

Inducing Grammars from Linguistic Universals and Realistic Amounts of Supervision

Dan Garrette

Collaboration with:

Jason Baldridge

UT

Chris Dyer

CMU

Noah Smith

CMU

Weakly-Supervised Learning

Motivation

Can we learn new parsers quickly?

Type-Level Supervision

Type-Level Supervision

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

Type-Level Supervision

Used for POS tagging for 20+ years

[Kupiec, 1992]
[Merialdo, 1994]

Type-Level Supervision

Good POS tagger performance
even with low supervision

[Ravi & Knight 2009]

[Das & Petrov 2011]

[Garrette & Baldridge 2013]

[Garrette et al. 2013]

Combinatory Categorial Grammar (CCG)

CCG

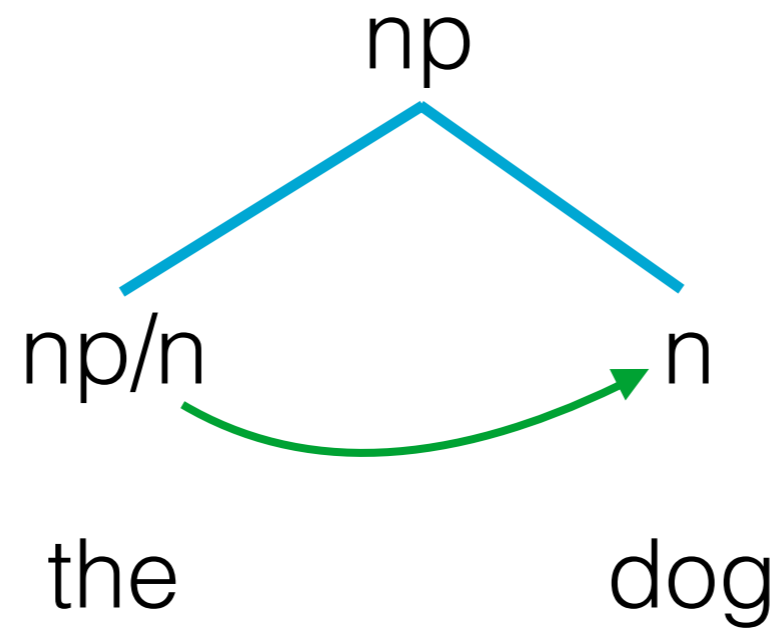
Every word token is associated with a **category**

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

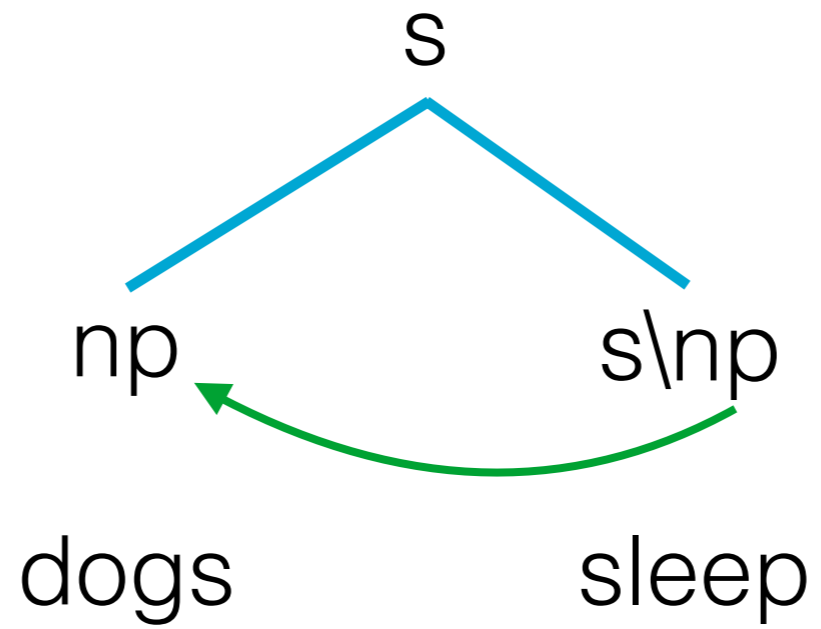
[Steedman, 2000]

[Steedman and Baldridge, 2011]

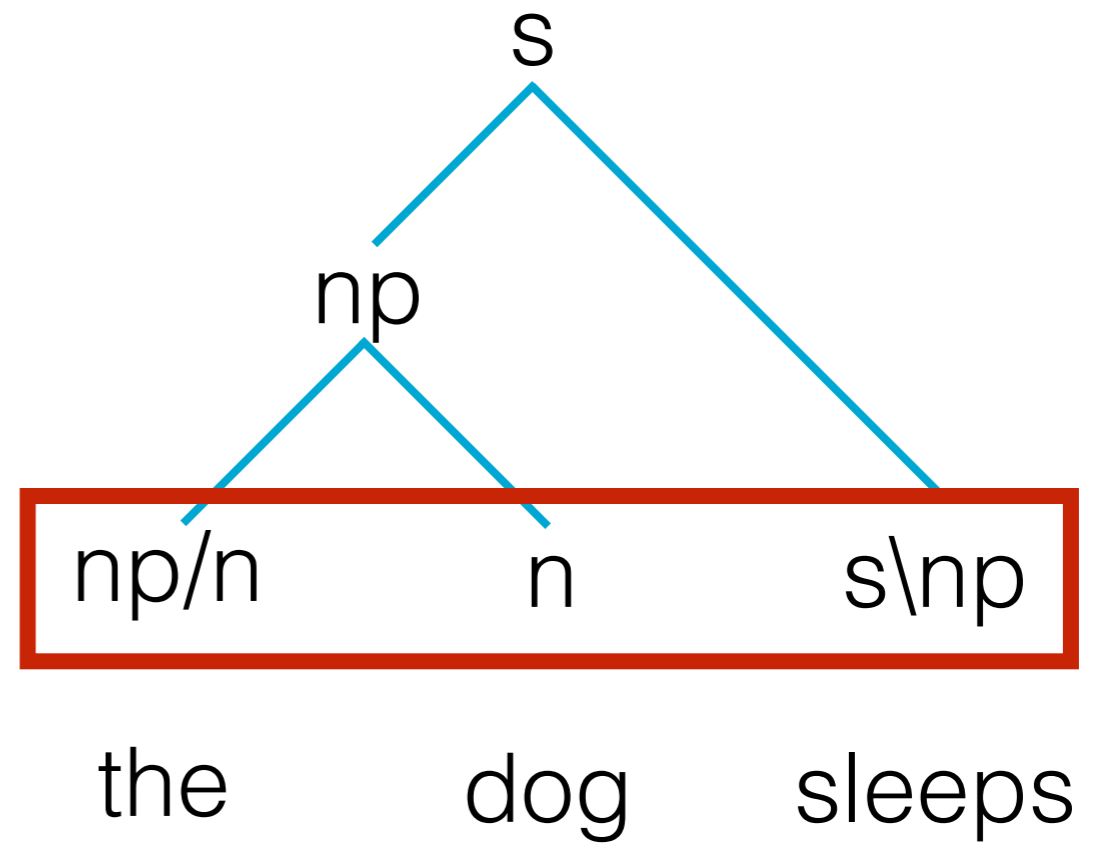
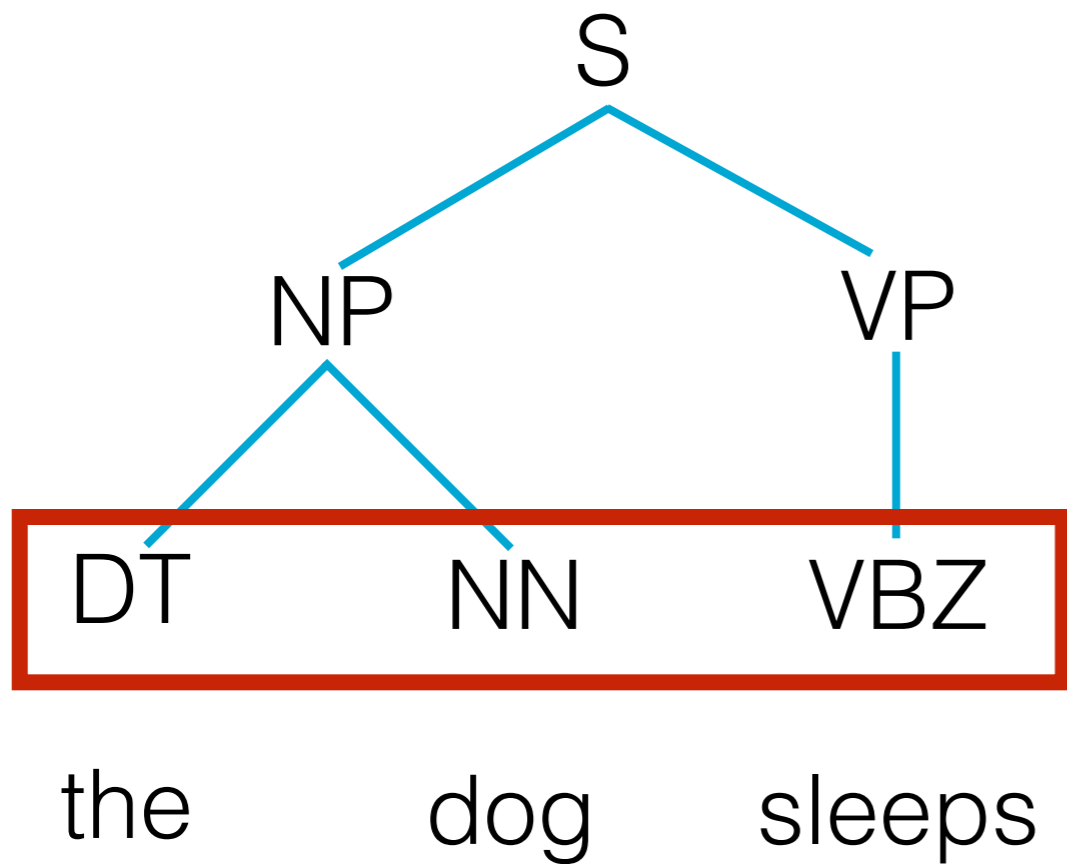
CCG



