

# Weakly-Supervised Grammar-Informed Bayesian CCG Parser Learning

Dan Garrette	UT-Austin
Chris Dyer	CMU
Jason Baldridge	UT-Austin
Noah A. Smith	CMU

# Motivation

Annotating parse trees *by hand*  
is extremely difficult.

# Motivation

Can we learn new parsers cheaply?

# Motivation

Can we learn new parsers cheaply?

(cheaper = less supervision)

# Motivation

When supervision is ***scarce***,  
we have to be ***smarter*** about data.

# Type-Level Supervision

# Type-Level Supervision

- Unannotated text
- Incomplete tag dictionary:  $\text{word} \mapsto \{\text{tags}\}$

# Type-Level Supervision

Used for part-of-speech tagging for 20+ years

[Kupiec, 1992]  
[Merialdo, 1994]



# Type-Level Supervision

Good tagger performance  
even with low supervision

[Ravi & Knight, 2009]  
[Das & Petrov, 2011]  
[Garrette & Baldridge, 2013]  
[Garrette *et al.*, 2013]

# Combinatory Categorial Grammar (CCG)

# CCG

[Steedman, 2000]

[Steedman and Baldrige, 2011]

# CCG

Every word token is associated with a **category**

[Steedman, 2000]

[Steedman and Baldridge, 2011]

# CCG

Every word token is associated with a **category**

Categories **combine** to form categories of larger constituents

[Steedman, 2000]

[Steedman and Baldridge, 2011]

# CCG

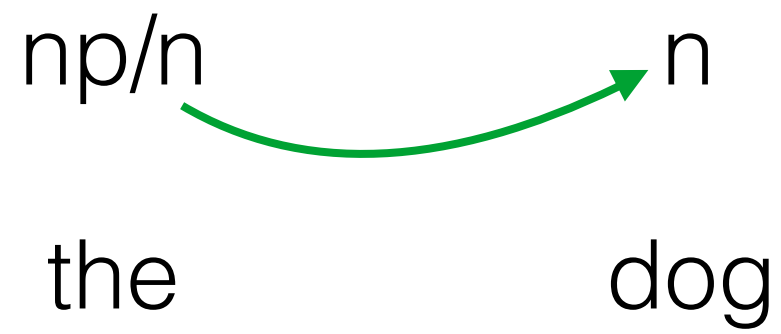
np/n

n

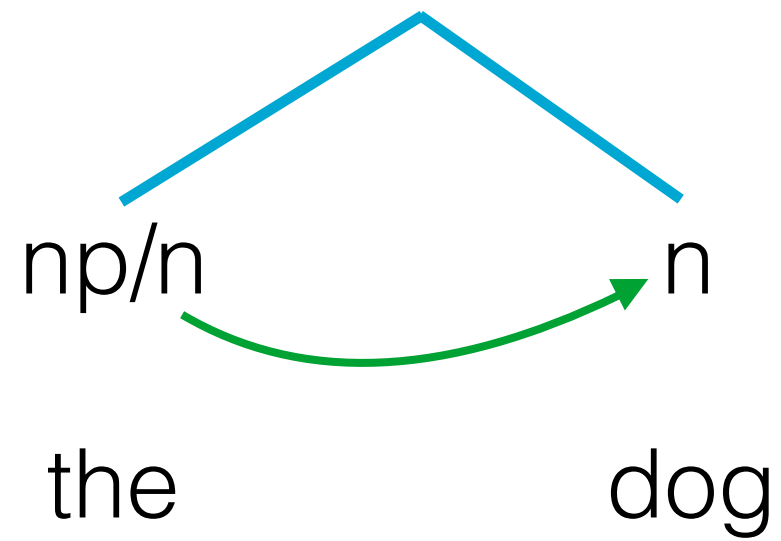
the

dog

# CCG

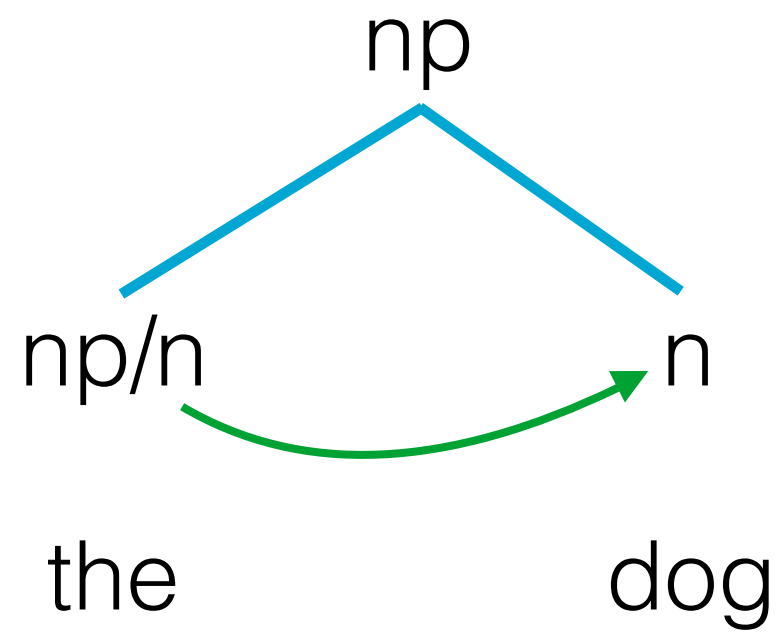


# CCG





# CCG



# CCG

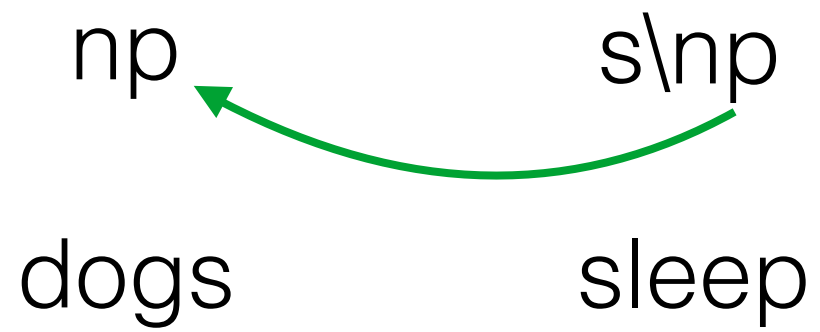
np

s\np

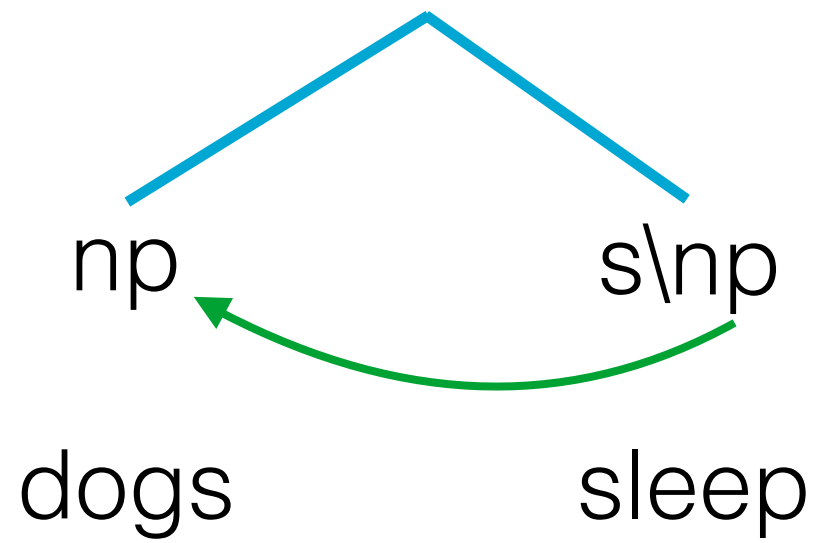
dogs

sleep

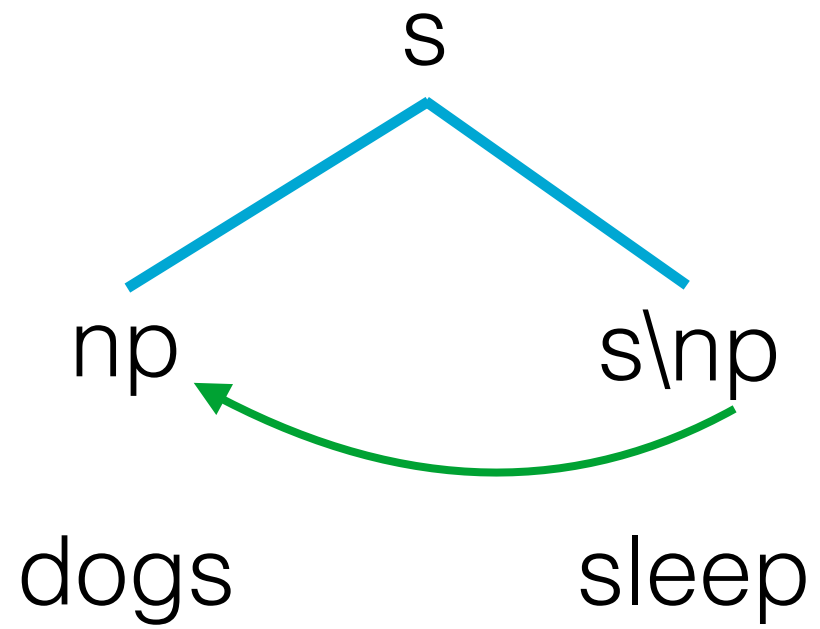
# CCG



# CCG



# CCG



# Type-Supervised CCG

**the**

**lazy**

**dogs**

**wander**

# Type-Supervised CCG

**the**

**lazy**

**dogs**

**wander**

np/n

# Type-Supervised CCG

**the**

**lazy**

**dogs**

**wander**

np/n

n/n

np



# Type-Supervised CCG

**the**

**lazy**

**dogs**

**wander**

np/n

n/n

n

np

np

(s\np)/np

# Type-Supervised CCG

<b>the</b>	<b>lazy</b>	<b>dogs</b>	<b>wander</b>
np/n	n/n	n	n
	np	np	n/n
		(s\np)/np	np/n
			s\np
			...

