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?

(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: word ↦ {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)

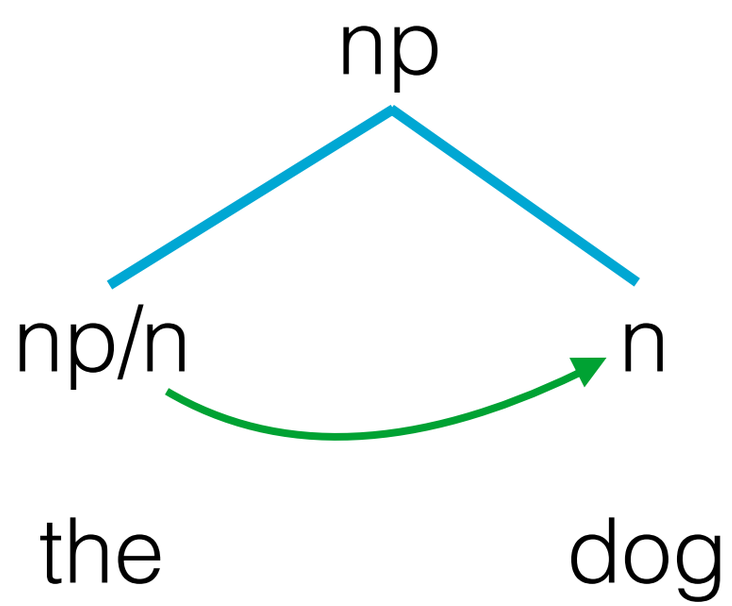
CCG

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

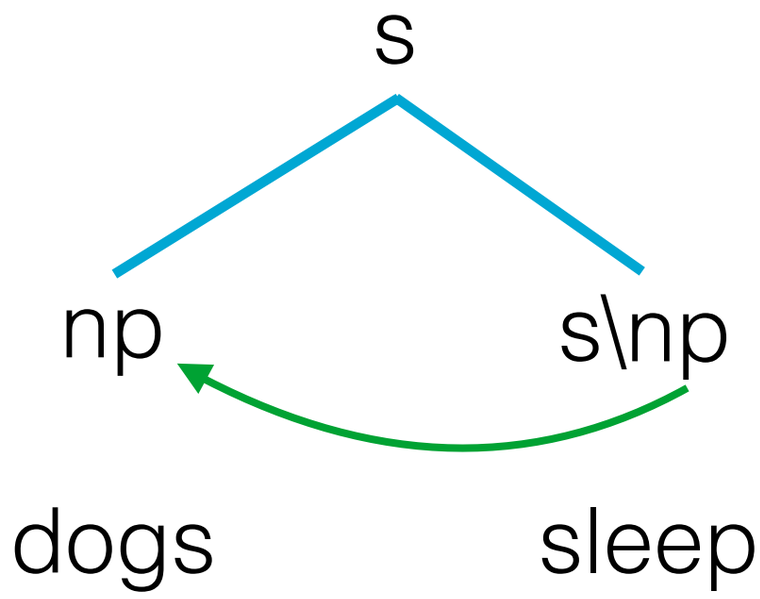
Categories **combine** to form categories of larger constituents

[Steedman, 2000]
[Steedman and Baldridge, 2011]