CCG



POS vs. Supertags



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

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

CCG Supertagging

CCG Supertagging

- Sequence tagging problem, like POS-tagging
- Building block for grammatical parsing

Supertagging

Type-supervised learning for supertagging
is much more difficult than for POS

Penn Treebank POS

Supertagging

Type-supervised learning for supertagging
is much more difficult than for POS

Penn Treebank POS

48 tags

CCGBank Supertags

1,239 tags

Supertagging

“almost parsing”

Why Supertagging?

np/n

the

n/n

lazy

n

dog

s\np

sleeps

Why Supertagging?

np/n

the

n/n

lazy

n

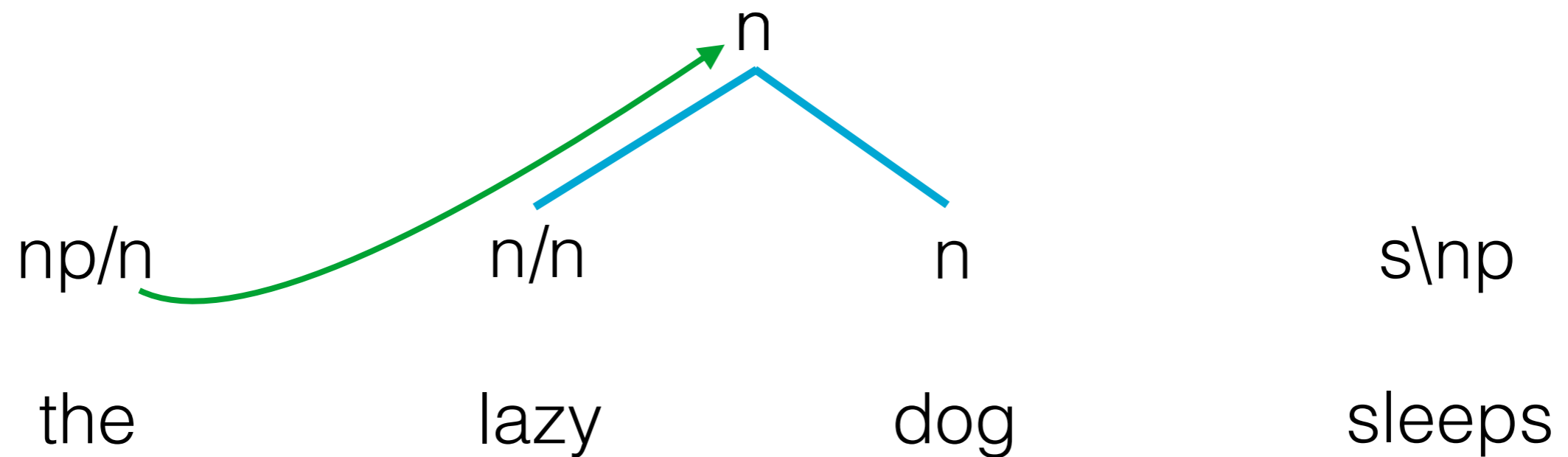
dog

s\np

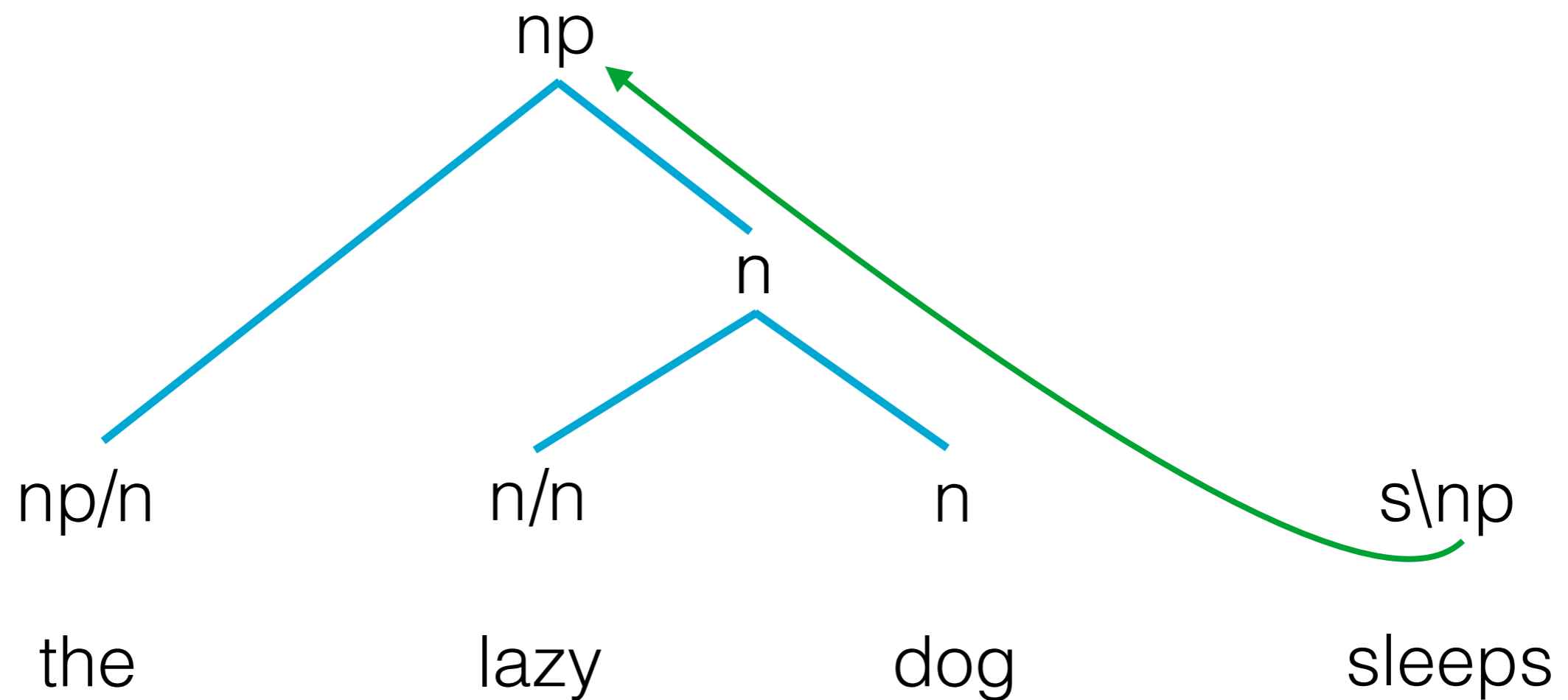
sleeps



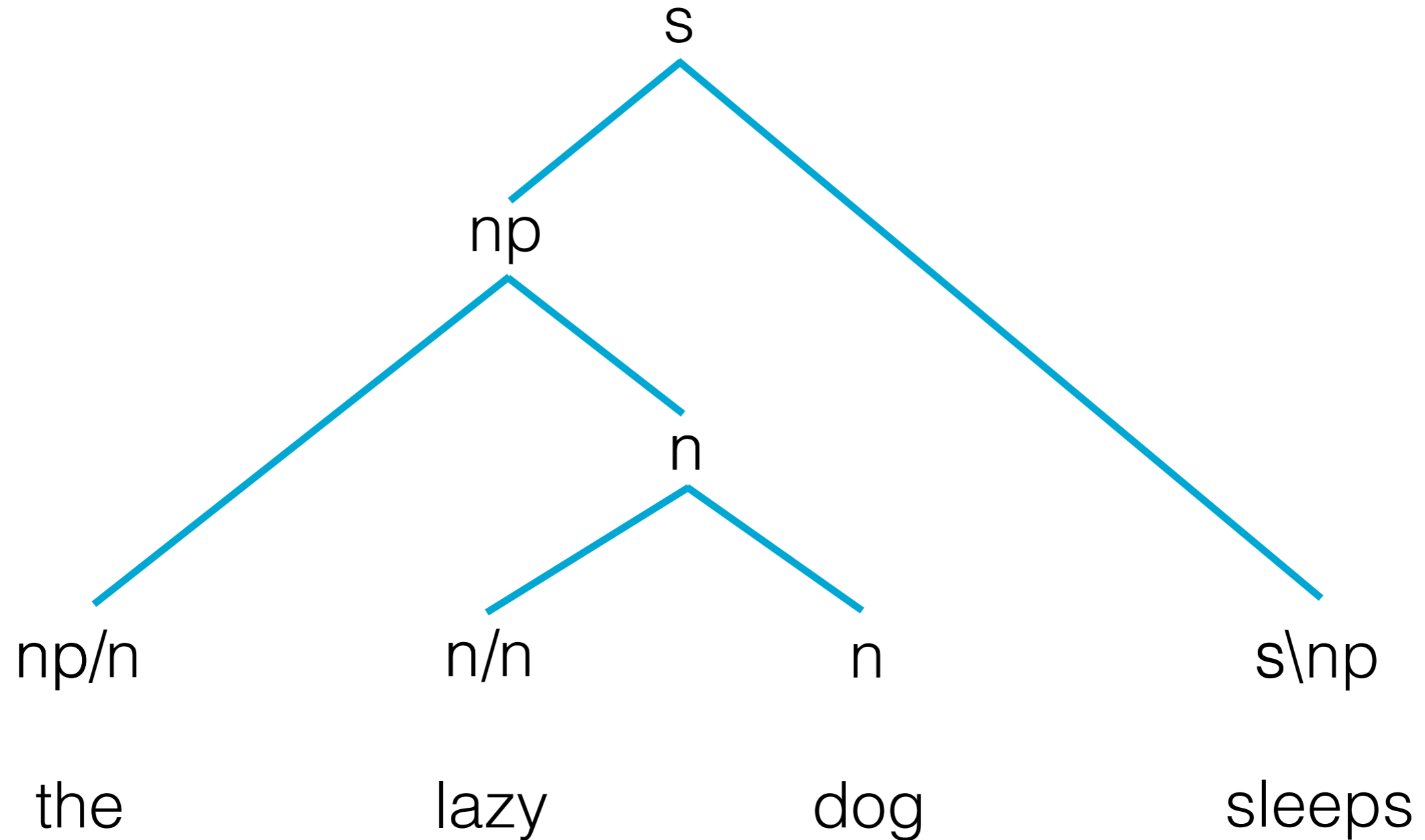
Why Supertagging?