# CCG Parsing

np/n

the

n/n

lazy

n

dogs

s\np

wander

# CCG Parsing

np/n

the

n/n

lazy

n

dogs

s\np

wander



# CCG Parsing

np/n

the

n/n

lazy

n

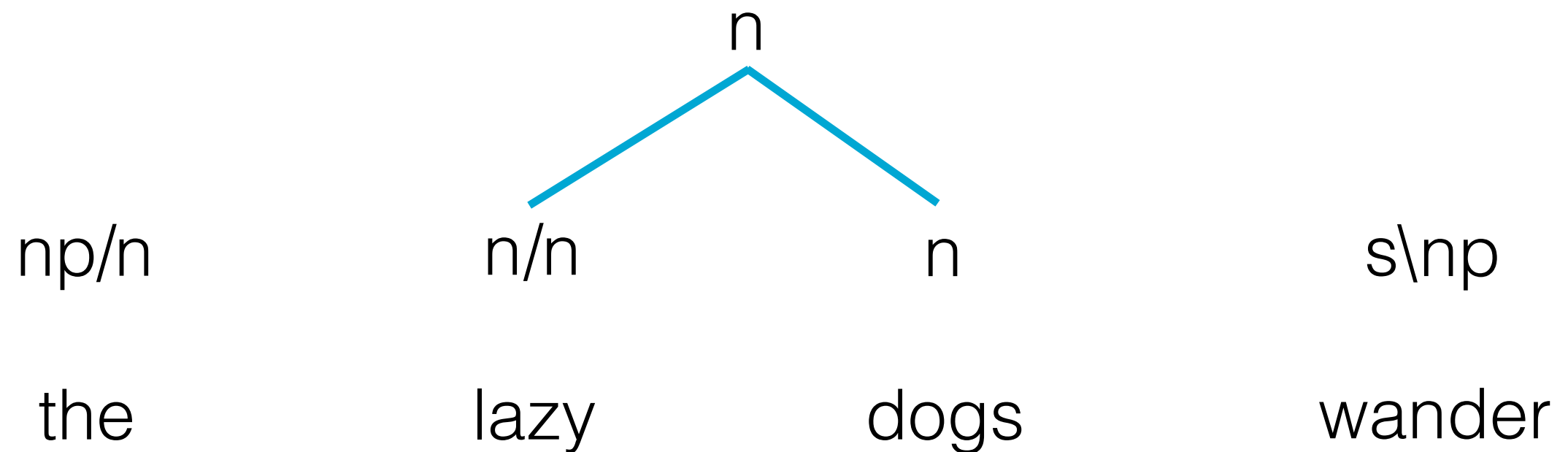
dogs

s\np

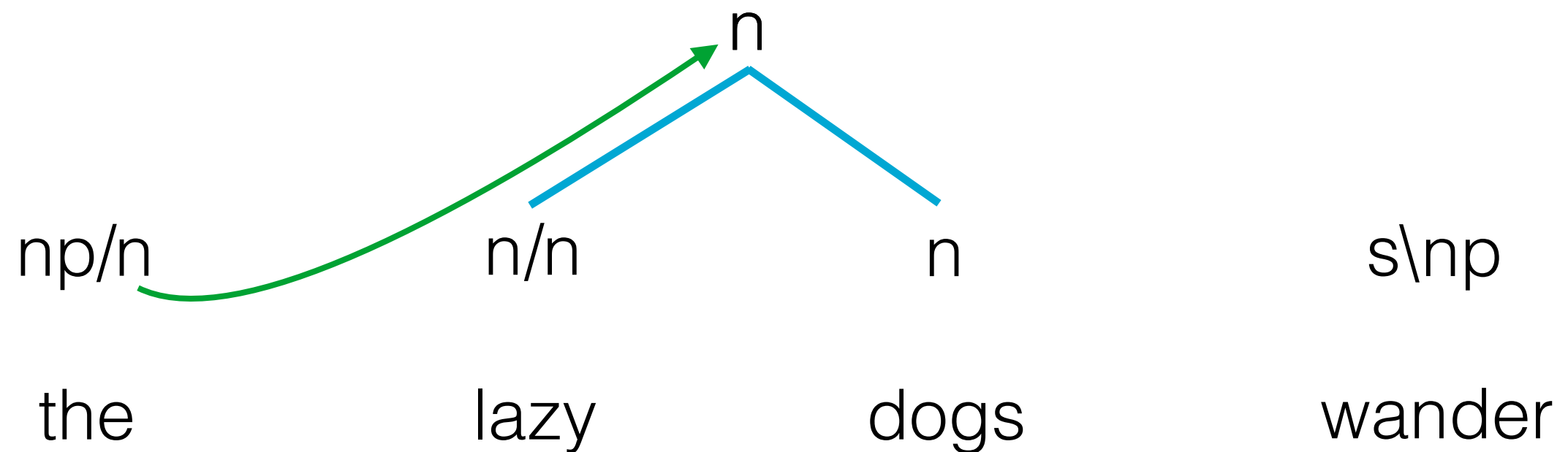
wander



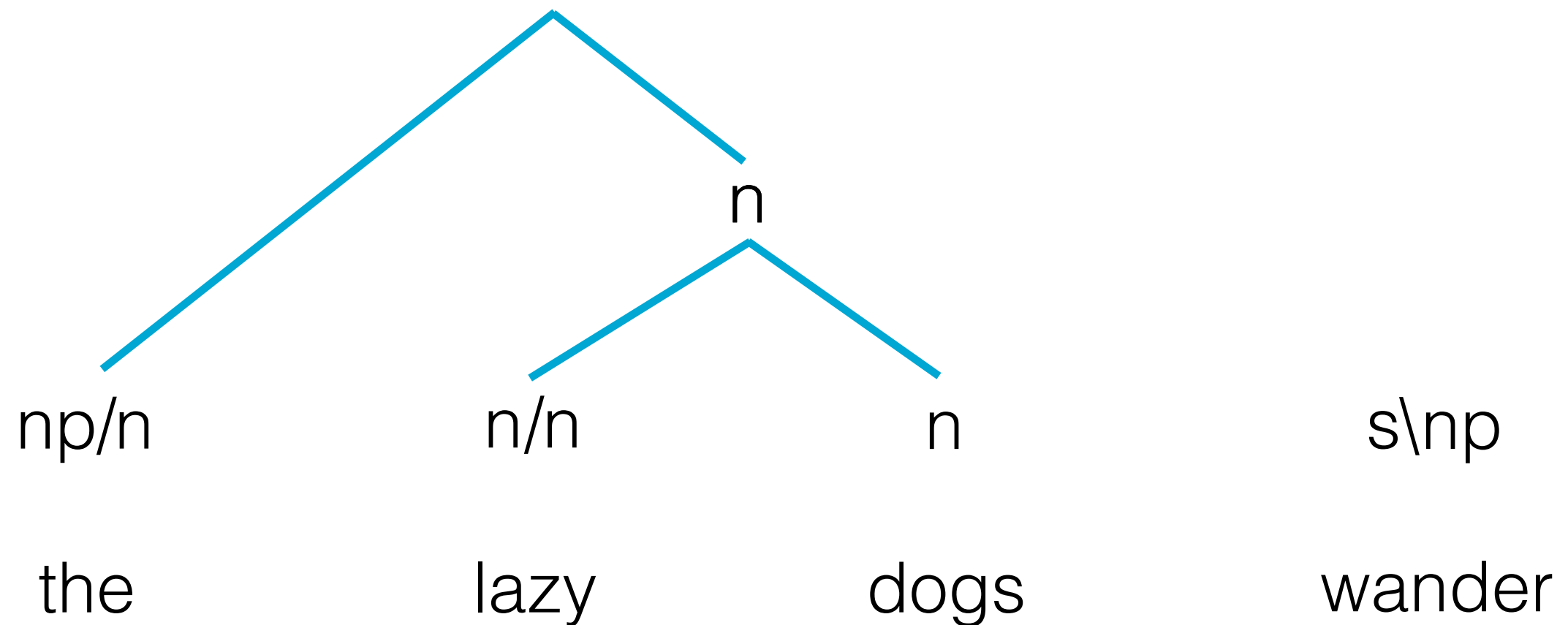
# CCG Parsing



# CCG Parsing

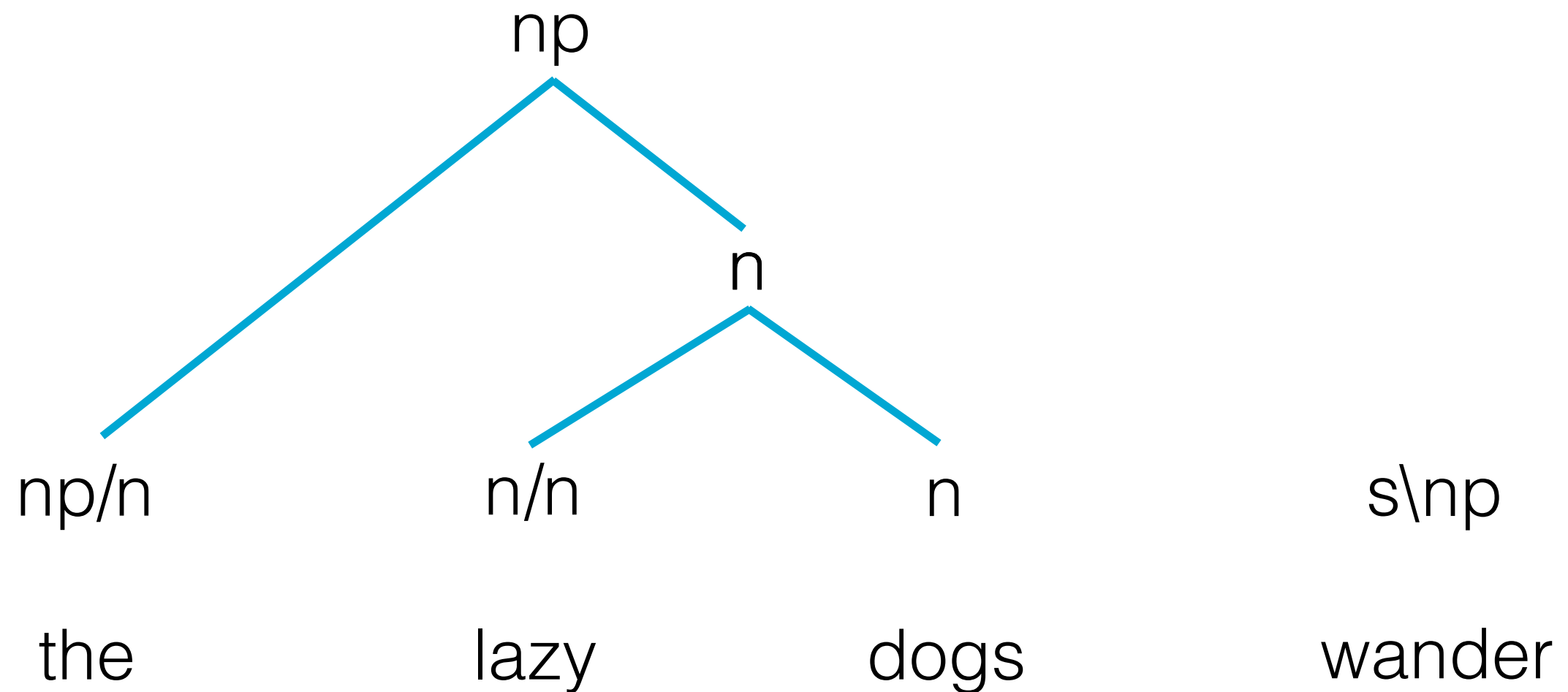


# CCG Parsing

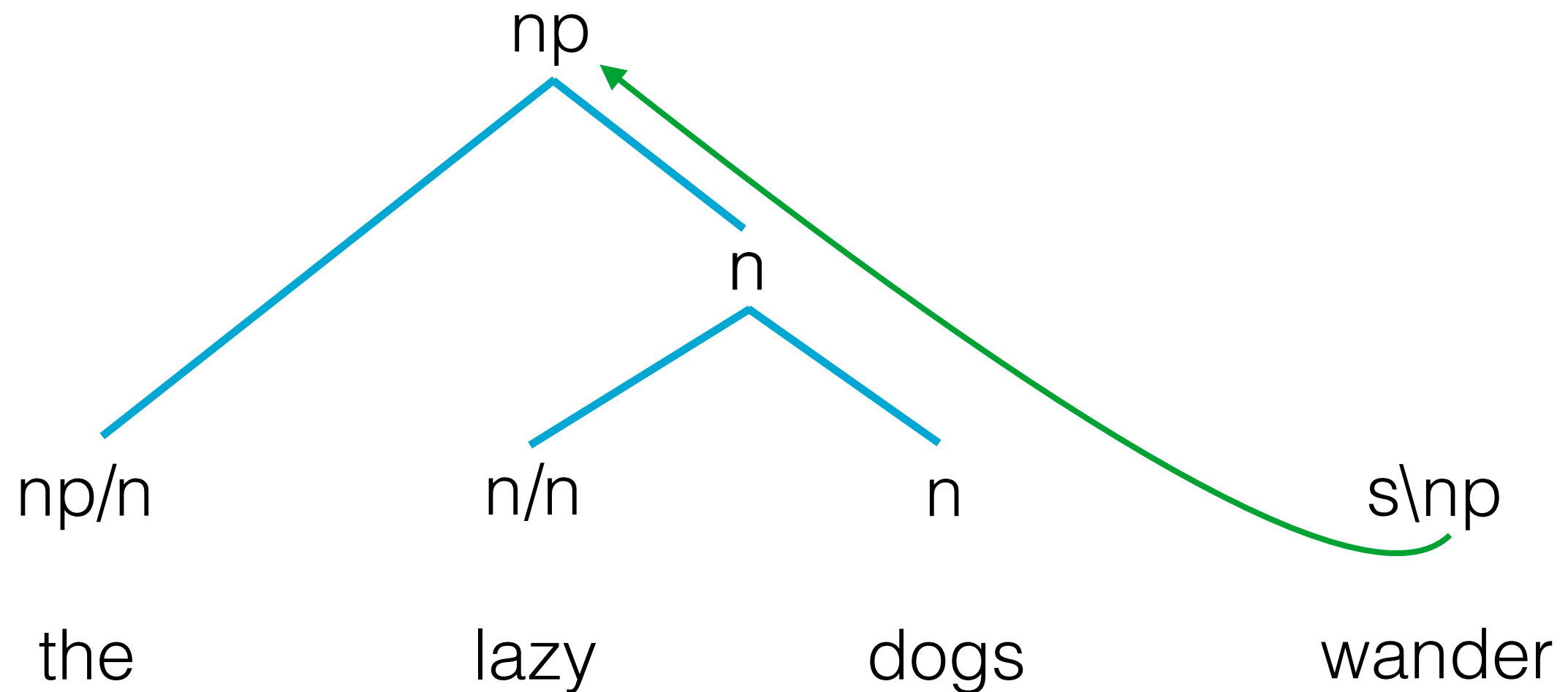




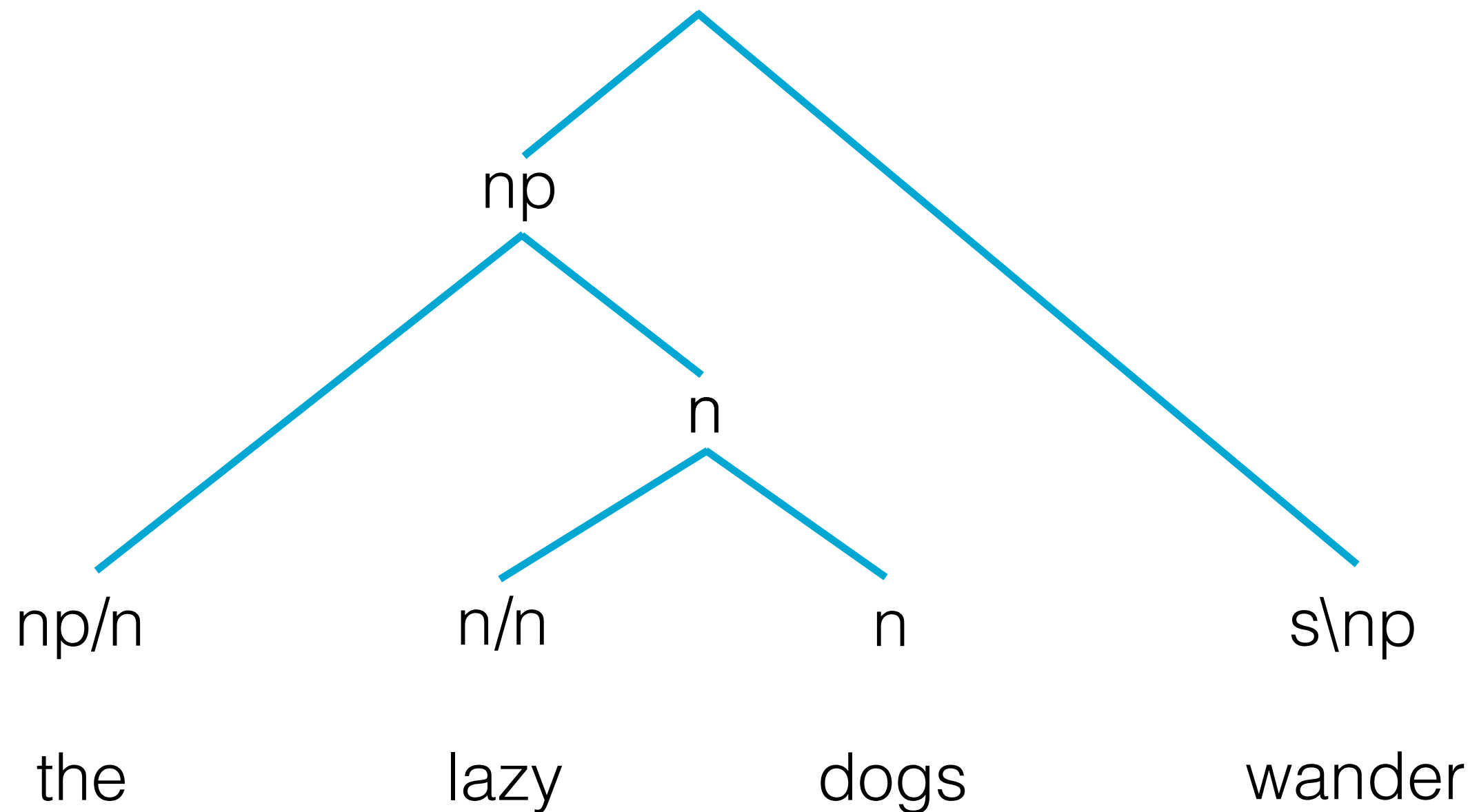
# CCG Parsing



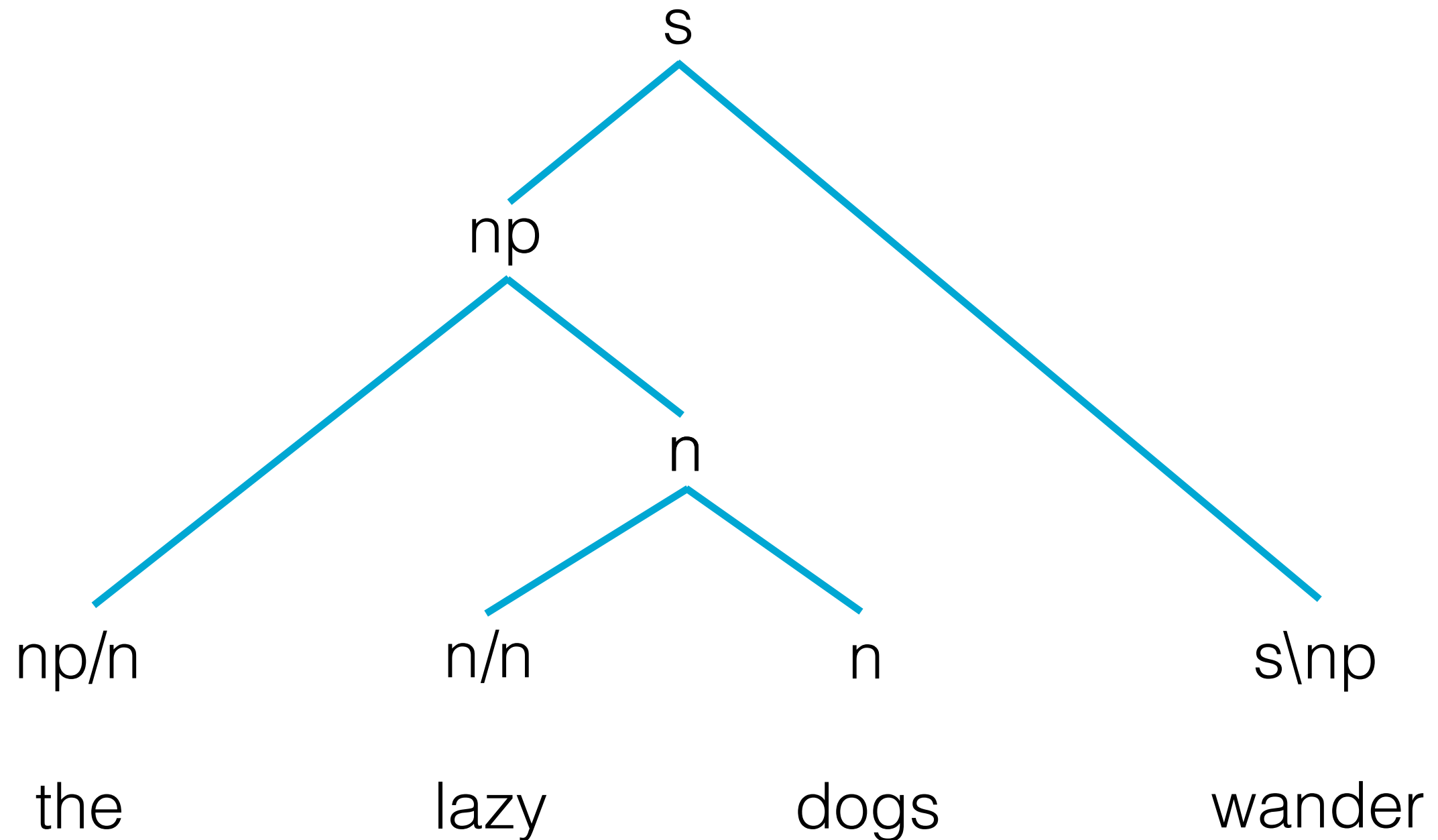
# CCG Parsing



# CCG Parsing



# CCG Parsing



# Why CCG?

## **Machine Translation**

[Weese, Callison-Burch, and Lopez, 2012]

