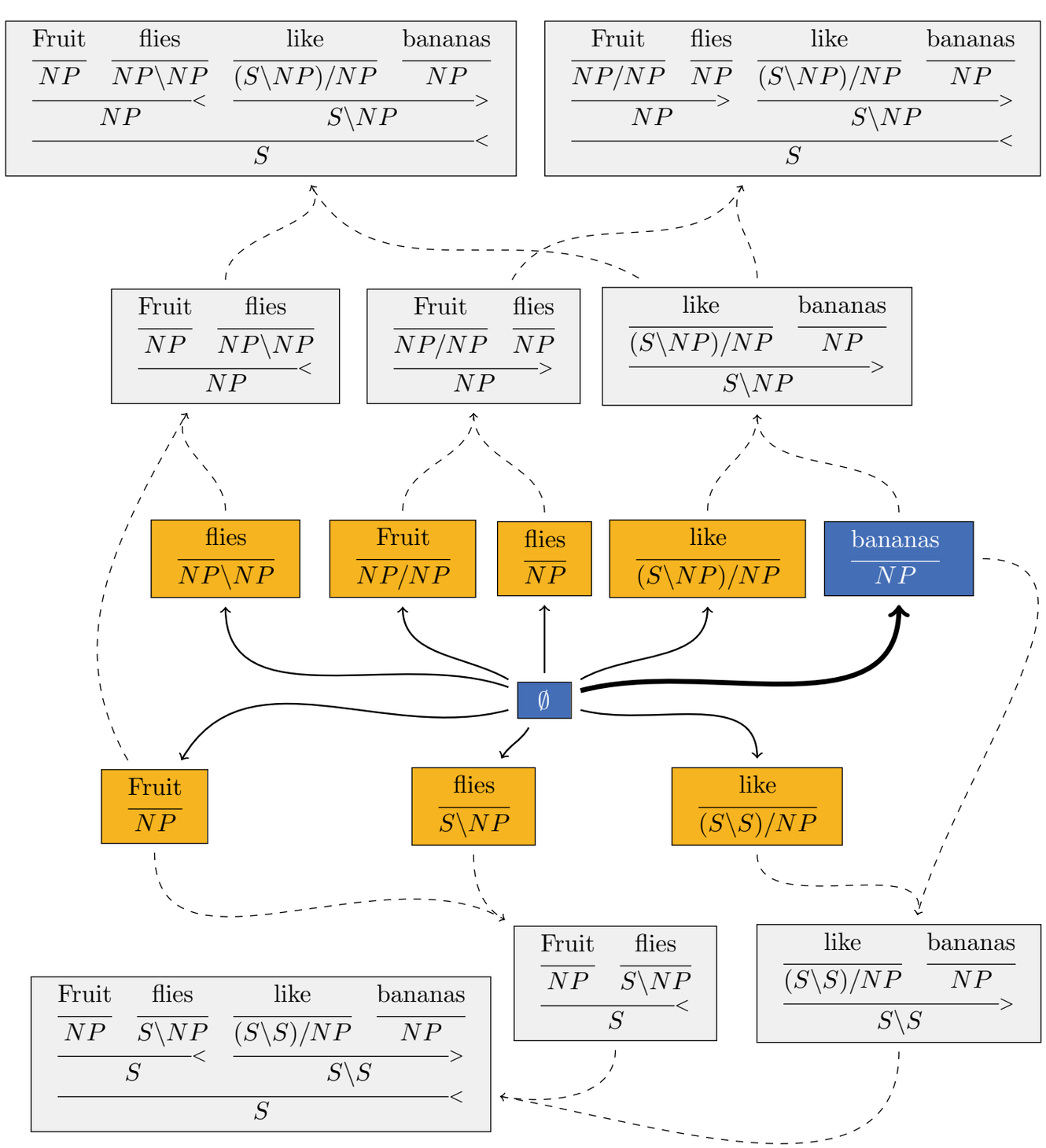
A* Parsing

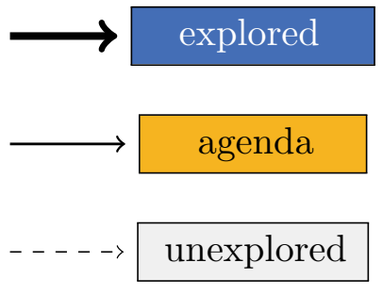




Agenda position	$f(y)$	y
1	4.5	bananas $\overline{NP}$
2	3.1	like $\overline{(S\backslash NP)/NP}$
3	1.9	Fruit $\overline{NP}$
4	-0.5	Fruit $\overline{NP/NP}$

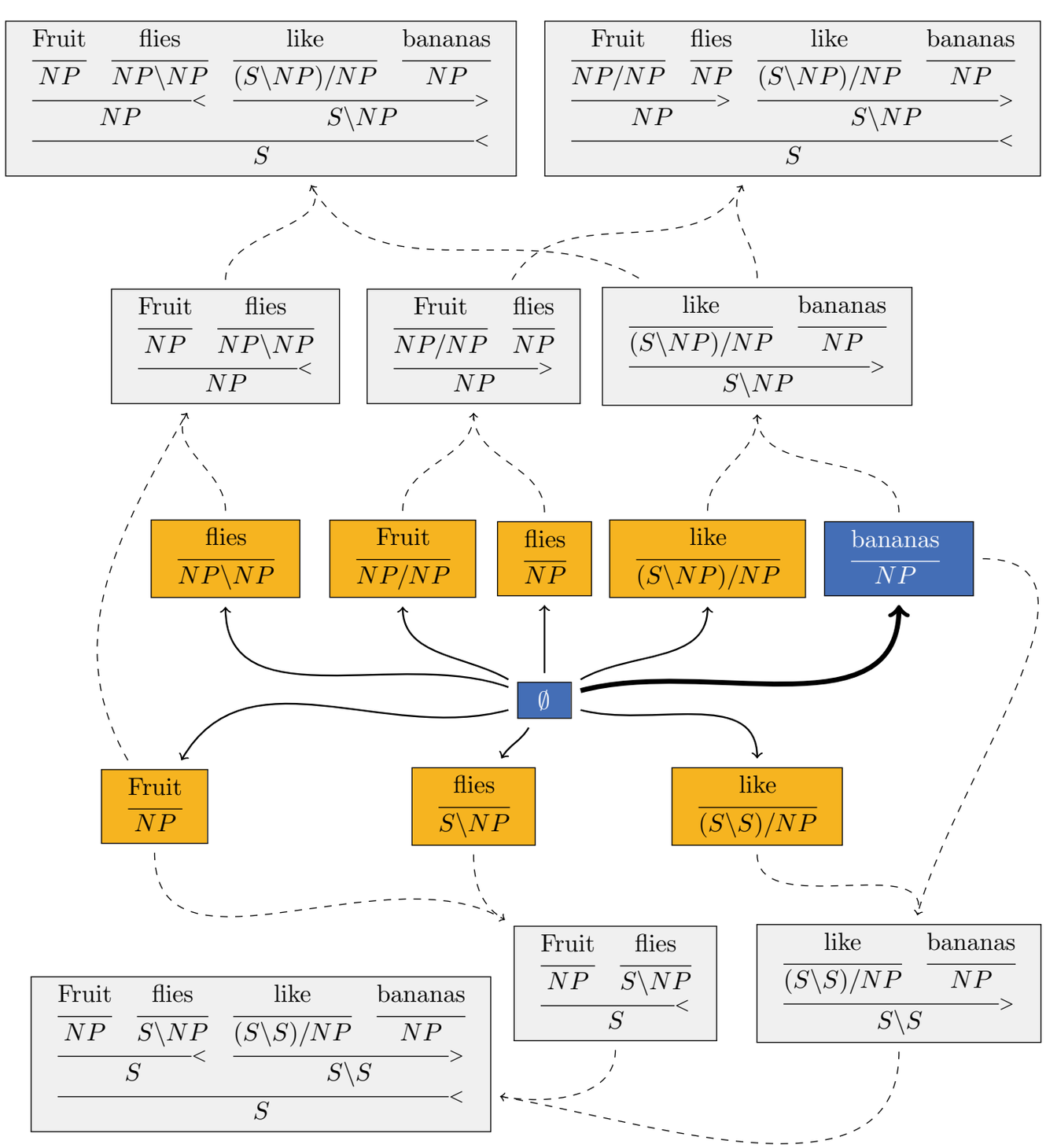
A* Parsing

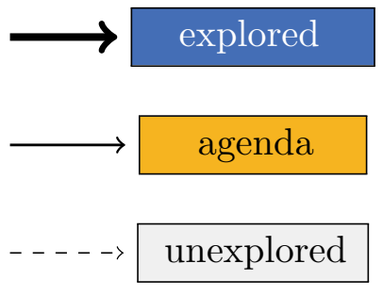




Agenda position	$f(y)$	y
2	3.1	like $\frac{}{(S\backslash NP)/NP}$
3	1.9	Fruit $\frac{}{NP}$
4	-0.5	Fruit $\frac{}{NP/NP}$