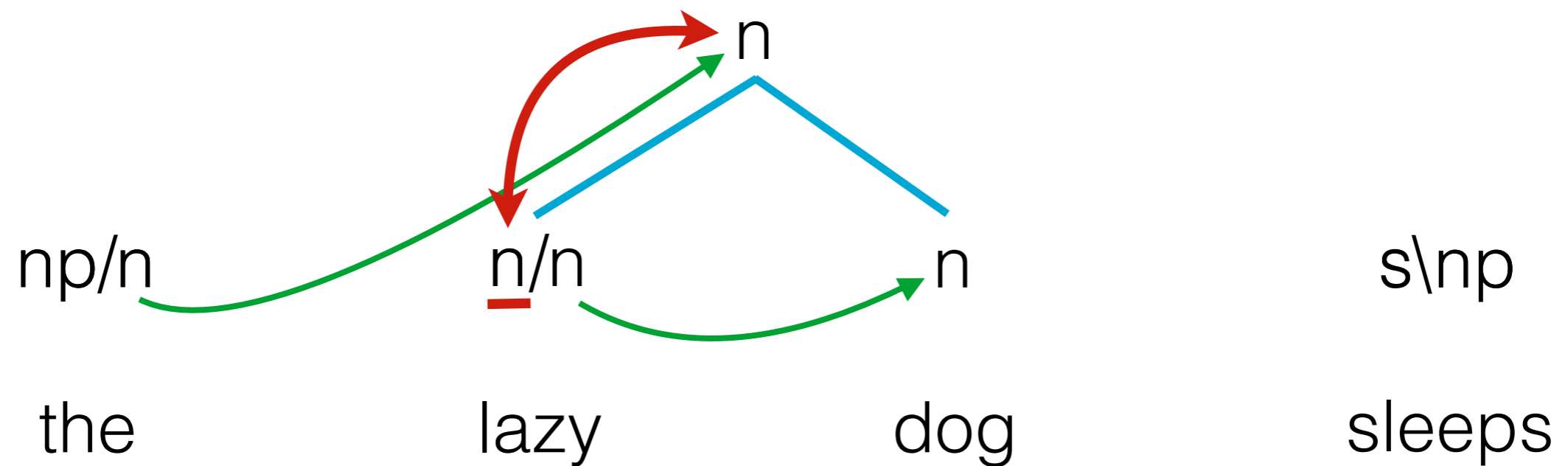
Why Supertagging?



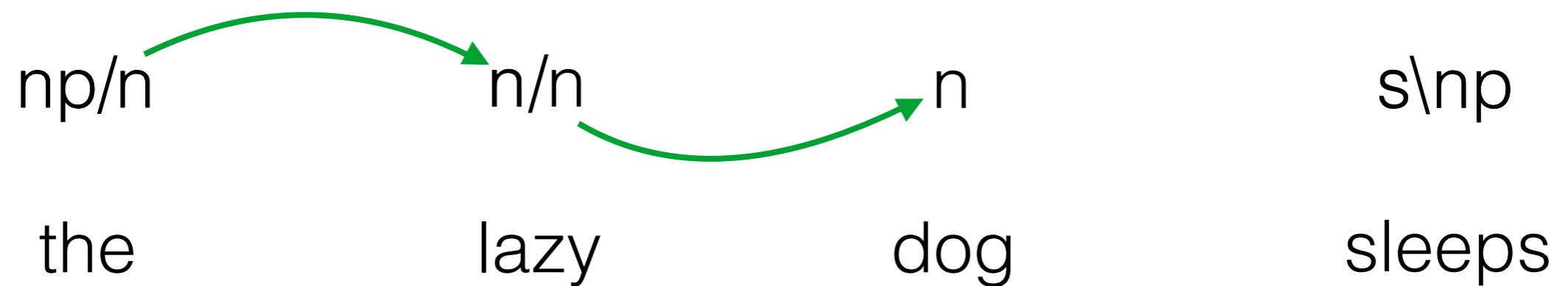
Why Supertagging?