## **Semantic Parsing**

[Zettlemoyer and Collins, 2005]

# Type-Supervised CCG

Type-supervised learning for CCG  
is highly *ambiguous*

# Type-Supervised CCG

Type-supervised learning for CCG  
is highly *ambiguous*

Penn Treebank  
parts-of-speech

# Type-Supervised CCG

Type-supervised learning for CCG  
is highly *ambiguous*

Penn Treebank  
parts-of-speech

**48 tags**



# Type-Supervised CCG

Type-supervised learning for CCG  
is highly *ambiguous*

Penn Treebank  
parts-of-speech

**48 tags**

CCGBank  
Categories

# Type-Supervised CCG

Type-supervised learning for CCG  
is highly *ambiguous*

Penn Treebank  
parts-of-speech

**48 tags**

CCGBank  
Categories

**1,300+ categories**

# Our Strategy

The grammar formalism *itself*  
can be used to guide learning

# Our Strategy

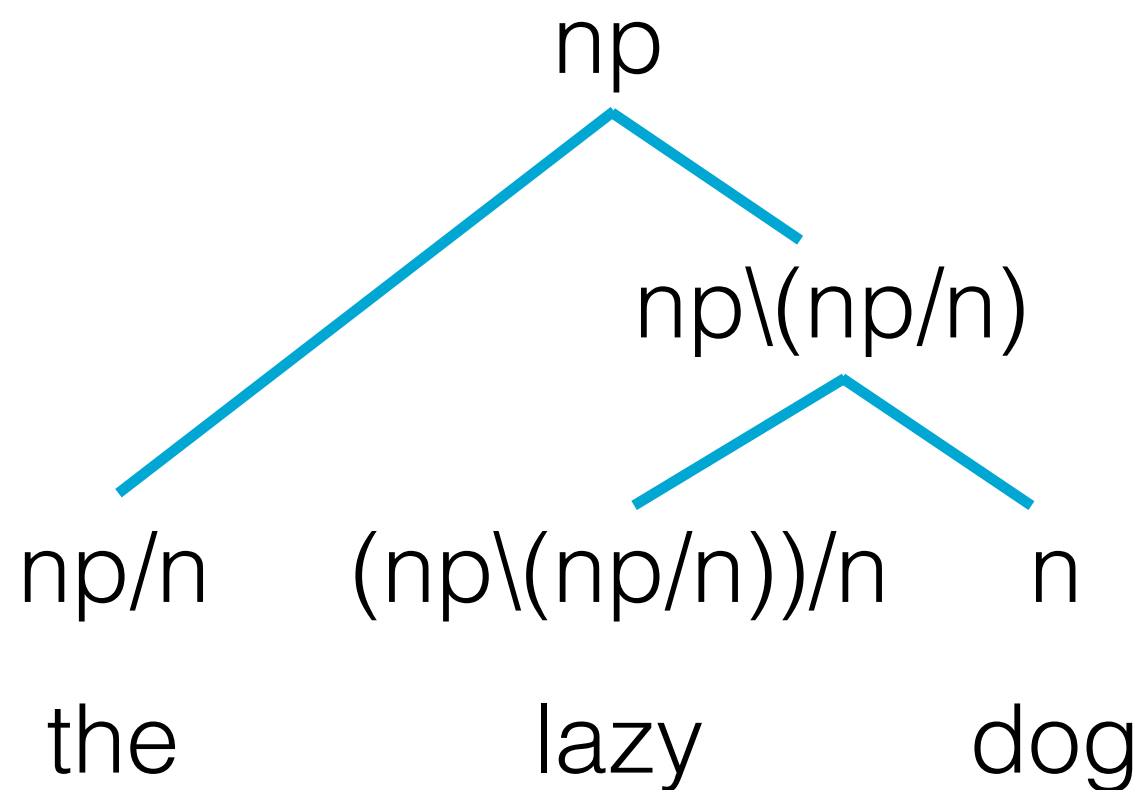
Incorporate *universal knowledge*  
about grammar into learning

Universal Knowledge

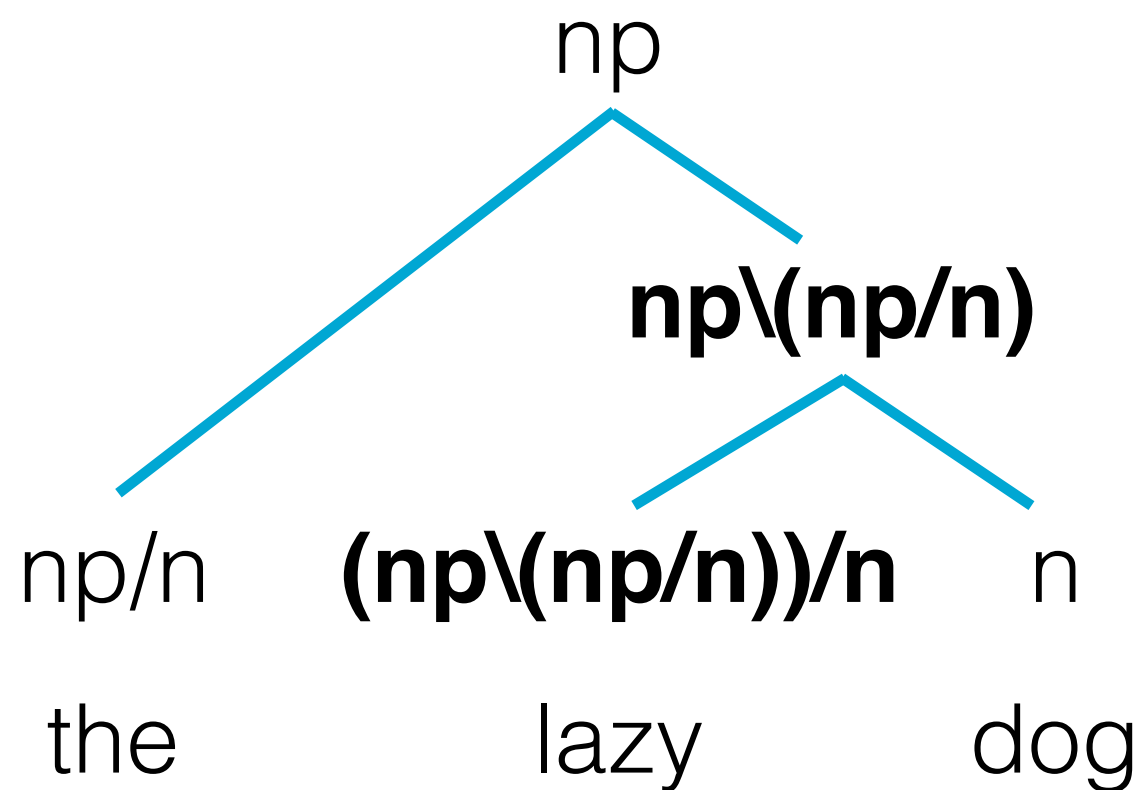
# Prefer Simpler Categories

the lazy dog

# Prefer Simpler Categories

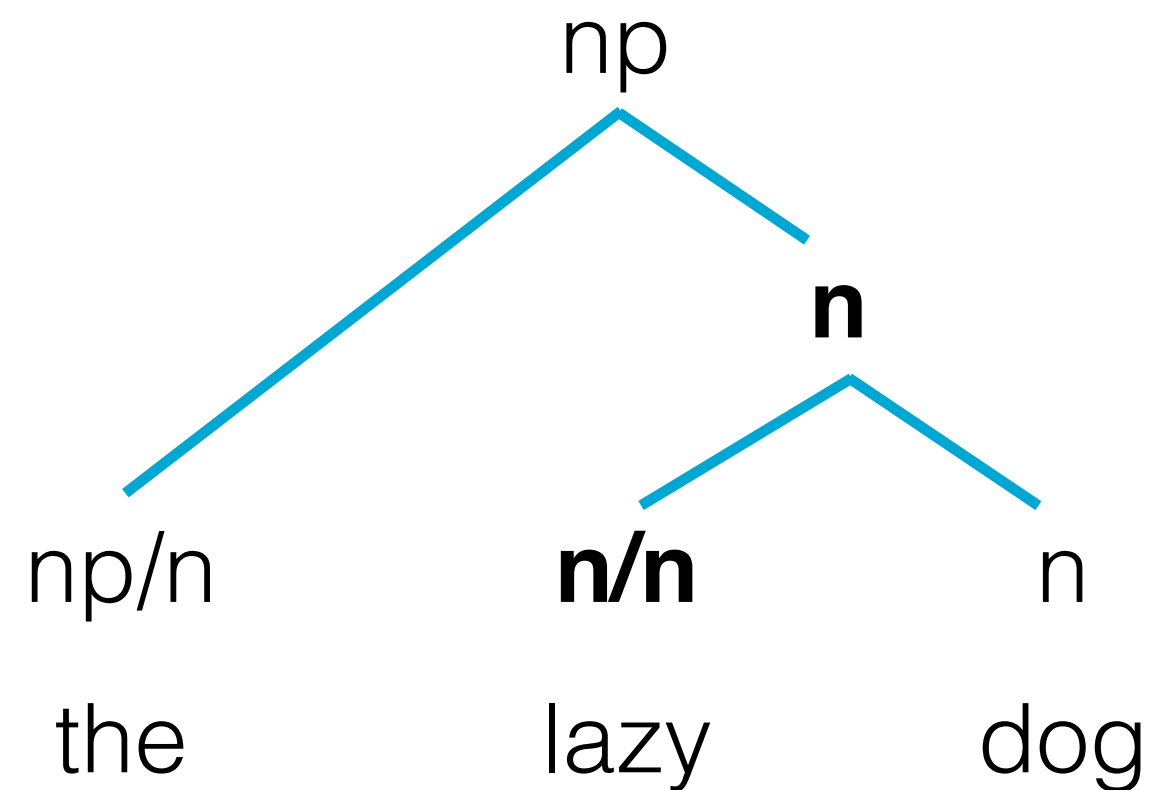
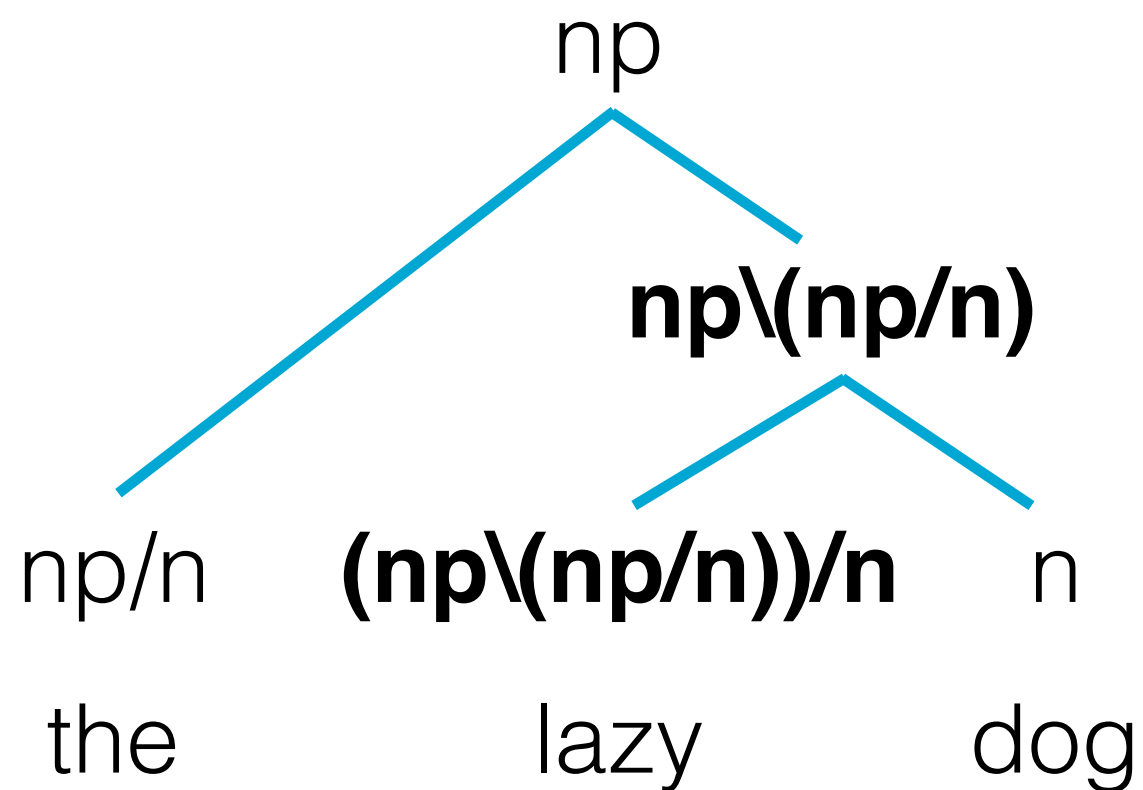