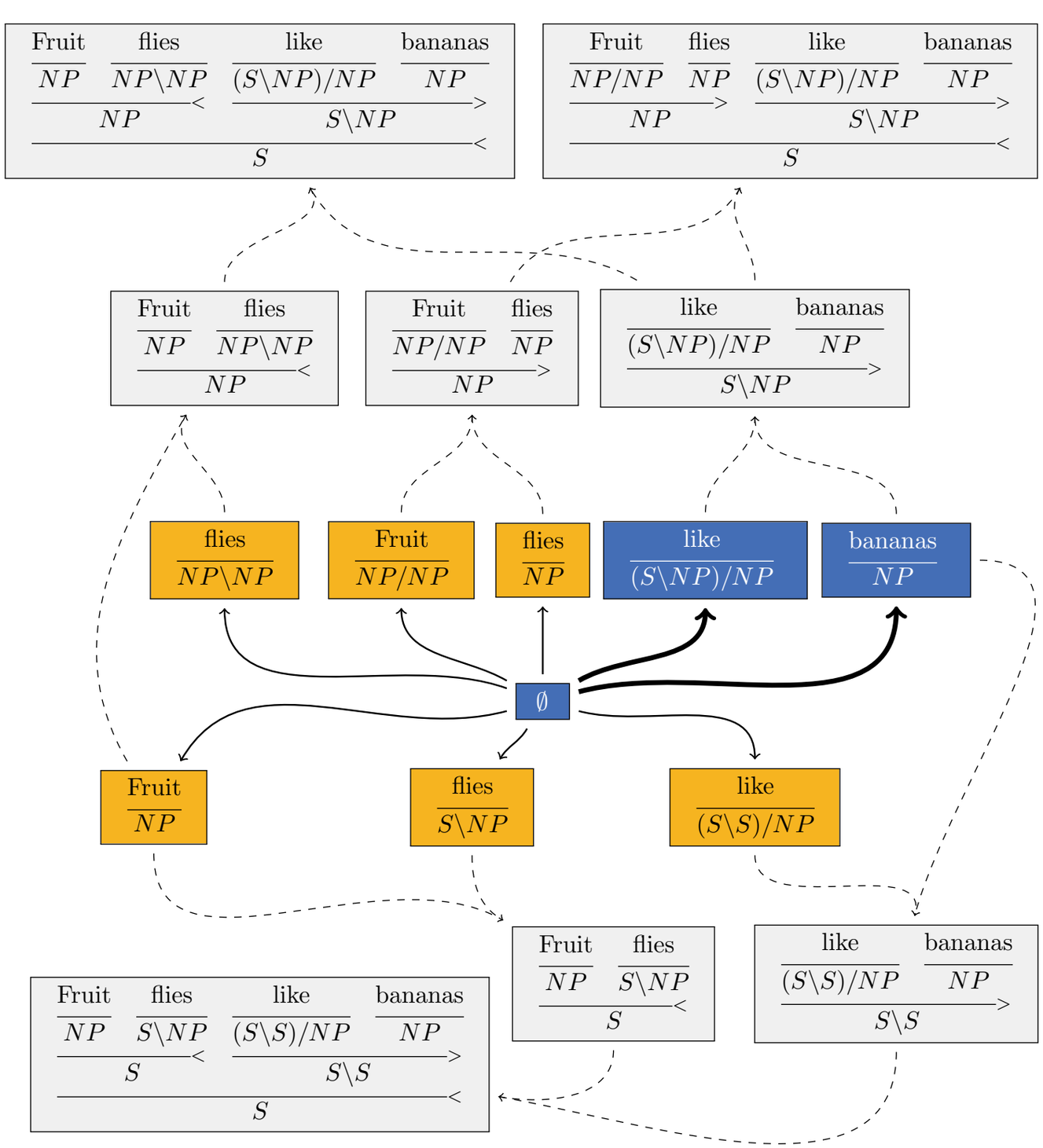
A* Parsing

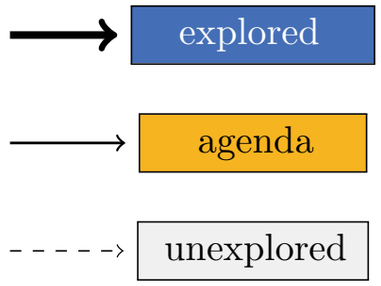




Agenda position	$f(y)$	y
1	3.1	like $\frac{}{(S\backslash NP)/NP}$
2	1.9	Fruit $\frac{}{NP}$
3	-0.5	Fruit $\frac{}{NP/NP}$
4	-1.3	flies $\frac{}{NP}$

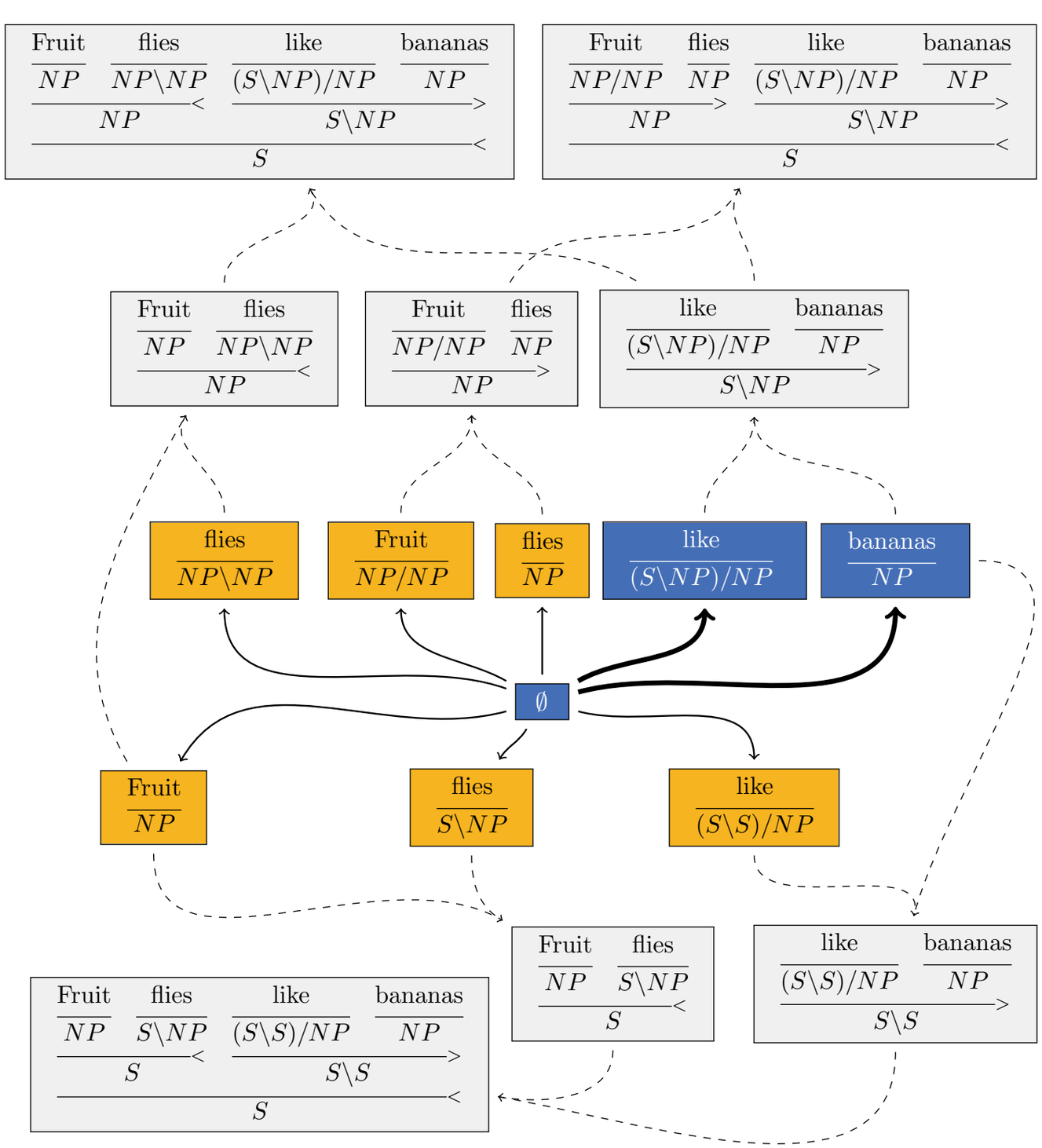
A* Parsing

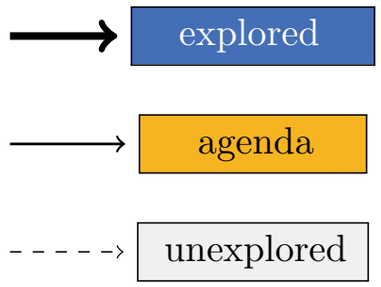




Agenda position	$f(y)$	y
1	3.1	like $\frac{}{(S\backslash NP)/NP}$
2	1.9	Fruit $\frac{}{NP}$
3	-0.5	Fruit $\frac{}{NP/NP}$
4	-1.3	flies $\frac{}{NP}$

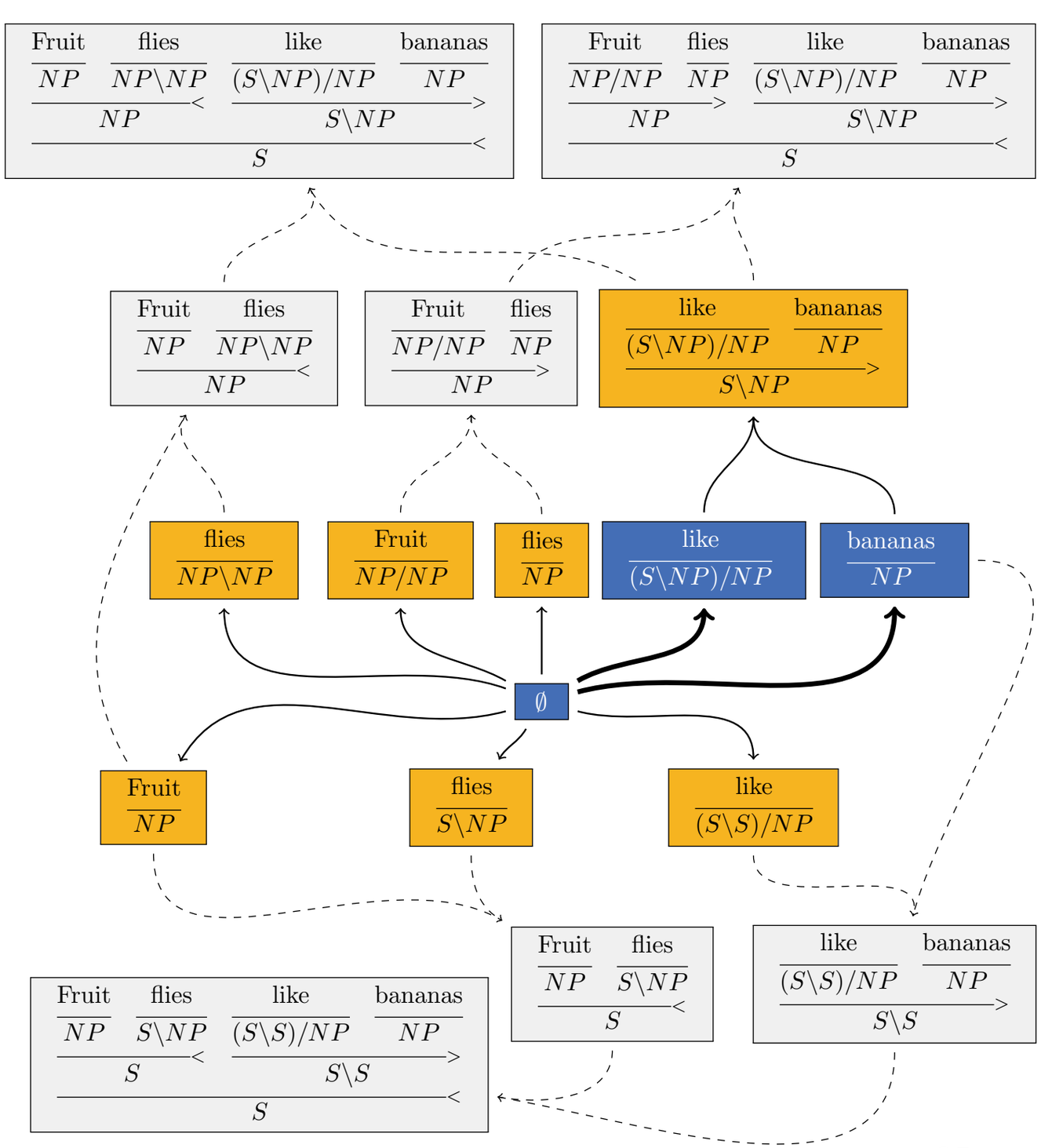
A* Parsing

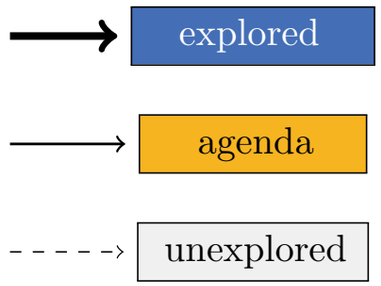




Agenda position	$f(y)$	y
2	1.9	Fruit $\overline{NP}$
3	-0.5	Fruit $\overline{NP/NP}$
4	-1.3	flies $\overline{NP}$