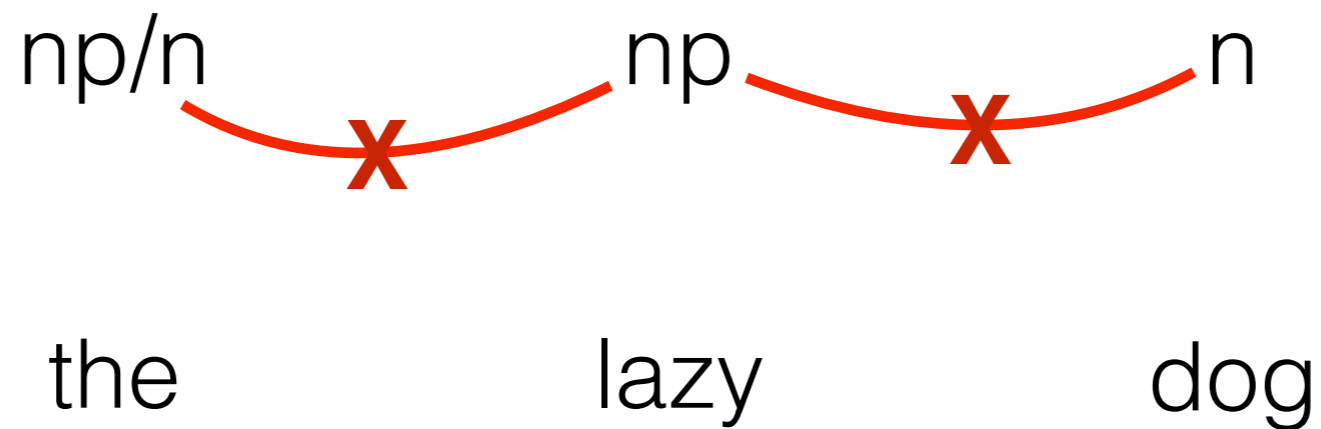
CCG Supertagging



CCG Supertagging

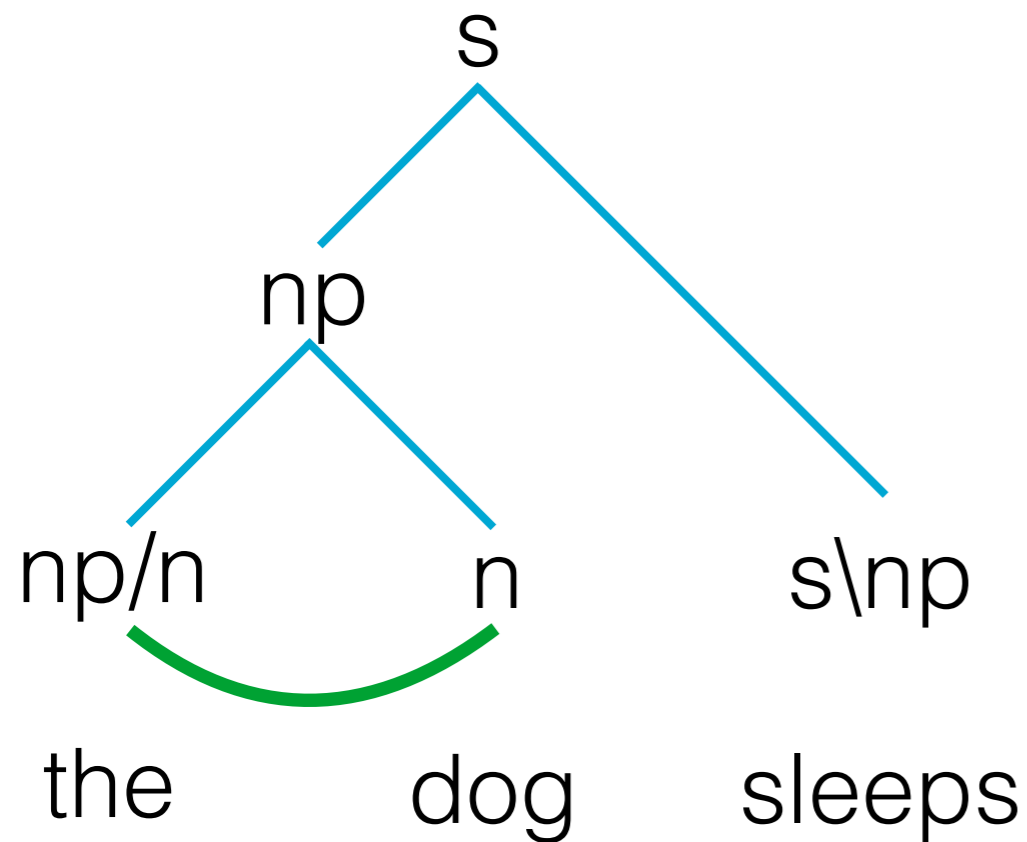


Principle #1

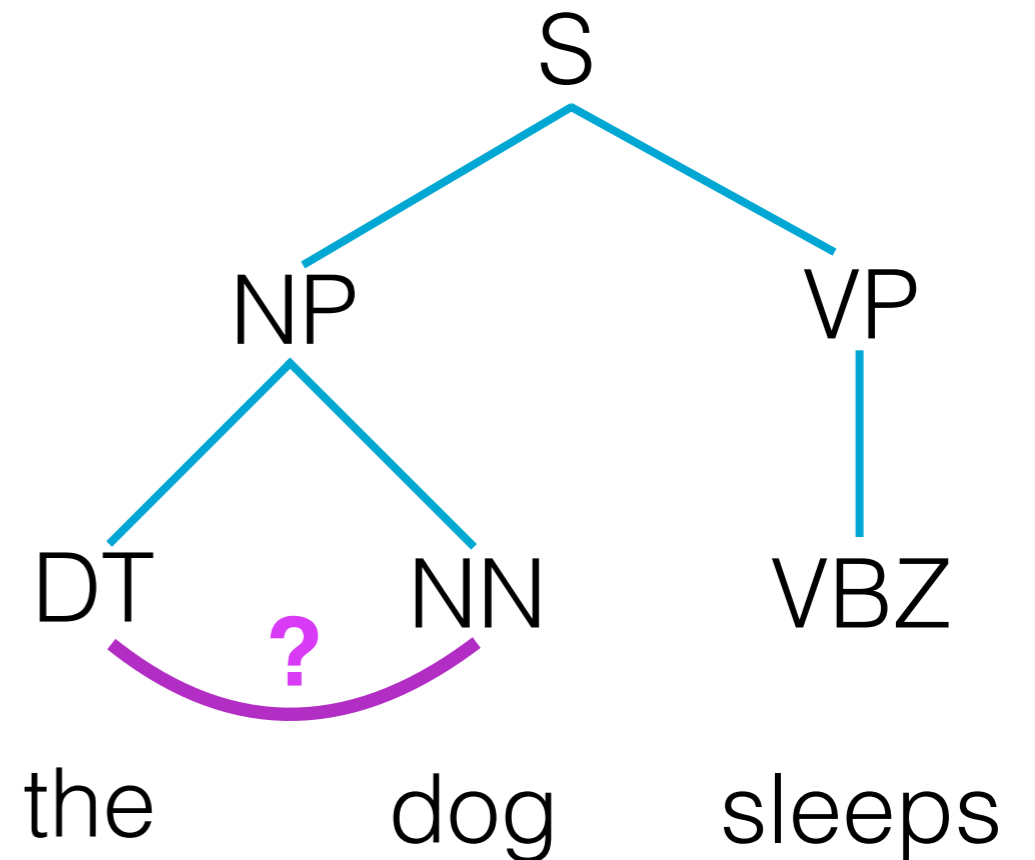


Prefer Connections

Supertags vs. POS



universal, intrinsic
grammar properties



all relationships
must be learned

Principle #2

np/n

(np\ \backslash (np/n))/n

n

the

lazy

dog

Avoid Complex Categories

CCG Supertagging

the

np/n

lazy

np

dog

n

$(np \backslash (np/n)) / n$

n/n

HMM Transition Prior

$$P(\mathbf{t} \rightarrow \mathbf{u}) = \lambda \cdot \underbrace{P(\mathbf{u})}_{\text{simple is good}} + (1-\lambda) \cdot \underbrace{P(\mathbf{t} \rightarrow \mathbf{u})}_{\text{connecting is good}}$$

Type-Supervised Learning

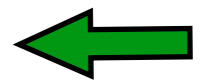
unlabeled corpus

tag dictionary



same as
POS tagging

universal properties of the CCG formalism

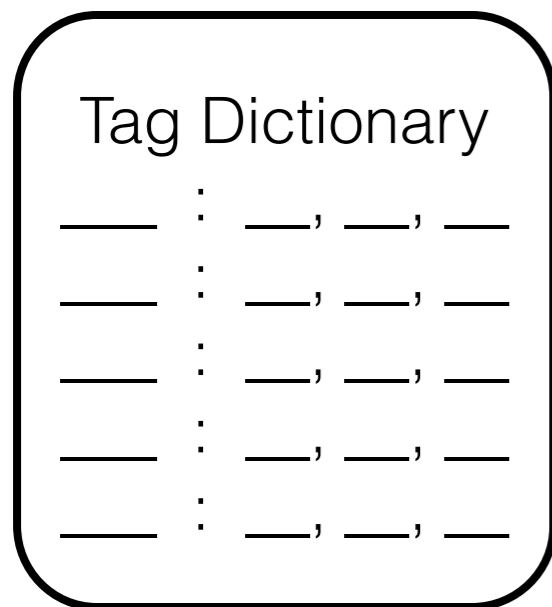
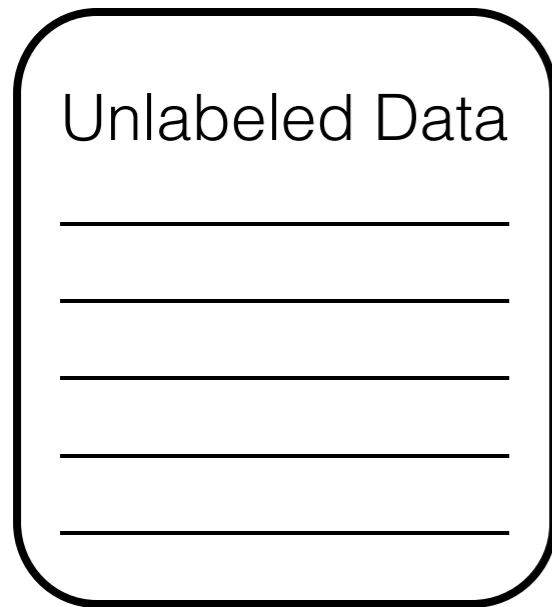


Training

Posterior Inference

Forward-Filter Backward-Sample (FFBS)