CCG



CCG



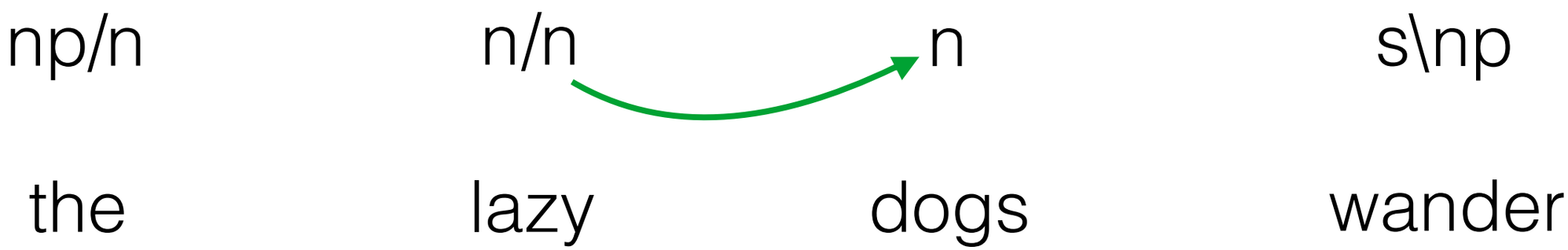
Type-Supervised CCG

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

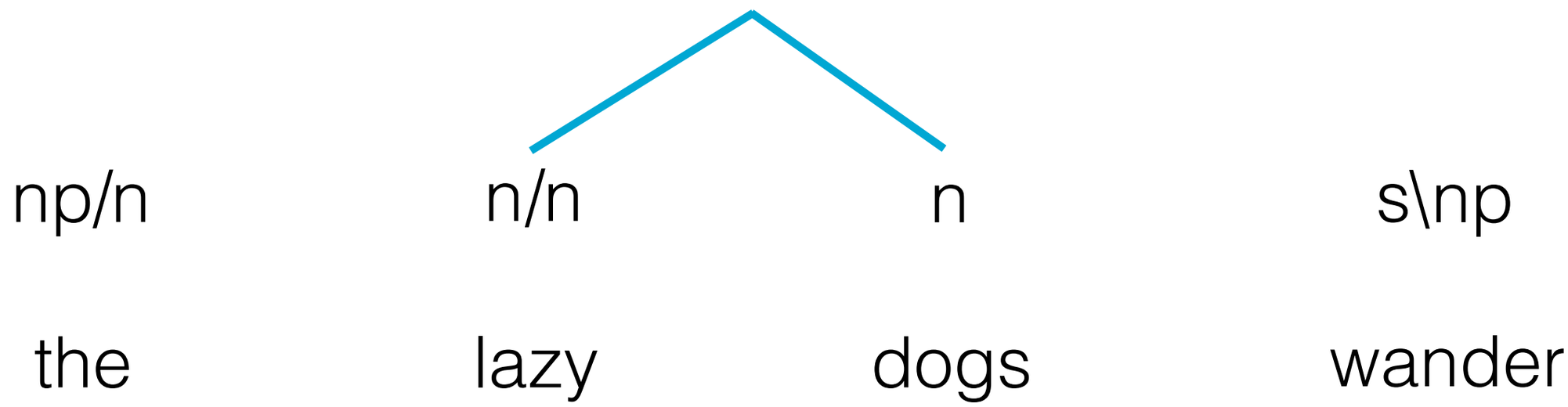
CCG Parsing

np/n	n/n	n	s\np
the	lazy	dogs	wander

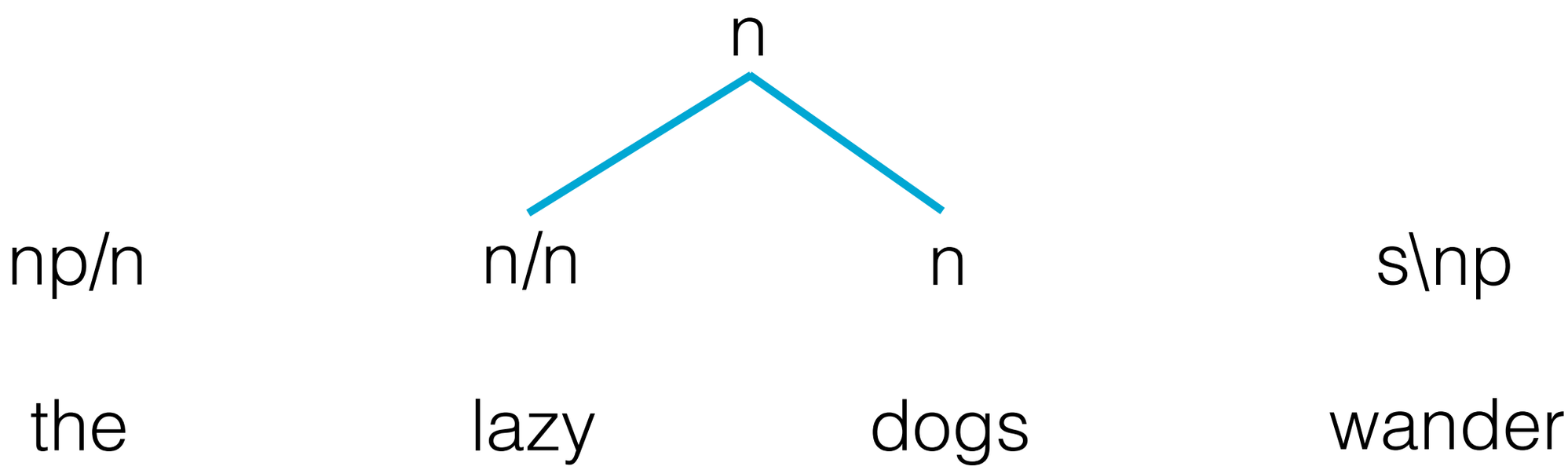
CCG Parsing



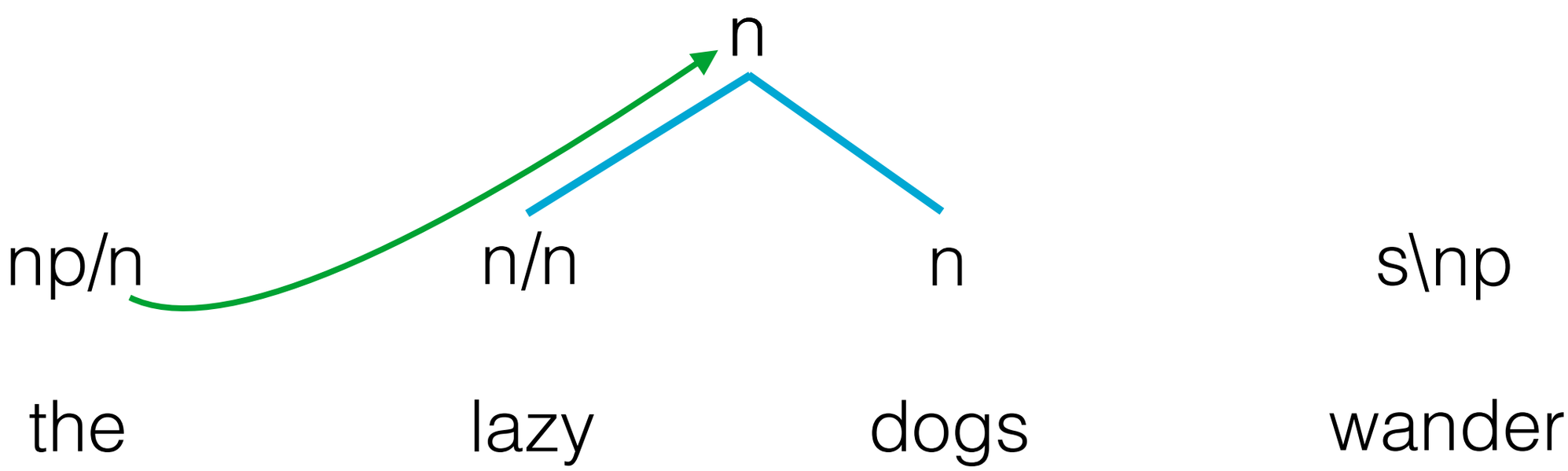
CCG Parsing



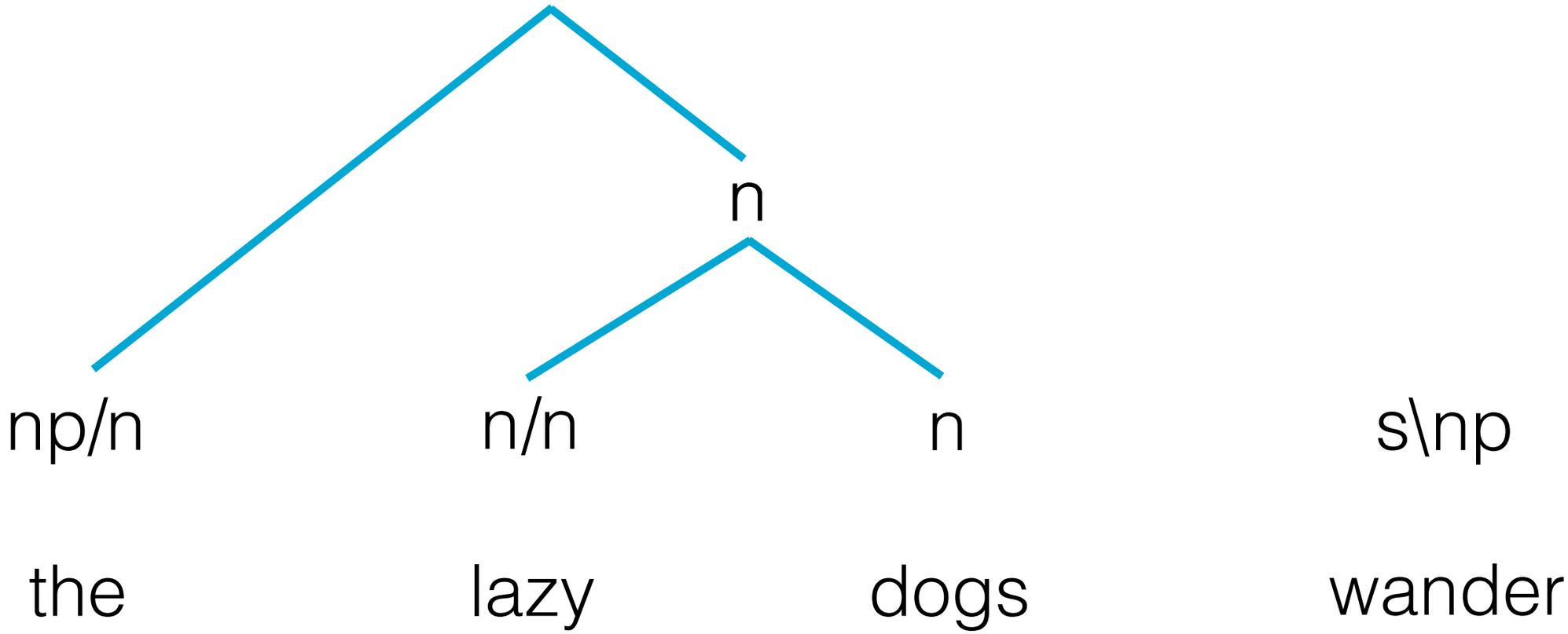
CCG Parsing



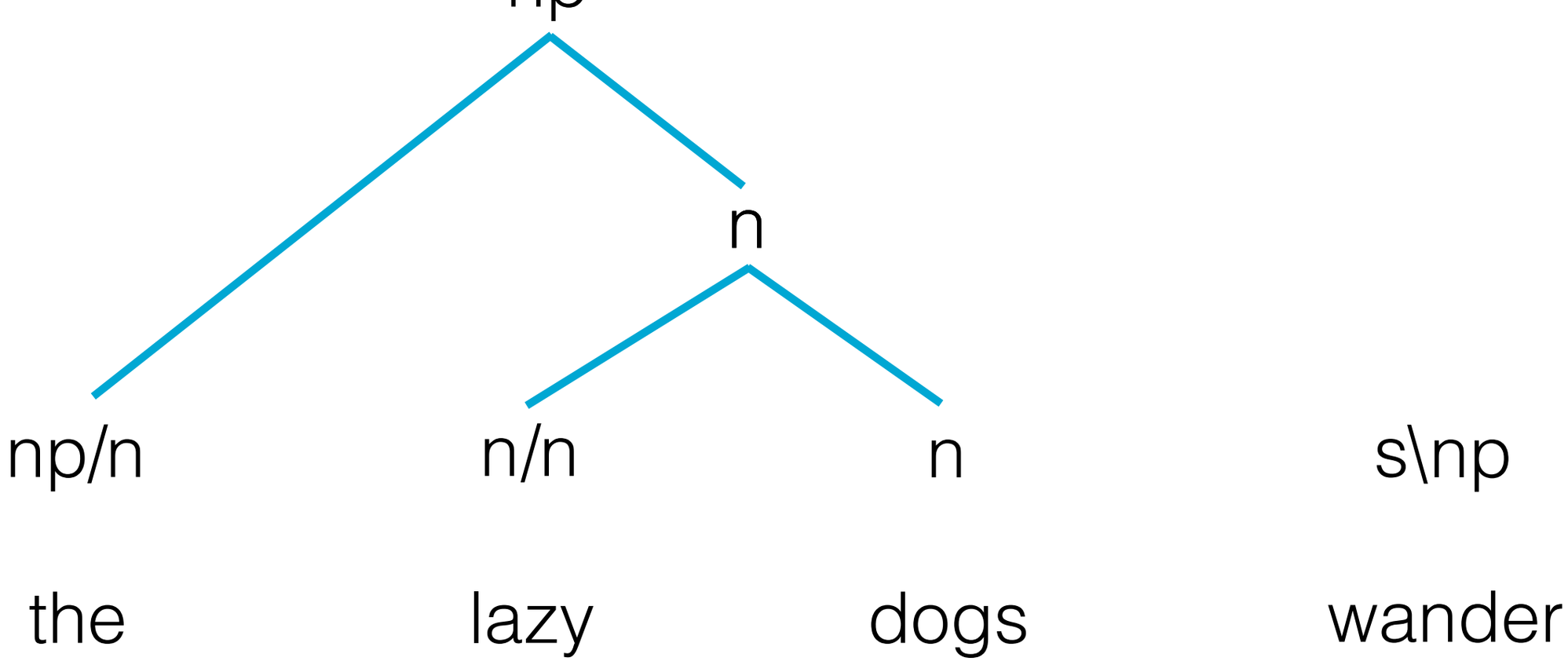
CCG Parsing



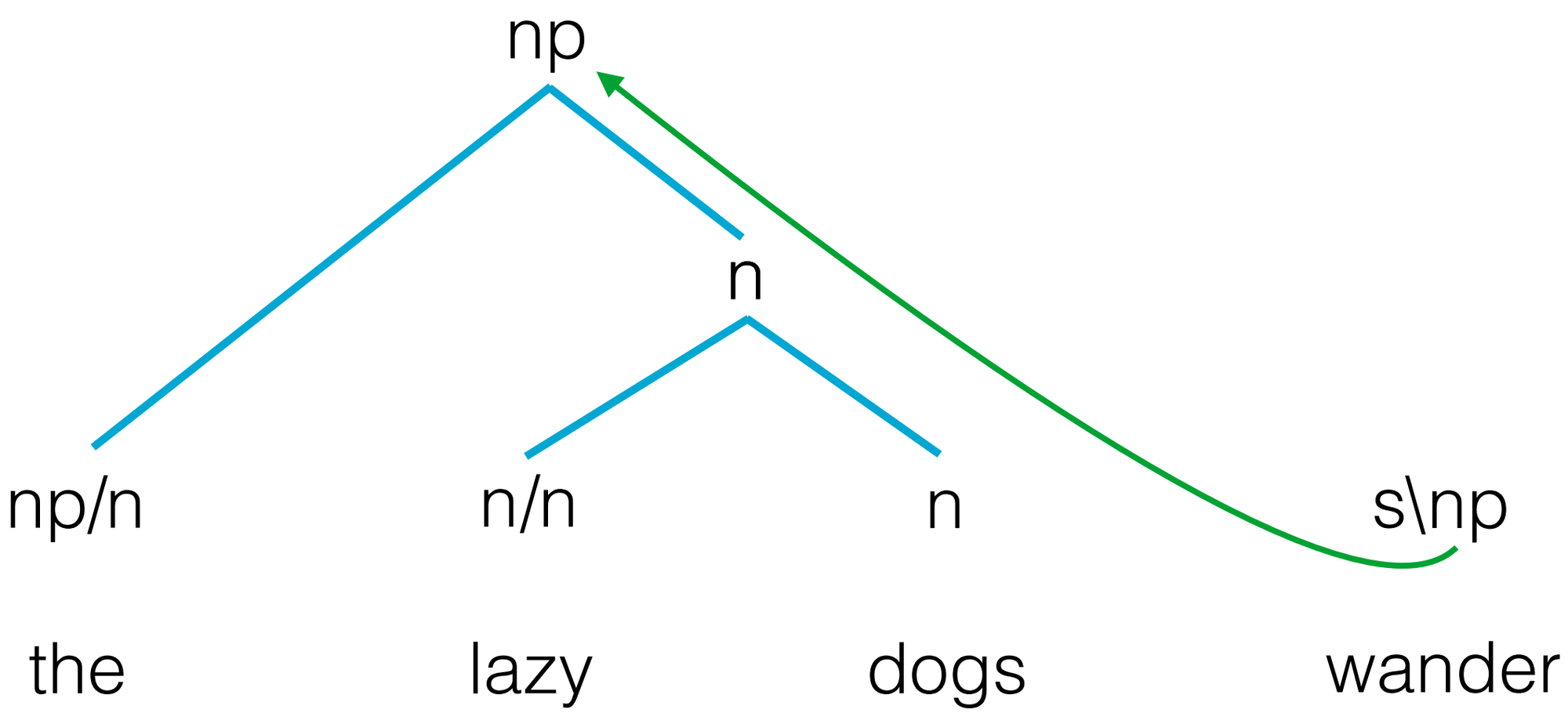
CCG Parsing



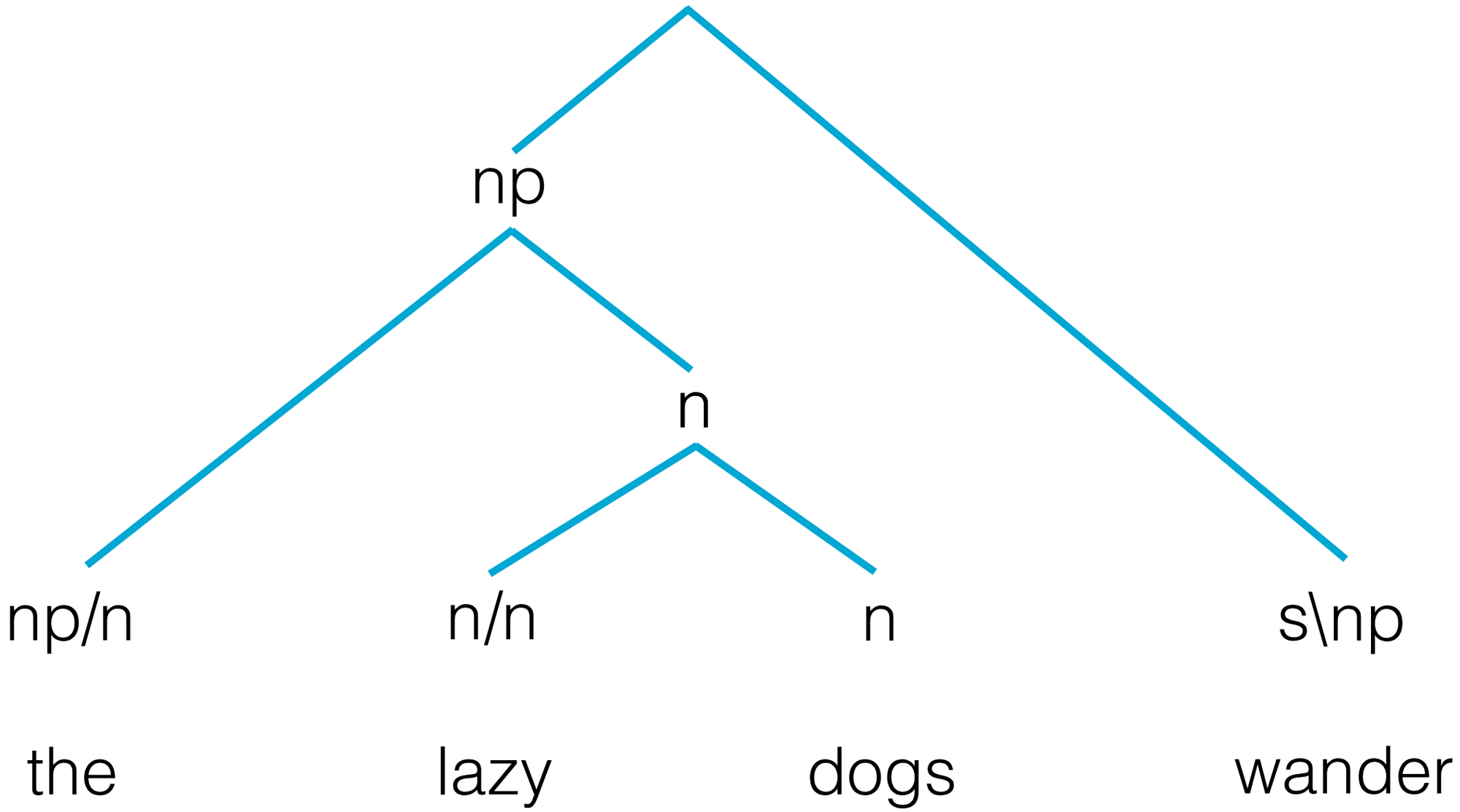
CCG Parsing



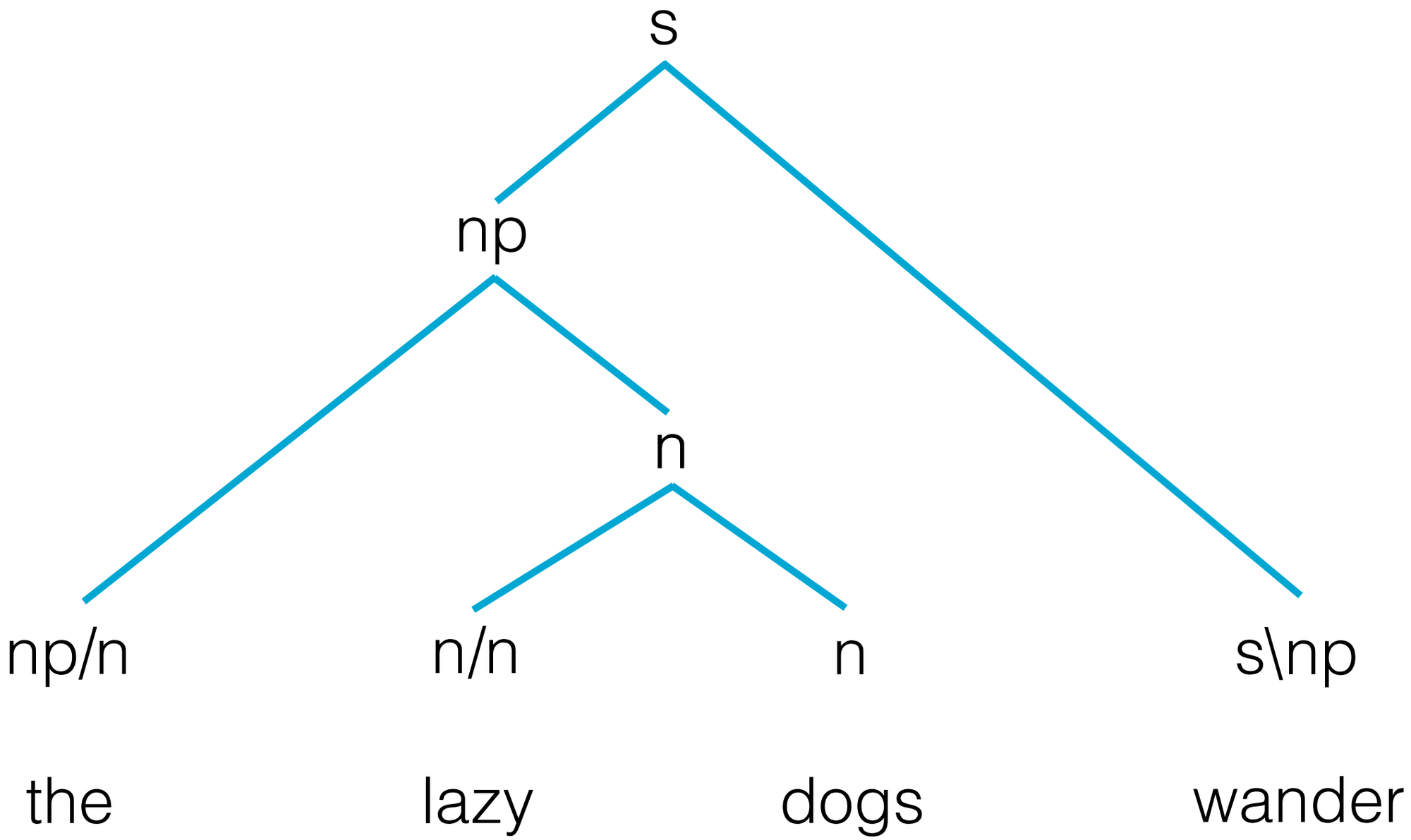
CCG Parsing



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*