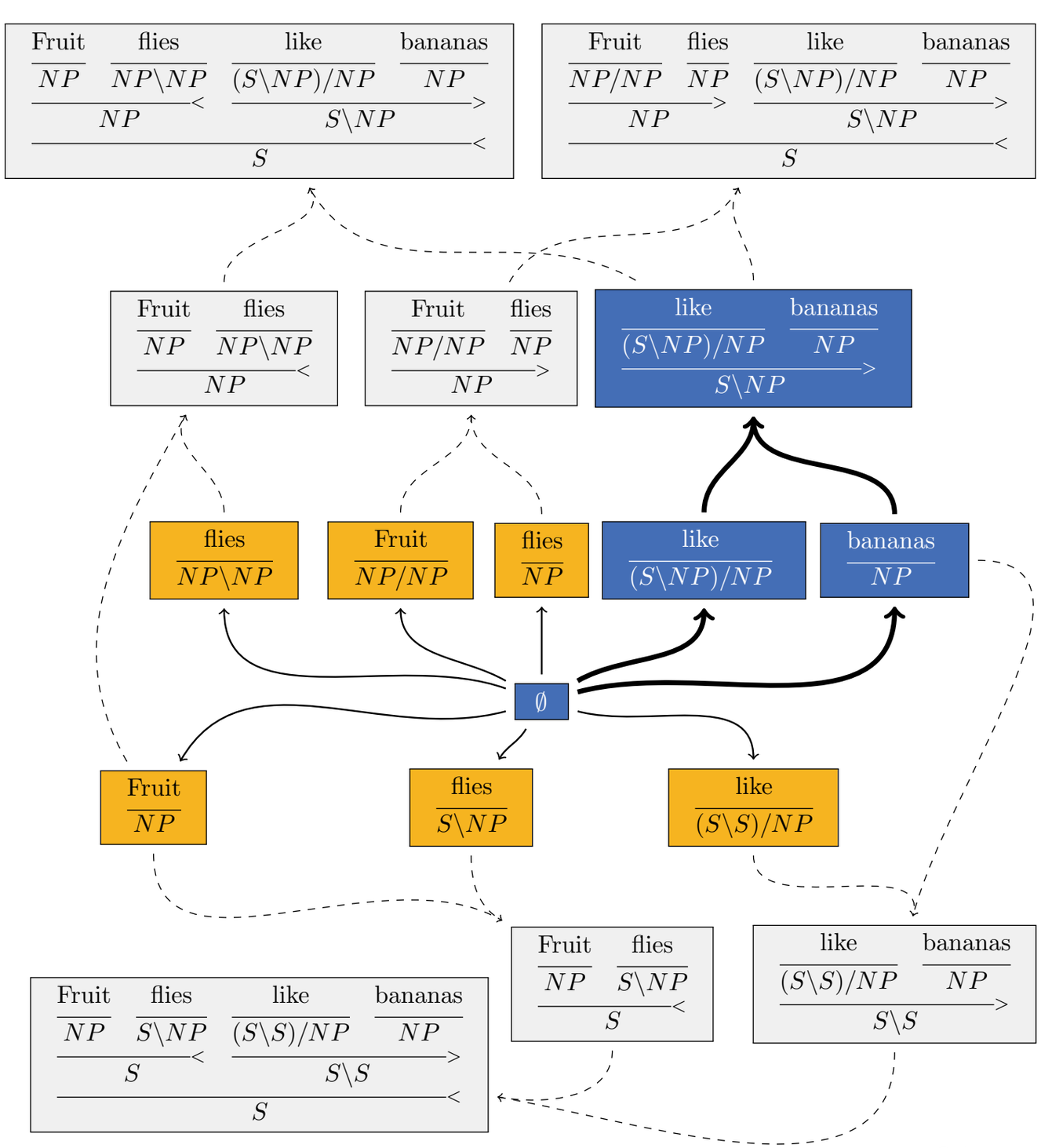
A* Parsing





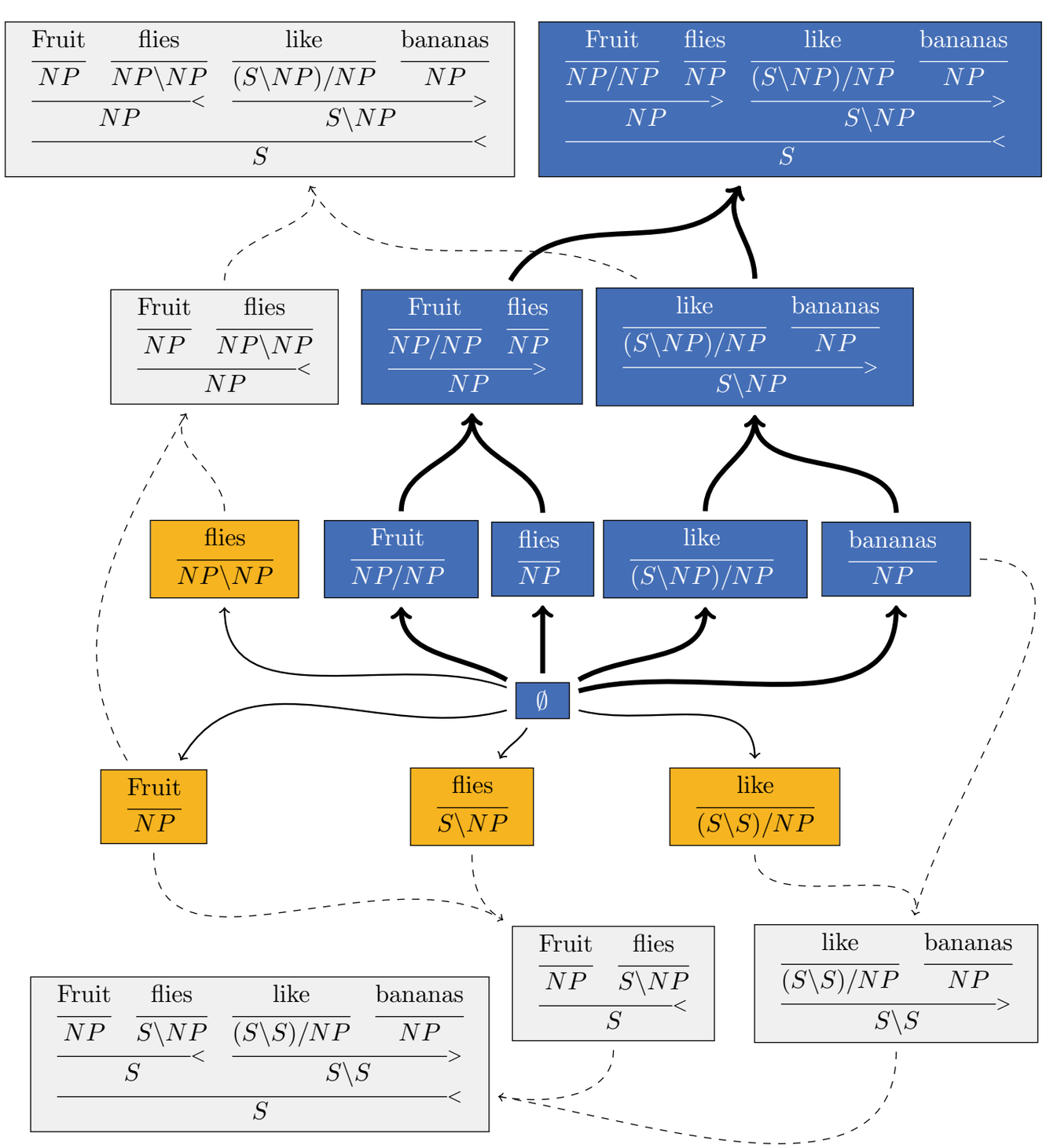
Agenda position	$f(y)$	y
1	2.1	like $\frac{}{(S\backslash NP)/NP}$ bananas $\frac{}{NP}$ → $S\backslash NP$
2	1.9	Fruit $\frac{}{NP}$
3	-0.5	Fruit $\frac{}{NP/NP}$
4	-1.3	flies $\frac{}{NP}$

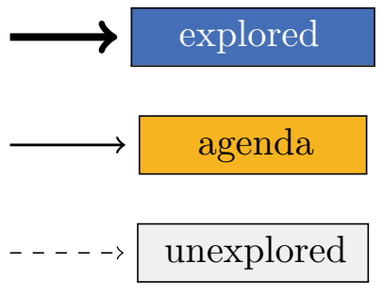
A* Parsing



Agenda position	$f(y)$	y
1	2.1	like $\frac{}{(S\backslash NP)/NP}$ bananas $\frac{}{NP}$ → $S\backslash NP$
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	1.9	Fruit $\overline{NP}$
2	-1.5	like $\overline{(S\backslash S)/NP}$
3	…	…
4	…	…