Posterior Inference



Posterior Inference

the

lazy

dogs

wander

Tag Dictionary

—	:	—,	—,	—
—	:	—,	—,	—
—	:	—,	—,	—
—	:	—,	—,	—
—	:	—,	—,	—

Posterior Inference

the

lazy

dogs

wander

np/n

n/n

n

np

np

(s\np)/np

Posterior Inference

the

lazy

dogs

wander

np/n

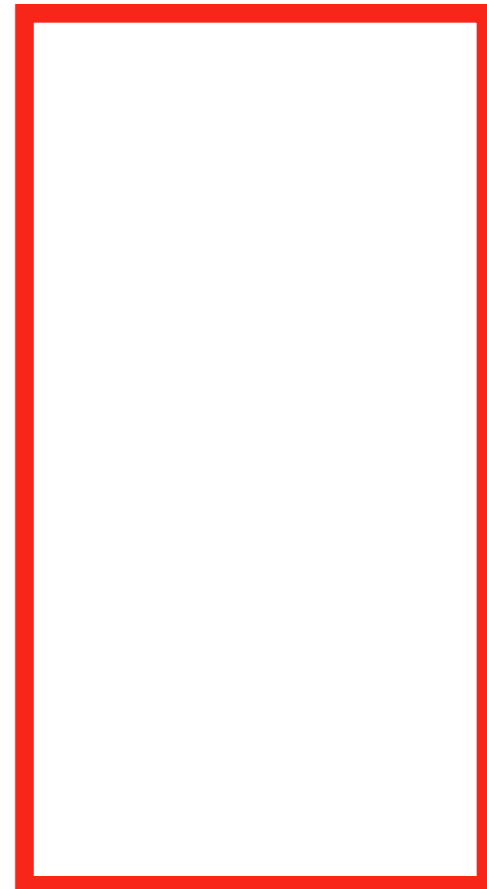
n/n

n

np

np

(s\np)/np



Posterior Inference

the

lazy

dogs

wander

np/n

n/n

n

n

np

np

n/n

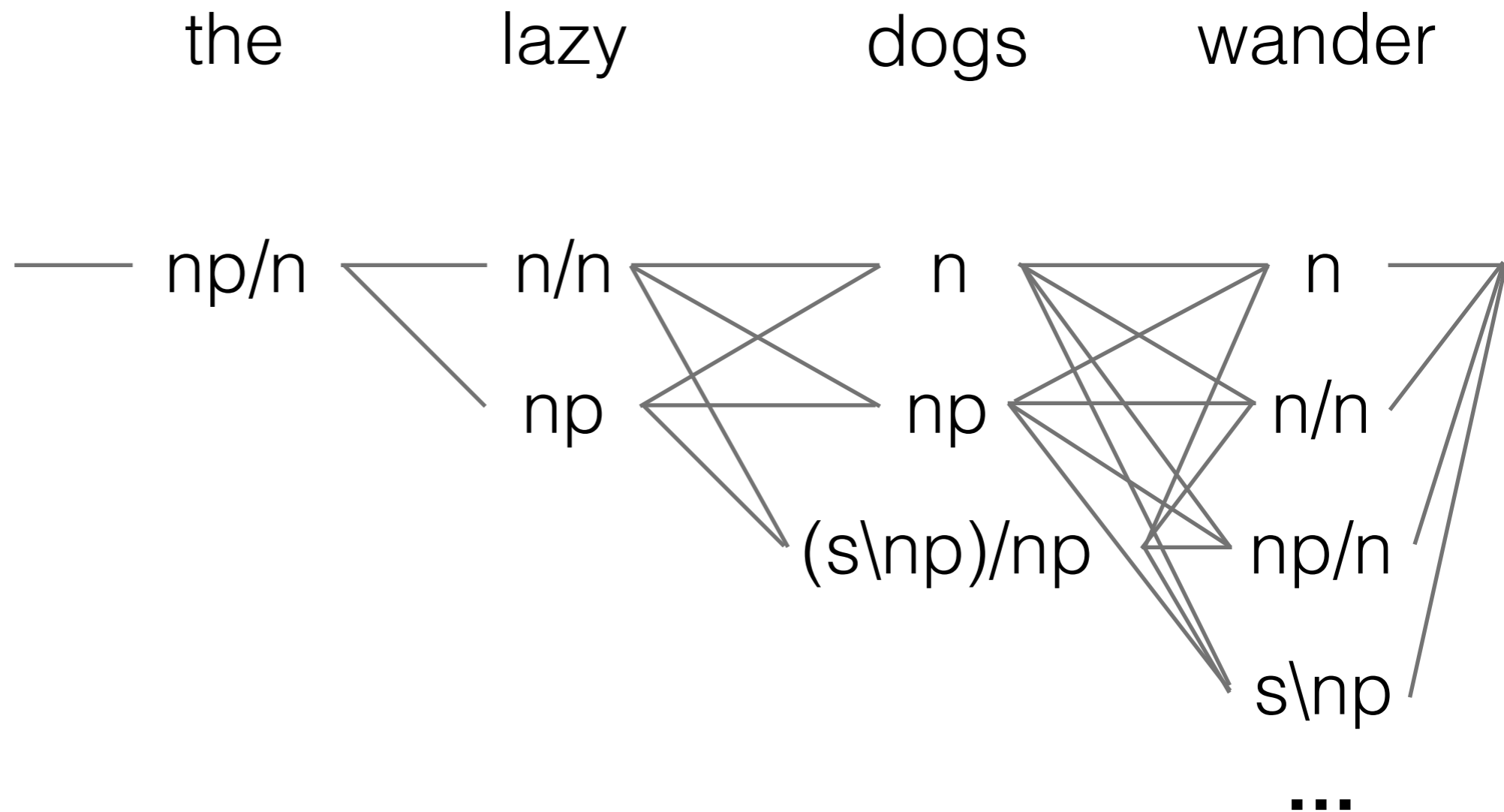
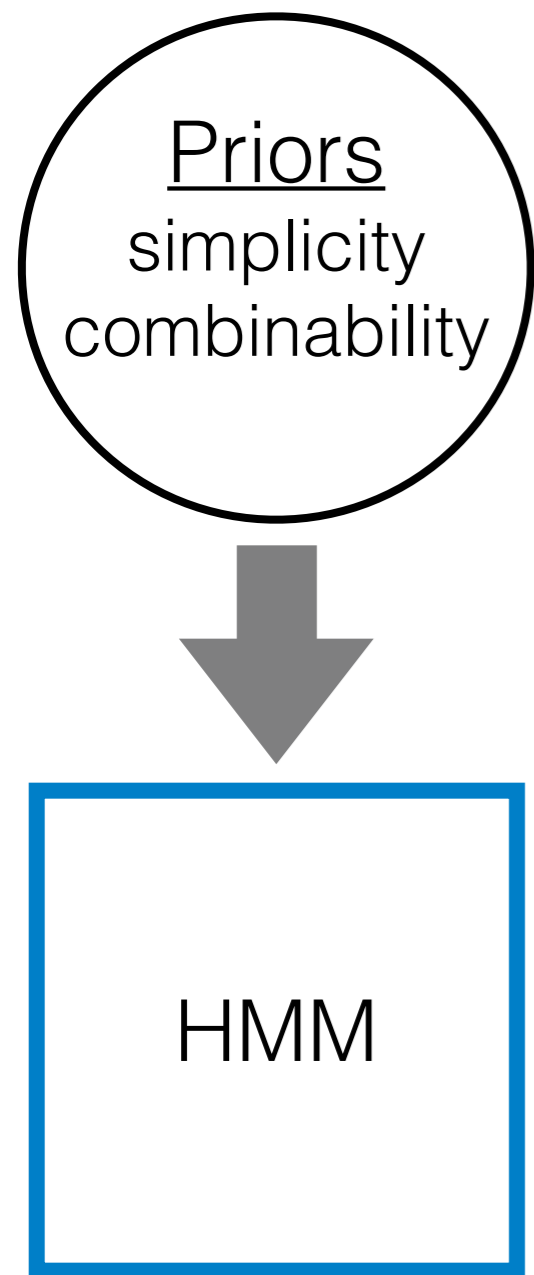
(s\np)/np

np/n

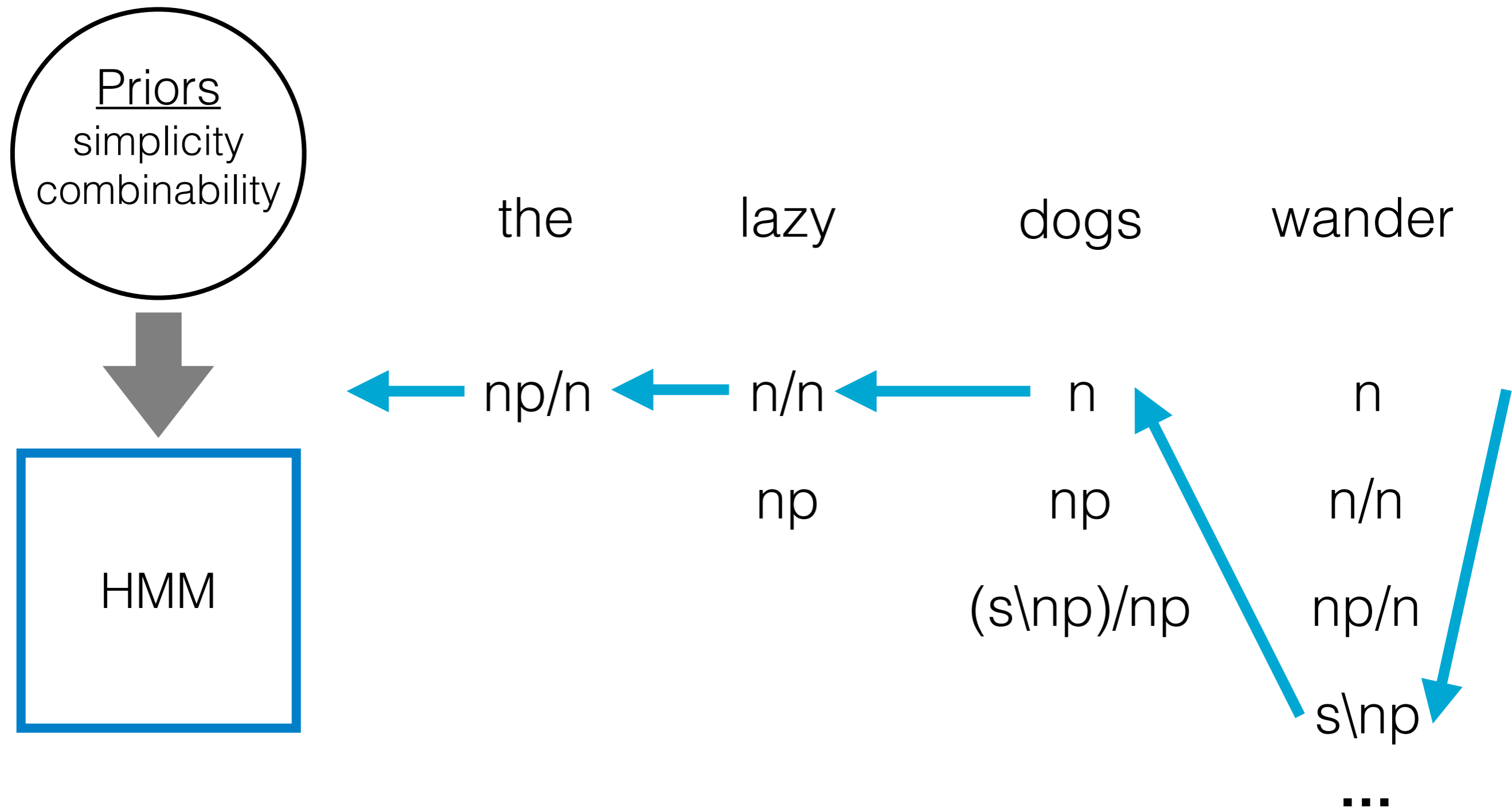
s\np

...