# Prefer Simpler Categories

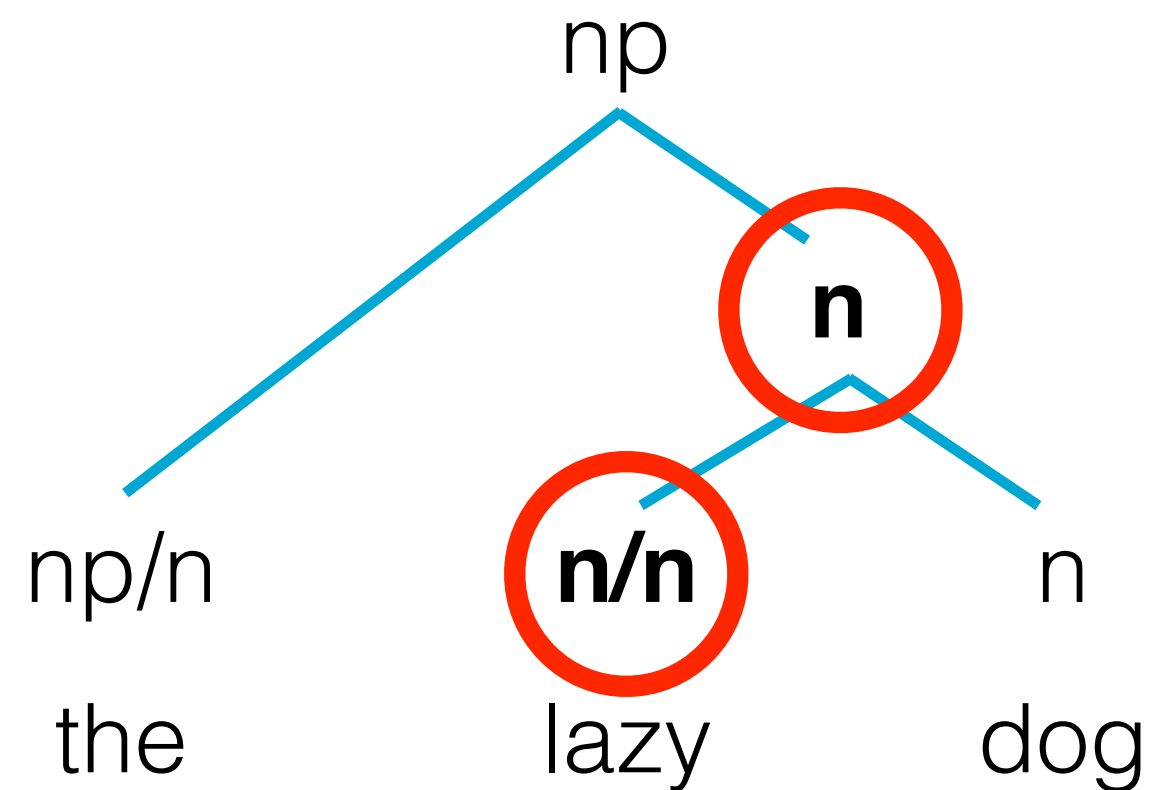
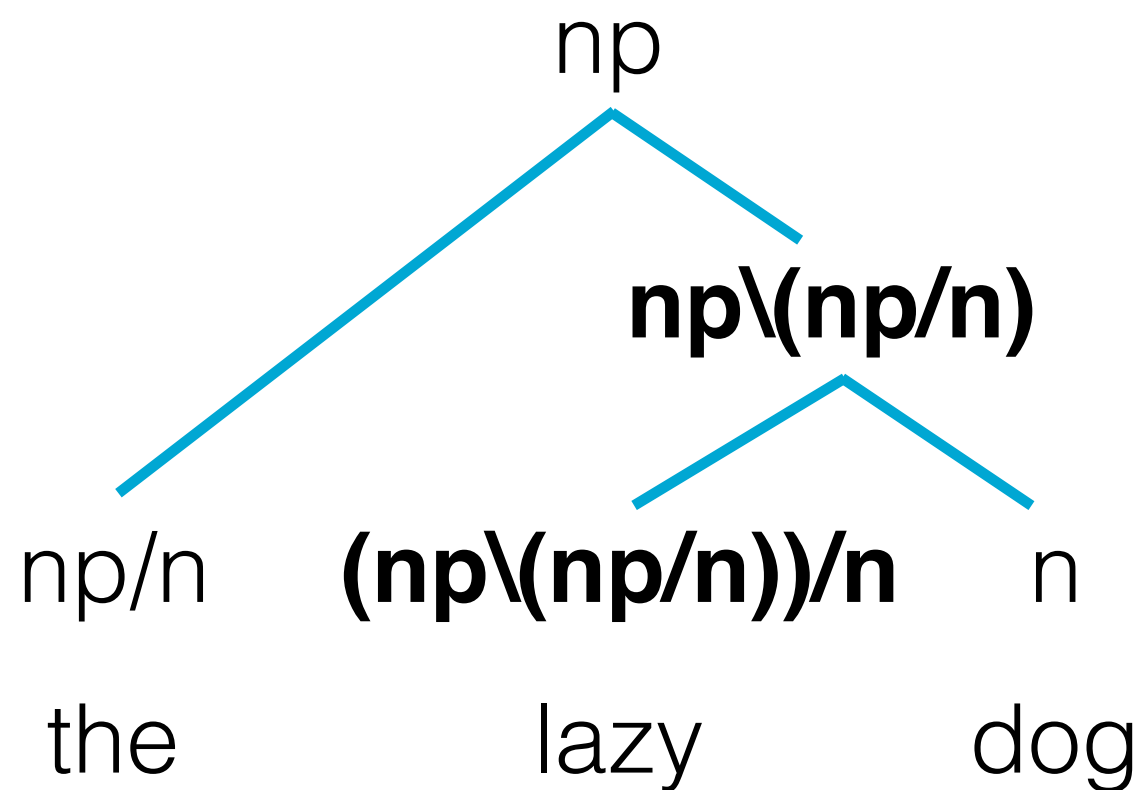




# Prefer Simpler Categories



# Prefer Simpler Categories



# Prefer Simpler Categories

$\text{buy} := (s_b \backslash np) / np$

e.g. “Opponents don't **buy** such arguments.”

# Prefer Simpler Categories

buy := (s<sub>b</sub>\np)/np                      appears **342** times in CCGbank

e.g. “Opponents don't **buy** such arguments.”

# Prefer Simpler Categories

buy := (s<sub>b</sub>\np)/np      appears **342** times in CCGbank

e.g. “Opponents don't **buy** such arguments.”

$$\text{buy} := (((s_b \backslash np) / \mathbf{pp}) / \mathbf{pp}) / np$$

“Tele-Communications agreed to **buy** half of Showtime Networks from Viacom for \$ 225 million.”  
pp pp

# Prefer Simpler Categories

buy := (s<sub>b</sub>\np)/np      appears **342** times in CCGbank

e.g. “Opponents don't **buy** such arguments.”

buy := (((s<sub>b</sub>\np)/**pp**)/**pp**)/np      appears **once**

“Tele-Communications agreed to **buy** half of Showtime Networks from Viacom for \$ 225 million.”  
pp pp

# Prefer Modifier Categories

# Prefer Modifier Categories

$((s_b \backslash np) / np) / ((s_b \backslash np) / np)$



# Prefer Modifier Categories

$((s_b \backslash np) / np)$  /  $((s_b \backslash np) / np)$

# Prefer Modifier Categories

$((s_b \backslash np) / np)$  /  $((s_b \backslash np) / np)$

# Prefer Modifier Categories

$(s_b \backslash np) / np$

$((s_b \backslash np) / np)$  /  $((s_b \backslash np) / np)$

# Prefer Modifier Categories

$(s_b \backslash np) / np$

*transitive verb*: (he) **hides** (the money)

$((s_b \backslash np) / np)$  /  $((s_b \backslash np) / np)$

# Prefer Modifier Categories

$(s_b \backslash np) / np$

*transitive verb*: (he) **hides** (the money)

$((s_b \backslash np) / np)$  /  $((s_b \backslash np) / np)$

*adverb*: (he) **quickly** (hides) (the money)

# Weighted Tag Grammar

$a \longrightarrow \{s, np, n, \dots\}$

$A \longrightarrow B / C$

$A \longrightarrow B \setminus C$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a)$$

$$A \longrightarrow B / C$$

$$A \longrightarrow B \setminus C$$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / C$$

$$A \longrightarrow B \setminus C$$



# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / C \quad \overline{p_{\text{term}}}$$

$$A \longrightarrow B \setminus C \quad \overline{p_{\text{term}}}$$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / C \quad \overline{p_{\text{term}}} \times p_{\text{fwd}}$$

$$A \longrightarrow B \setminus C \quad \overline{p_{\text{term}}}$$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / C \quad \overline{p_{\text{term}}} \times p_{\text{fwd}}$$

$$A \longrightarrow B \setminus C \quad \overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}}$$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / B \quad \overline{p_{\text{term}}} \times p_{\text{fwd}} \times p_{\text{mod}}$$

$$A \longrightarrow B / C \quad \overline{p_{\text{term}}} \times p_{\text{fwd}}$$

$$A \longrightarrow B \setminus B \quad \overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}} \times p_{\text{mod}}$$

$$A \longrightarrow B \setminus C \quad \overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}}$$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / B \quad \overline{p_{\text{term}}} \times p_{\text{fwd}} \times p_{\text{mod}}$$

$$A \longrightarrow B / C \quad \overline{p_{\text{term}}} \times p_{\text{fwd}} \times \overline{p_{\text{mod}}}$$

$$A \longrightarrow B \setminus B \quad \overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}} \times p_{\text{mod}}$$

$$A \longrightarrow B \setminus C \quad \overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}} \times \overline{p_{\text{mod}}}$$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / B \quad \overline{p_{\text{term}}} \times p_{\text{fwd}} \times p_{\text{mod}}$$

$$A \longrightarrow B / C \quad \overline{p_{\text{term}}} \times p_{\text{fwd}} \times \overline{p_{\text{mod}}}$$

$$A \longrightarrow B \setminus B \quad \overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}} \times p_{\text{mod}}$$

$$A \longrightarrow B \setminus C \quad \overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}} \times \overline{p_{\text{mod}}}$$

# Weighted Tag Grammar

$$a \longrightarrow \{s, np, n, \dots\} \quad p_{\text{atom}}(a) \times p_{\text{term}}$$

$$A \longrightarrow B / B$$

$$\overline{p_{\text{term}}} \times p_{\text{fwd}} \times p_{\text{mod}} \\ +$$

$$A \longrightarrow B / C$$

$$\overline{p_{\text{term}}} \times p_{\text{fwd}} \times \overline{p_{\text{mod}}}$$

$$A \longrightarrow B \setminus B$$

$$\overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}} \times p_{\text{mod}} \\ +$$

$$A \longrightarrow B \setminus C$$

$$\overline{p_{\text{term}}} \times \overline{p_{\text{fwd}}} \times \overline{p_{\text{mod}}}$$

