An HDP Model for Inducing Combinatory Categorial Grammars

Yonatan Bisk & Julia Hockenmaier
University of Illinois at Urbana-Champaign

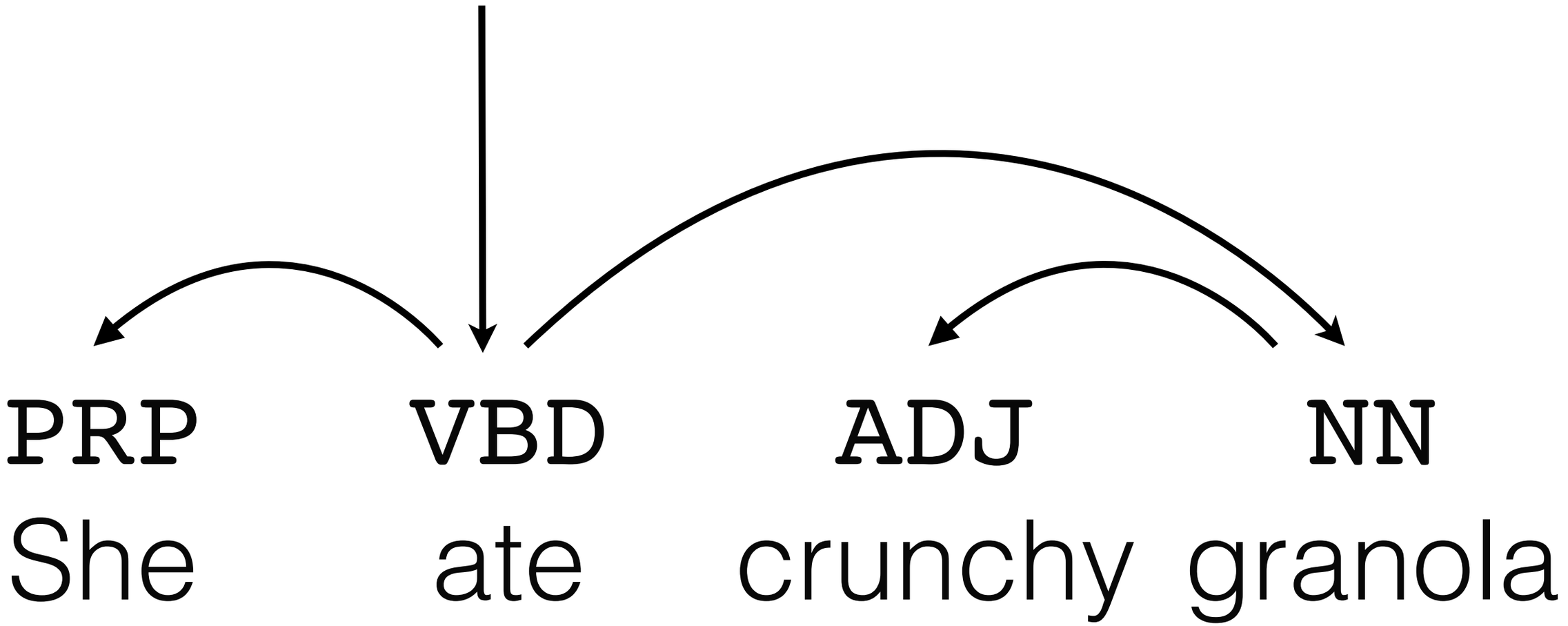
TACL Vol 1(2013):75–88

PRP	VBD	ADJ	NN
She	ate	crunchy	granola

Dependency Grammar Induction

PRP	VBD	ADJ	NN
She	ate	crunchy	granola

Dependency Grammar Induction



Dependency Grammar Induction

Problem for unsupervised Dependency Grammar learner:

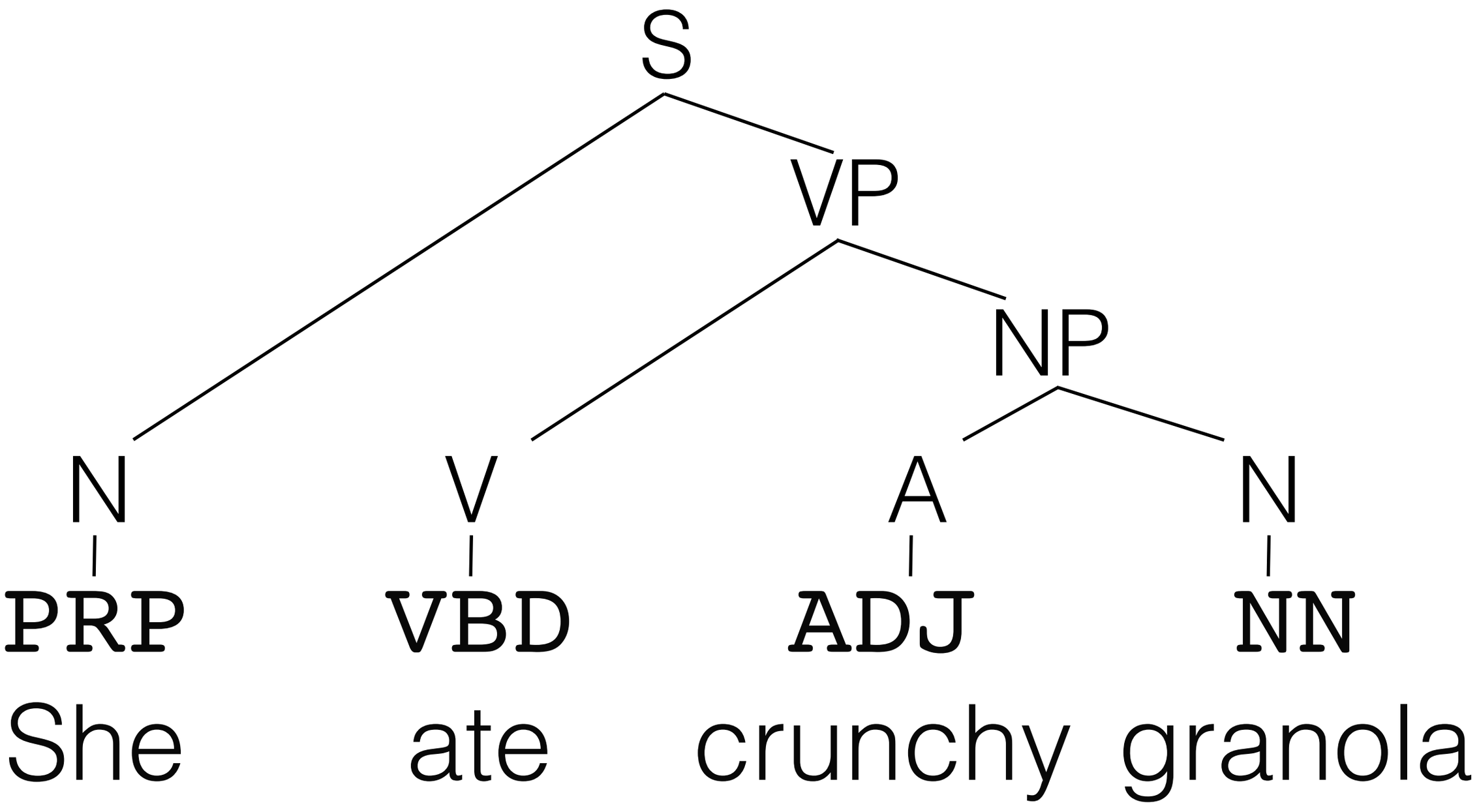
Unlabeled dependencies provide no explicit structure

PI N
Sh hola

CFG Induction

PRP	VBD	ADJ	NN
She	ate	crunchy	granola

CFG Induction

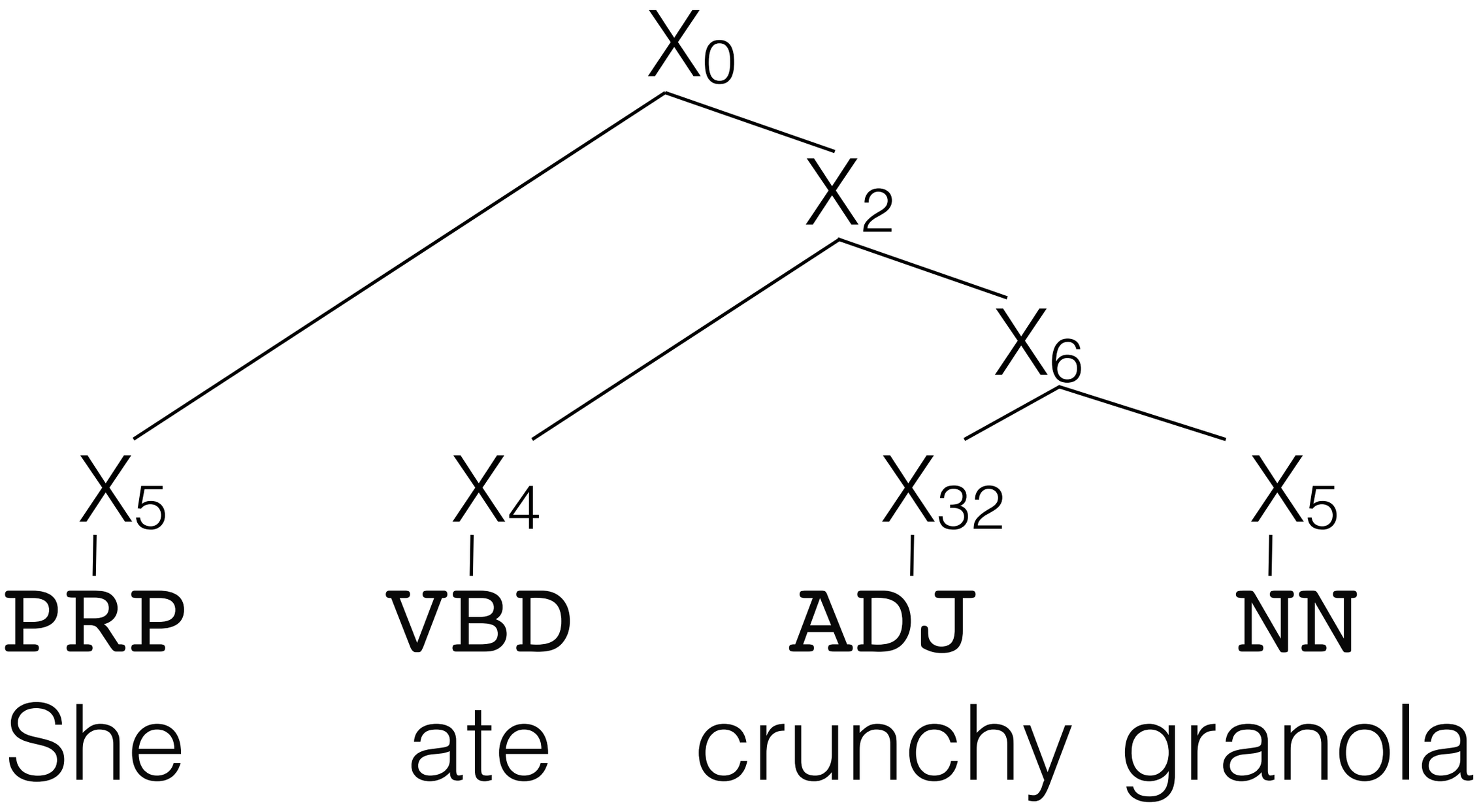


CFG Induction

Problem for unsupervised CFG learner:

CFG symbols and rewrite rules are arbitrary

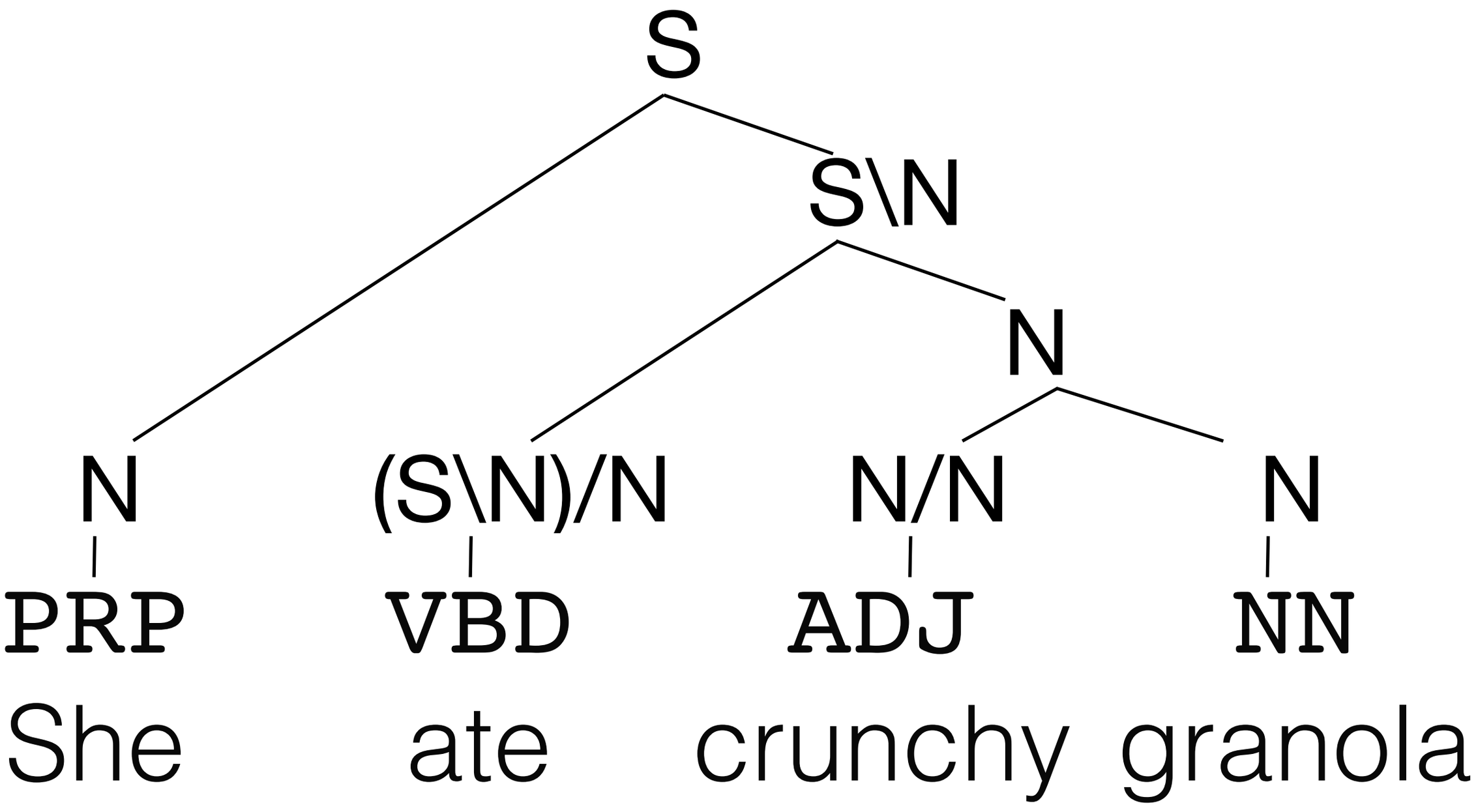
CFG Induction in Practice



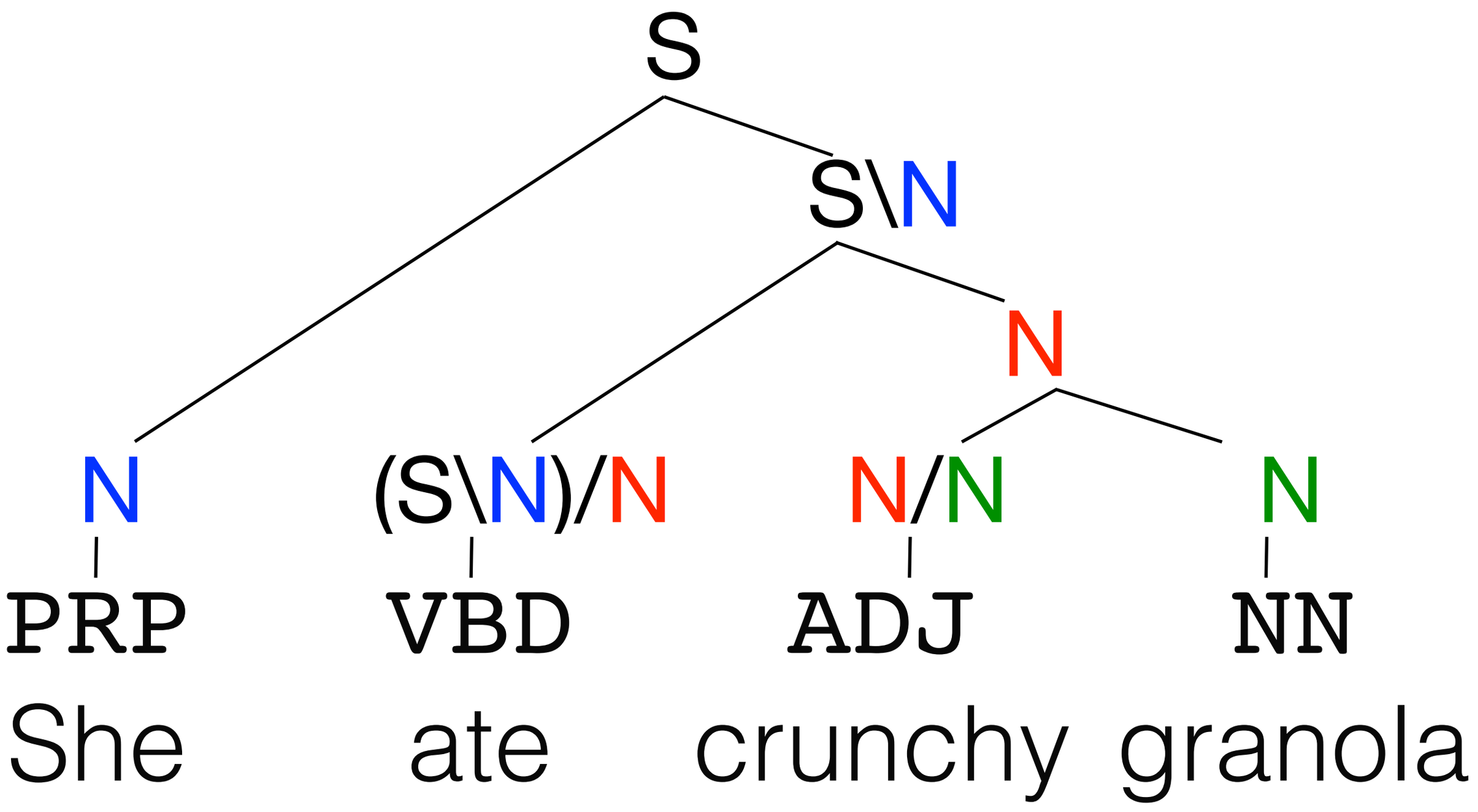
CFG Induction in Practice

What kind of grammatical representation is suitable for unsupervised induction?