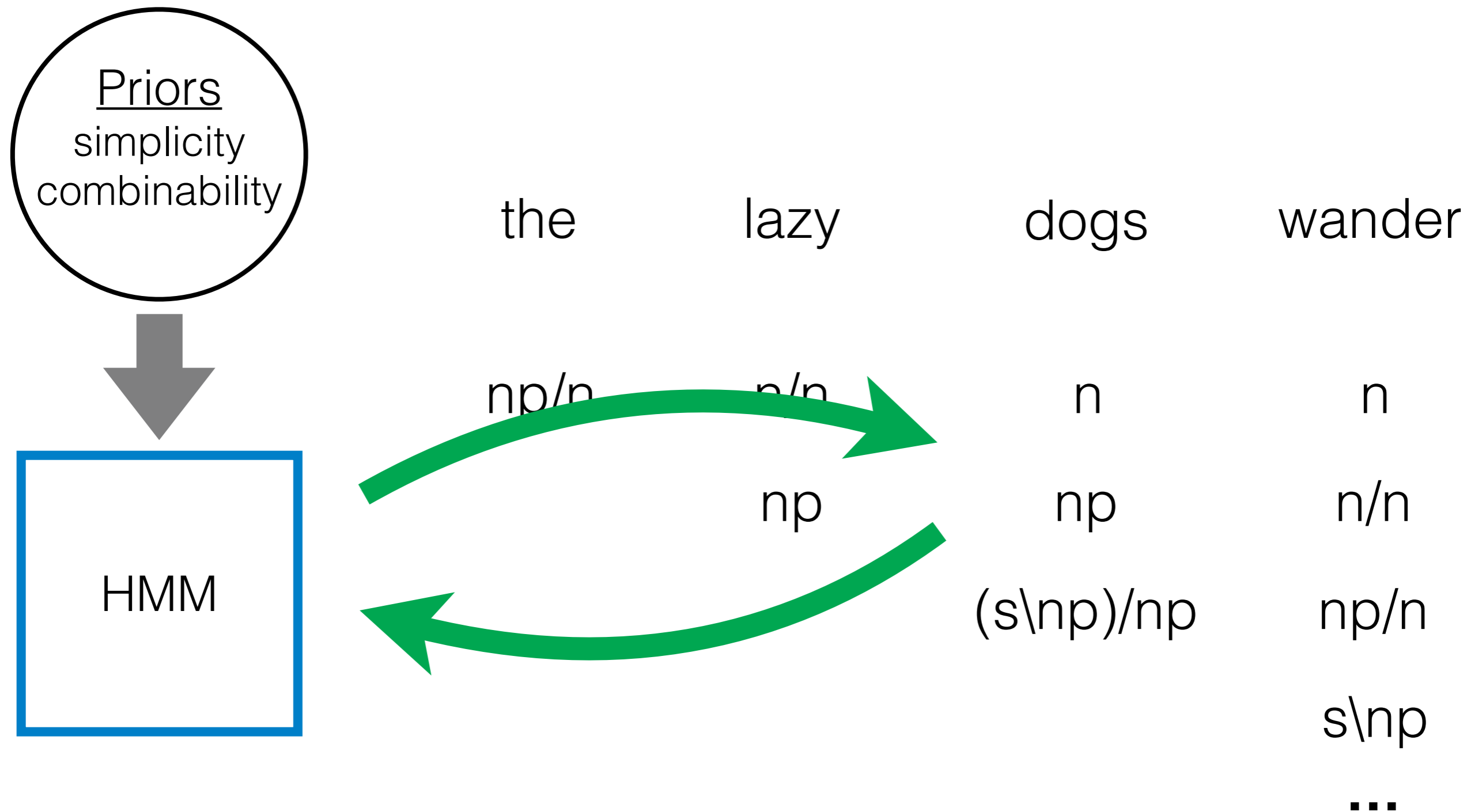
Posterior Inference



Posterior Inference

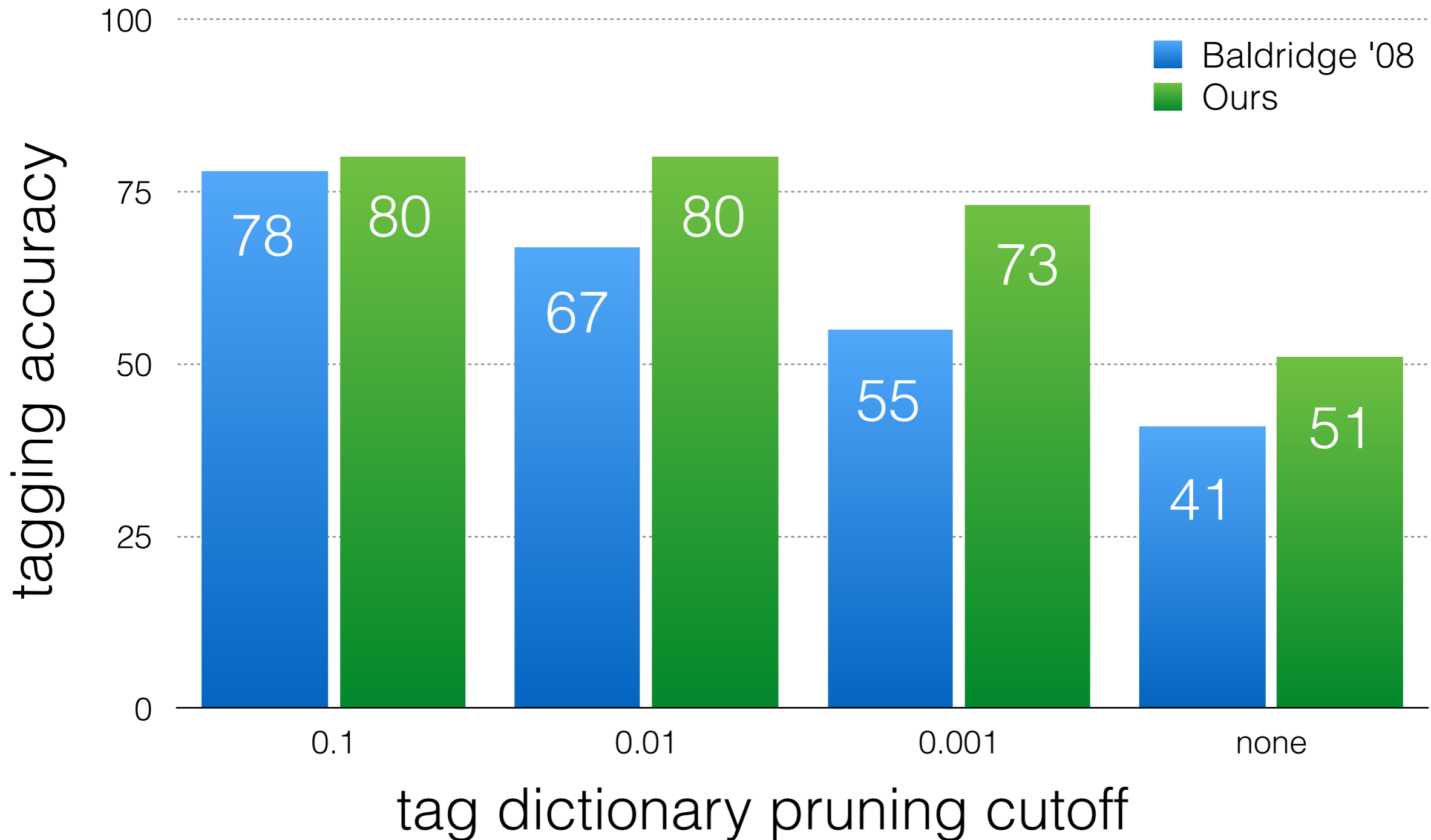


Posterior Inference

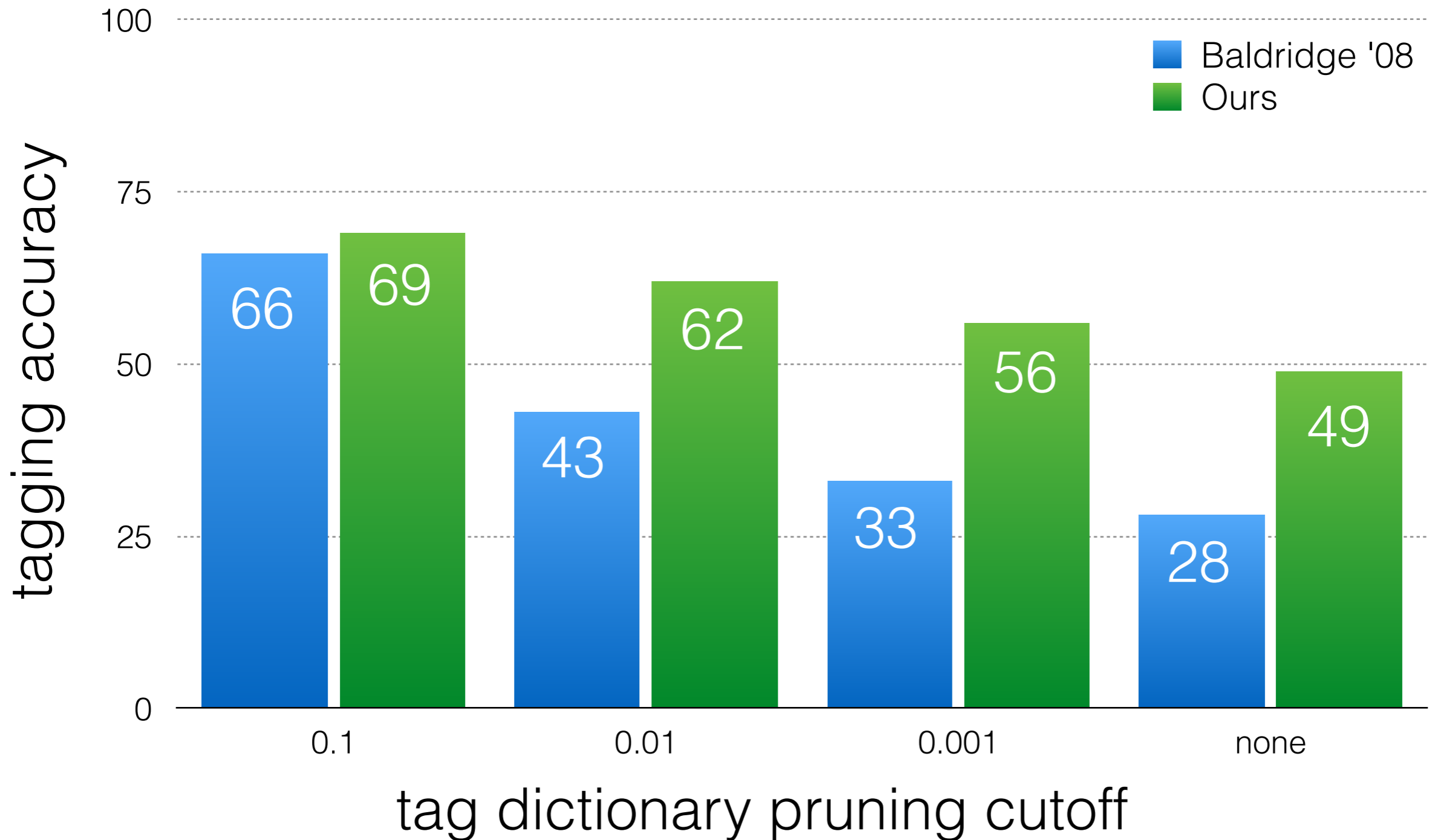


Supertagging Experiments

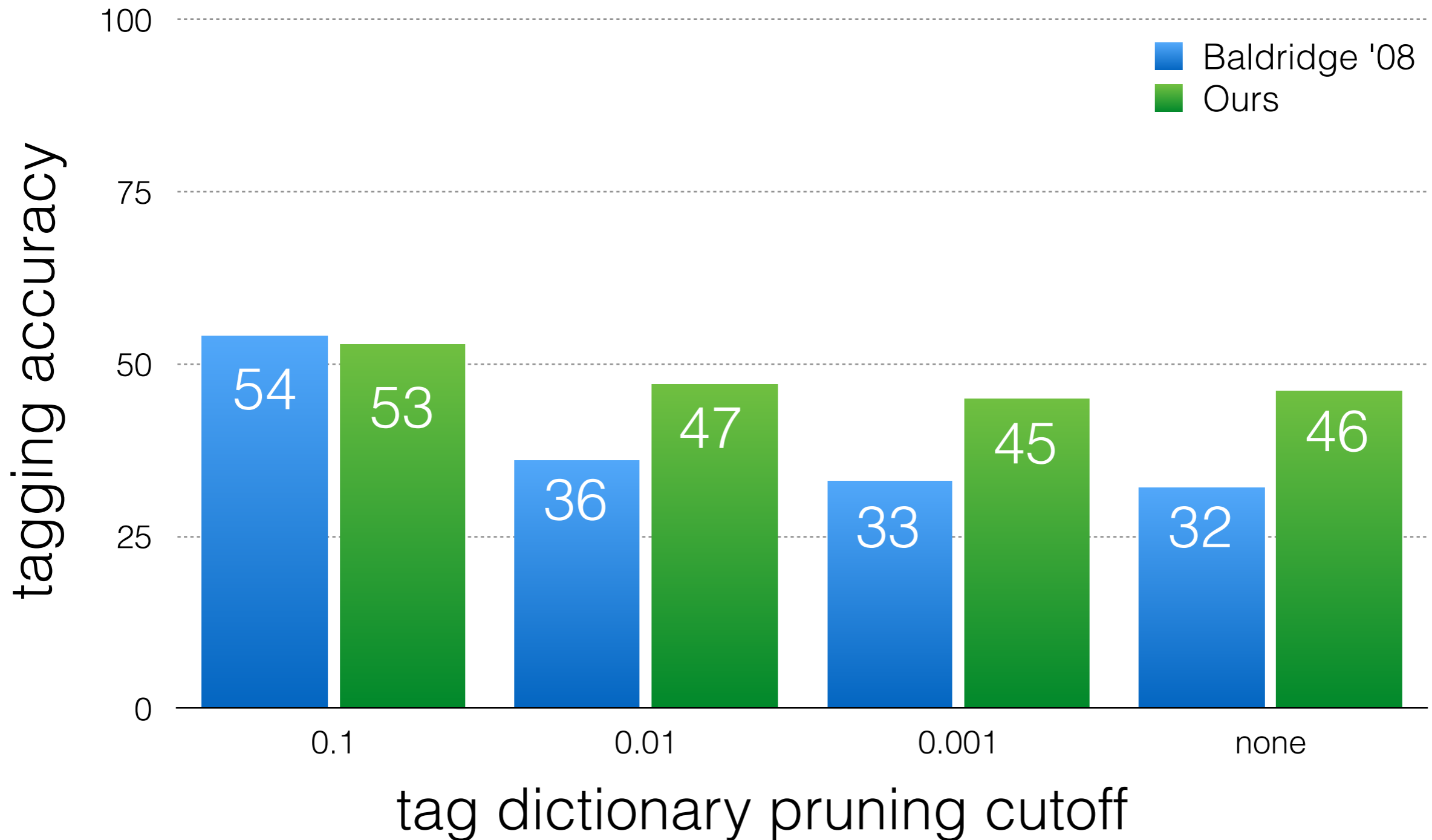