Penn Treebank
parts-of-speech

48 tags

CCGBank
Categories

1,300+ categories

Our Strategy

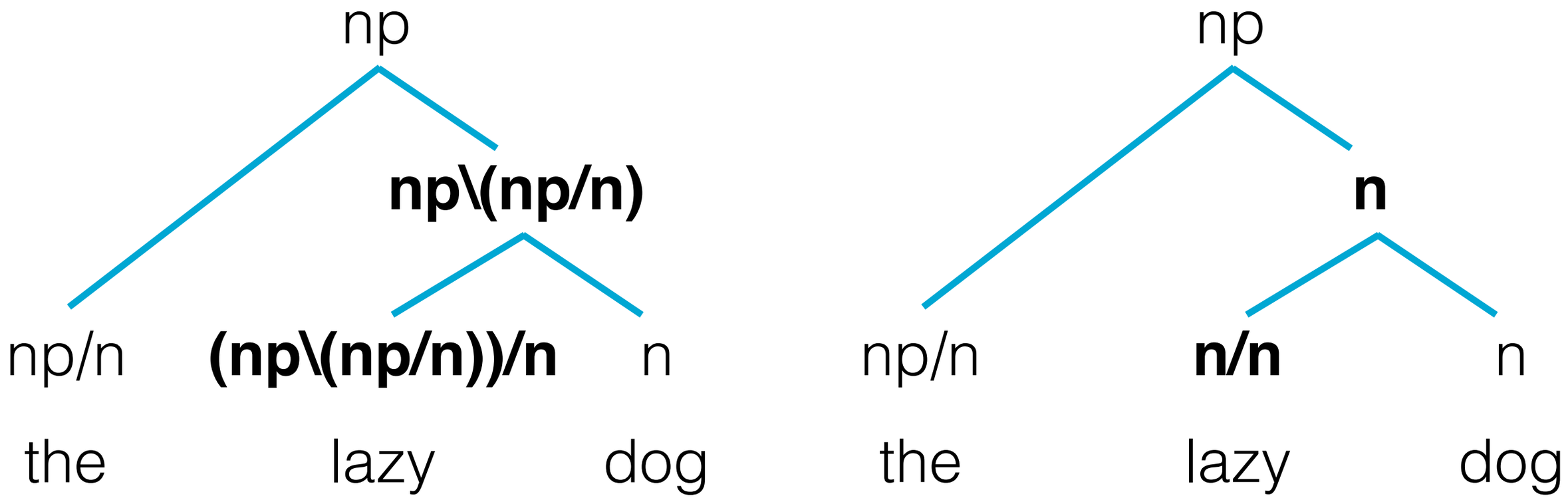
The grammar formalism *itself*
can be used to guide learning

Our Strategy

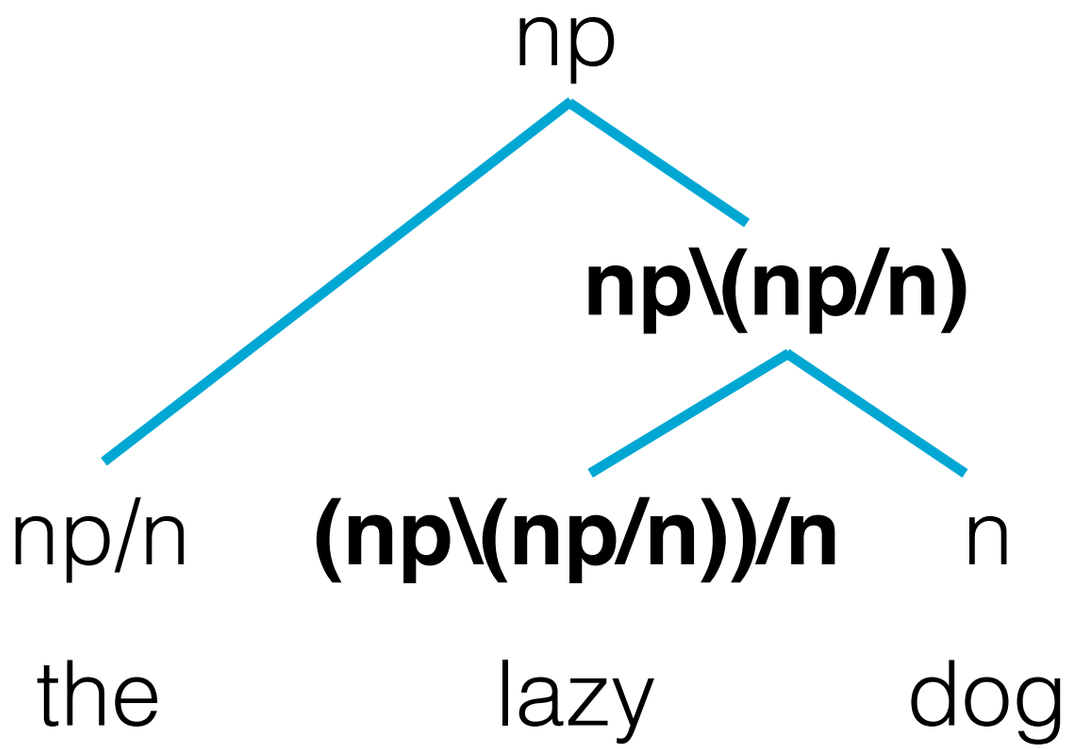
Incorporate *universal knowledge*
about grammar into learning

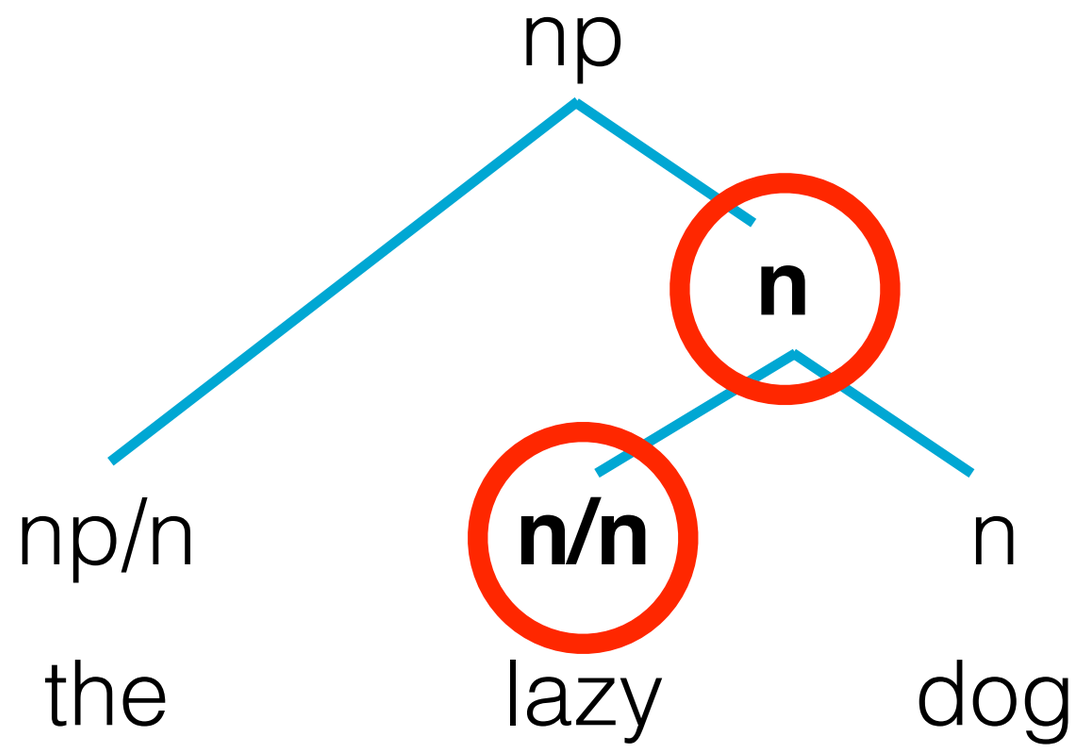
Universal Knowledge

Prefer Simpler Categories



Prefer Simpler Categories





Prefer Simpler Categories

buy := $(s_b\backslash np)/np$ appears **342** times in CCGbank

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

buy := $(((s_b\backslash np)/\mathbf{pp})/\mathbf{pp})/np$ appears **once**

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

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 Category Grammar

a	⟶	{s, np, n,…}	$p_{atom}(a) \times p_{term}$
A	⟶	B / B	$\overline{p_{term}} \times p_{fwd} \times p_{mod}$
A	⟶	B / C	$\overline{p_{term}} \times p_{fwd} \times \overline{p_{mod}}$
A	⟶	B \ B	$\overline{p_{term}} \times \overline{p_{fwd}} \times p_{mod}$
A	⟶	B \ C	$\overline{p_{term}} \times \overline{p_{fwd}} \times \overline{p_{mod}}$

Weighted Category Grammar

$a \longrightarrow \{s, np, n, \ldots\} \qquad p_{atom}(a) \times p_{term}$

$A \longrightarrow B \,/\, B \qquad \overline{p_{term}} \times p_{fwd} \times p_{mod}$