# Prefer Likely Categories

np/n

the

n/n

lazy

n

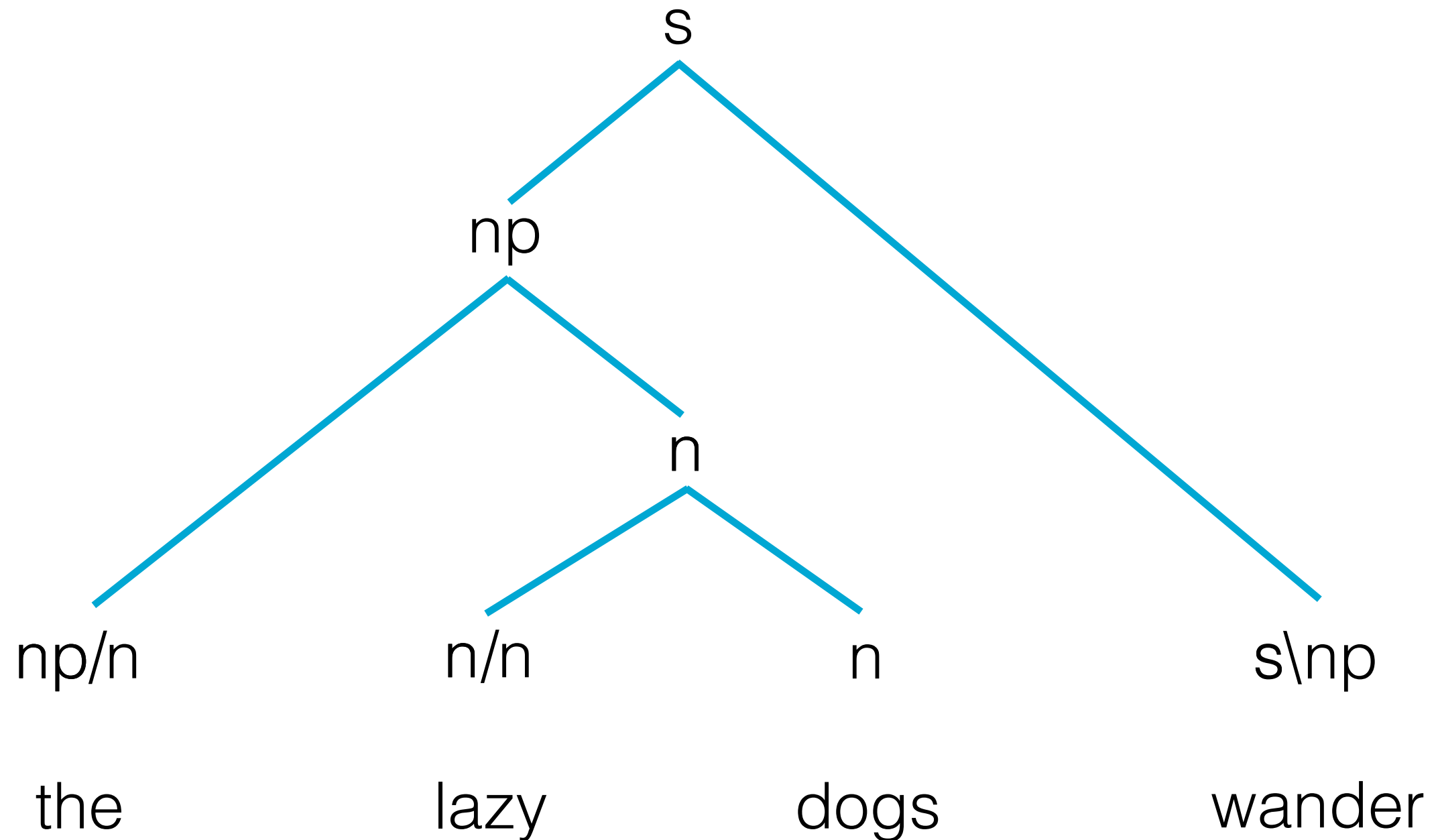
dogs

s\np

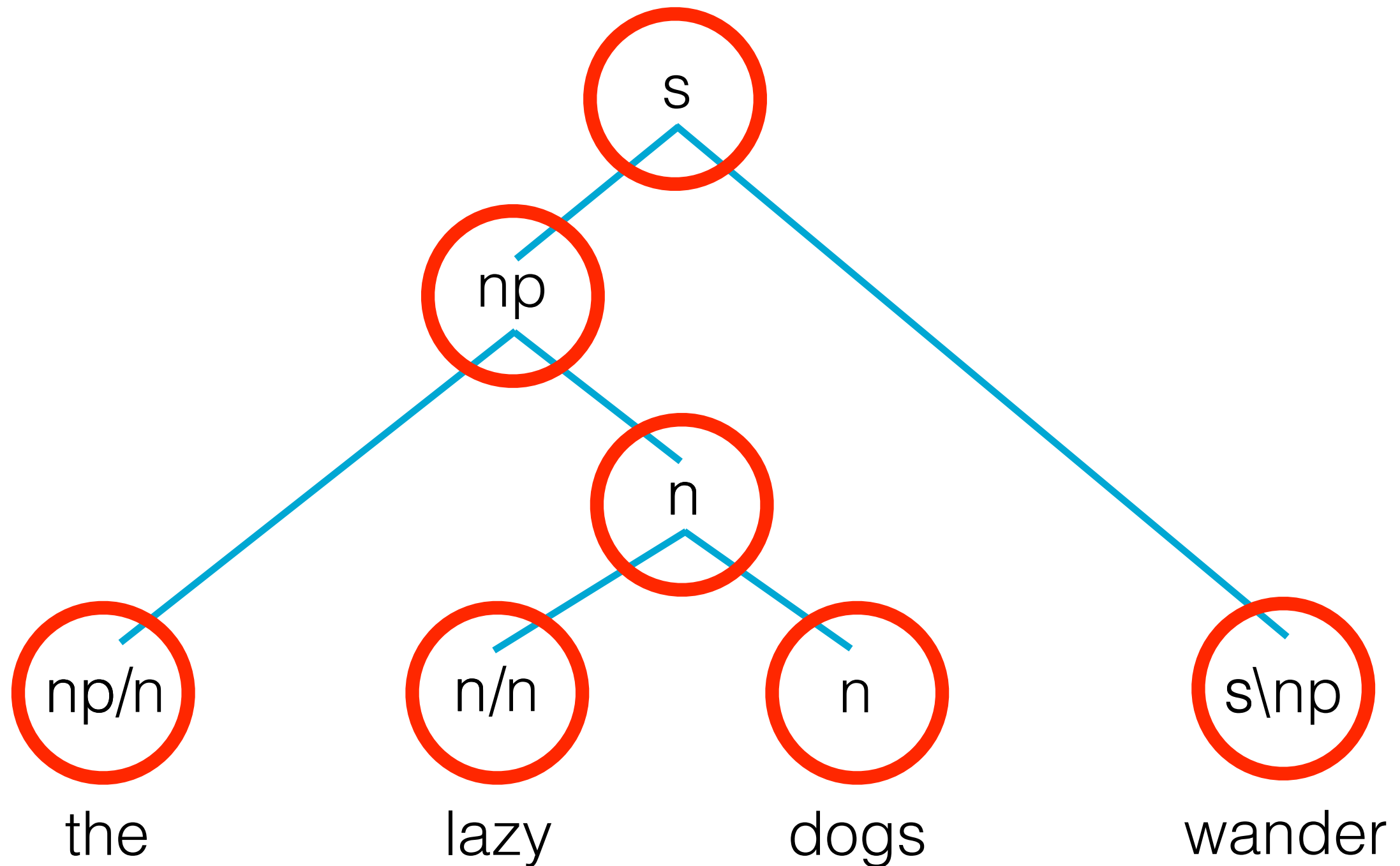
wander



# Prefer Likely Categories



# Prefer Likely Categories



# Type-Supervised Learning

unlabeled corpus

tag dictionary

# Type-Supervised Learning

unlabeled corpus

tag dictionary



same as  
POS tagging

# Type-Supervised Learning

unlabeled corpus  
tag dictionary

same as  
POS tagging

universal properties of the CCG formalism ←

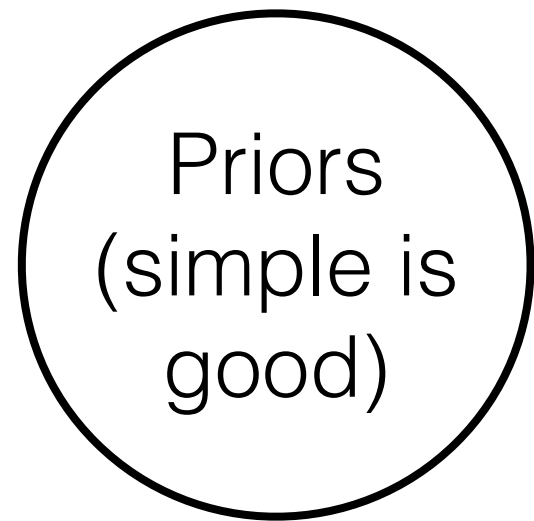
# Posterior Inference

# Posterior Inference



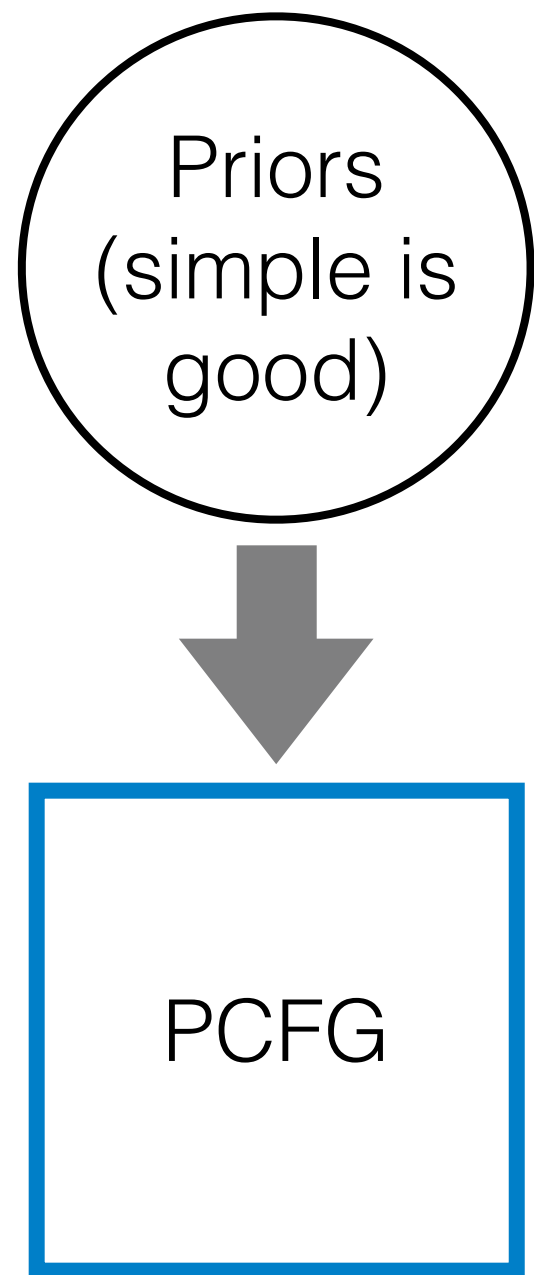
PCFG

# Posterior Inference

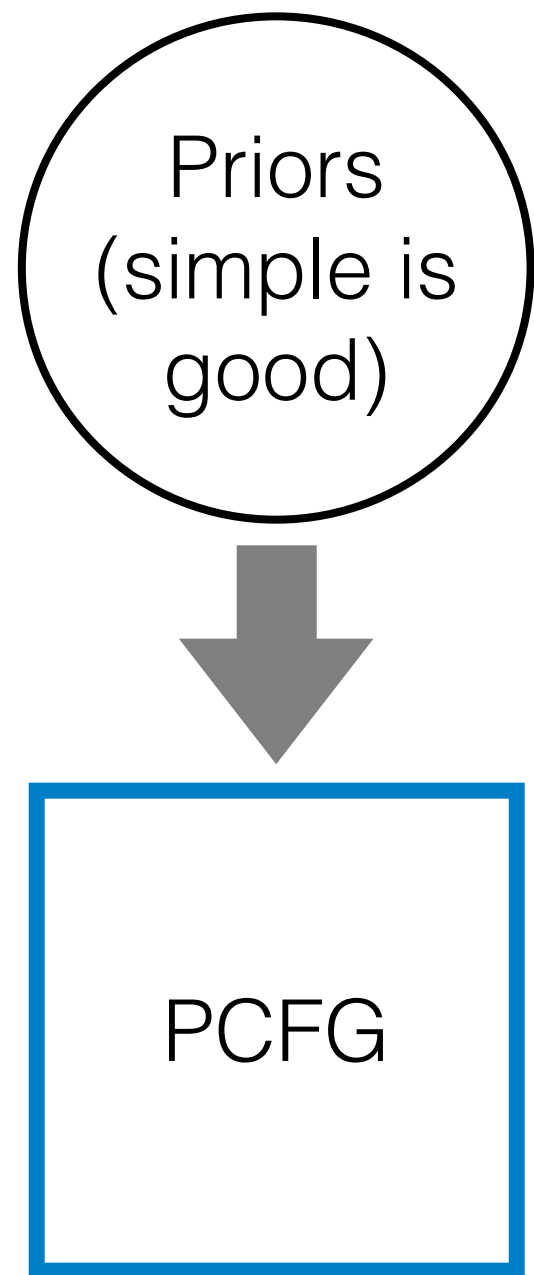




# Posterior Inference

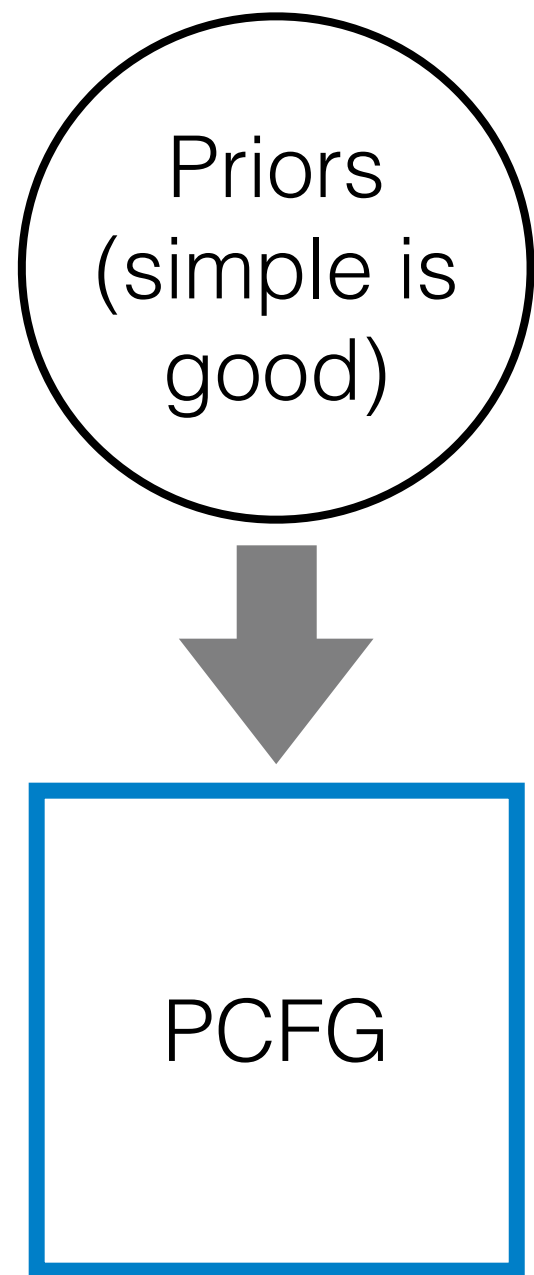


# Posterior Inference



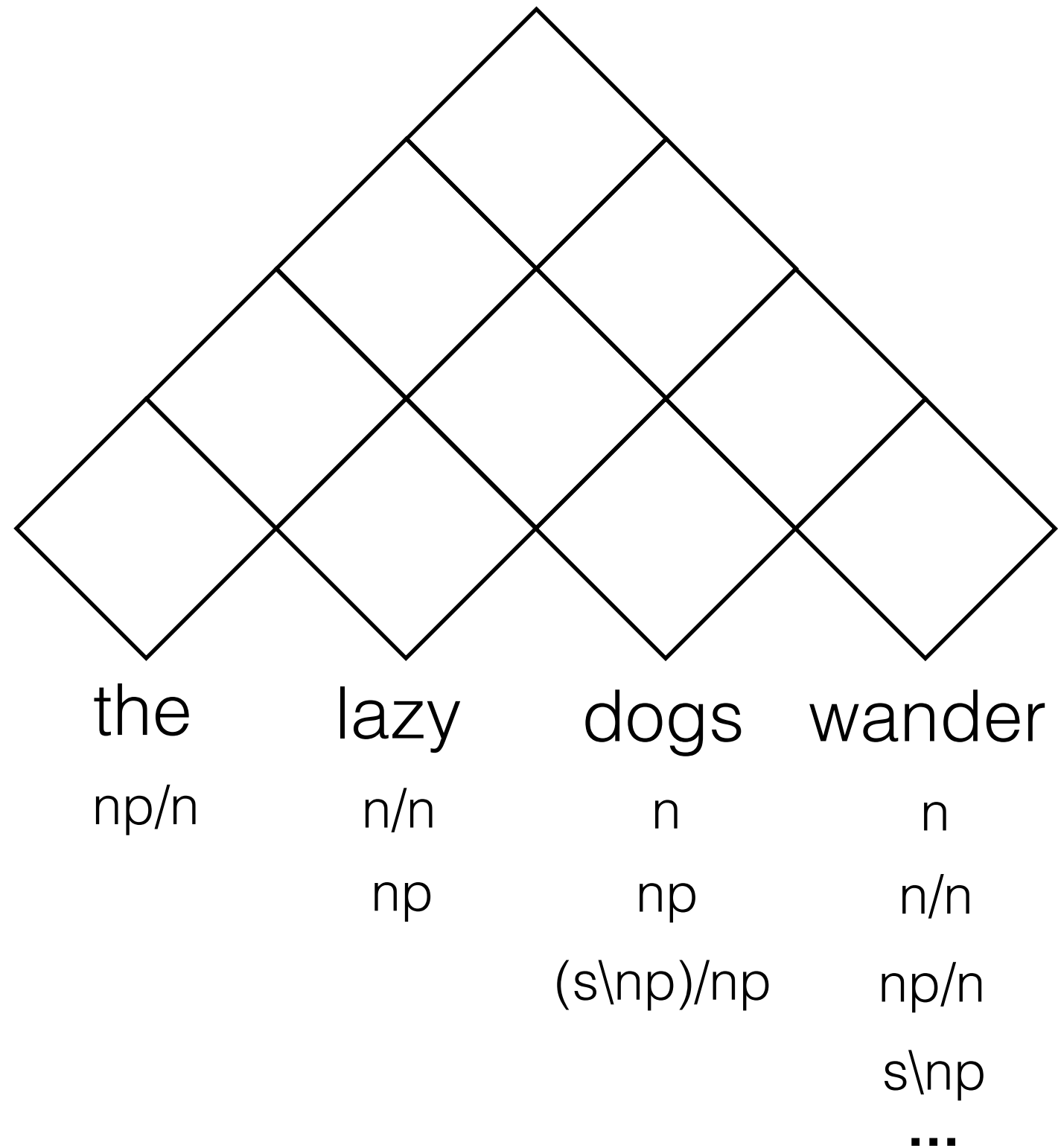
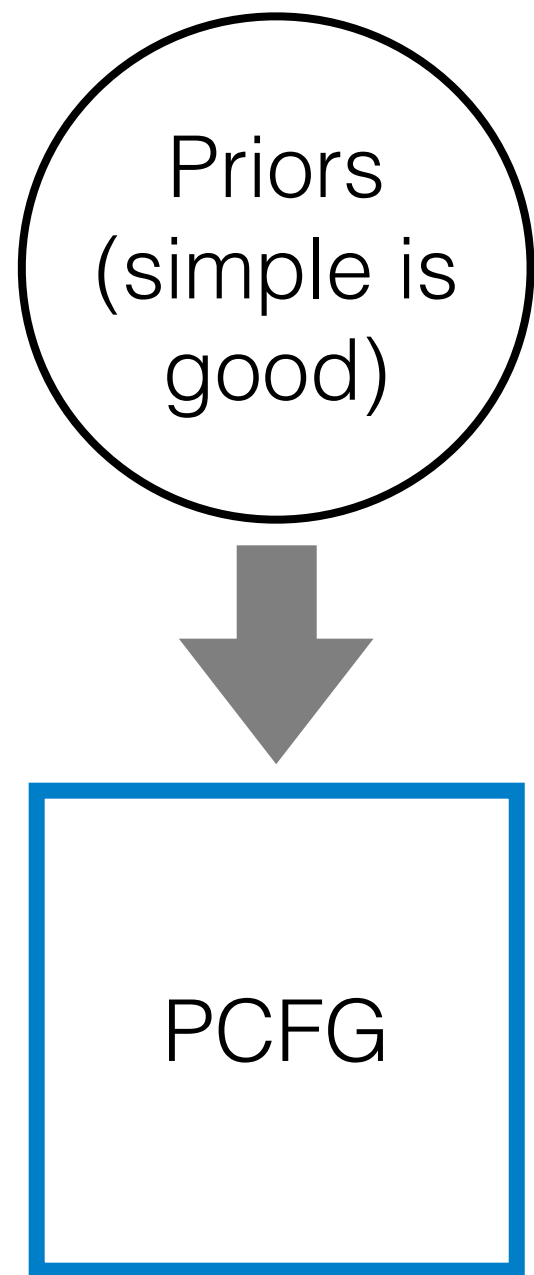
the lazy dogs wander