English Supertagging



Chinese Supertagging



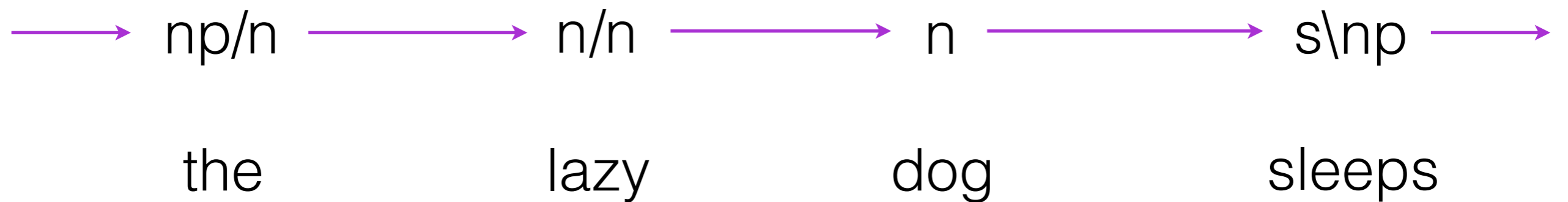
Italian Supertagging



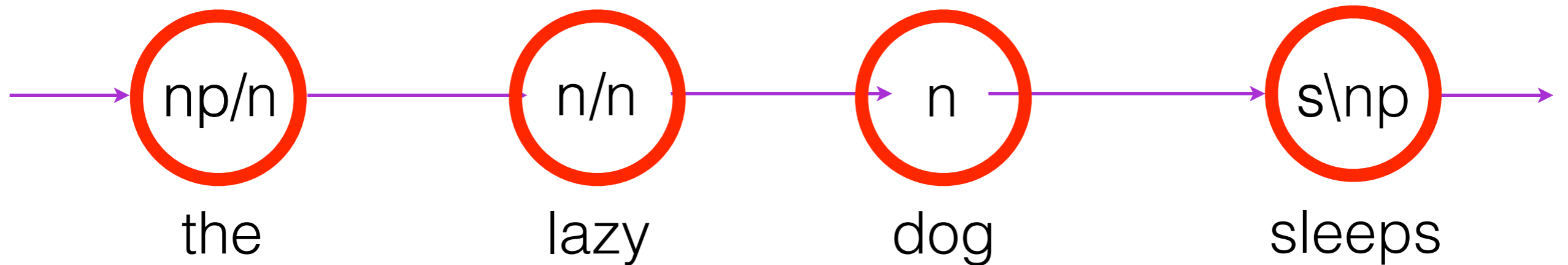
CCG Parsing with Grammatically-Informed Priors

[Garrette, Dyer, Baldridge, and Smith, 2015]