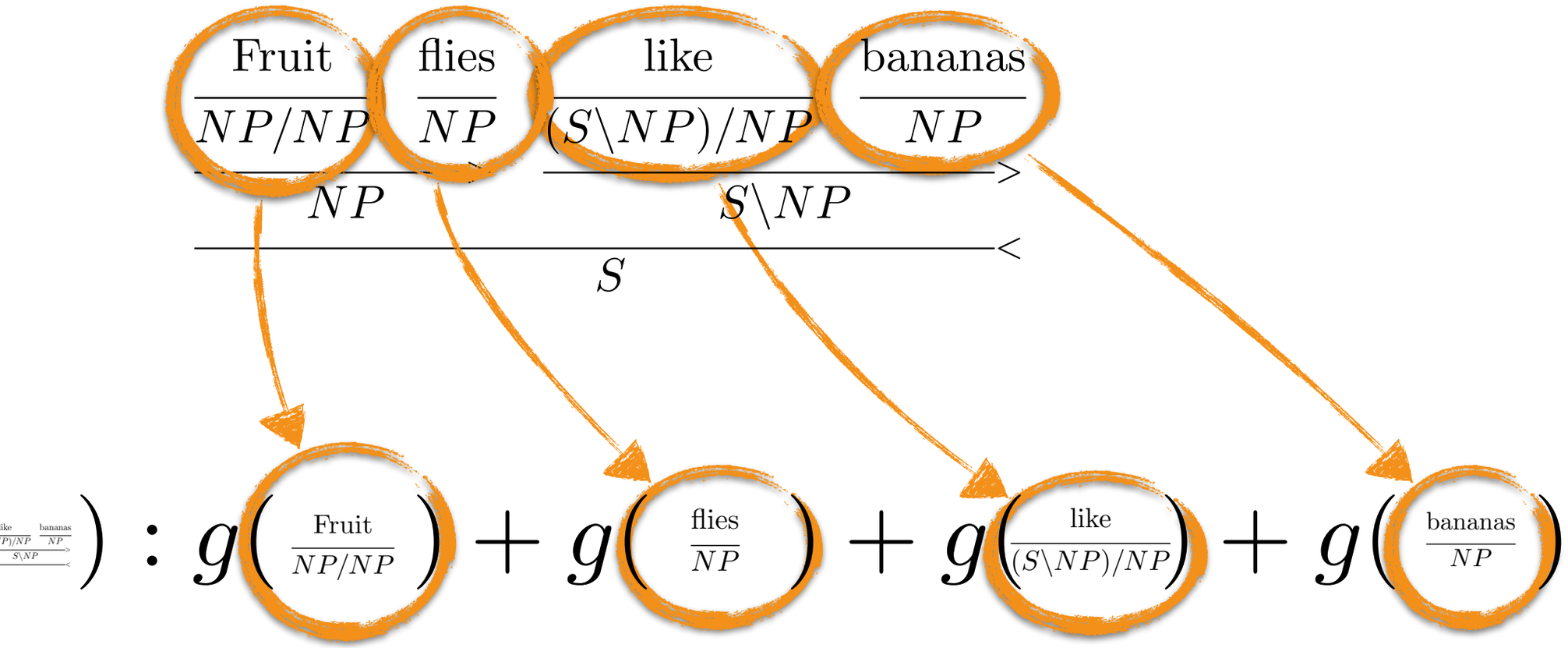
Locally Factored Model

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

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

Locally Factored Model

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



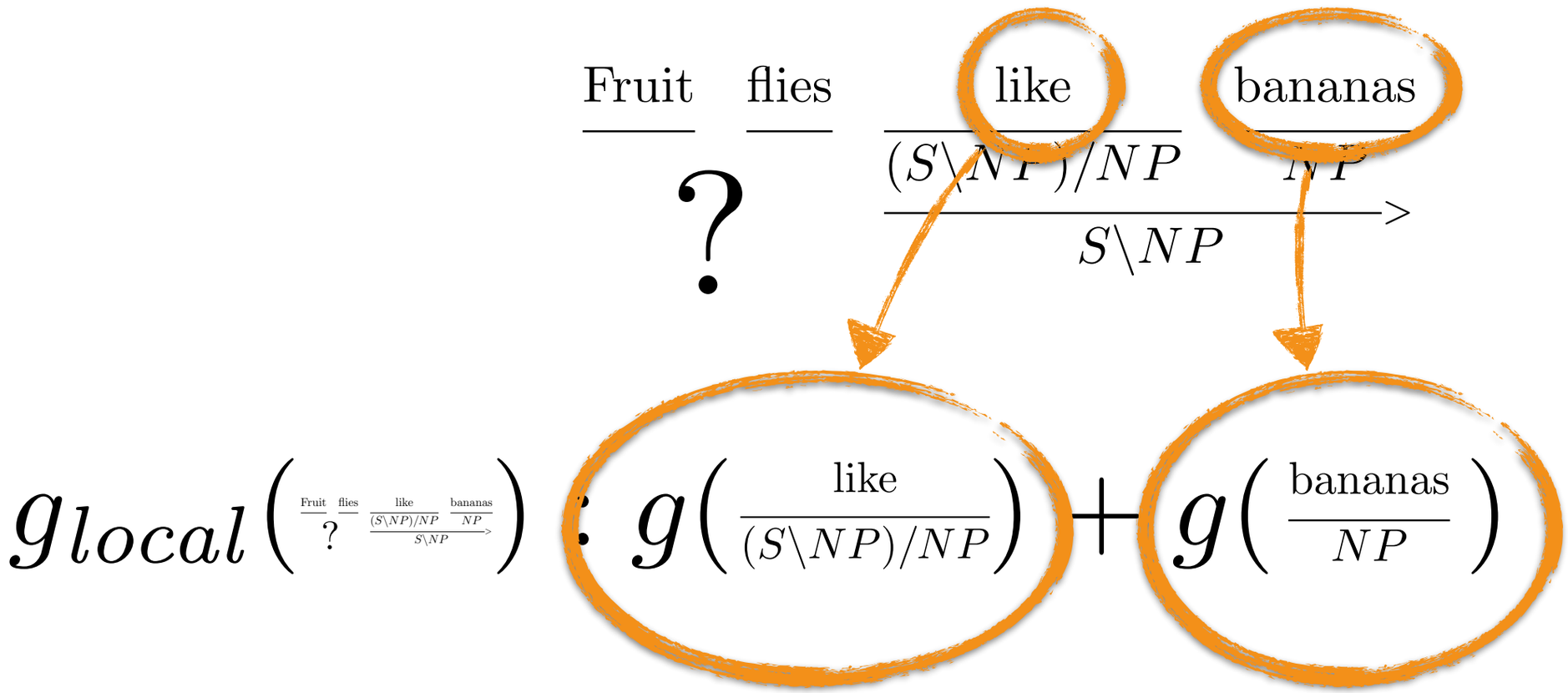
Locally Factored Model

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

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

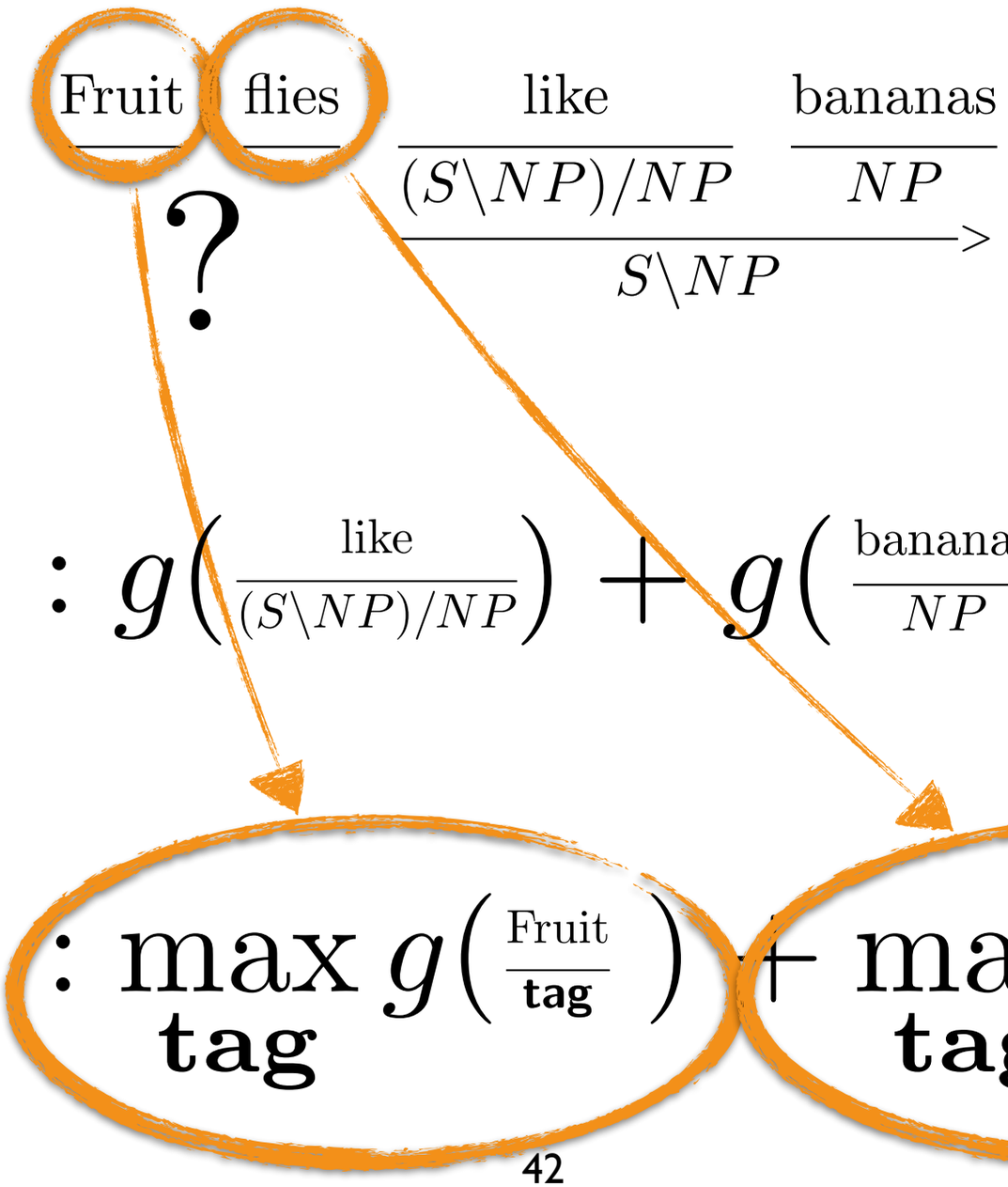
Locally Factored Model

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



Locally Factored Model

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



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

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

Outline

- Background: A* parsing
- **Combined global and local parsing model**
- Learning to search accurately and efficiently
- Experiments on CCGBank