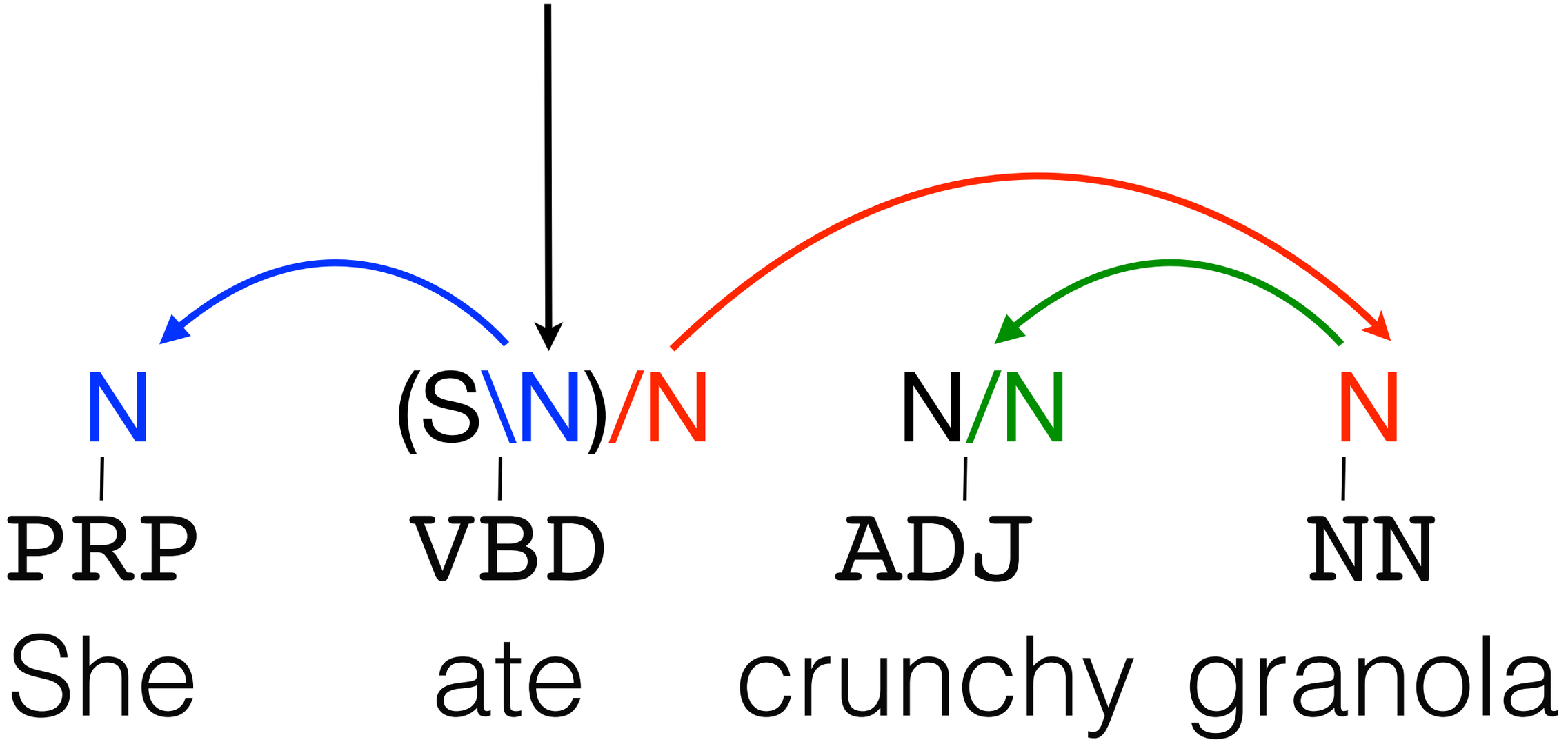
Categorial Grammar Induction



Categorial Grammar Induction



Categorial Grammar Induction



Features of CCG

Features of CCG

- Linguistically motivated symbolic representation:

Features of CCG

- Linguistically motivated symbolic representation:

  CCG captures core dependencies
  CCG captures basic word order

Features of CCG

- Linguistically motivated symbolic representation:

  CCG captures core dependencies
  CCG captures basic word order

- Rules and categories are heavily constrained:

Features of CCG

- Linguistically motivated symbolic representation:

  CCG captures core dependencies
  CCG captures basic word order

- Rules and categories are heavily constrained:

  CCG categories are functions
  CCG rules = function application & composition

Advantages of CCG

- Linguistically motivated symbolic representation:

- Rules and categories are heavily constrained:

Advantages of CCG

- Linguistically motivated symbolic representation:

  CCG is more robust than DG on longer sentences

  CCG returns linguistically interpretable parses

- Rules and categories are heavily constrained:

Advantages of CCG

- Linguistically motivated symbolic representation:

  CCG is more robust than DG on longer sentences

  CCG returns linguistically interpretable parses

- Rules and categories are heavily constrained:

  CCG has a simpler probability model than CFGs

  CCG allows fast variational inference

Categorial Grammar

CCG categories are functions

CCG categories are functions

CCG has **two atomic categories:**

CCG categories are functions

CCG has **two atomic categories:**

$$S, N$$

CCG categories are functions

CCG has **two atomic categories:**

S, N

All other CCG categories are **functions:**

CCG categories are functions

CCG has **two atomic categories:**

S, N

All other CCG categories are **functions:**

CCG categories are functions

CCG has **two atomic categories:**

S, N

All other CCG categories are **functions:**

S

Result

CCG categories are functions

CCG has **two atomic categories:**

S, N

All other CCG categories are **functions:**

S Result

N Argument

CCG categories are functions

CCG has **two atomic categories:**

S, N

All other CCG categories are **functions:**

S / N

Result Dir. Argument

Rules: Function application

Rules: Function application

S/N

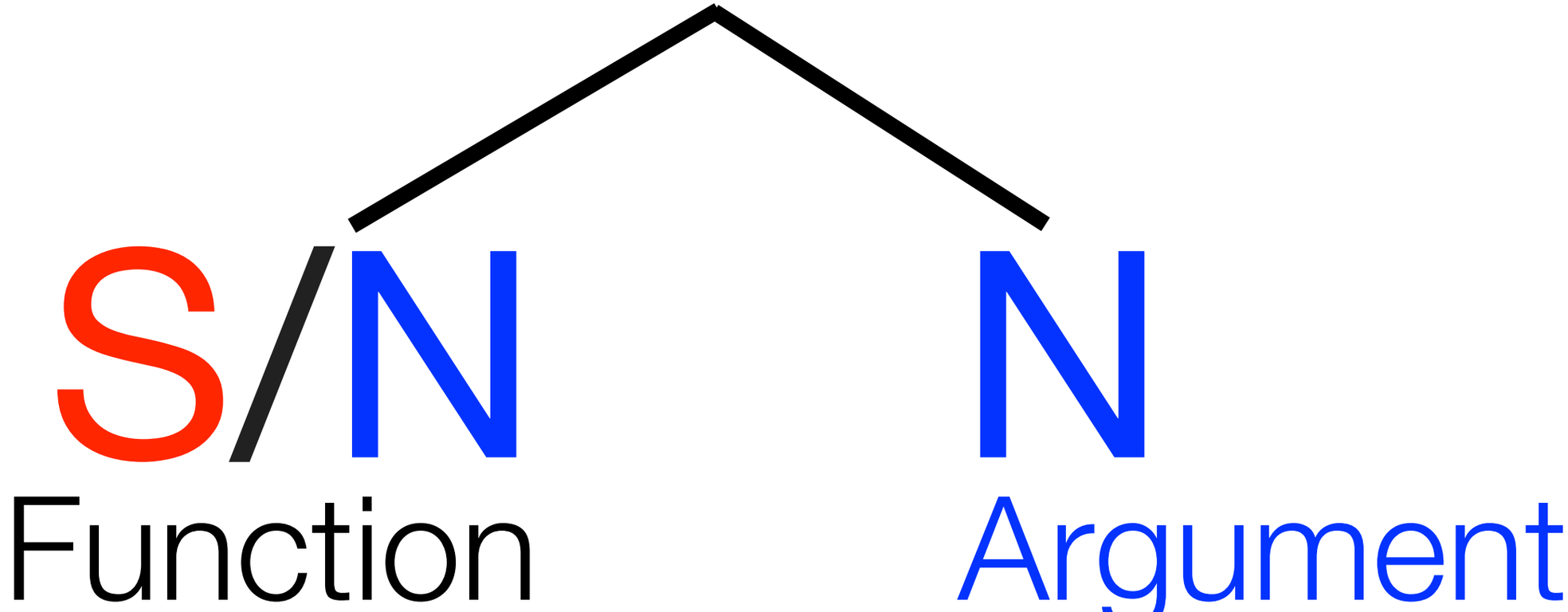
Function

Rules: Function application

Rules: Function application

Result

S



Rules: Function application

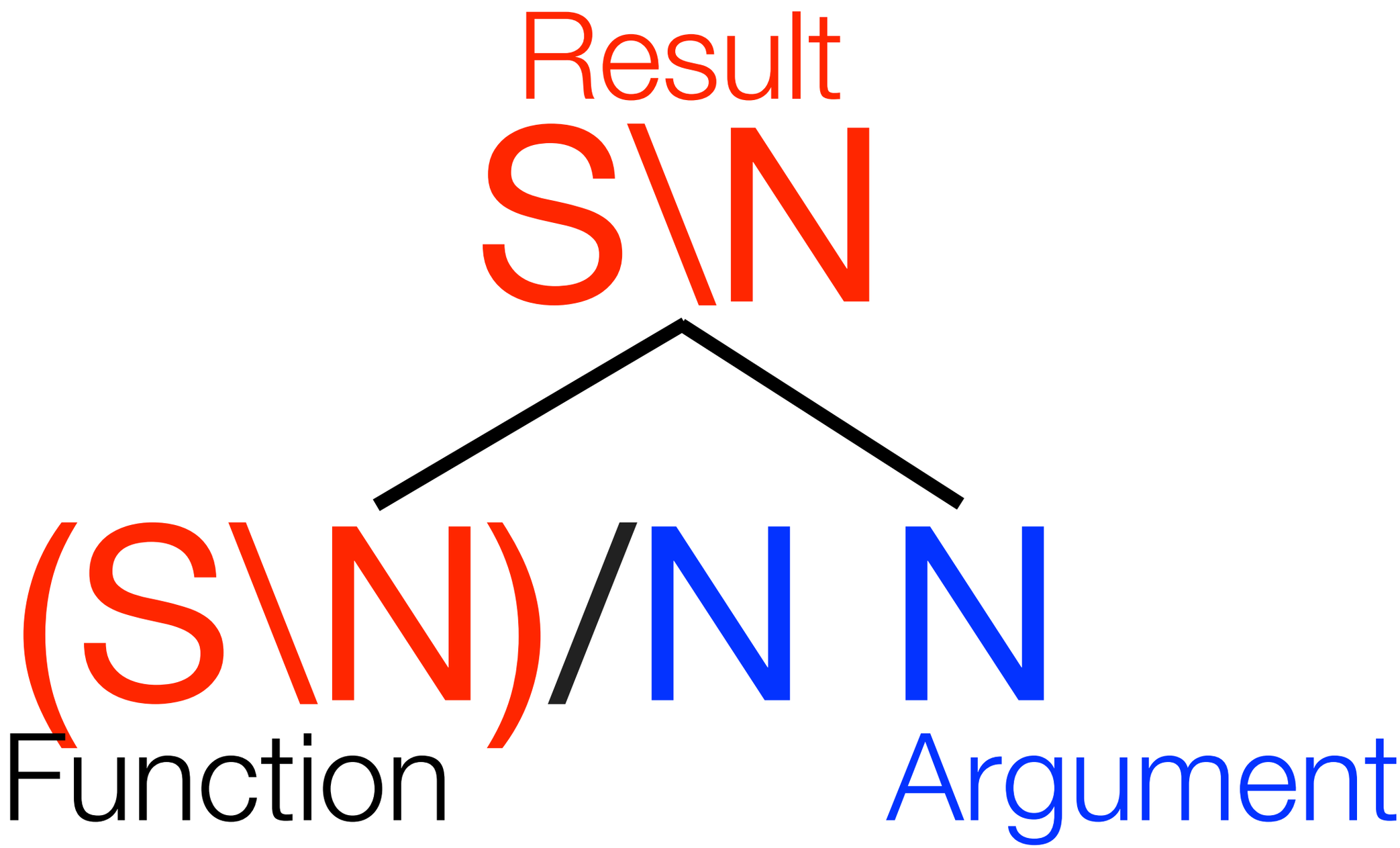
Rules: Function application

(S\N)/N

Function

Rules: Function application

Rules: Function application



Inducing CCGs

Bisk & Hockenmaier,
AAAI 2012

Seed knowledge: Atoms

Seed knowledge: Atoms

Atomic CCG category	Part-of-speech tag class

Seed knowledge: Atoms

Atomic CCG category	Part-of-speech tag class
S	`Verb`

Seed knowledge: Atoms

Atomic CCG category	Part-of-speech tag class
S	`Verb`
N	`Det, Noun, Pron, Num`

Seed knowledge: Atoms

Atomic CCG category	Part-of-speech tag class
S	`Verb`
N	`Det, Noun, Pron, Num`
conj	`Conj`