$+$

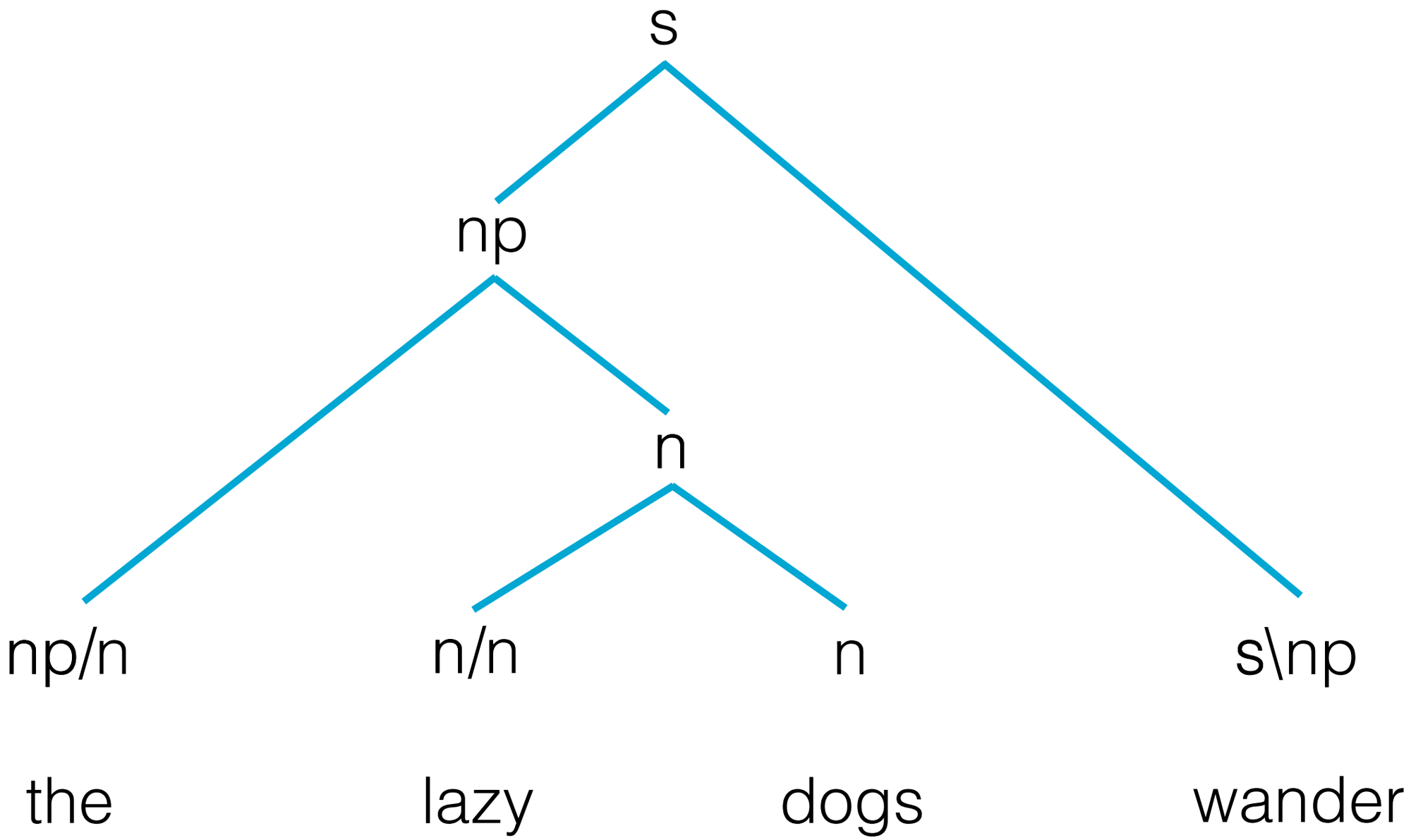
$A \longrightarrow B \,/\, C \qquad \overline{p_{term}} \times p_{fwd} \times \overline{p_{mod}}$

$A \longrightarrow B \,\backslash\, B \qquad \overline{p_{term}} \times \overline{p_{fwd}} \times p_{mod}$

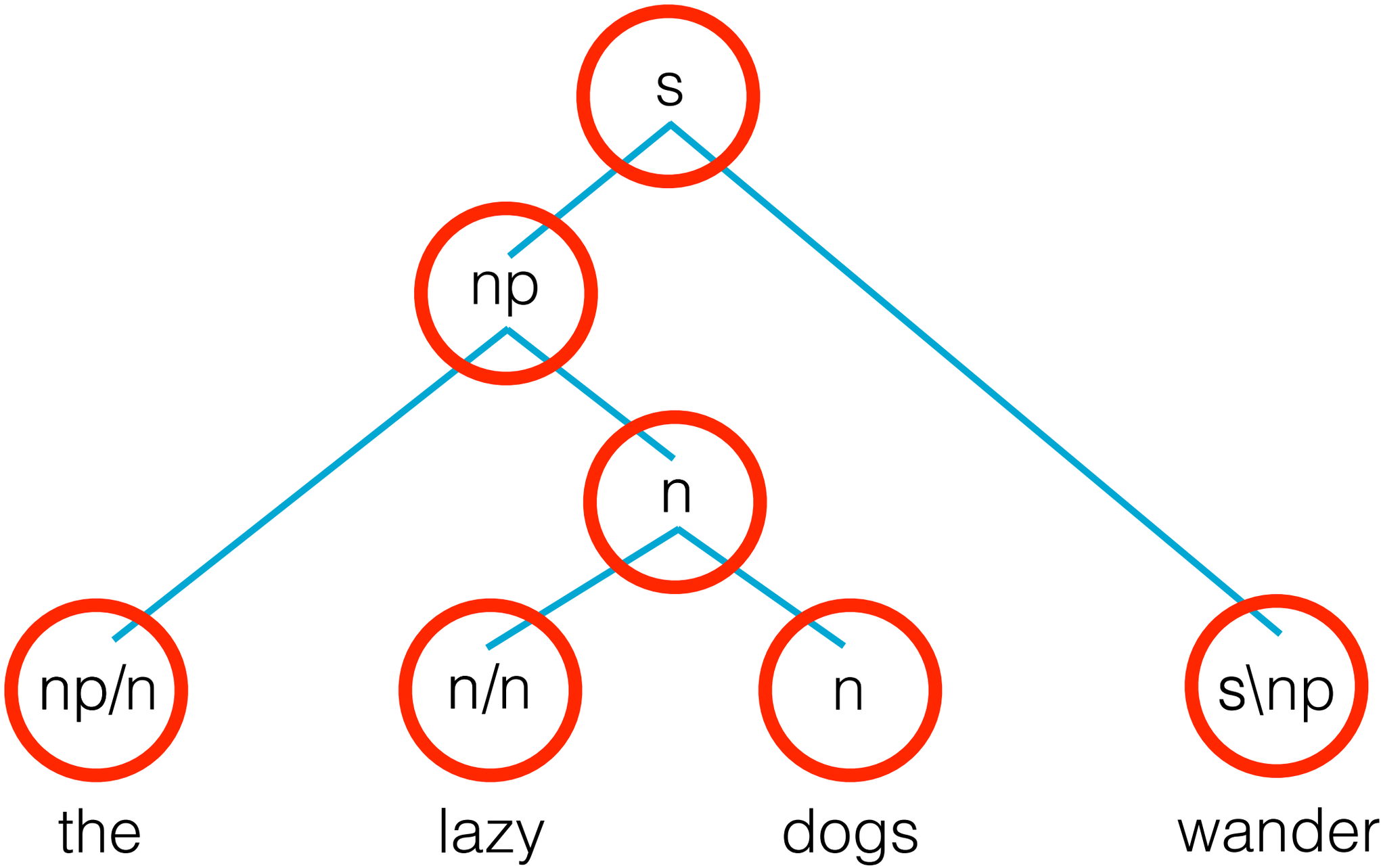
$+$

$A \longrightarrow B \,\backslash\, C \qquad \overline{p_{term}} \times \overline{p_{fwd}} \times \overline{p_{mod}}$