# Posterior Inference

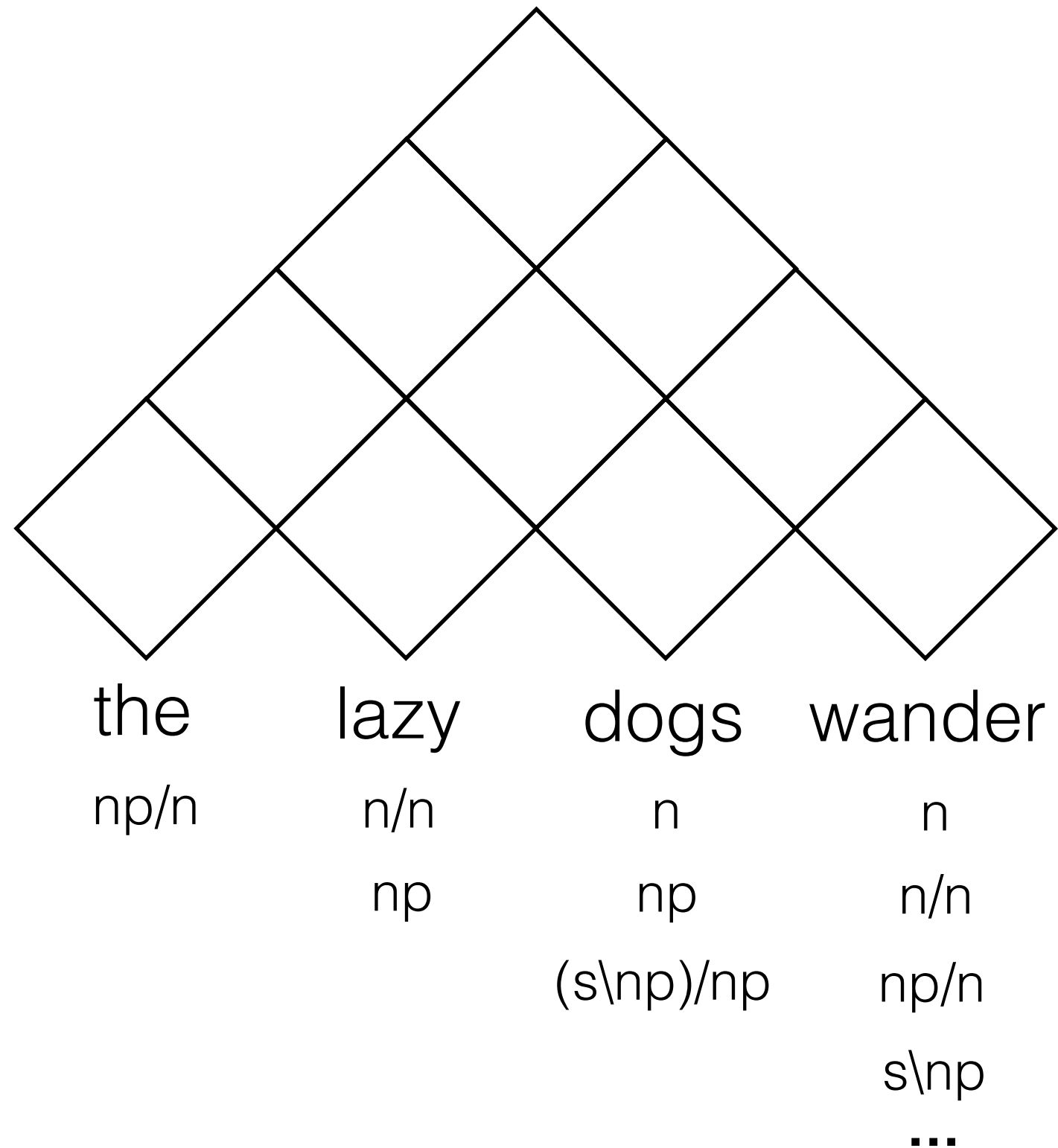
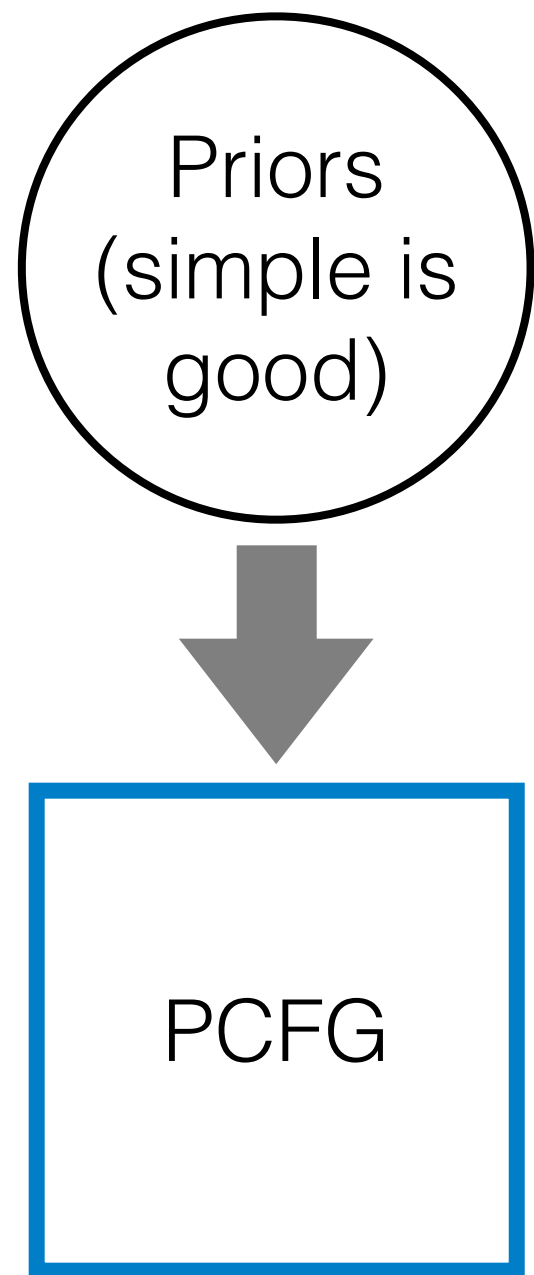


the	lazy	dogs	wander
np/n	n/n	n	n
	np	np	n/n
		(s\np)/np	np/n
			s\np
			...

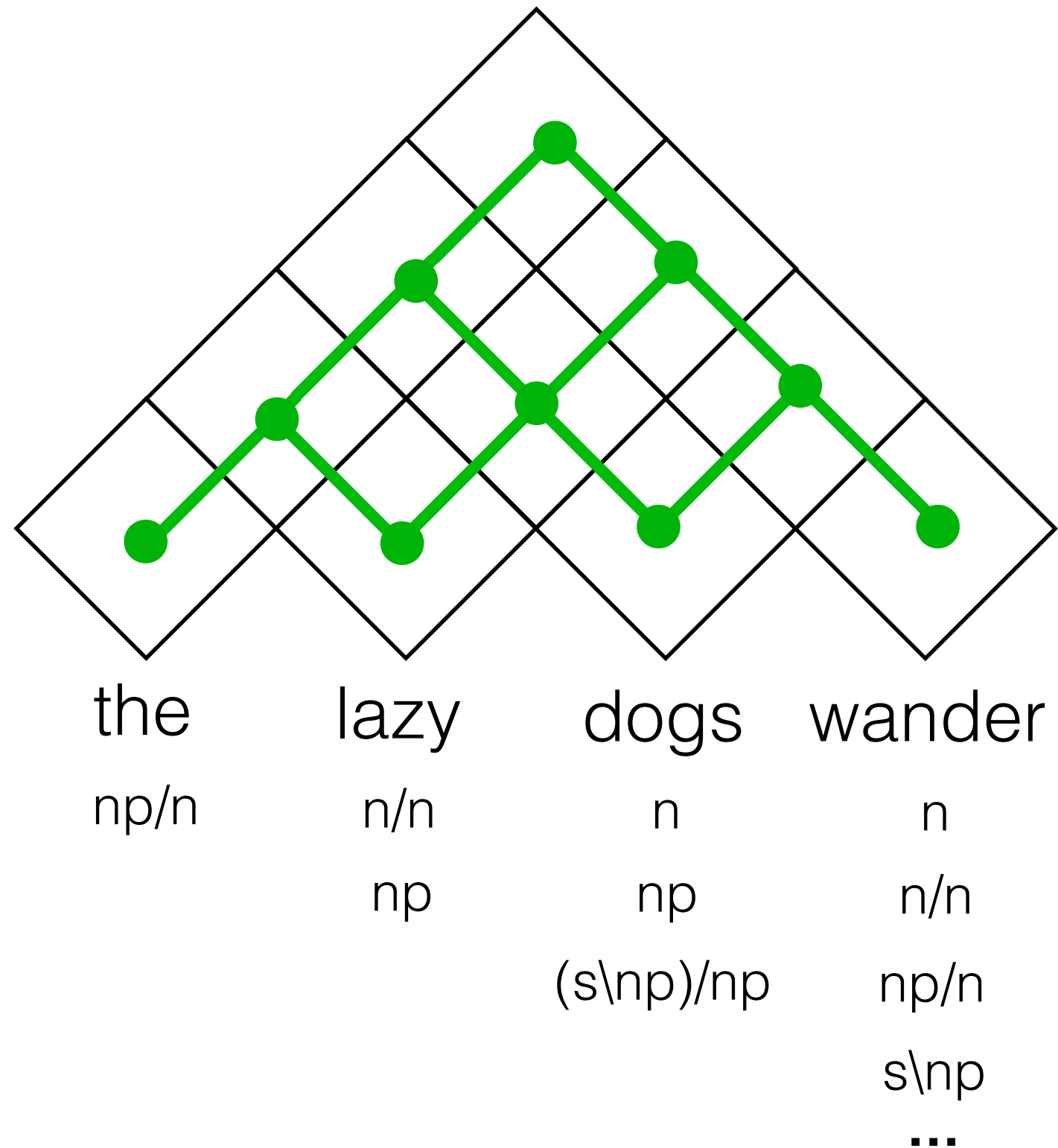
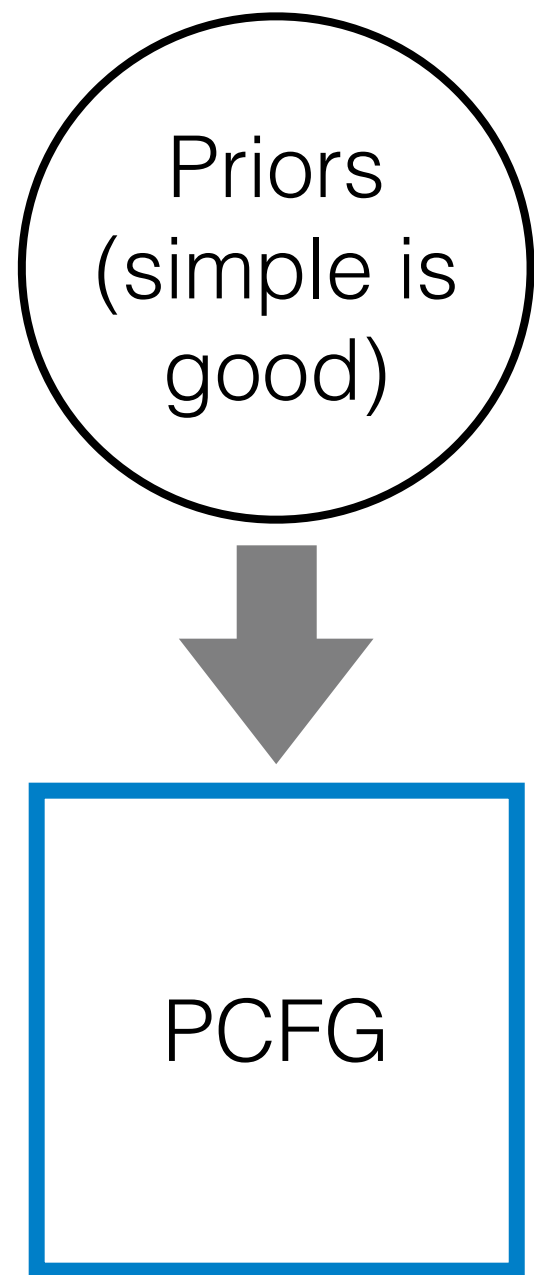
# Posterior Inference



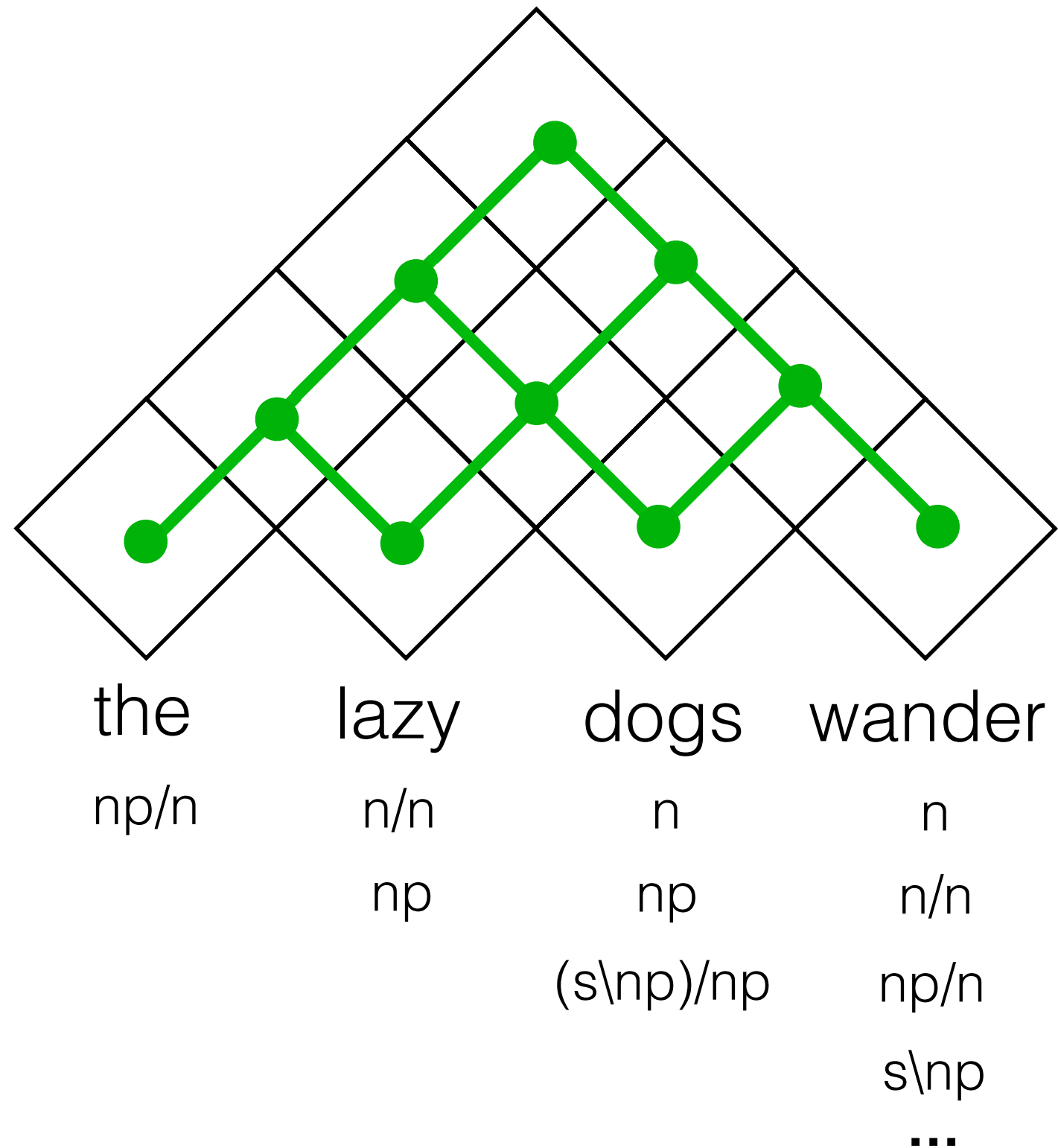
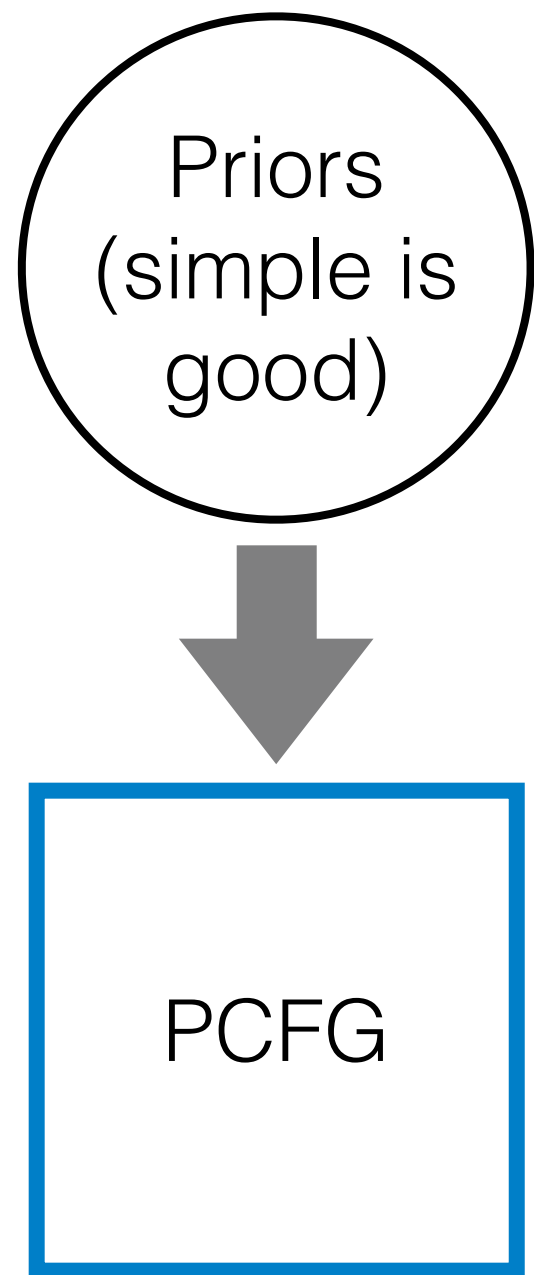
# Posterior Inference