Inducing complex categories

The man ate quickly

N S

Inducing complex categories

The man ate quickly

N S

S\N

Inducing complex categories

The	man	ate	quickly
?	N	S	
		S\N	

Inducing complex categories

The	man	ate	quickly
?	N	S	?
		S\N	

Inducing complex categories

The	man	ate	quickly
	N	S	?
		S\N	

Inducing complex categories

The	man	ate	quickly
N/N	N	S	?

S\N

Inducing complex categories

The	man	ate	quickly
N/N	N	S	
		S\N	

Inducing complex categories

The	man	ate	quickly
N/N	N	S	S\S

S\N

Inducing complex categories

The	man	ate	quickly
N/N	N	S	S\S

S/S S\N

N\N

...

An HDP Model for CCG

Hierarchical Dirichlet Process

Hierarchical Dirichlet Process

Nonparametric Bayesian model

Hierarchical Dirichlet Process

Nonparametric Bayesian model

We do not need to fix the category inventory in advance

Hierarchical Dirichlet Process

Nonparametric Bayesian model

We do not need to fix the category inventory in advance

Hierarchical model

Hierarchical Dirichlet Process

Nonparametric Bayesian model

We do not need to fix the category inventory in advance

Hierarchical model

All distributions share a common base

Hierarchical Dirichlet Process

Nonparametric Bayesian model

We do not need to fix the category inventory in advance

Hierarchical model

All distributions share a common base

Parameter tying (smoothing)