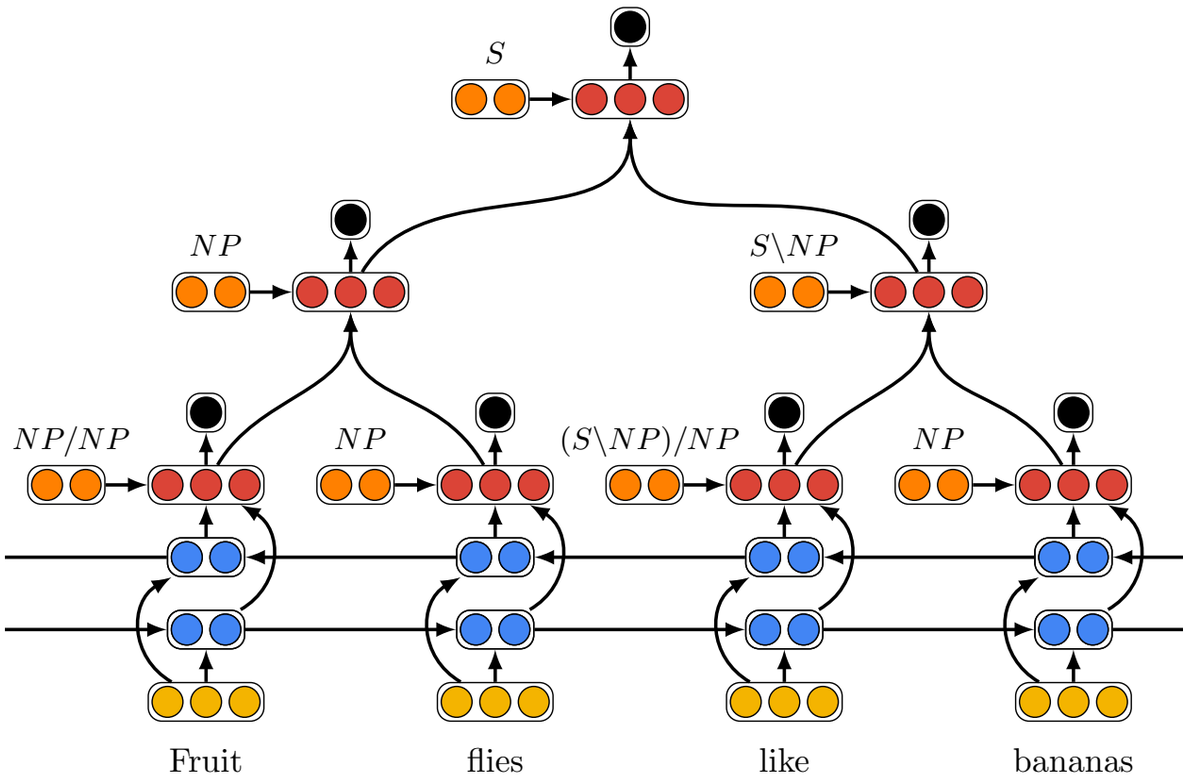
Global A* Parsing

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

- 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

$g_{global}(y)$:



$h_{global}(y)$:

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

Modeling Global Structure

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

Non-positive global model

$$h(y) = 0$$

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

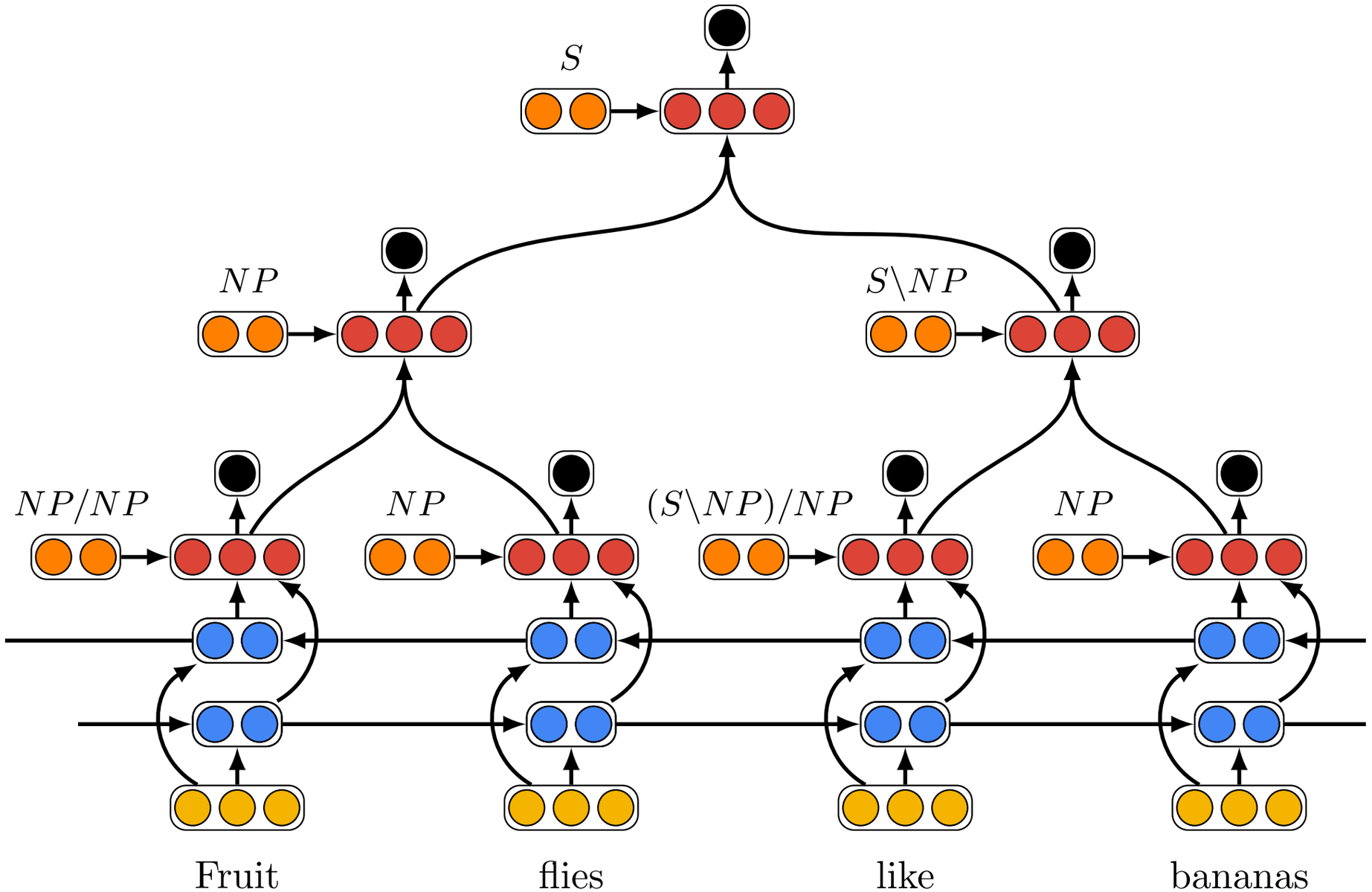
Global Model: $g_{global}(y)$

Parse Scores

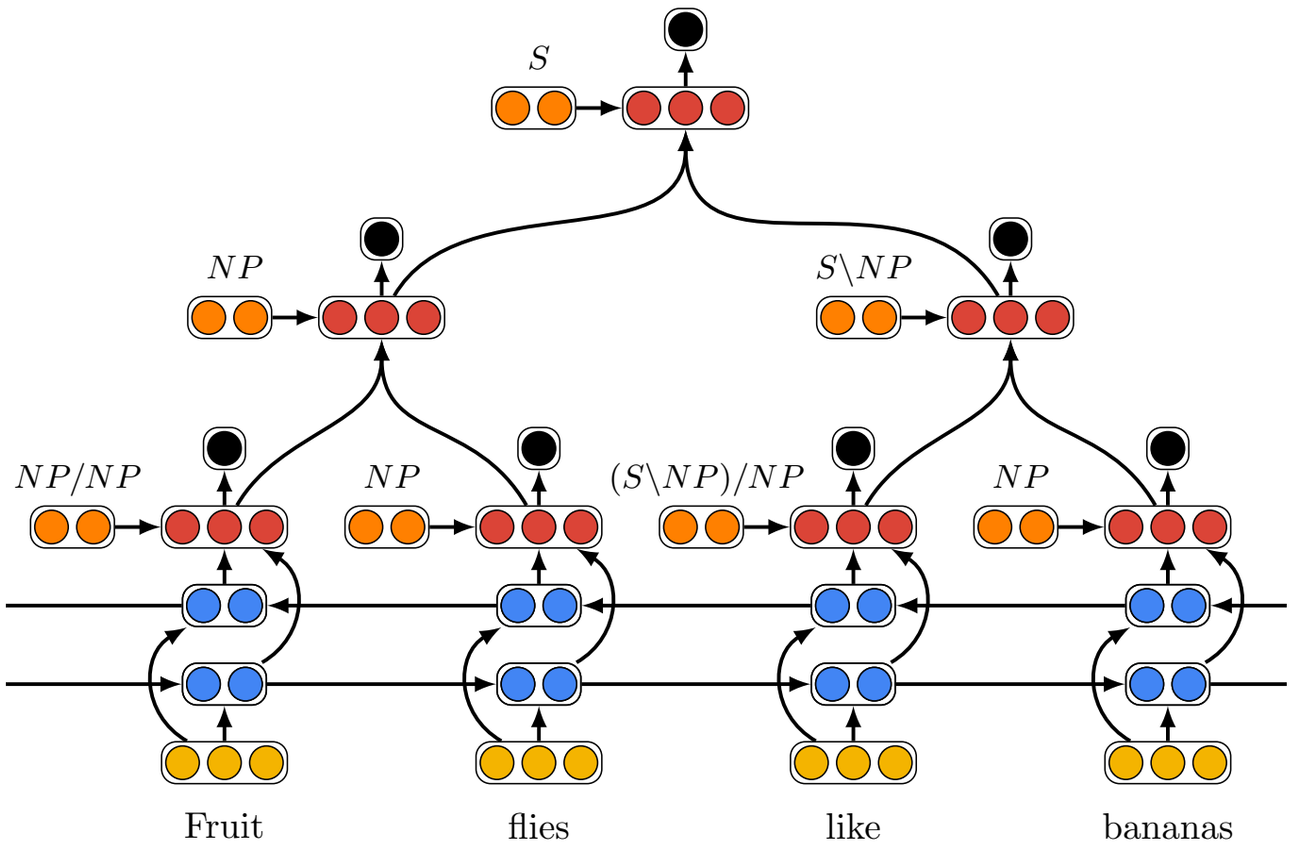
Tree-LSTM

Bidirectional LSTM

Word embeddings



Non-positive Global Model



Log-probability of a logistic regression layer

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

Division of Labor

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

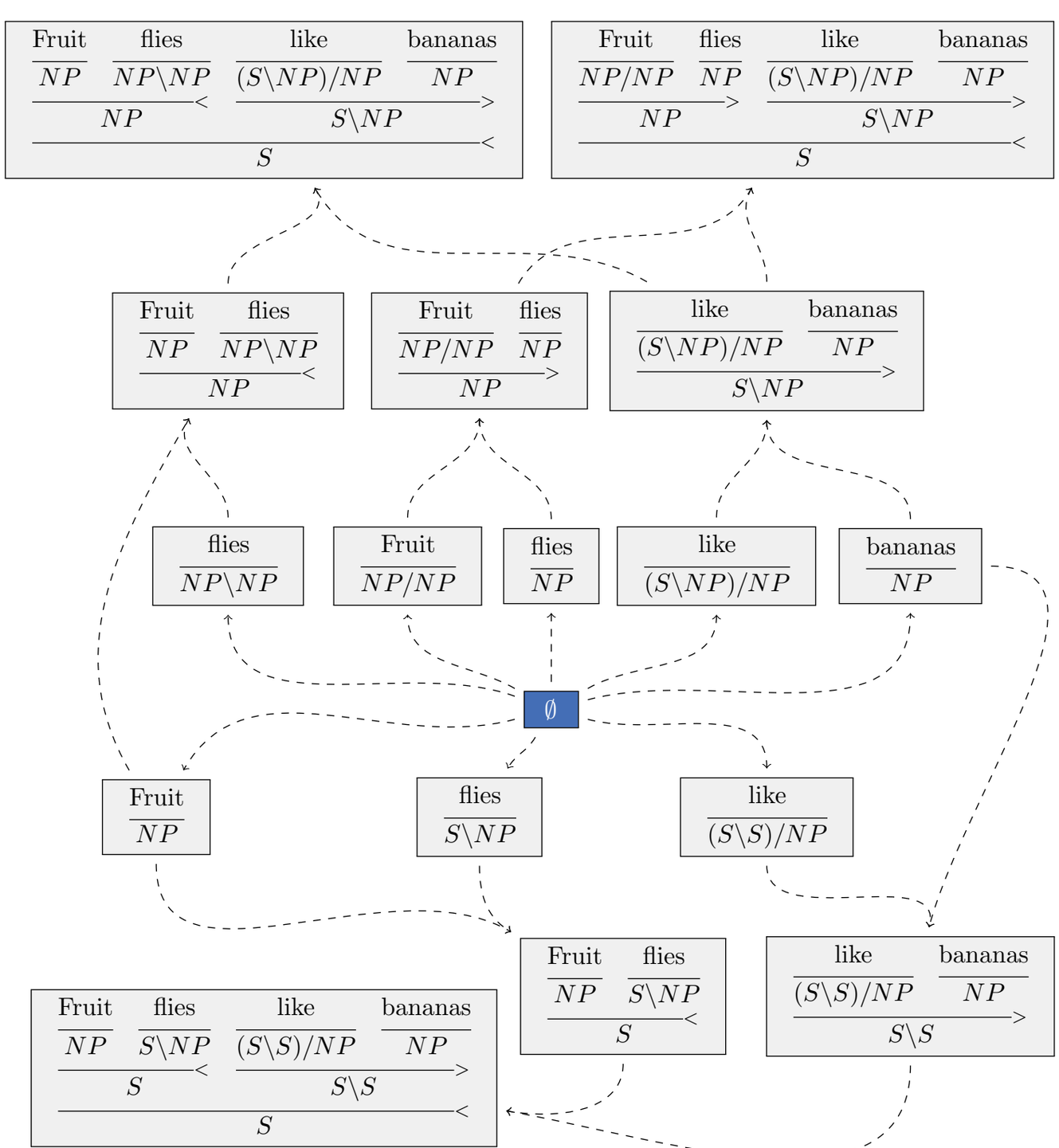
- Limited expressivity
- Provides guidance with an A^* heuristic

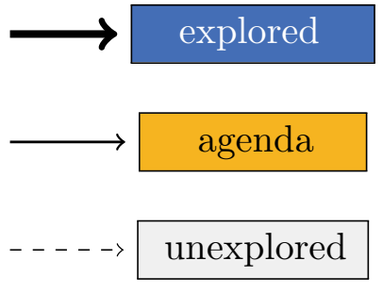
- Global expressivity
- **Discriminative only when necessary**

Outline

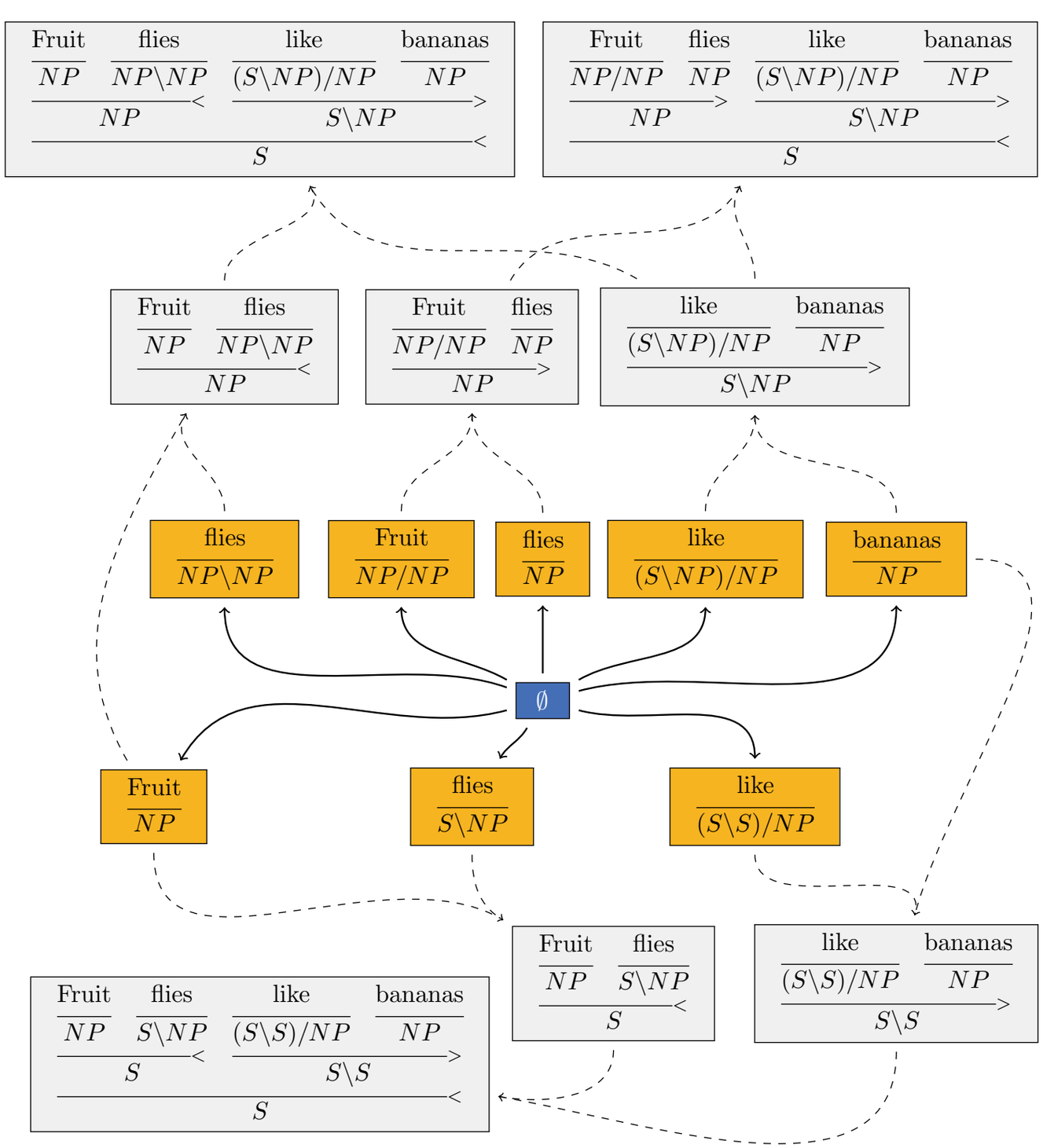
- Background: A* parsing
- Combined global and local parsing model
- **Learning to search accurately and efficiently**
- Experiments on CCGBank

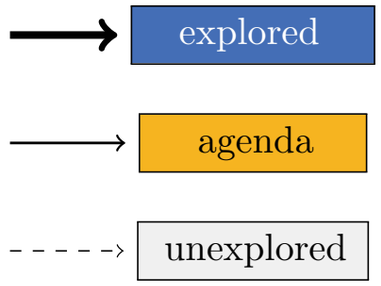
Learning with A*





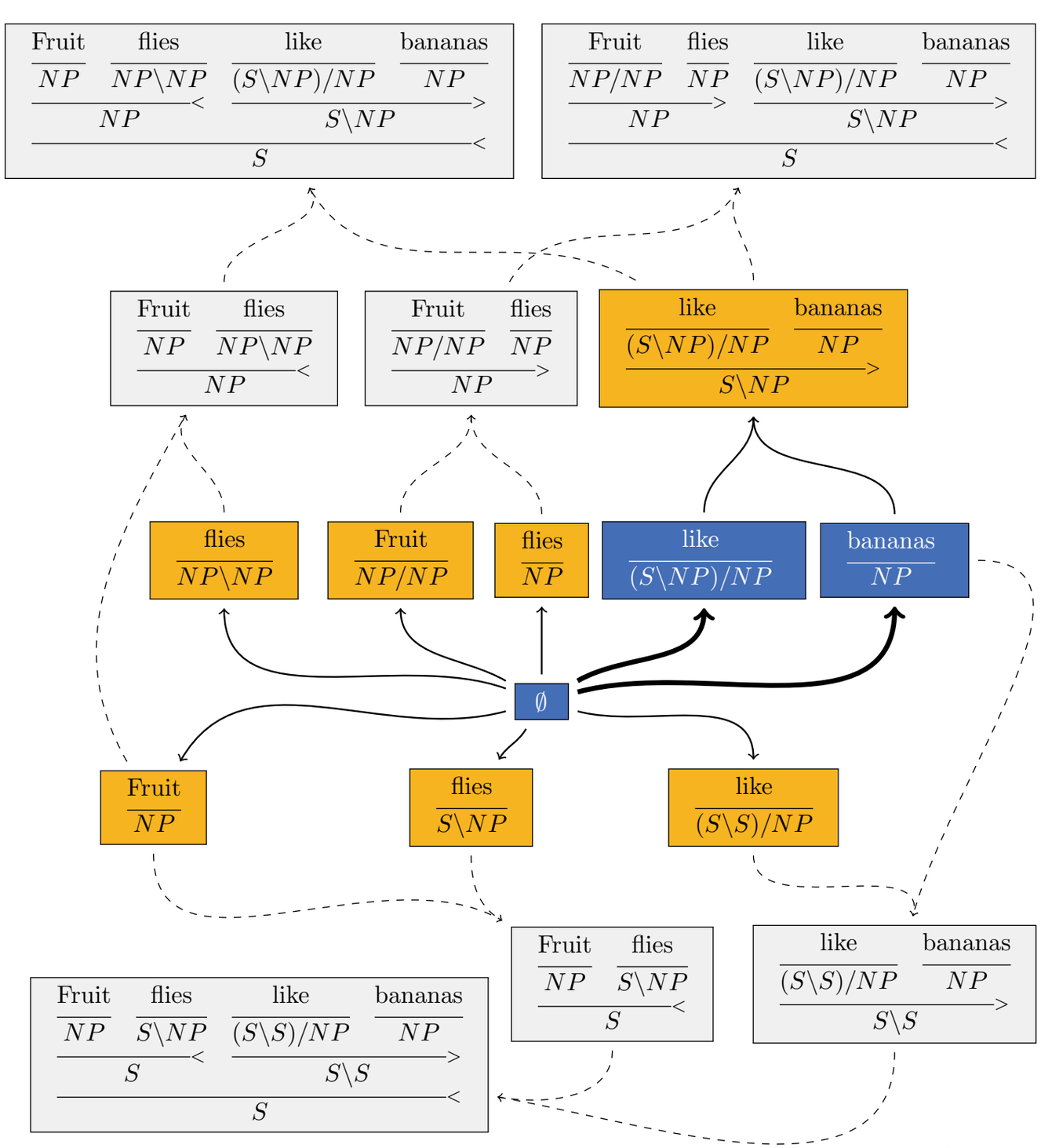
Learning with A*

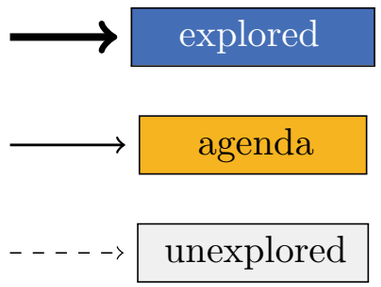




Agenda position	$f(y)$	y	Is correct?
1	4.5	bananas / NP	✓
2	3.1	like / $(S\backslash NP)/NP$	✓
3	1.9	Fruit / NP	✗
4	-0.5	Fruit / NP/NP	✓