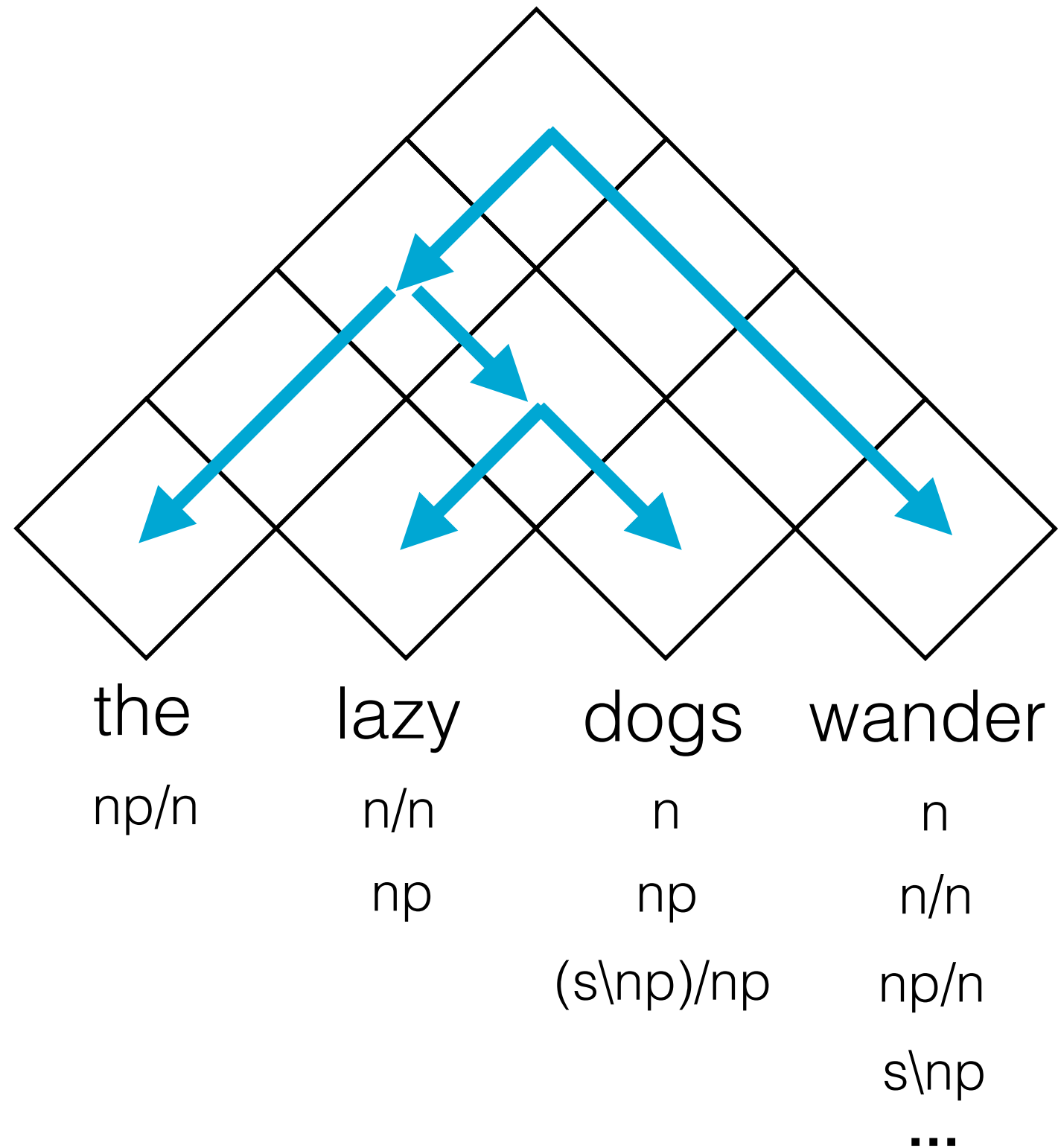
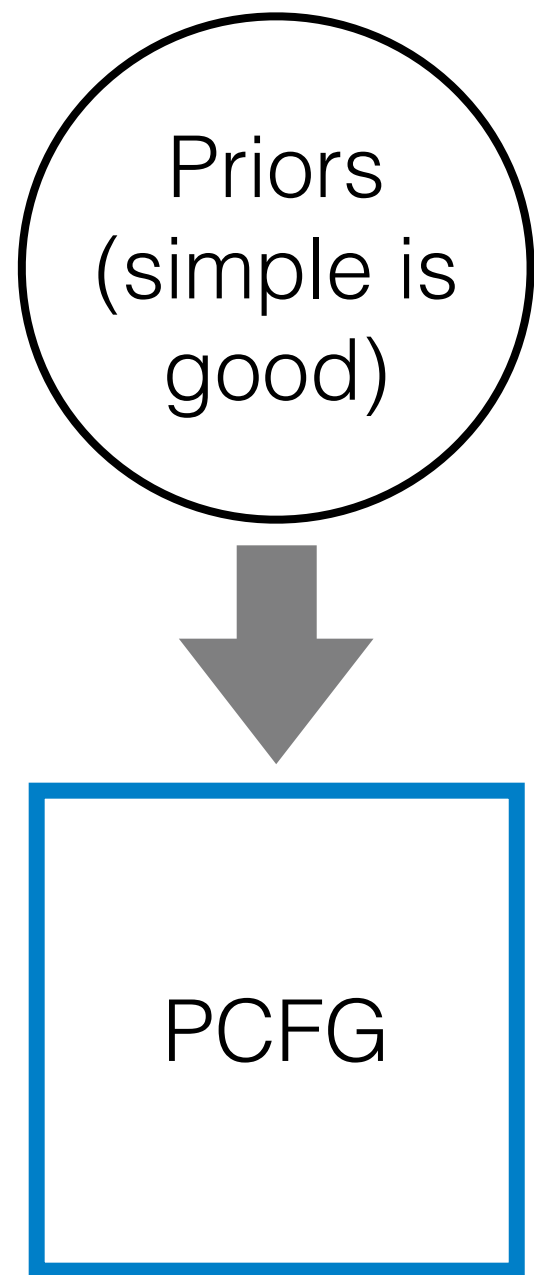
# Posterior Inference



# Posterior Inference

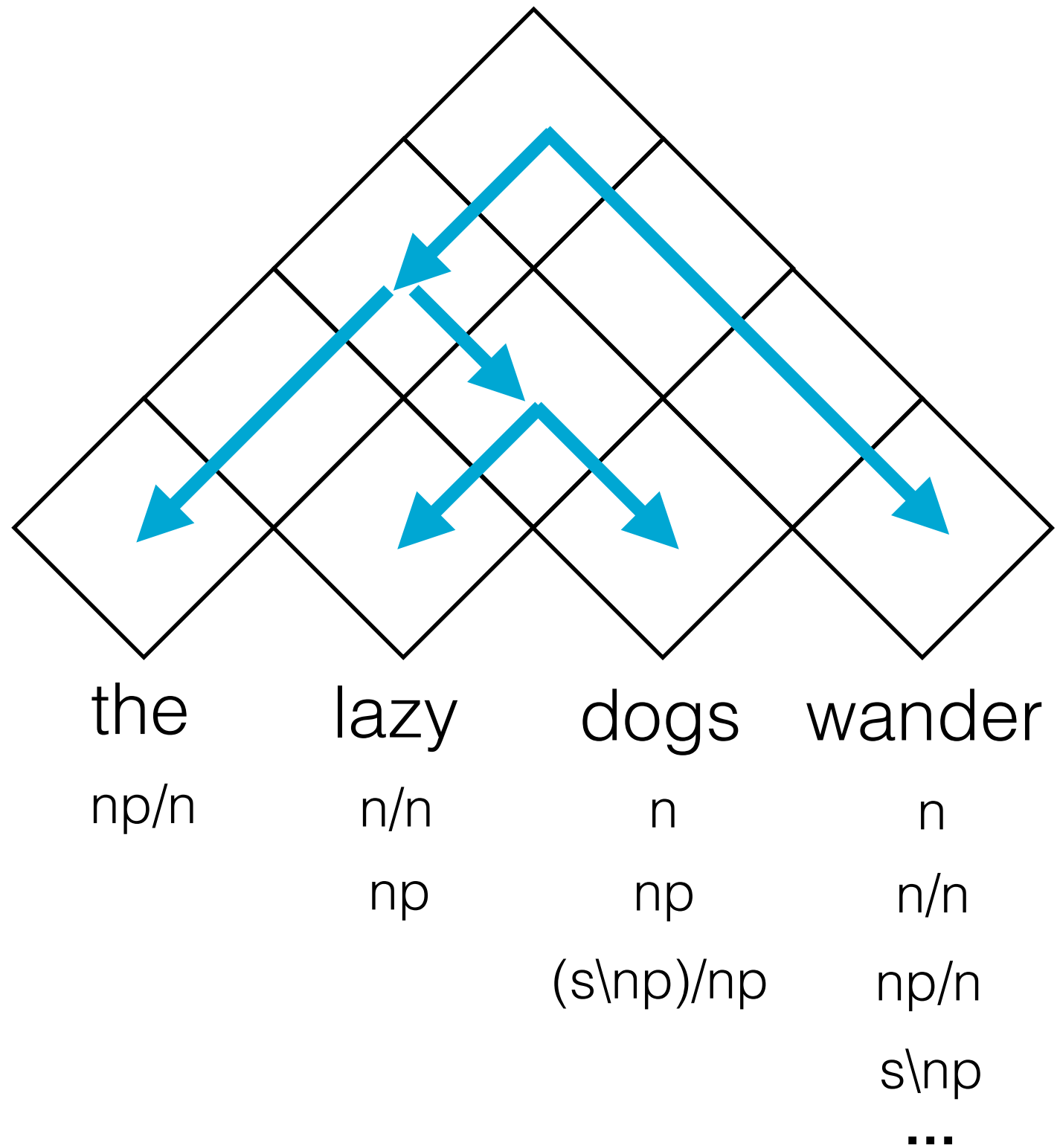
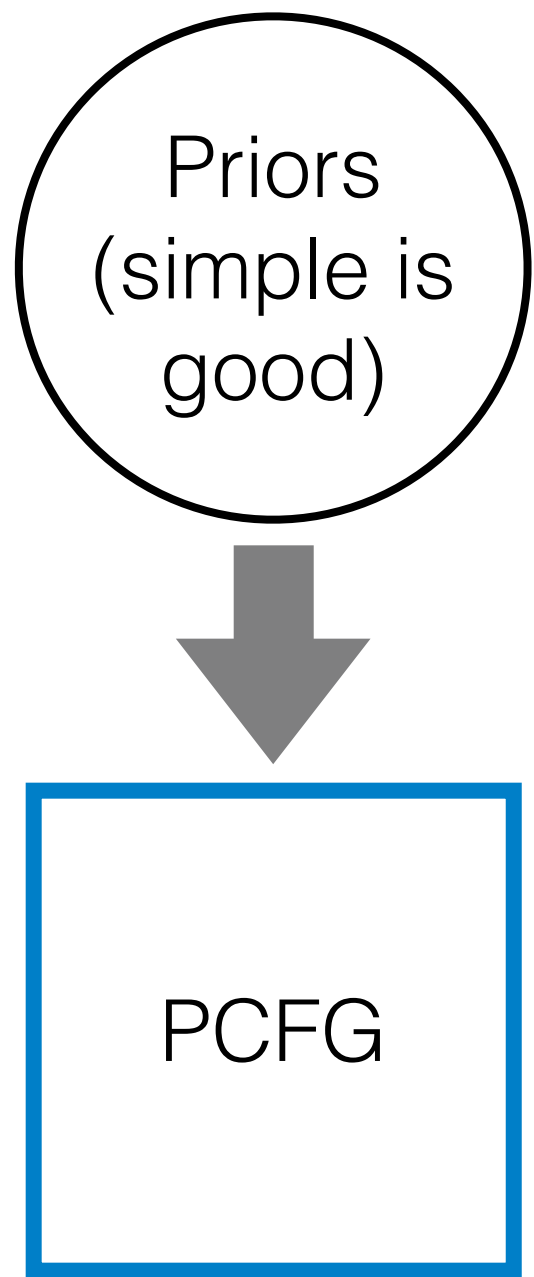