HDPs for CFGs

Liang et al. 2009

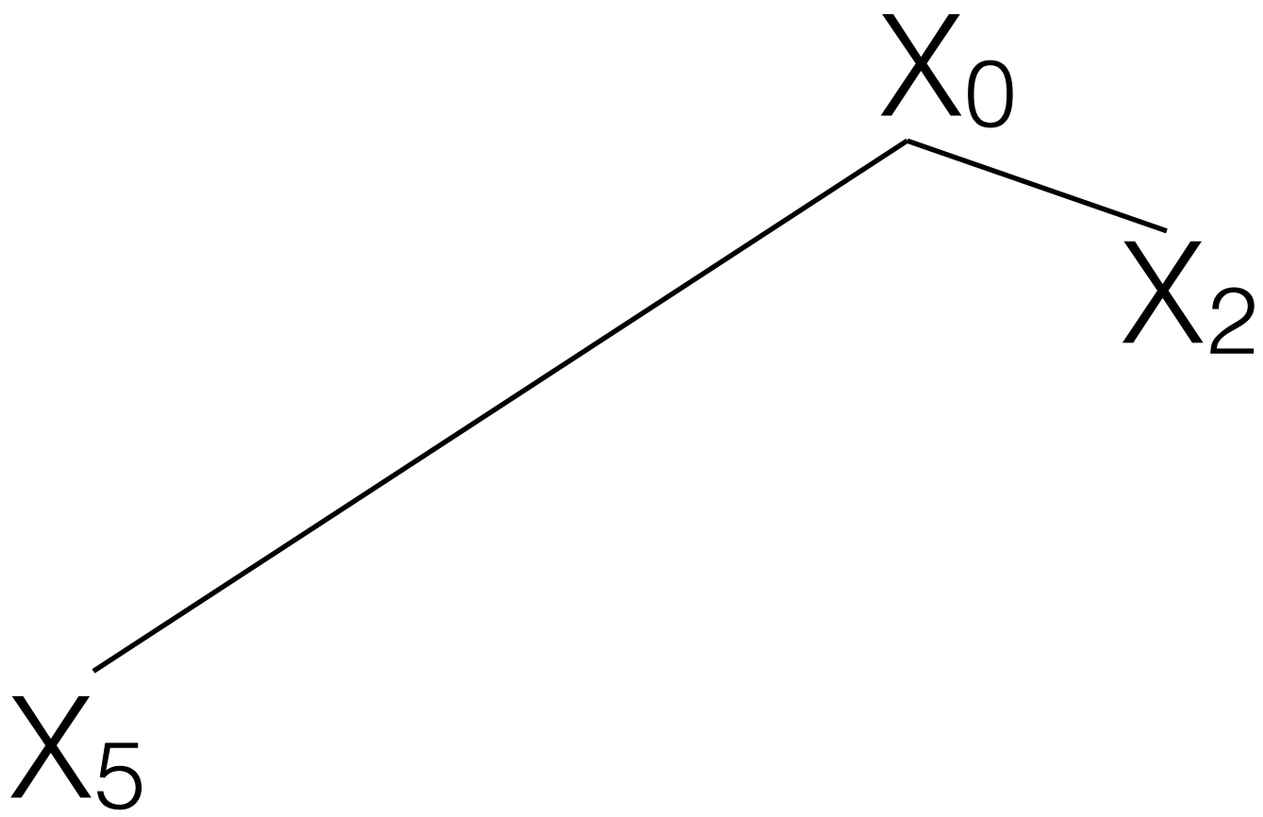
HDPs for CFGs

Liang et al. 2009

X_0

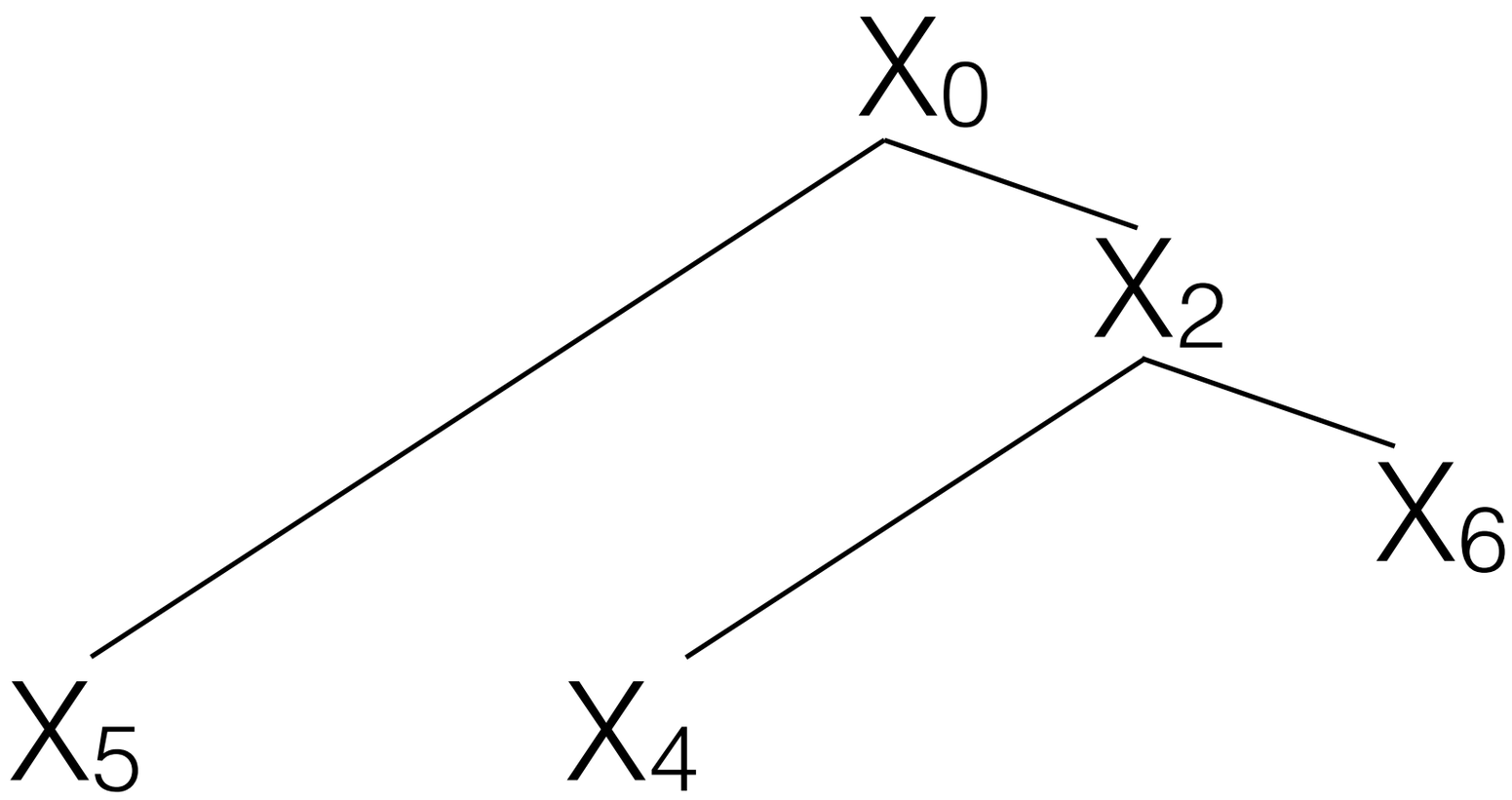
HDPs for CFGs

Liang et al. 2009



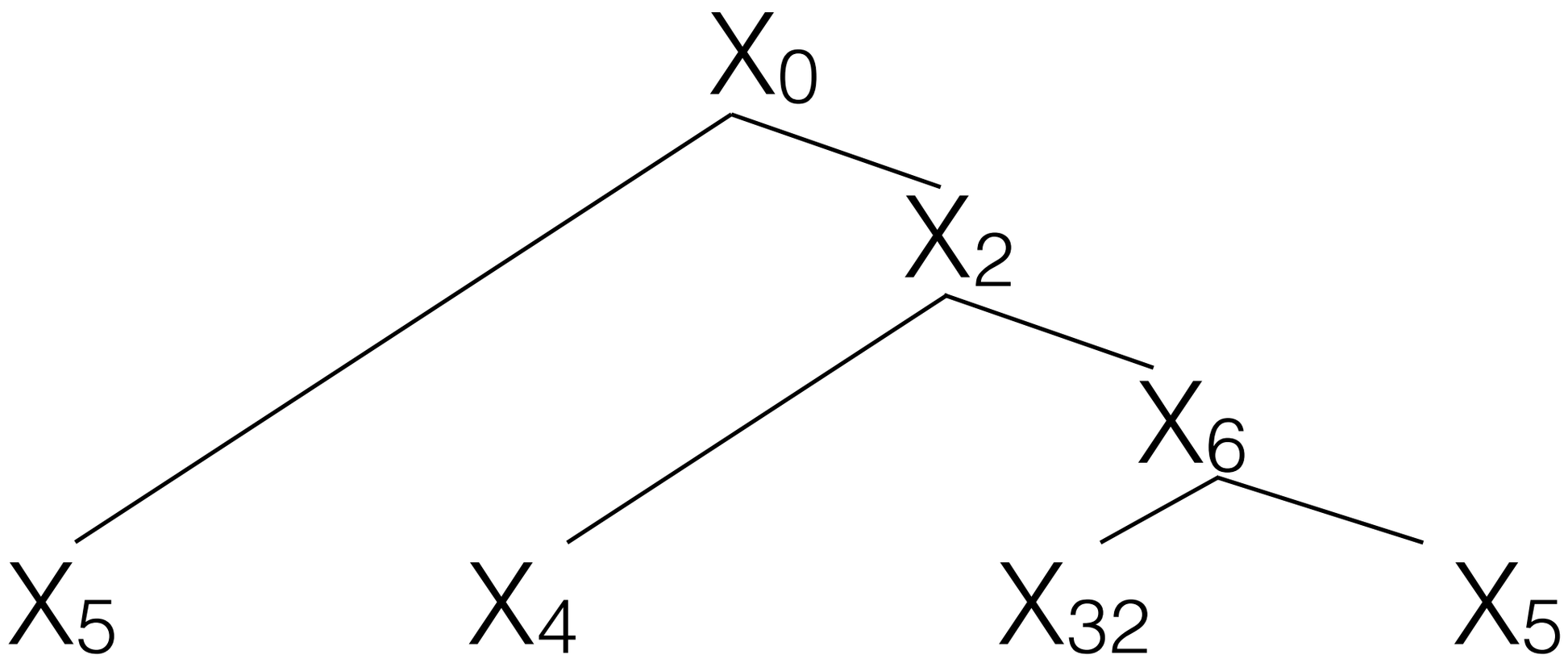
HDPs for CFGs

Liang et al. 2009



HDPs for CFGs

Liang et al. 2009



Parameters for $X_i \rightarrow X_j \ X_k$

Parameters for $X_i \rightarrow X_j \ X_k$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	...
X_1										
X_2										
X_3										
X_4										
X_5										
X_6										
X_7										
X_8										
X_9										
...										

Parameters for $X_i \rightarrow X_j \ X_k$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	...
X_1	?	?	?	?	?	?	?	?	?	?
X_2										
X_3										
X_4										
X_5										
X_6										
X_7										
X_8										
X_9										
...										

Parameters for $X_i \rightarrow X_j \ X_k$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	...
X_1	?	?	?	?	?	?	?	?	?	?
X_2	?	?	?	?	?	?	?	?	?	?
X_3										
X_4										
X_5										
X_6										
X_7										
X_8										
X_9										
...										

Parameters for $X_i \rightarrow X_j \ X_k$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	...
X_1	?	?	?	?	?	?	?	?	?	?
X_2	?	?	?	?	?	?	?	?	?	?
X_3	?	?	?	?	?	?	?	?	?	?
X_4	?	?	?	?	?	?	?	?	?	?
X_5	?	?	?	?	?	?	?	?	?	?
X_6	?	?	?	?	?	?	?	?	?	?
X_7	?	?	?	?	?	?	?	?	?	?
X_8	?	?	?	?	?	?	?	?	?	?
X_9	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for $X_i \rightarrow X_j \; X_k$

Problem for nonparametric PCFG models:

Each LHS nonterminal X_i is allowed a **doubly infinite cross-product** of RHS children X_j, X_k

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S	?	?	?	?	?	?	?	?	?	?
N	?	?	?	?	?	?	?	?	?	?
S/S	?	?	?	?	?	?	?	?	?	?
S\S	?	?	?	?	?	?	?	?	?	?
S/N	?	?	?	?	?	?	?	?	?	?
S\N	?	?	?	?	?	?	?	?	?	?
(S\N)/N	?	?	?	?	?	?	?	?	?	?
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N	?	?	?	?	?	?	?	?	?	?
S/S	?	?	?	?	?	?	?	?	?	?
S\S	?	?	?	?	?	?	?	?	?	?
S/N	?	?	?	?	?	?	?	?	?	?
S\N	?	?	?	?	?	?	?	?	?	?
(S\N)/N	?	?	?	?	?	?	?	?	?	?
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N									?	
S/S	?	?	?	?	?	?	?	?	?	?
S\S	?	?	?	?	?	?	?	?	?	?
S/N	?	?	?	?	?	?	?	?	?	?
S\N	?	?	?	?	?	?	?	?	?	?
(S\N)/N	?	?	?	?	?	?	?	?	?	?
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N									?	
S/S						?				
S\S	?	?	?	?	?	?	?	?	?	?
S/N	?	?	?	?	?	?	?	?	?	?
S\N	?	?	?	?	?	?	?	?	?	?
(S\N)/N	?	?	?	?	?	?	?	?	?	?
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N									?	
S/S						?				
S\S										
S/N	?	?	?	?	?	?	?	?	?	?
S\N	?	?	?	?	?	?	?	?	?	?
(S\N)/N	?	?	?	?	?	?	?	?	?	?
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N									?	
S/S						?				
S\S										
S/N										
S\N	?	?	?	?	?	?	?	?	?	?
(S\N)/N	?	?	?	?	?	?	?	?	?	?
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N									?	
S/S						?				
S\S										
S/N										
S\N				?						
(S\N)/N	?	?	?	?	?	?	?	?	?	?
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N									?	
S/S						?				
S\S										
S/N										
S\N				?						
(S\N)/N		?								
(S\N)\S	?	?	?	?	?	?	?	?	?	?
(S\N)\N	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

Parameters for **S\N → ... ...**

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...
S								?		
N									?	
S/S						?				
S\S										
S/N										
S\N				?						
(S\N)/N		?								
(S\N)\S										
(S\N)\N										
...	?	?	?	?	?	?	?	?	?	?

CCG rules

CCG rules

Parent	Combinator	Left	Right

CCG rules

Parent	Combinator	Left	Right
(S\N)/N	>B_0	((S\N)/N)/Y	Y

CCG rules

Parent	Combinator	Left	Right
(S\N)/N	>B_0	((S\N)/N)/Y	Y
(S\N)/N	>B_1	(S\N)/Y	Y/N

CCG rules

Parent	Combinator	Left	Right
(S\N)/N	>B_0	((S\N)/N)/Y	Y
(S\N)/N	>B_1	(S\N)/Y	Y/N
(S\N)/N	>B_2	S\Y	(Y\N)/N

CCG rules

Parent	Combinator	Left	Right
(S\N)/N	$>B_0$	((S\N)/N)/Y	Y
(S\N)/N	$>B_1$	(S\N)/Y	Y/N
(S\N)/N	$>B_2$	S\Y	(Y\N)/N
(S\N)/N	$<B_0$	Y	((S\N)/N)\Y

CCG rules

Parent	Combinator	Left	Right
(S\N)/N	>B_0	((S\N)/N)/Y	Y
(S\N)/N	>B_1	(S\N)/Y	Y/N
(S\N)/N	>B_2	S\Y	(Y\N)/N
(S\N)/N	<B_0	Y	((S\N)/N)\Y
(S\N)/N	<B_1	Y/N	(S\N)\Y

CCG rules

Parent	Combinator	Left	Right
(S\N)/N	>B_0	((S\N)/N)/Y	Y
(S\N)/N	>B_1	(S\N)/Y	Y/N
(S\N)/N	>B_2	S\Y	(Y\N)/N
(S\N)/N	<B_0	Y	((S\N)/N)\Y
(S\N)/N	<B_1	Y/N	(S\N)\Y
(S\N)/N	<B_2	(Y\N)/N	S\Y

CCG rules

CCG rules

Parent
(S\N)/N
(S\N)/N
(S\N)/N
(S\N)/N
(S\N)/N
(S\N)/N

CCG rules

Parent	Y
(S\N)/N	S
(S\N)/N	S
(S\N)/N	S
(S\N)/N	S
(S\N)/N	S
(S\N)/N	S

CCG rules

Parent	Y	Combinator
(S\N)/N	S	$>B_0$
(S\N)/N	S	$>B_1$
(S\N)/N	S	$>B_2$
(S\N)/N	S	$<B_0$
(S\N)/N	S	$<B_1$
(S\N)/N	S	$<B_2$

CCG rules

Parent	Y	Combinator	Left	Right
(S\N)/N	S	>B_0	((S\N)/N)/S	S
(S\N)/N	S	>B_1	(S\N)/S	S/N
(S\N)/N	S	>B_2	S\S	(S\N)/N
(S\N)/N	S	<B_0	S	((S\N)/N)\S
(S\N)/N	S	<B_1	S/N	(S\N)\S
(S\N)/N	S	<B_2	(S\N)/N	S\S

CCG rules

Pa	Right
(S	S
(S	S/N
(S	S\N)/N
(S	N)/N)\S
(S	S\N)\S
(S	S\S

CCG rules are heavily constrained:

For a **given parent** category, the **Y category** and **combinator** determine both children

HDPs for CCGs

HDPs for CCGs

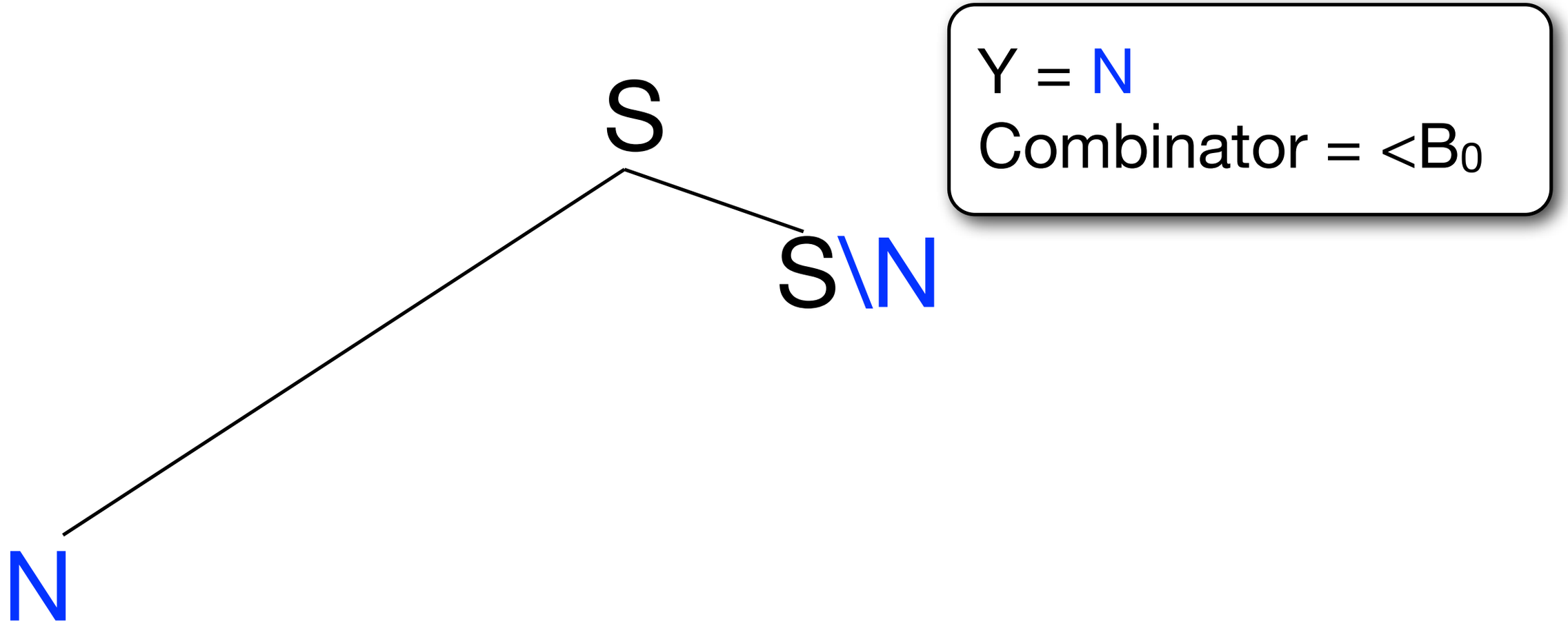
s

HDPs for CCGs

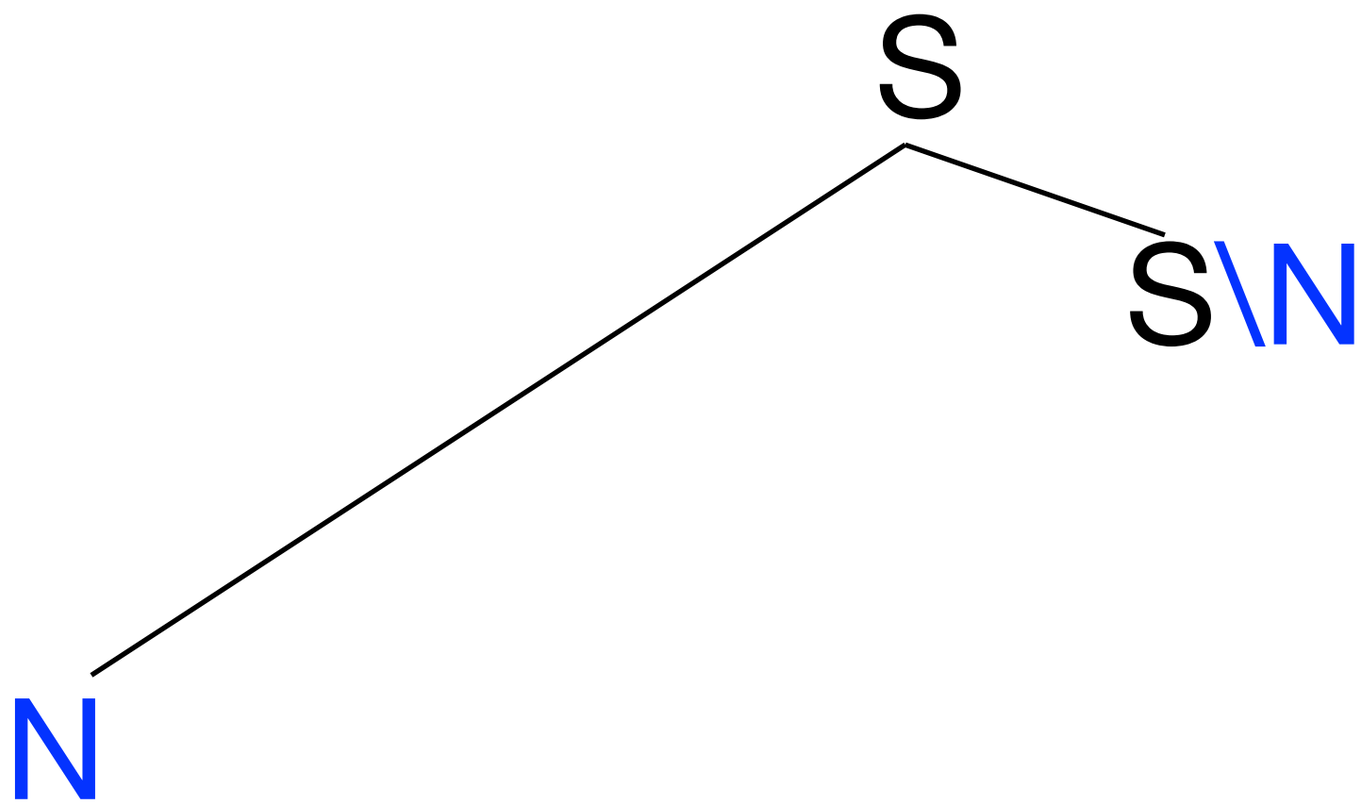
S

Y = N
Combinator = <B_0

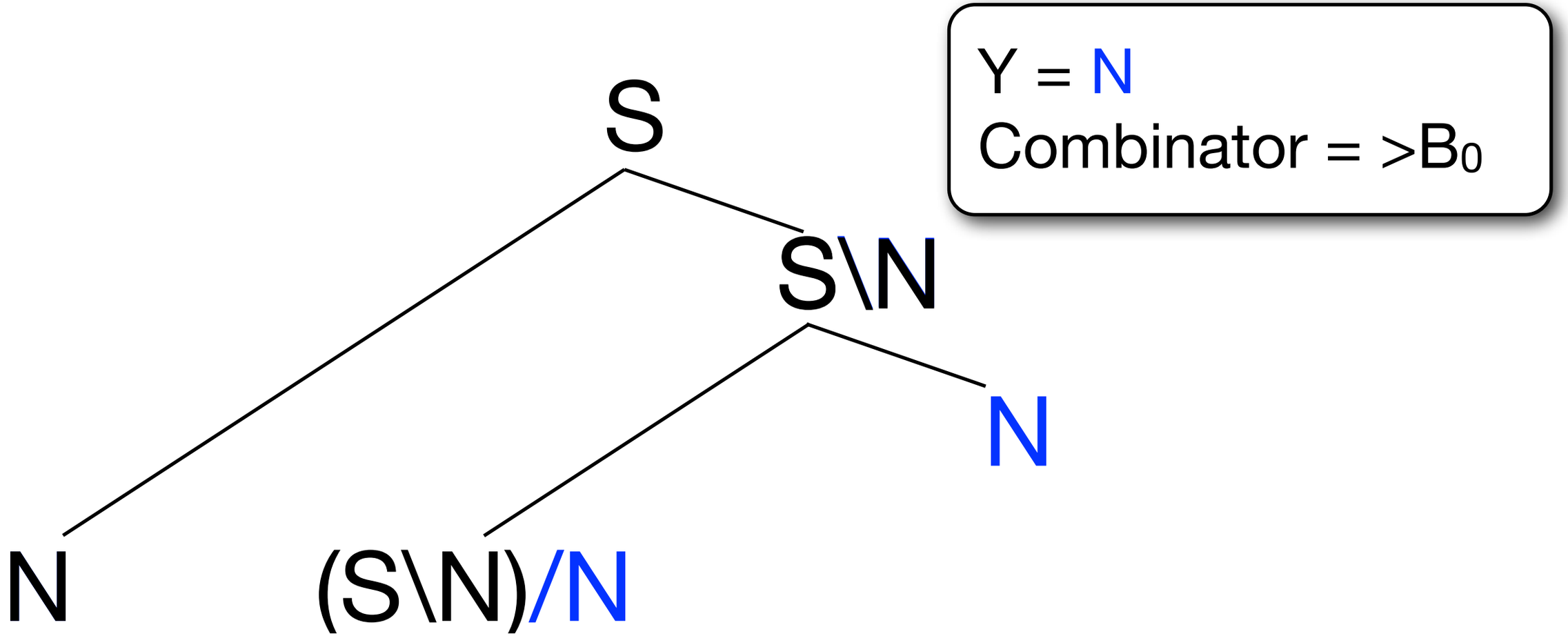
HDPs for CCGs



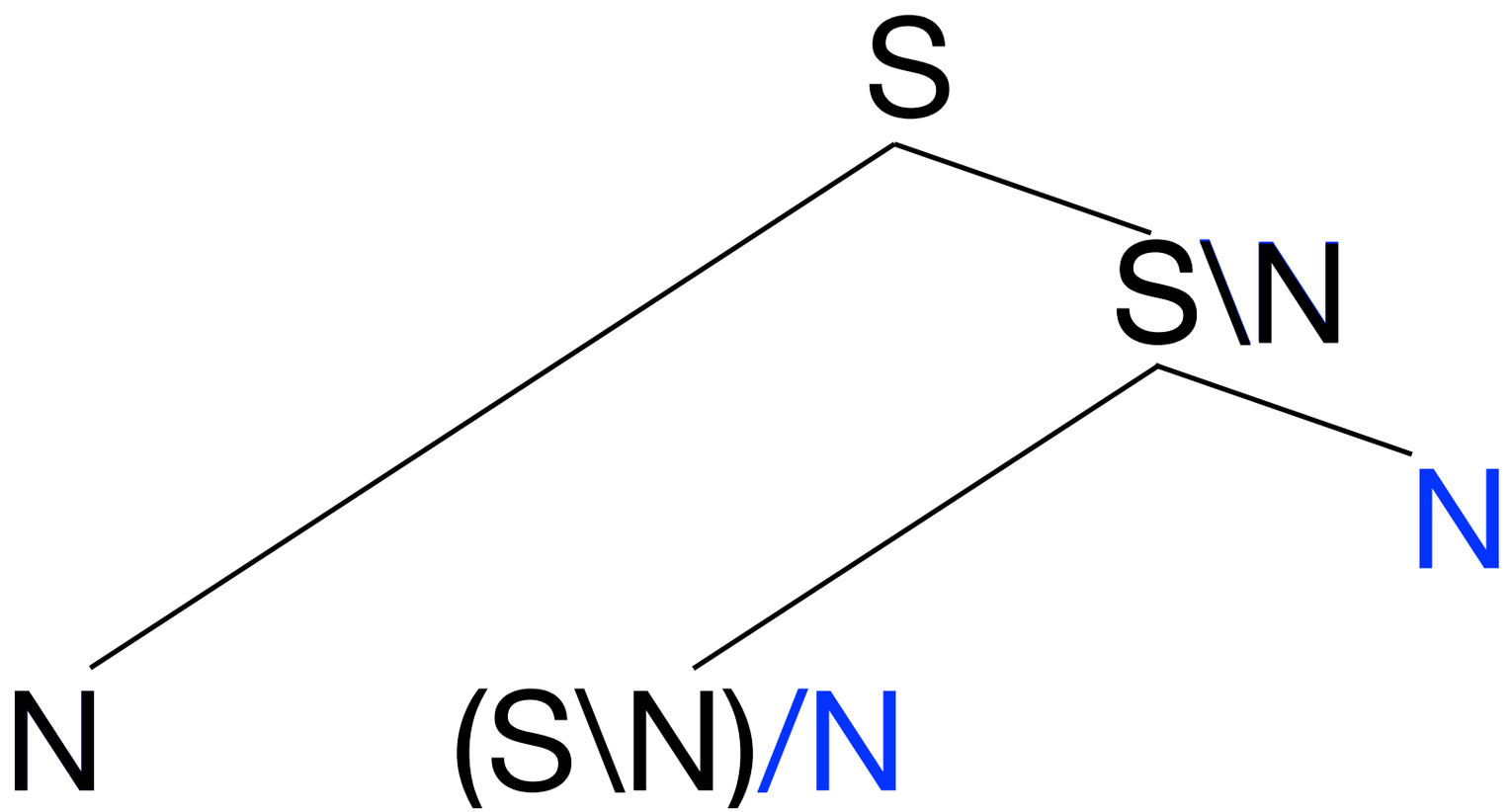
HDPs for CCGs



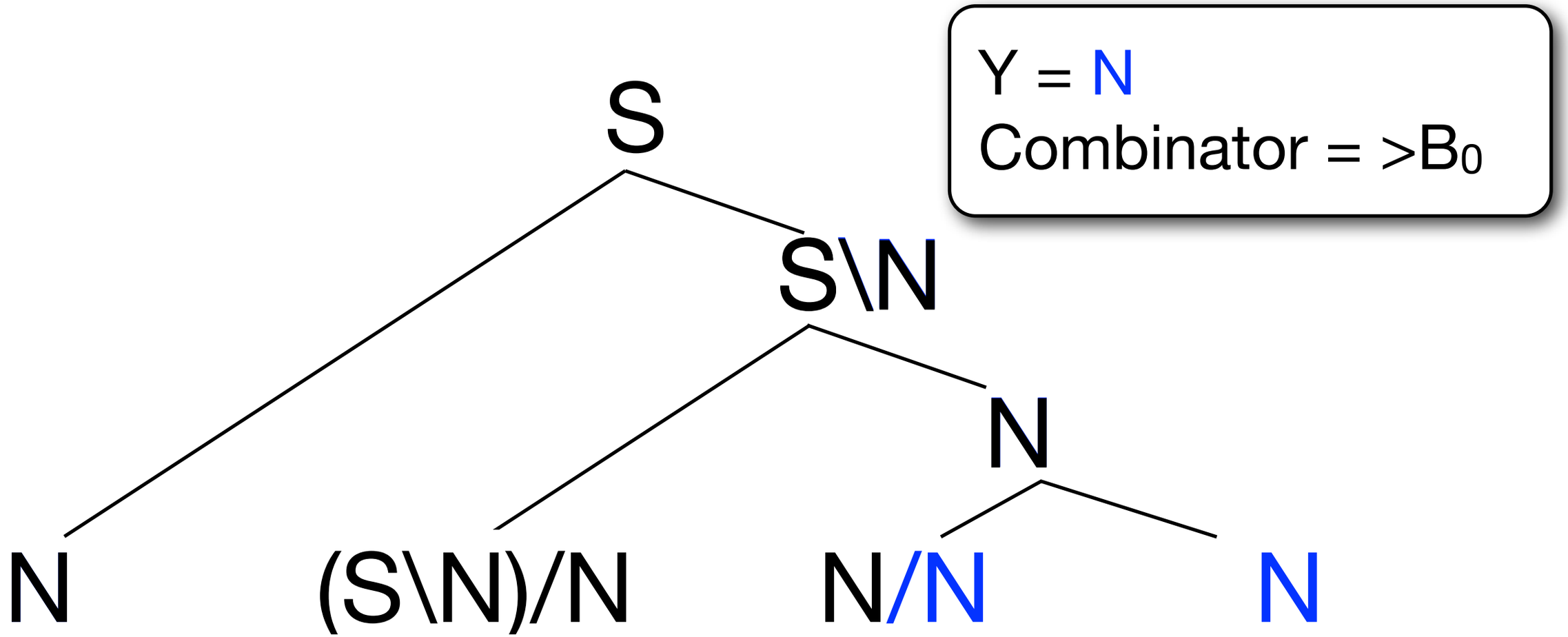
HDPs for CCGs



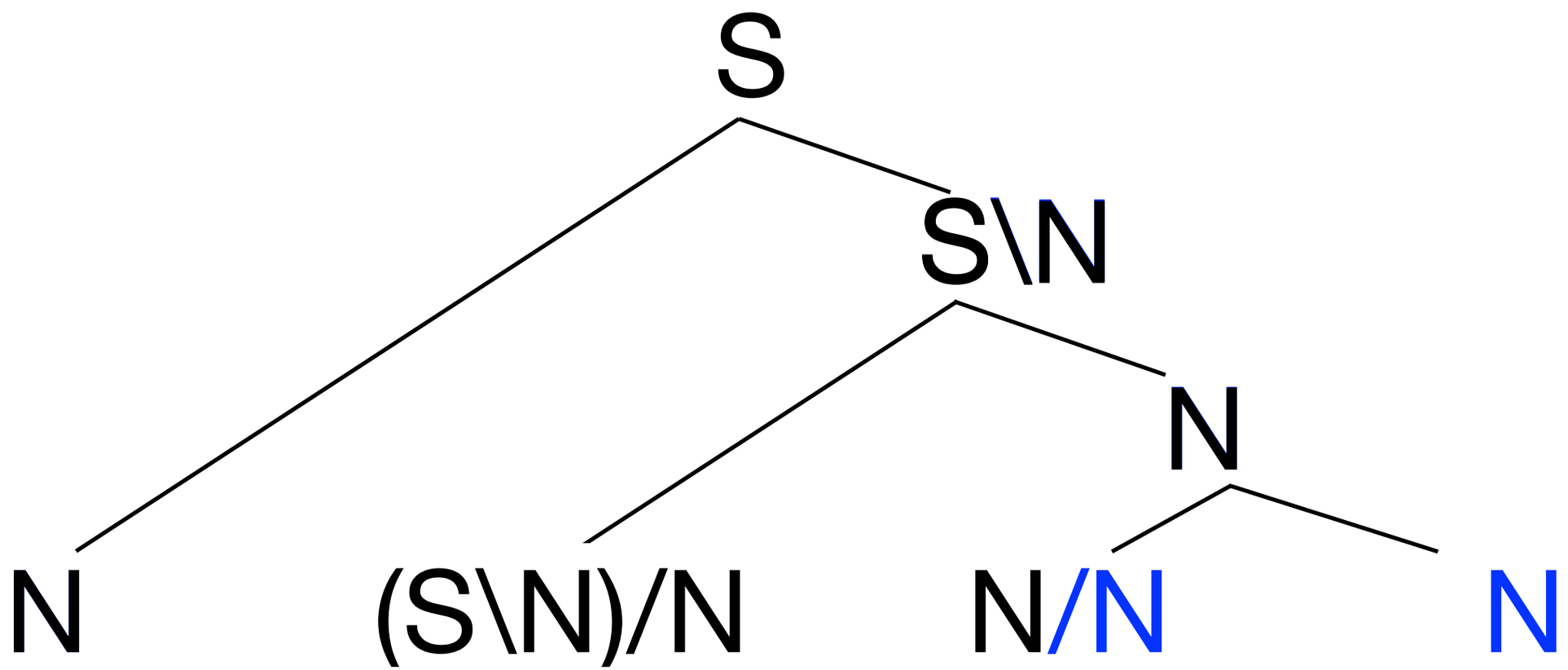
HDPs for CCGs



HDPs for CCGs



HDPs for CCGs



HDP-CFG vs HDP-CCG

HDP-CFG vs HDP-CCG

CFG: **doubly infinite** $P(X_i \rightarrow X_j\ X_k | X_i)$

HDP-CFG vs HDP-CCG

CFG: **doubly infinite** $P(X_i \rightarrow X_j\ X_k | X_i)$

	X1	X2	X3	X4	X5	X6	X7	X8	X9	...
X1	?	?	?	?	?	?	?	?	?	?
X2	?	?	?	?	?	?	?	?	?	?
X3	?	?	?	?	?	?	?	?	?	?
X4	?	?	?	?	?	?	?	?	?	?
X5	?	?	?	?	?	?	?	?	?	?
X6	?	?	?	?	?	?	?	?	?	?
X7	?	?	?	?	?	?	?	?	?	?
X8	?	?	?	?	?	?	?	?	?	?
X9	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