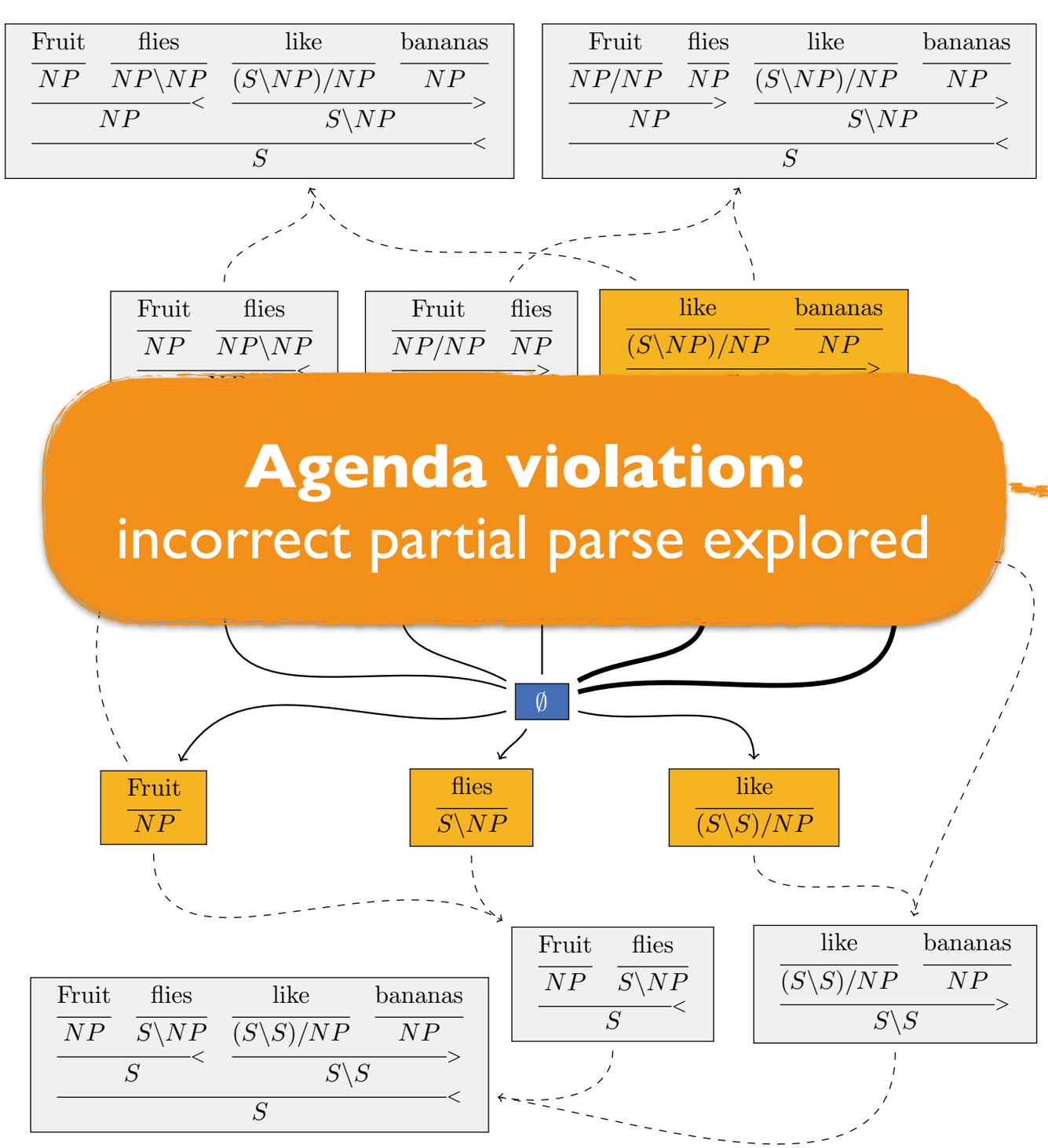
Learning with A*





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

Learning with A*



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

. . .

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

. . .

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

. . .]

$$L(\mathcal{A}) = \sum_{t=1}^{T} \max_{y \in \mathcal{A}_t} f(y) - \max_{y \in \text{GOLD}(\mathcal{A}_t)} f(y)$$

Top of agenda

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	…	…	…
4	…	…	…

Explicitly optimize for search efficiency!

Outline

- Background: A* parsing
- Combined global and local parsing model
- Learning to search accurately and efficiently
- **Experiments on CCGBank**

Experimental Setup

- $g_{local}(y)$: supertag-factored model from Lewis et al. (2016)
- Evaluate on CCGBank (Hockenmaier & Steedman, 2007)
- Comparisons:

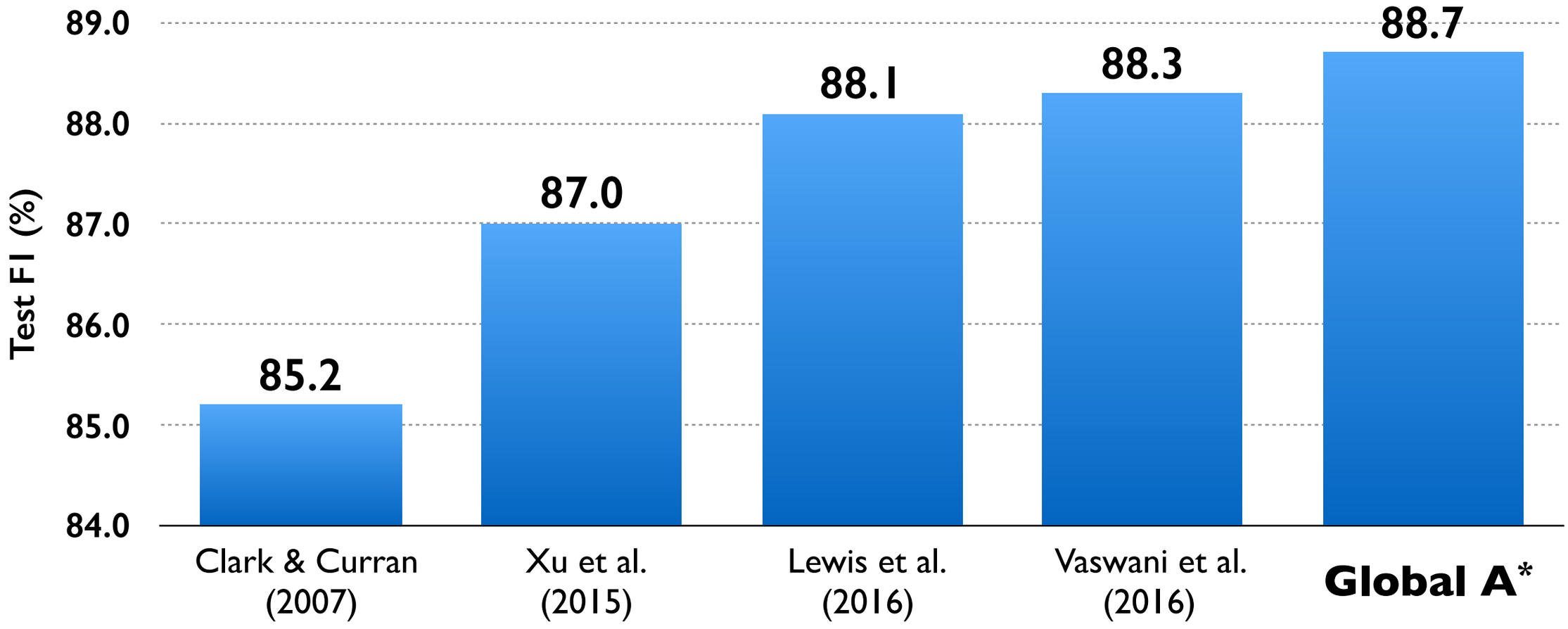
	Clark & Curran (2007)	Xu et al. (2015)	Lewis et al. (2016)	Vaswani et al. (2016)
Is global?		✓		✓
Is exact?			✓	

Experimental Setup

- $g_{local}(y)$: supertag-factored model from Lewis et al. (2016)
- Evaluate on CCGBank (Hockenmaier & Steedman, 2007)
- Comparisons:

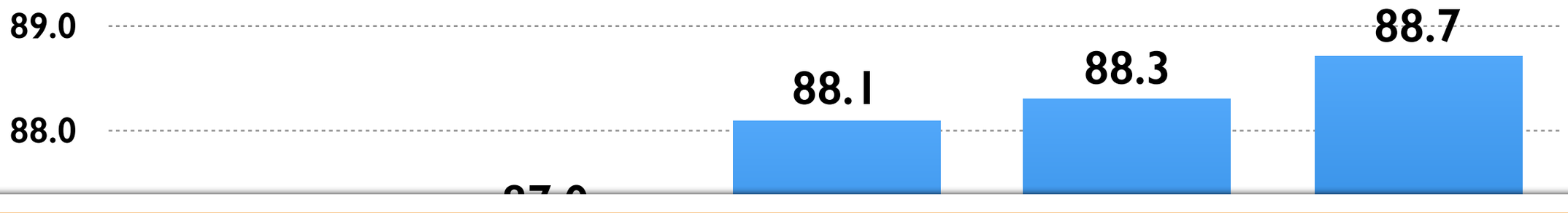
	Clark & Curran (2007)	Xu et al. (2015)	Lewis et al. (2016)	Vaswani et al. (2016)	**Global A***
Is global?		✓		✓	✓
Is exact?			✓		✓