Prefer Likely Categories



Prefer Likely Categories

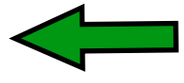


Type-Supervised Learning

unlabeled corpus

tag dictionary

same as
POS tagging

universal properties of the CCG formalism 

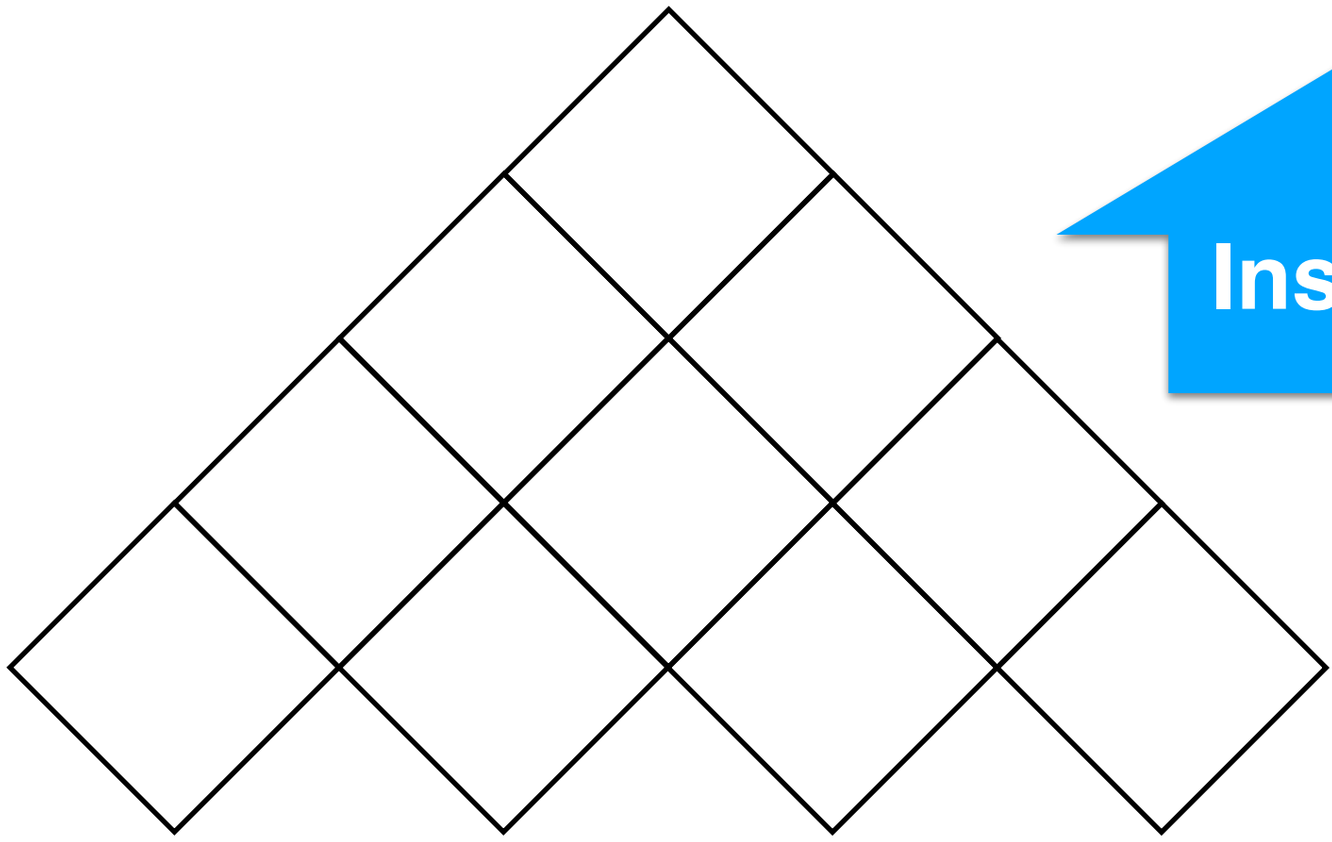
Posterior Inference

[Johnson, Griffiths, and Goldwater, 2007]

Posterior Inference

Priors (simple is good)

PCFG



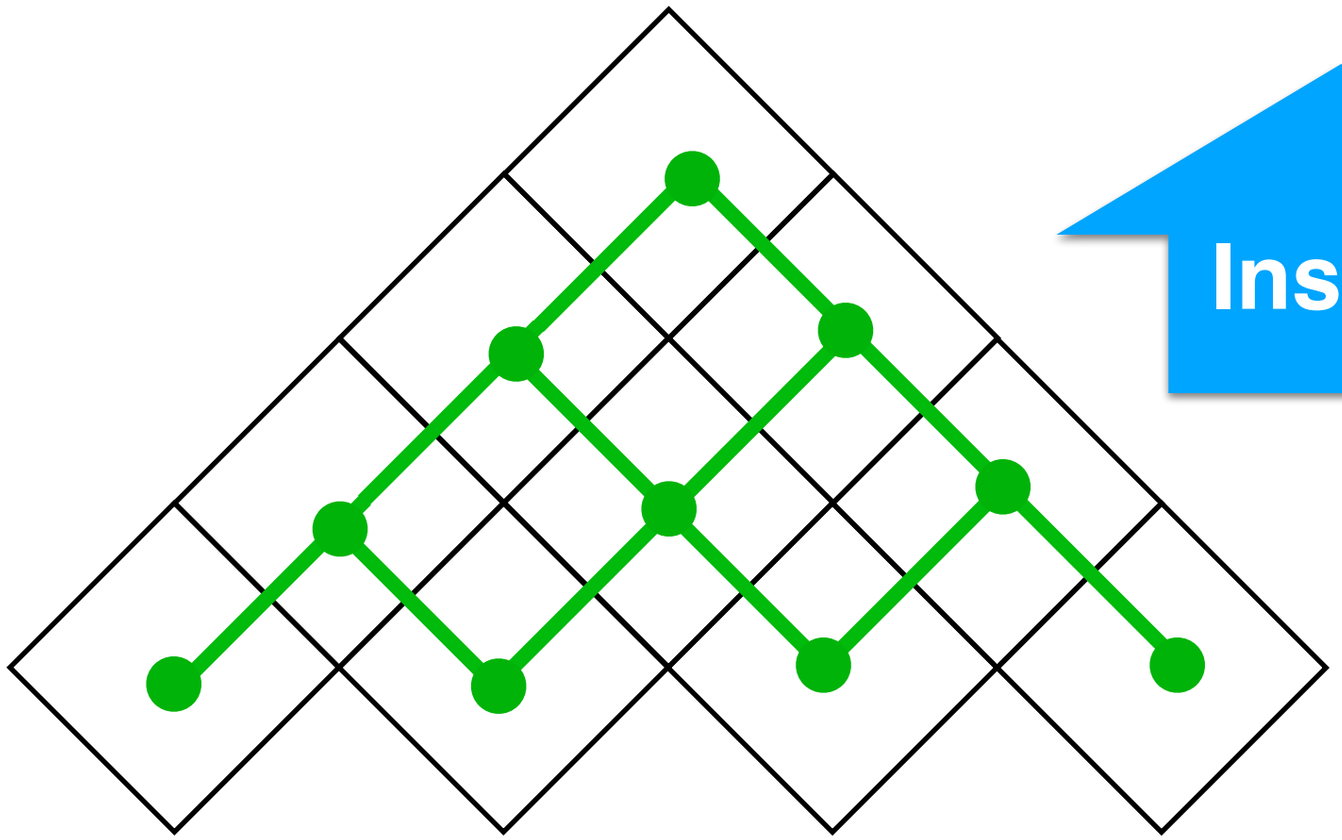
Inside

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

Posterior Inference

Priors (simple is good)

PCFG

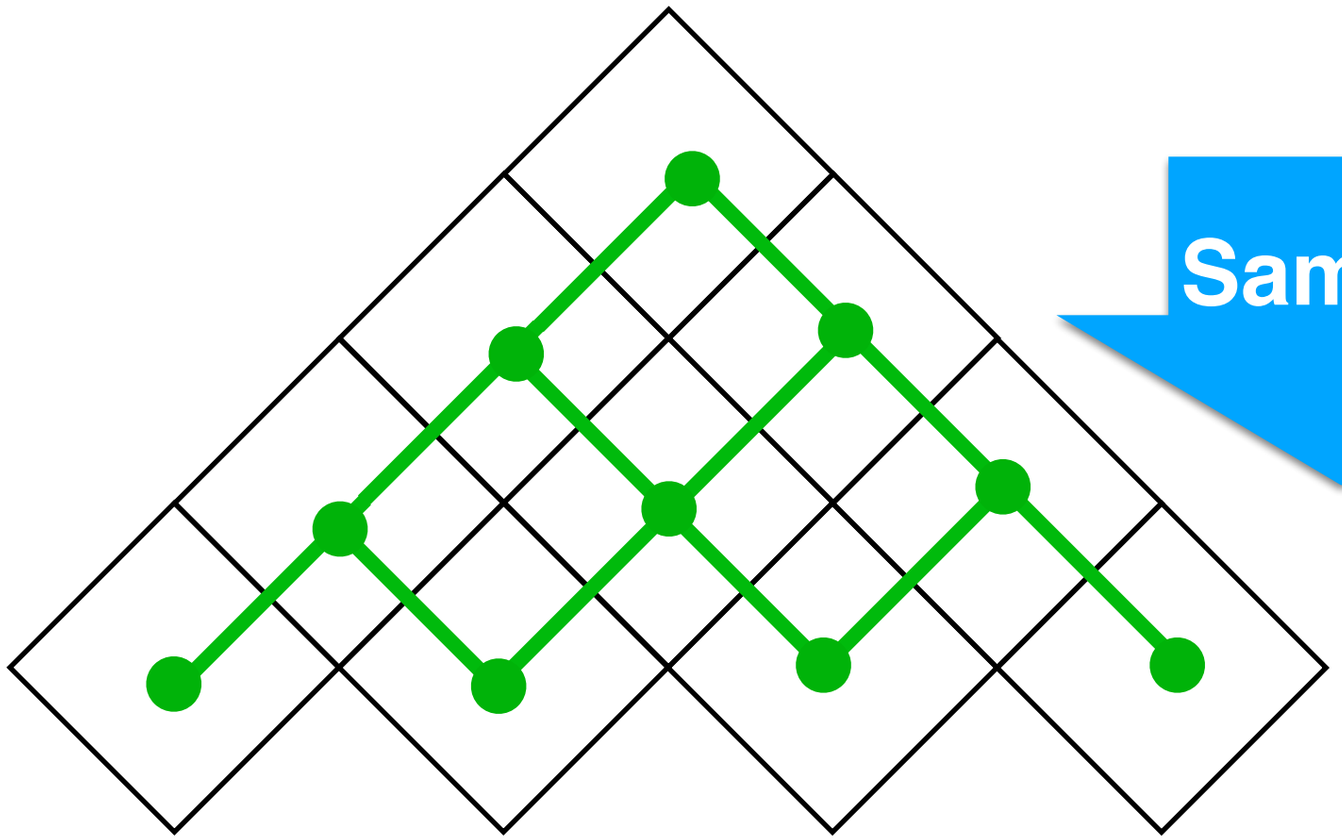


the	lazy	dogs	wander
np/n	n/n	n	n
	np	np	n/n
		(s\np)/np	np/n
			s\np
			...