HDP-CFG vs HDP-CCG

CFG: **doubly infinite** $P(X_i \rightarrow X_j\ X_k | X_i)$

	X1	X2	X3	X4	X5	X6	X7	X8	X9	...
X1	?	?	?	?	?	?	?	?	?	?
X2	?	?	?	?	?	?	?	?	?	?
X3	?	?	?	?	?	?	?	?	?	?
X4	?	?	?	?	?	?	?	?	?	?
X5	?	?	?	?	?	?	?	?	?	?
X6	?	?	?	?	?	?	?	?	?	?
X7	?	?	?	?	?	?	?	?	?	?
X8	?	?	?	?	?	?	?	?	?	?
X9	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

CCG: **infinite** $P(Y | X_i)$ and **finite** $P(c | Y, X_i)$

HDP-CFG vs HDP-CCG

CFG: **doubly infinite** $P(X_i \rightarrow X_j\ X_k | X_i)$

	X1	X2	X3	X4	X5	X6	X7	X8	X9	...
X1	?	?	?	?	?	?	?	?	?	?
X2	?	?	?	?	?	?	?	?	?	?
X3	?	?	?	?	?	?	?	?	?	?
X4	?	?	?	?	?	?	?	?	?	?
X5	?	?	?	?	?	?	?	?	?	?
X6	?	?	?	?	?	?	?	?	?	?
X7	?	?	?	?	?	?	?	?	?	?
X8	?	?	?	?	?	?	?	?	?	?
X9	?	?	?	?	?	?	?	?	?	?
...	?	?	?	?	?	?	?	?	?	?

CCG: **infinite** $P(Y | X_i)$ and **finite** $P(c | Y, X_i)$

	S	N	S/S	S\S	S/N	S\N	(S\N)/N	(S\N)\S	(S\N)\N	...

HDP-CFG vs HDP-CCG

CFG: doubly-infinite $P(X \rightarrow Y, Y | Y)$

CCG: $P(X \rightarrow Y, Y | Y, X_i)$

The **HDP-CFG** base measure requires $\beta\beta^T$

The **HDP-CCG** base measure is the standard $\beta \sim \mathbf{GEM}(\alpha)$ (akin to e.g. HDP-HMMs)

Variational EM for HDP-CCGs

Variational EM for HDP-CCGs

Computation parallels Inside-Outside:

Variational EM for HDP-CCGs

Computation parallels Inside-Outside:

$$\mathbf{W}_{\mathrm{P}}(Y) = \Psi(C(P,Y) + \alpha^{P}\beta_{Y}) - \Psi(C(P,*) + \alpha^{P})$$

Variational EM for HDP-CCGs

Computation parallels Inside-Outside:

$$\mathbf{W}_{\mathrm{P}}(Y) = \Psi(C(P,Y) + \alpha^{P}\beta_{Y}) - \Psi(C(P,*) + \alpha^{P})$$

Trivially parallelizeable; efficient

Variational EM for HDP-CCGs

Computation parallels Inside-Outside:

$$\mathrm{W_P}(Y) = \Psi(C(P,Y) + \alpha^P \beta_Y) - \Psi(C(P,*) + \alpha^P)$$

Trivially parallelizeable; efficient

- Experiments in paper:
  1 min – 4 hrs

Results

Impact of longer sentences

Impact of longer sentences

WSJ comparison with Naseem et al. 2010's Universal dependency grammar

Impact of longer sentences

	Trained and tested on	
	≤ 10	≤ 20

WSJ comparison with Naseem et al. 2010's Universal dependency grammar

Impact of longer sentences

	Trained and tested on	
	≤ 10	≤ 20
Naseem et al.	**71.9**	50.4

WSJ comparison with Naseem et al. 2010's Universal dependency grammar

Impact of longer sentences

	Trained and tested on	
	≤ 10	≤ 20
Naseem et al.	**71.9**	50.4
HDP-CCG	68.2	**64.2**

WSJ comparison with Naseem et al. 2010's Universal dependency grammar

Impact of longer sentences

	Trained and tested on	
	≤ 10	≤ 20
Naseem et al.	**71.9**	50.4
HDP-CCG	68.2	**64.2**

Impact of longer sentences

	Trained and tested on	
	≤ 10	≤ 20
Naseem et al.	**71.9**	50.4
HDP-CCG	68.2	**64.2**

Can long sentences help performance on short sentences?

Impact of longer sentences

	Trained and tested on	
	≤ 10	≤ 20
Naseem et al.	**71.9**	50.4
HDP-CCG	68.2	**64.2**

Can long sentences help performance on short sentences?

Yes! HDP-CCG achieves **71.9** on ≤10 if trained on ≤20

Impact of longer sentences

	Trained and tested on	
	≤ 10	≤ 20
Naseem et al.	**71.9**	50.4
HDP-CCG	68.2	**64.2**

Can long sentences help performance on short sentences?

Yes! HDP-CCG achieves **71.9** on ≤10 if trained on ≤20

Multilingual performance

* Max over all best performing systems (extra data, tuning, etc.)

Multilingual performance

NAACL WILS Shared Task 2012

* Max over all best performing systems (extra data, tuning, etc.)

Multilingual performance

NAACL WILS Shared Task 2012

Average ≤10 accuracy on 10 languages

* Max over all best performing systems (extra data, tuning, etc.)

Multilingual performance

NAACL WILS Shared Task 2012

Average ≤10 accuracy on 10 languages
(Arabic, Danish, Slovene, Swedish, Dutch, Basque, Portuguese, WSJ, CHILDES, Czech)

* Max over all best performing systems (extra data, tuning, etc.)

Multilingual performance

NAACL WILS Shared Task 2012

Average ≤10 accuracy on 10 languages

(Arabic, Danish, Slovene, Swedish, Dutch, Basque, Portuguese, WSJ, CHILDES, Czech)

Dependencies		CCG
Blunsom & Cohn 2010	State of the Art*	Bisk & Hockenmaier 2012
55.2	62.3	54.2

* Max over all best performing systems (extra data, tuning, etc.)

Multilingual performance

NAACL WILS Shared Task 2012

Average ≤10 accuracy on 10 languages

(Arabic, Danish, Slovene, Swedish, Dutch, Basque, Portuguese, WSJ, CHILDES, Czech)

Dependencies		CCG	CCG: new model
Blunsom & Cohn 2010	State of the Art*	Bisk & Hockenmaier 2012	MLE
55.2	62.3	54.2	**50.9**

* Max over all best performing systems (extra data, tuning, etc.)

Multilingual performance

NAACL WILS Shared Task 2012

Average ≤10 accuracy on 10 languages

(Arabic, Danish, Slovene, Swedish, Dutch, Basque, Portuguese, WSJ, CHILDES, Czech)

Dependencies		CCG	CCG: new model	
Blunsom & Cohn 2010	State of the Art*	Bisk & Hockenmaier 2012	MLE	HDP-CCG
55.2	62.3	54.2	**50.9**	**64.5**

* Max over all best performing systems (extra data, tuning, etc.)

Induced Lexicons: Adjectives

English

Adj Obj

Big	Ball
N/N	N

Induced Lexicons: Adjectives

English

Adj Obj

Big	Ball
N/N	N

Arabic

Obj Adj

كرة	كبيرة
N	**N\N**
(ball)	(big)

Induced Lexicons: Verbs

English

S V O

The man	wrote	a letter
N	**(S\N)/N**	N

Induced Lexicons: Verbs

English

S V O

The man	wrote	a letter
N	**(S\N)/N**	N

Child Directed Speech

∅ V O

∅	write	a letter
	S/N	N

Induced Lexicons: Verbs

English

S V O

The man	wrote	a letter
N	**(S\N)/N**	N

Child Directed Speech

∅ V O

∅	write	a letter
	S/N	N

Arabic

V S O

كتب	الرجال	رسالة
(S/N)/N	N	N
(wrote)	(the man)	(a letter)

Induced Lexicons: Adpositions

English

V ADP O

ran	on	beach
(S\N)/N	**(S\S)/N**	N

Induced Lexicons: Adpositions

English

V ADP O

ran	on	beach
(S\N)/N	**(S\S)/N**	N

Japanese

O ADP V

浜	を	走った
N	**(S/S)\N**	(S\N)/N
(beach)	(on)	(ran)

Summary of contributions

Summary of contributions

A new probability model for CCG

Summary of contributions

A new probability model for CCG

- Exploits CCG's functional constraints

Summary of contributions

A new probability model for CCG

- Exploits CCG's functional constraints
- Yields fast variational inference

Summary of contributions

A new probability model for CCG

- Exploits CCG's functional constraints
- Yields fast variational inference

State-of-the-Art accuracy

Summary of contributions

A new probability model for CCG

- Exploits CCG's functional constraints
- Yields fast variational inference

State-of-the-Art accuracy

- Performs well on 15 languages

Summary of contributions

A new probability model for CCG

- Exploits CCG's functional constraints
- Yields fast variational inference

State-of-the-Art accuracy

- Performs well on 15 languages
- Can harness longer sentences

Summary of contributions

A new probability model for CCG

- Exploits CCG's functional constraints
- Yields fast variational inference

State-of-the-Art accuracy

- Performs well on 15 languages
- Can harness longer sentences
- Induces linguistically informative lexicons

Work in progress

Work in progress

- Performance is robust
  beyond context-free CCG fragment

Work in progress

- Performance is robust
  beyond context-free CCG fragment
- Performance improves when
  generating words (not just POS tags)

Work in progress

- Performance is robust
  beyond context-free CCG fragment
- Performance improves when
  generating words (not just POS tags)
- Remove dependence on POS tags

Work in progress

- Performance is robust
  beyond context-free CCG fragment
- Performance improves when
  generating words (not just POS tags)
- Remove dependence on POS tags

Thank you!