CCG Parsing Results



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

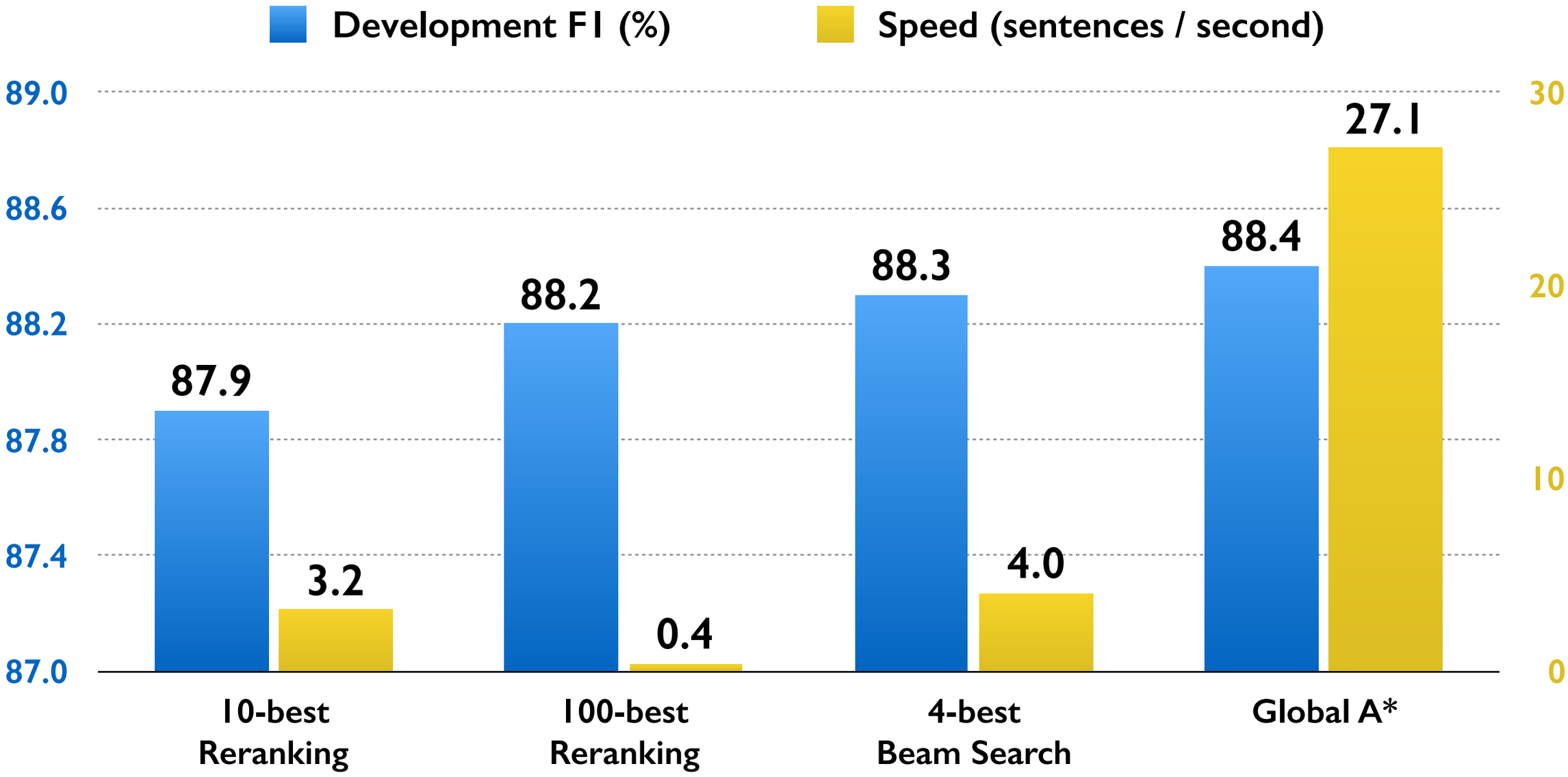
CCG Parsing Results



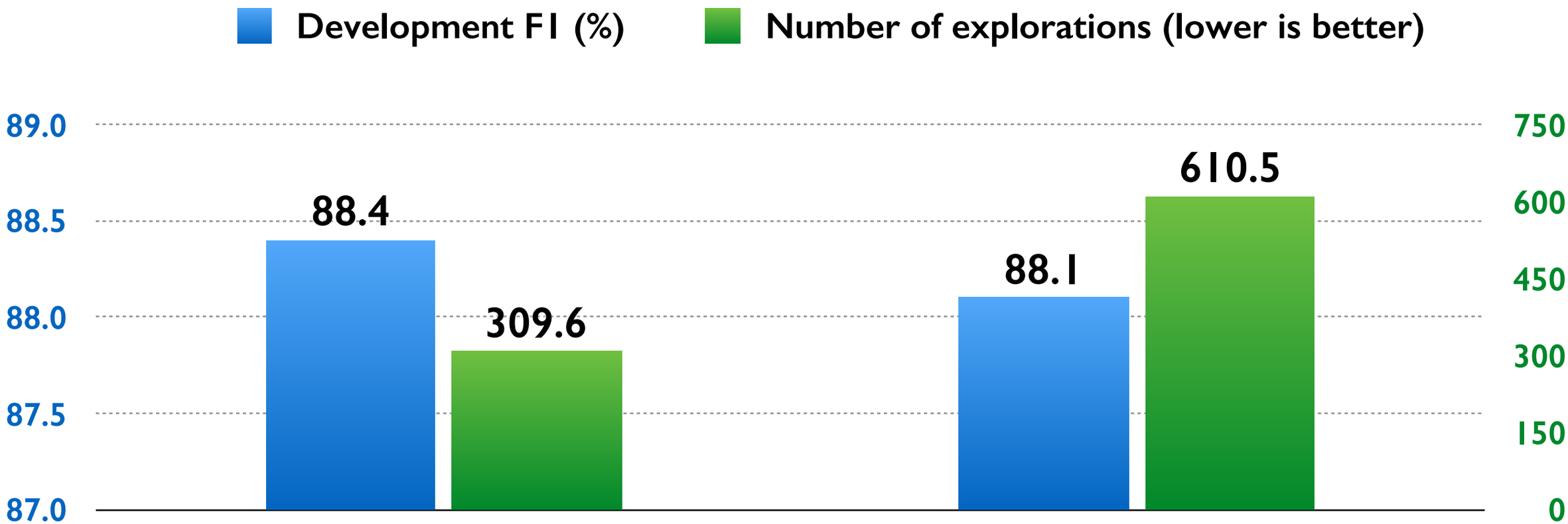
- Optimal parse found for 99.9% of sentences
- Explores only 190 partial parses on average

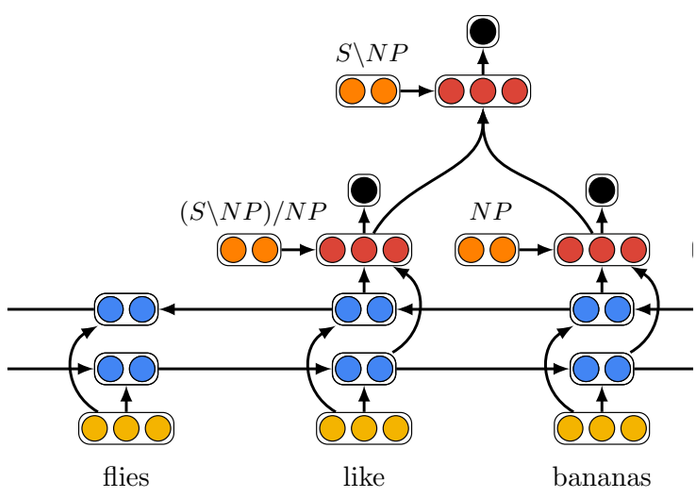
Is global?		✓		✓	✓
Is exact?			✓		✓

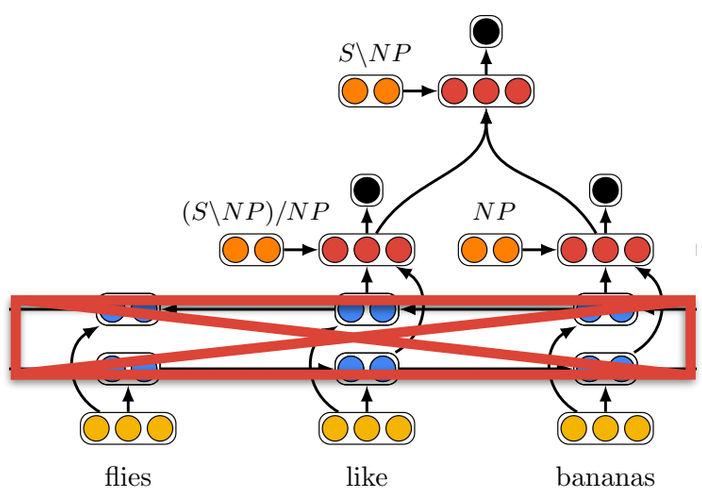
Decoder Comparisons



Context Ablation

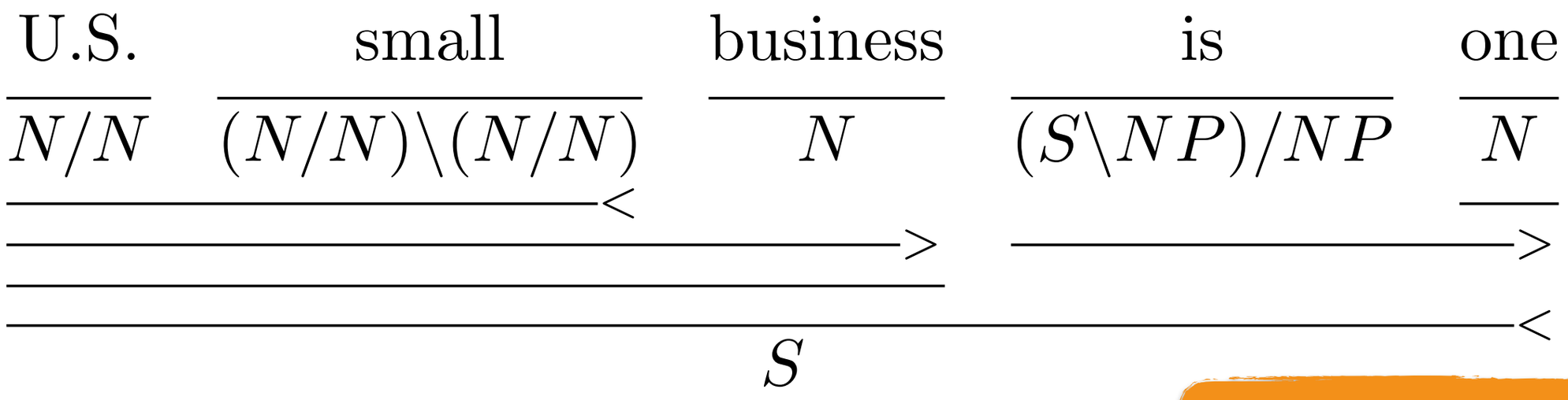






Garden Paths

Incorrect partial parse (syntactically plausible in isolation):



Input sentence:

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

Conclusion

- Combining local and global models enables **exact inference** with **global features**
- Efficient decoding by learning to search
- State of the art for CCG parsing
  - Applicable to other structured prediction tasks