Posterior Inference

Priors (simple is good)

PCFG



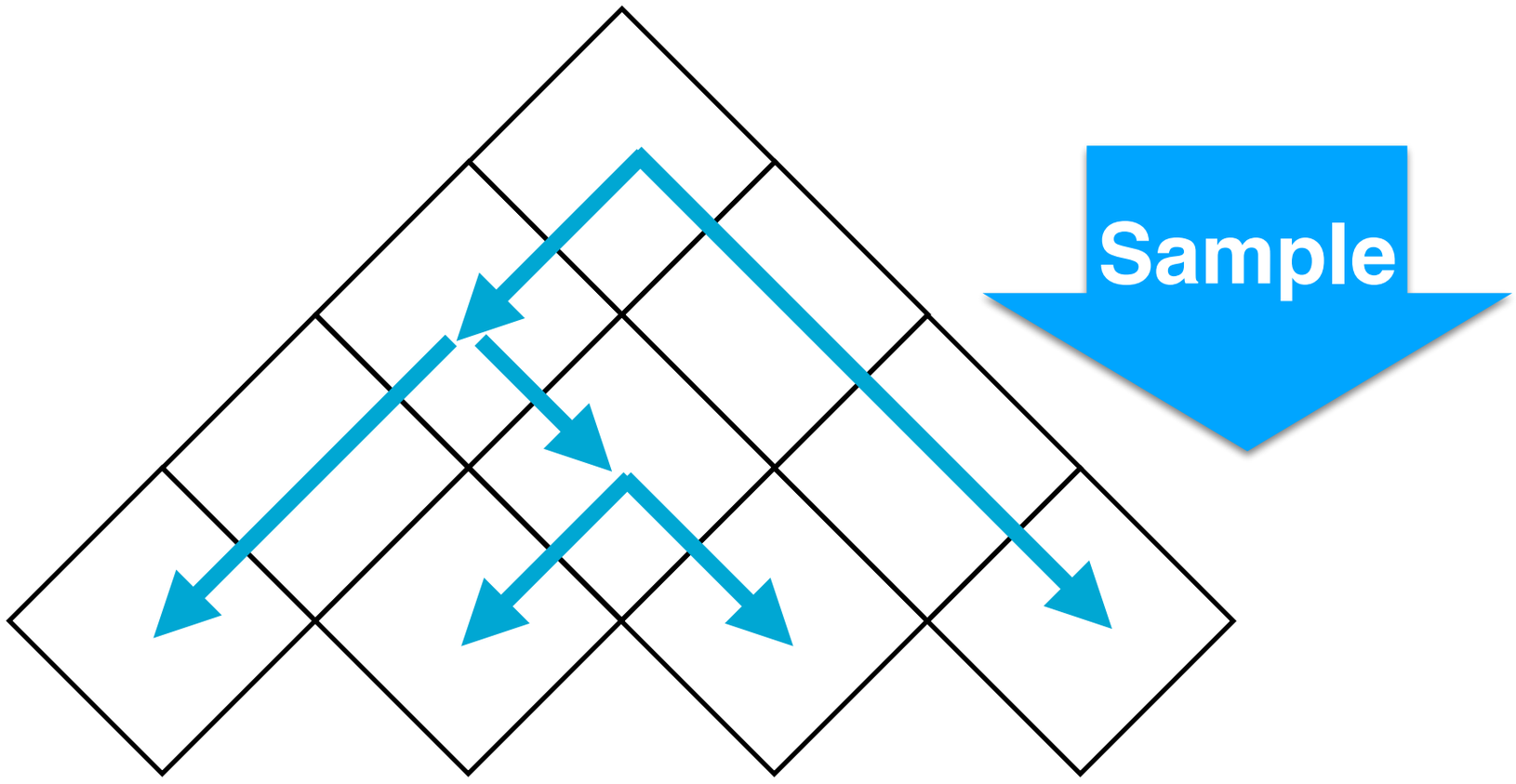
Sample

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

Posterior Inference

Priors (simple is good)

PCFG

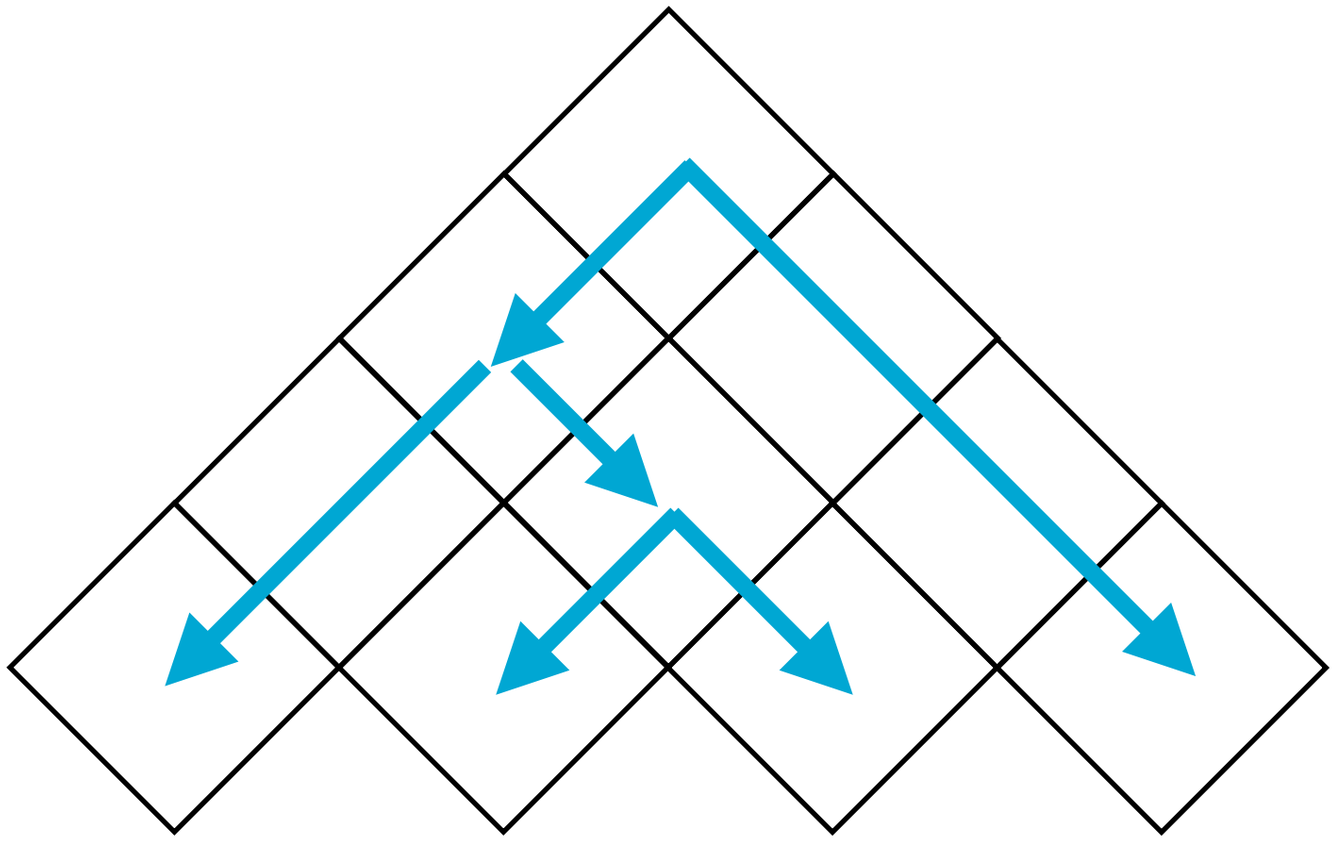


the	lazy	dogs	wander
np/n	n/n	n	n
	np	np	n/n
		(s\np)/np	np/n
			s\np
			...

Posterior Inference

Priors (simple is good)