# Posterior Inference

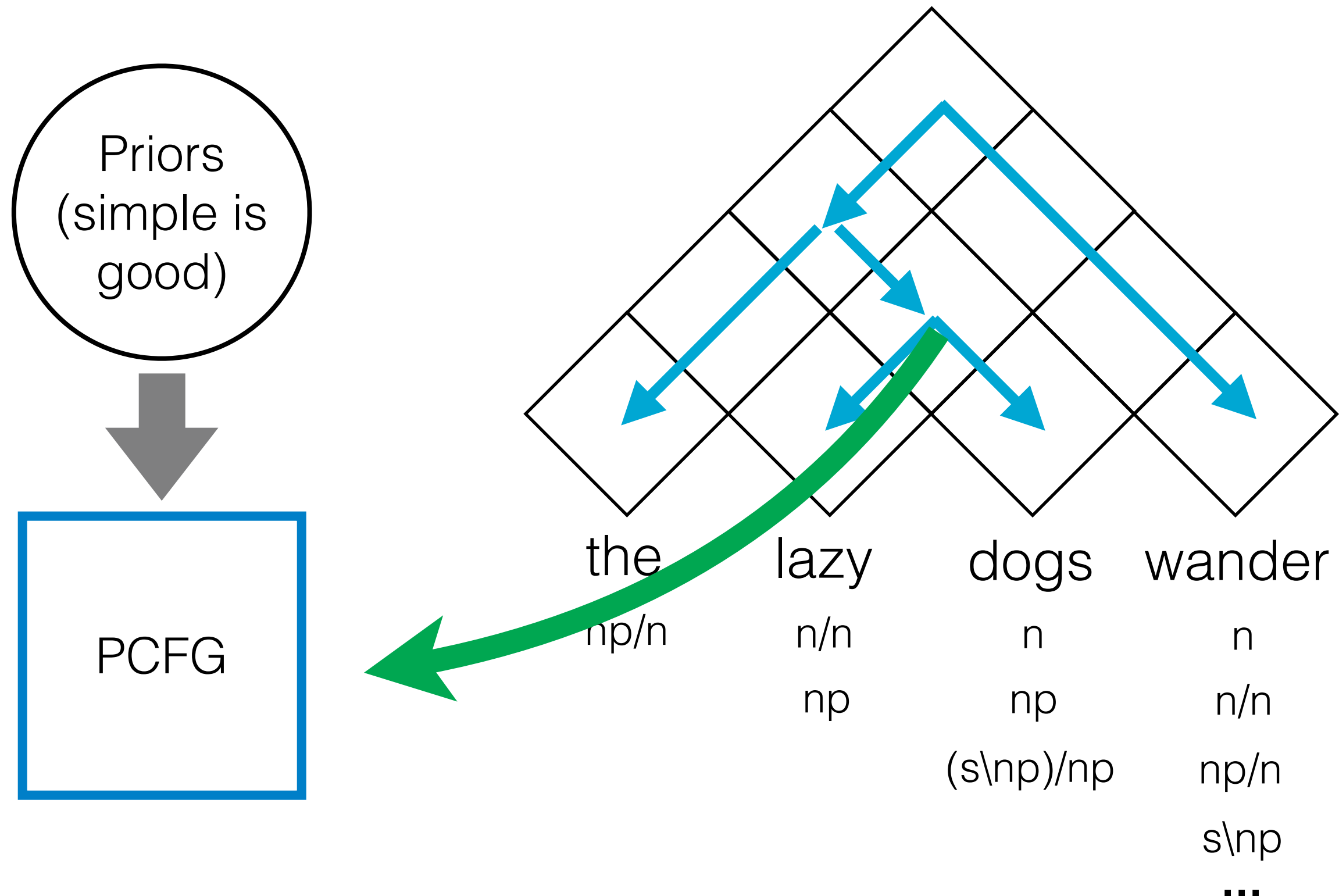




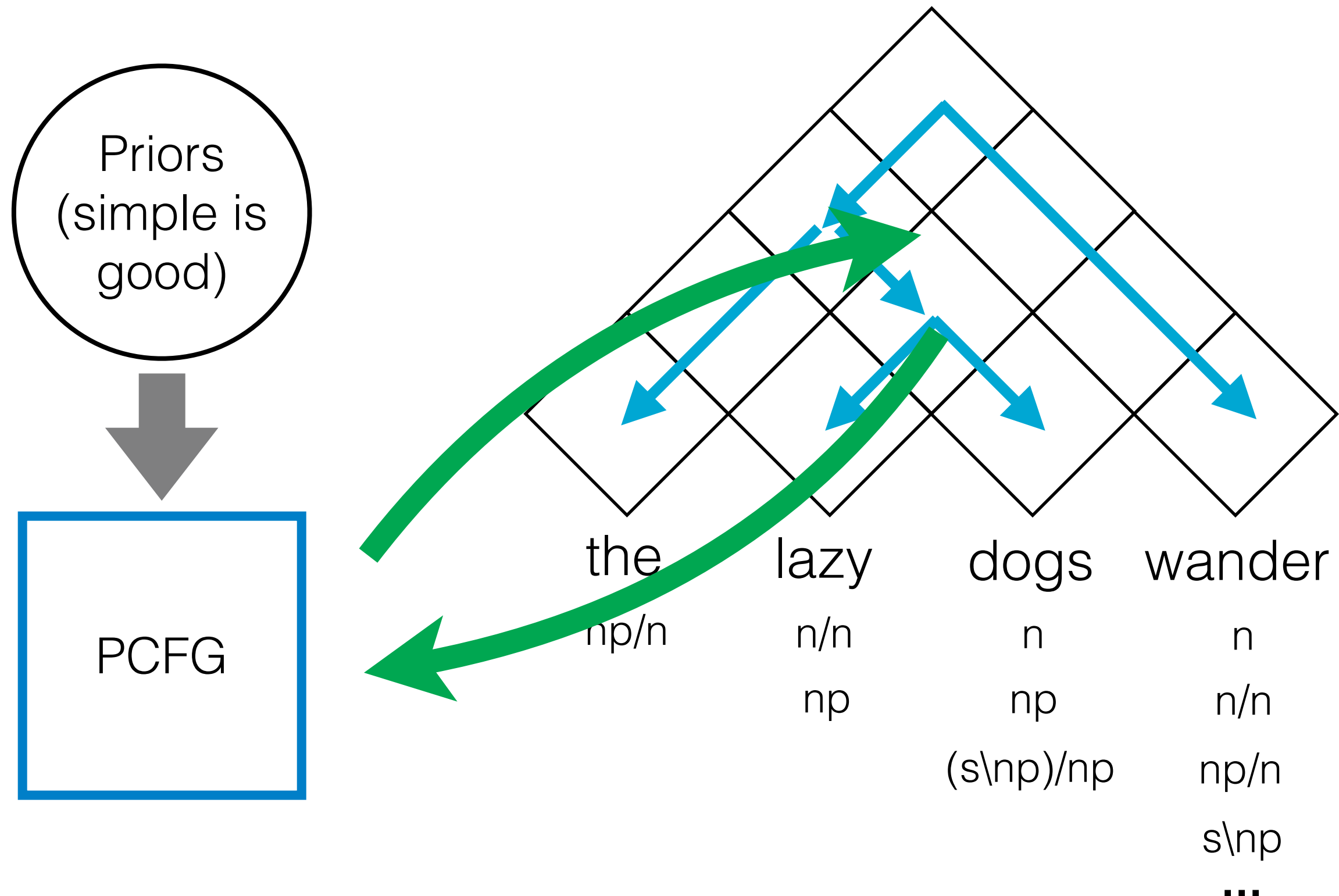
# Posterior Inference



# Posterior Inference

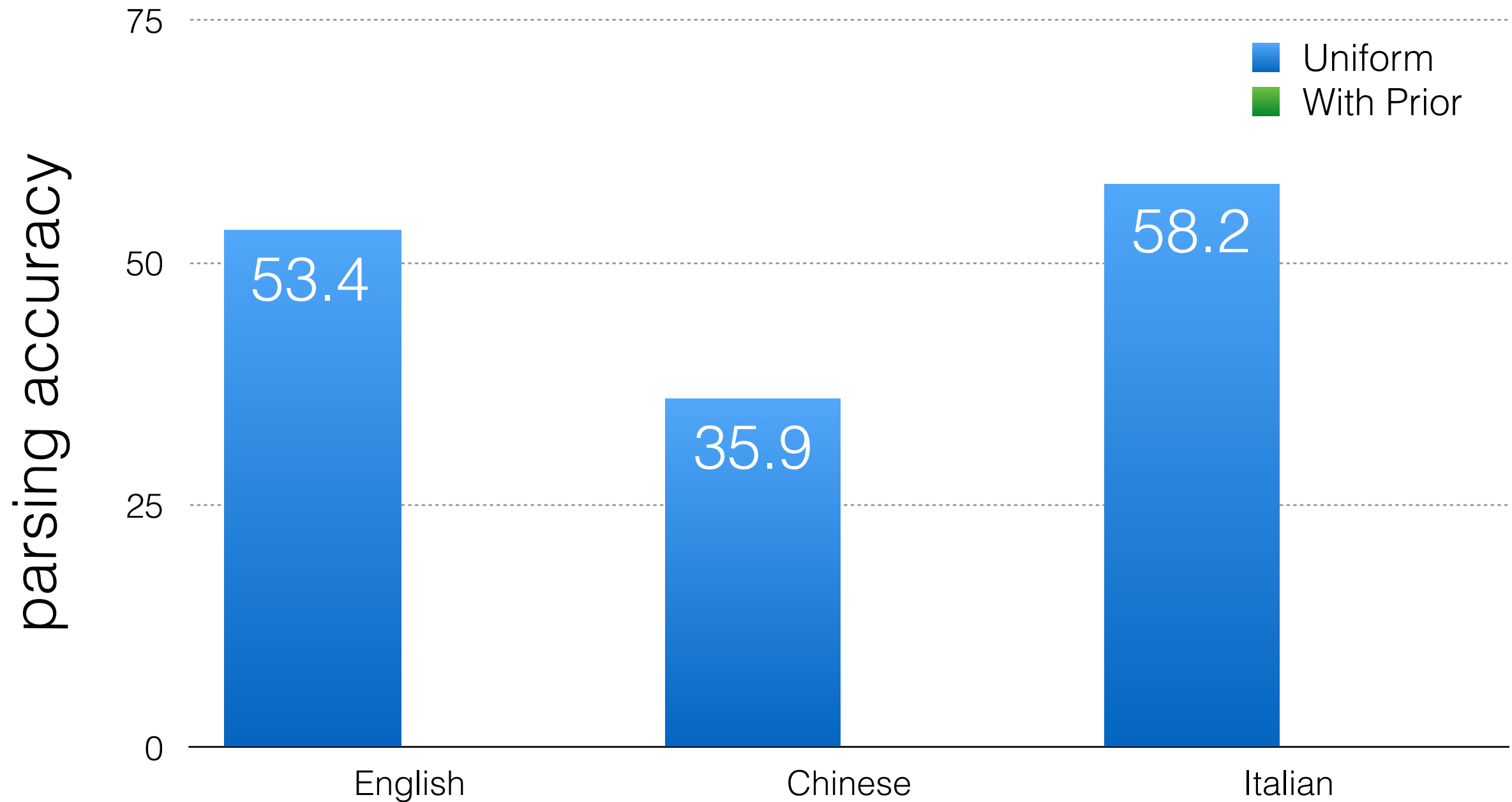


# Posterior Inference

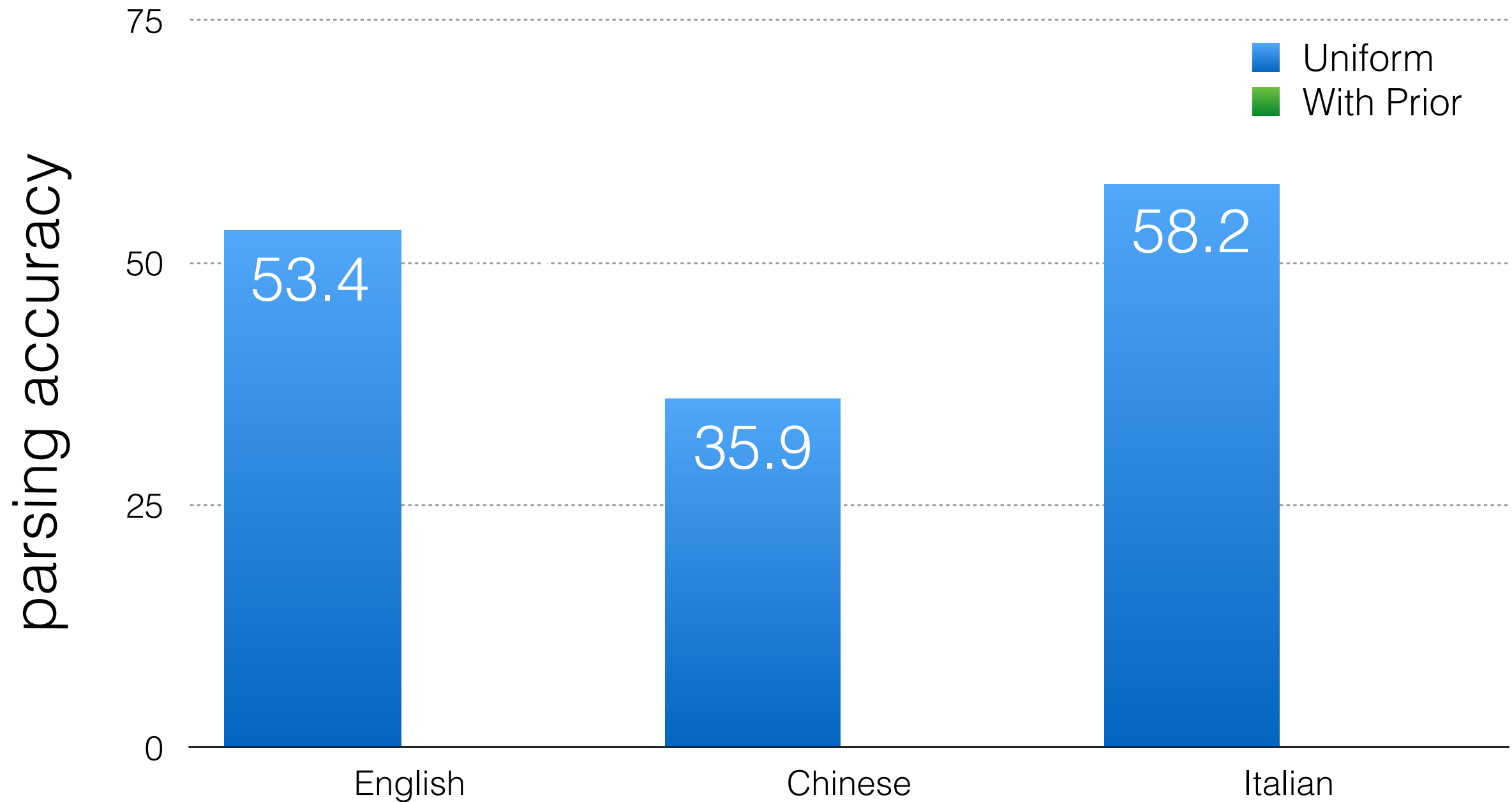


# Results

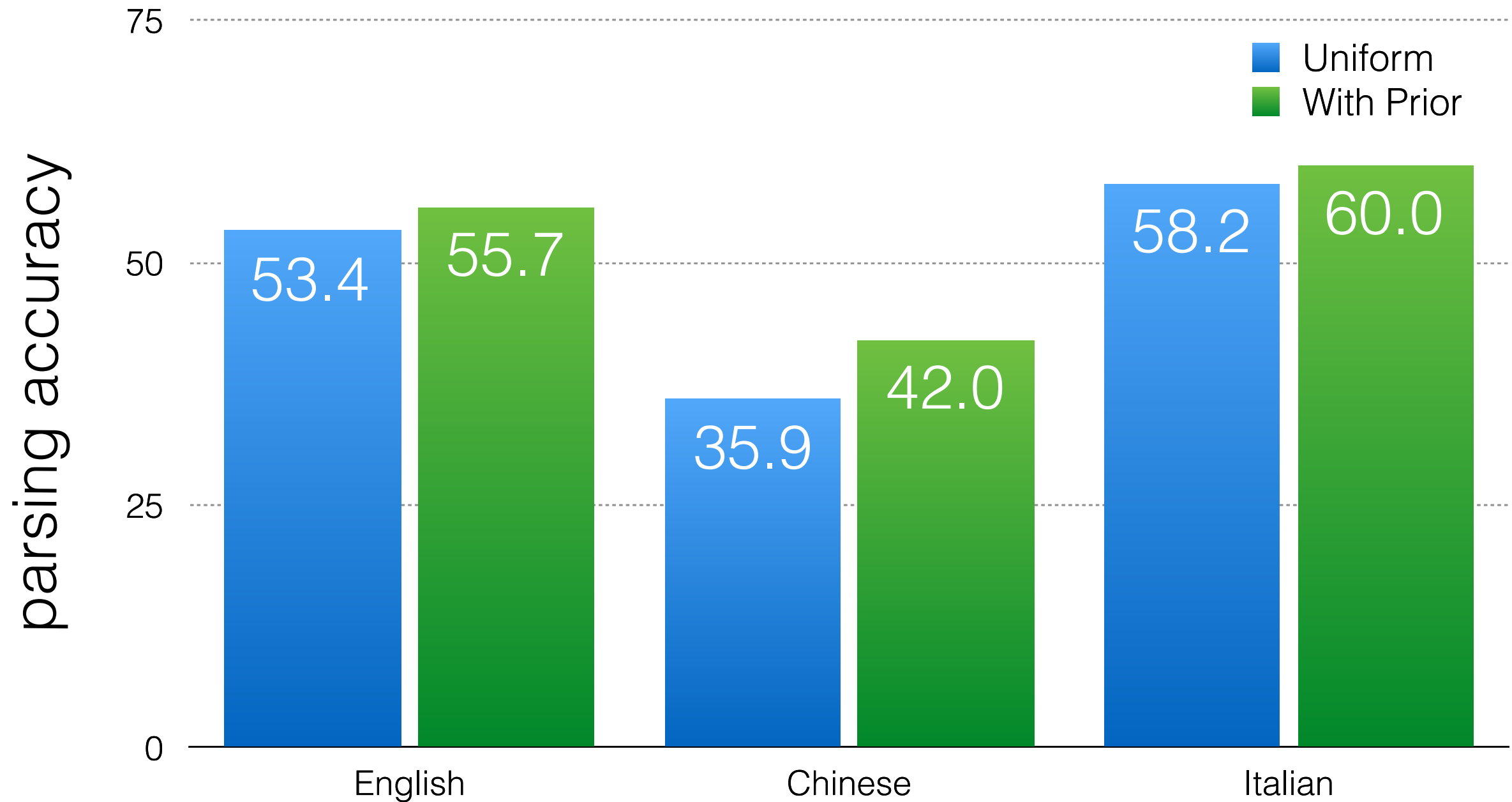
# CCG Parsing Results



# CCG Parsing Results



# CCG Parsing Results



# Conclusion

Using **universal grammatical knowledge**  
can make better use of weak supervision