PCFG

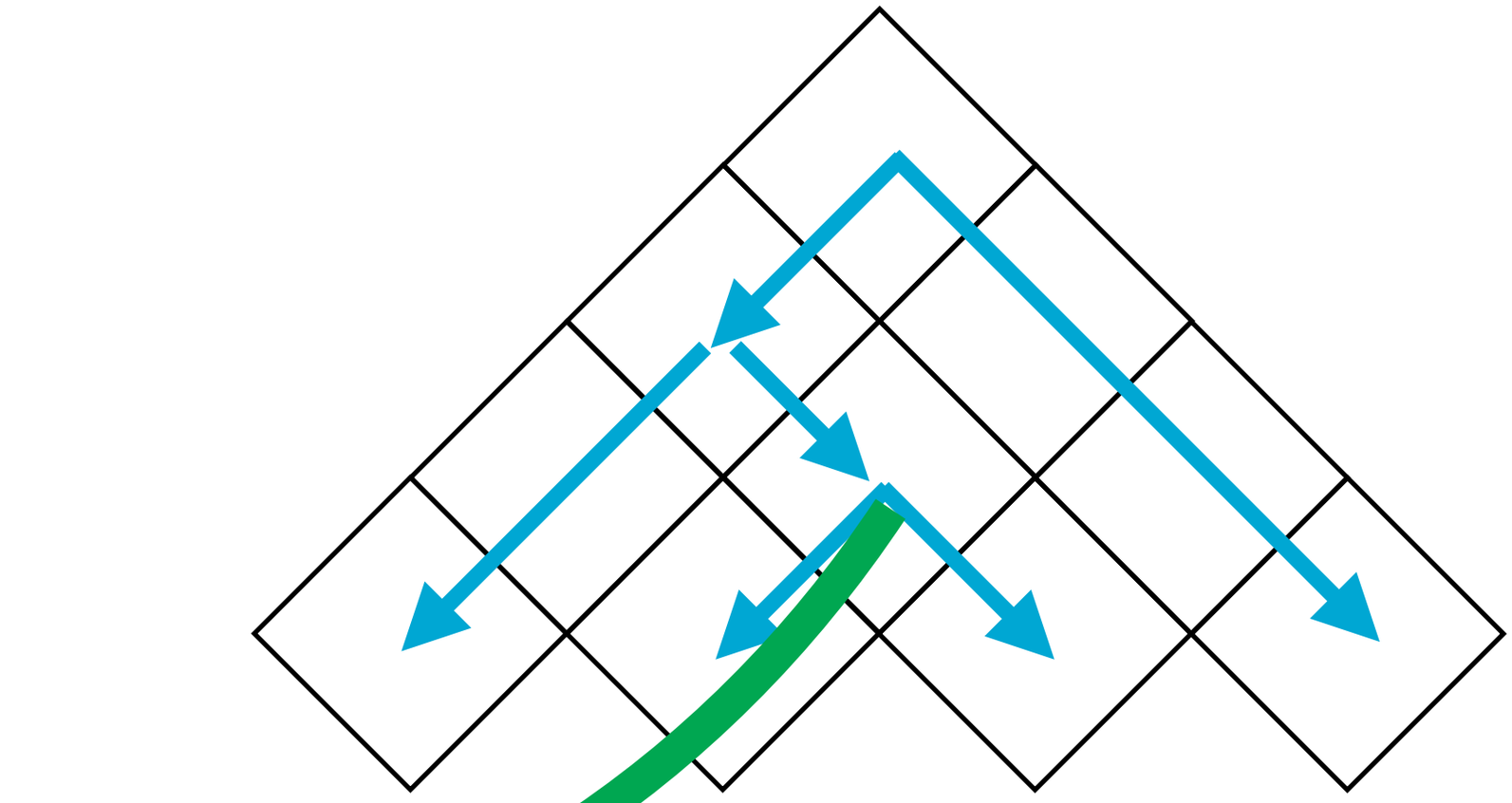


the	lazy	dogs	wander
np/n	n/n	n	n
	np	np	n/n
		(s\np)/np	np/n
			s\np
			...

Posterior Inference

Priors (simple is good)

PCFG

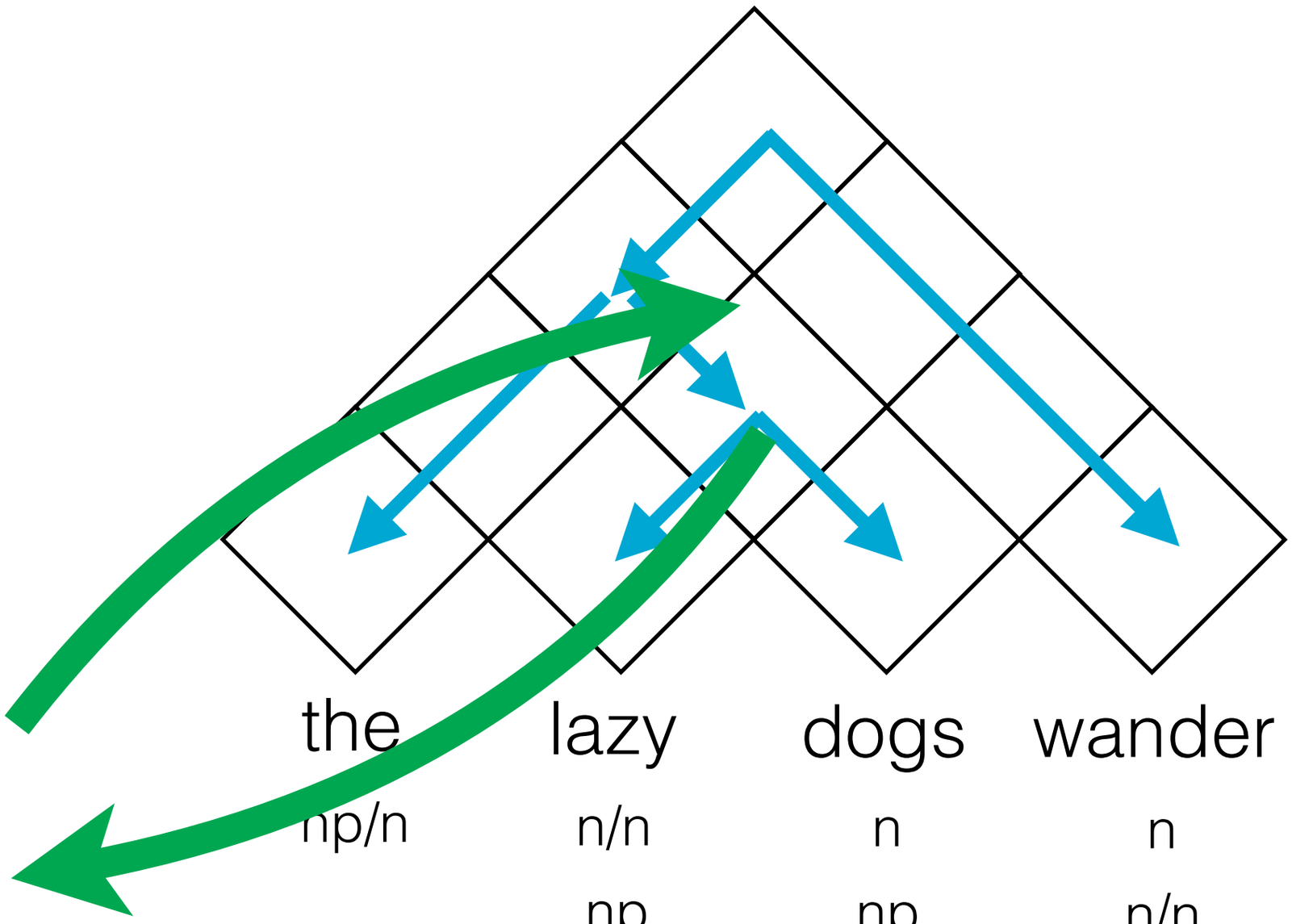


the	lazy	dogs	wander
np/n	n/n	n	n
	np	np	n/n
		(s\np)/np	np/n
			s\np
			...

Posterior Inference

Priors (simple is good)

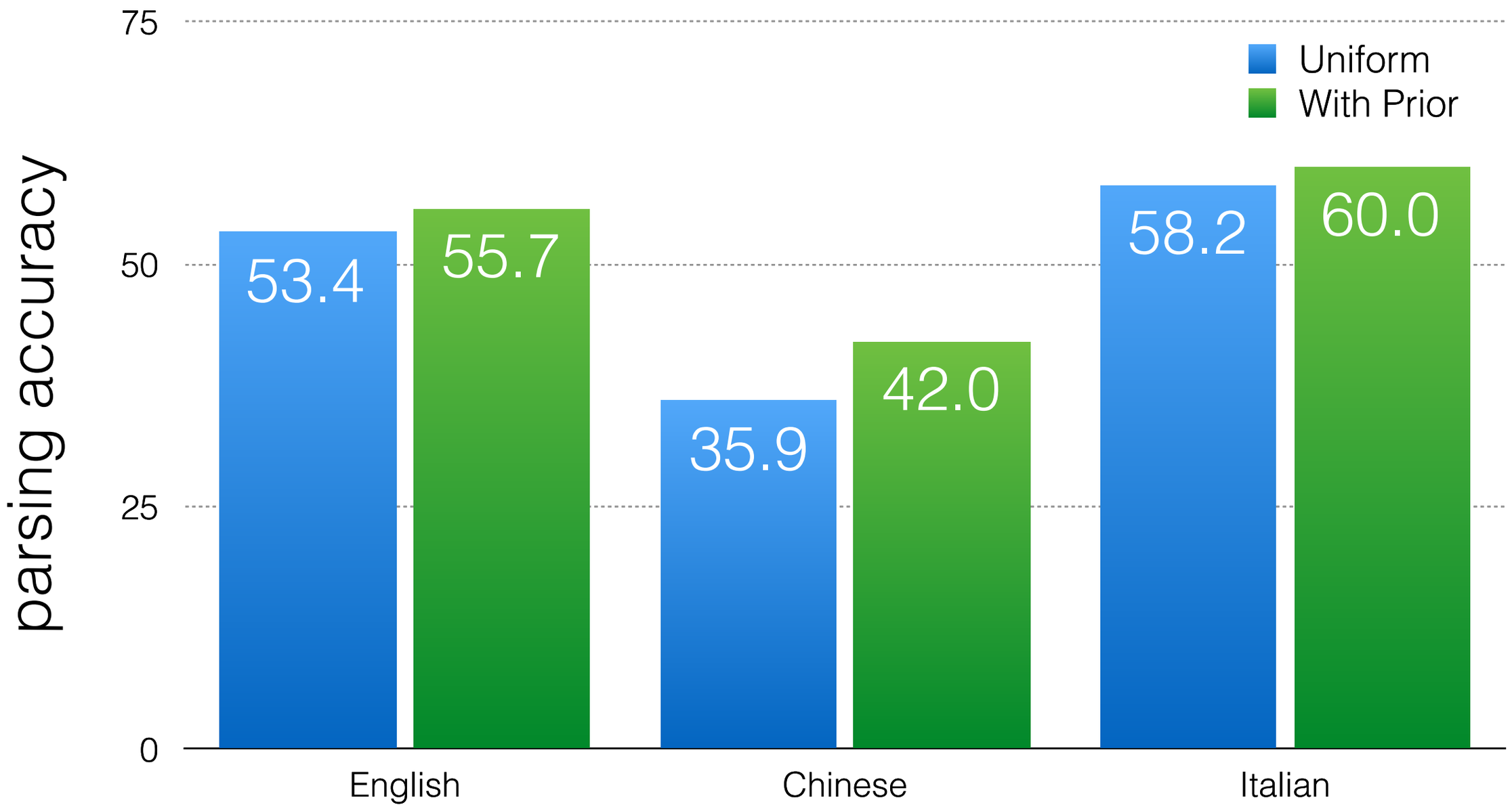
PCFG



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

Results

CCG Parsing Results



Conclusion

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