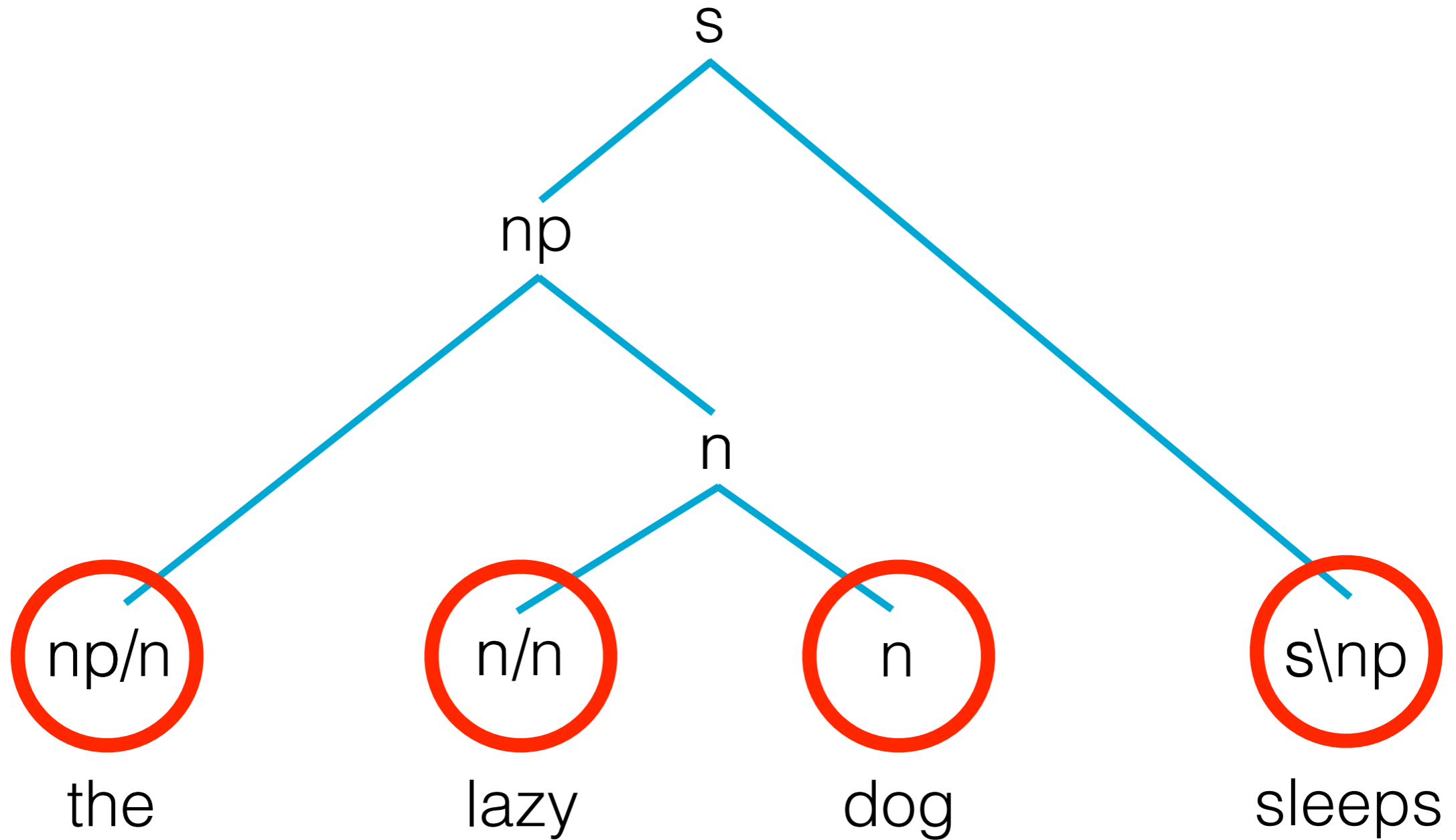
Prefer Plausible Categories



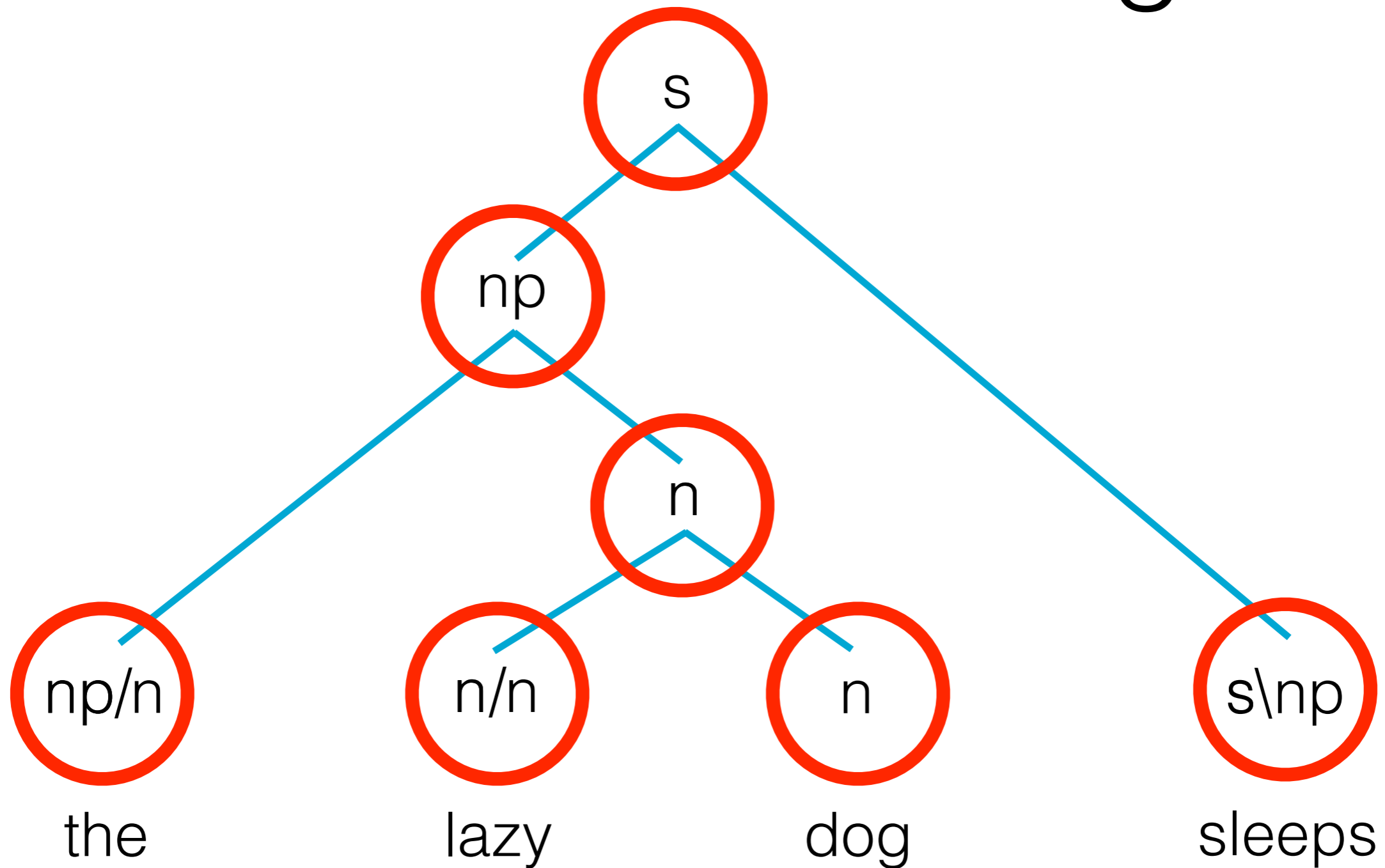
Prefer Plausible Categories



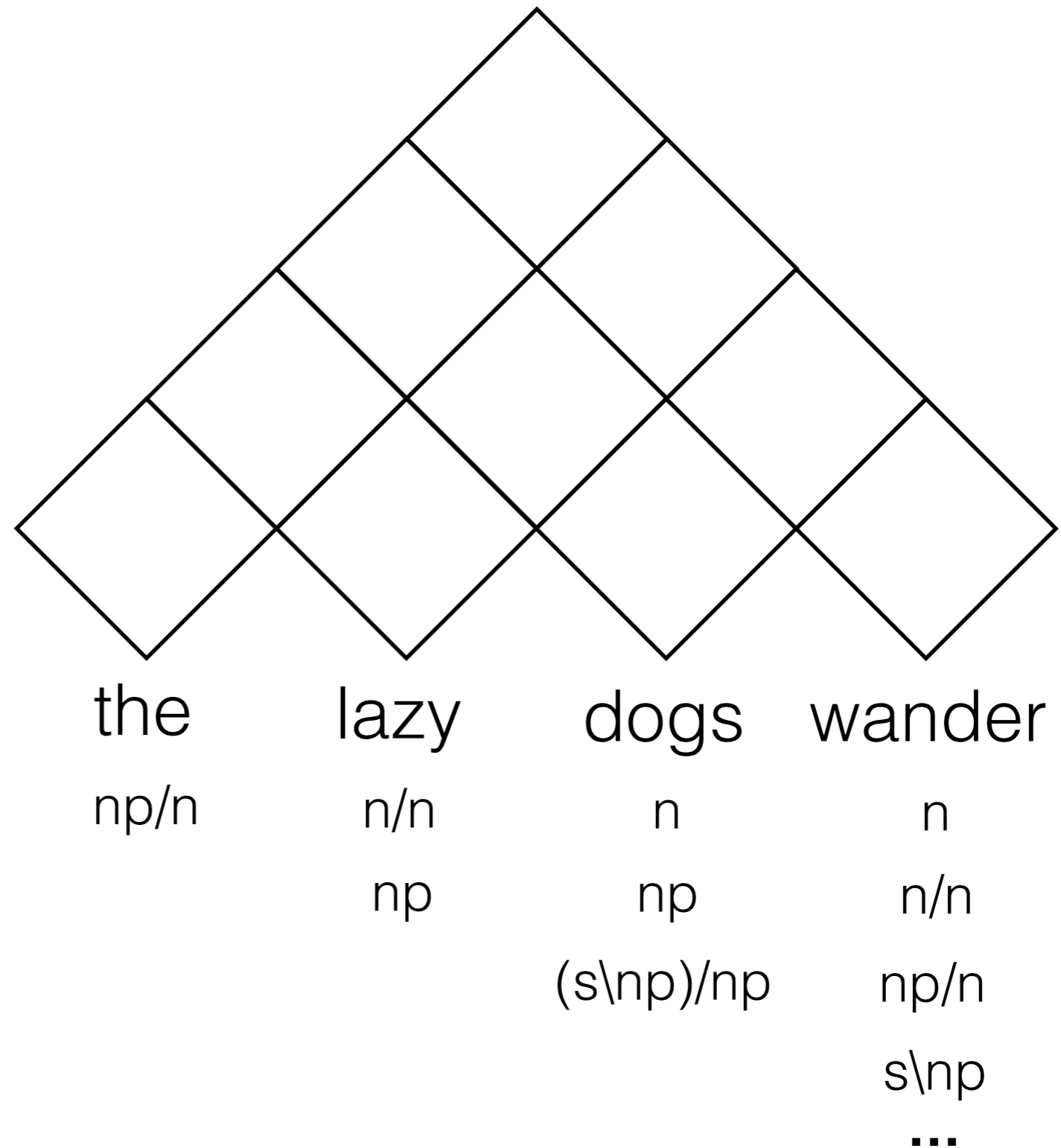
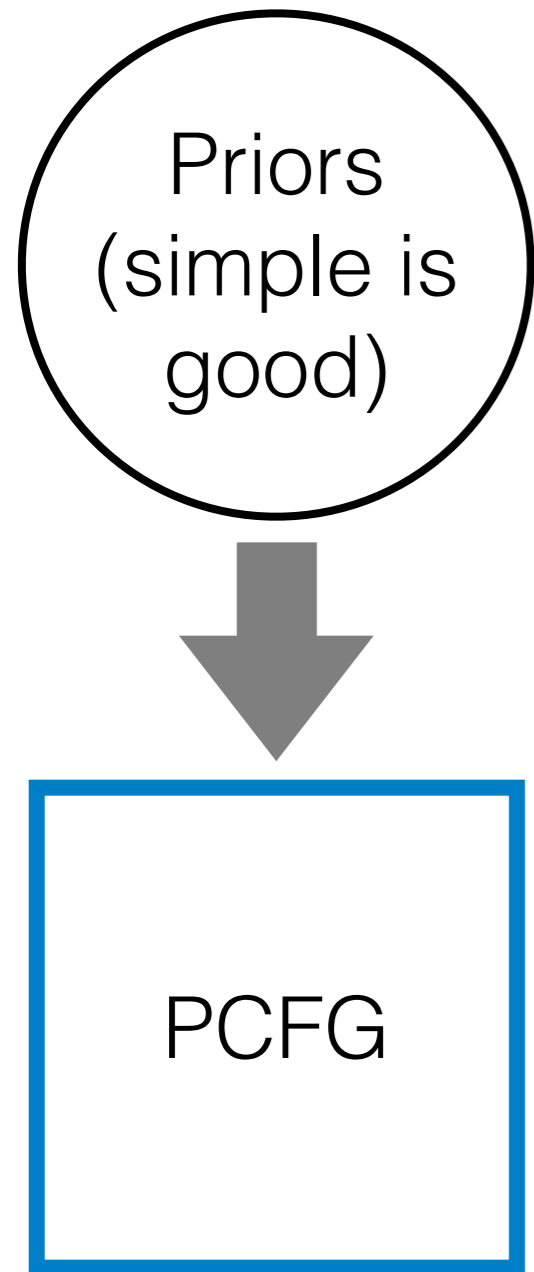
Prefer Plausible Categories



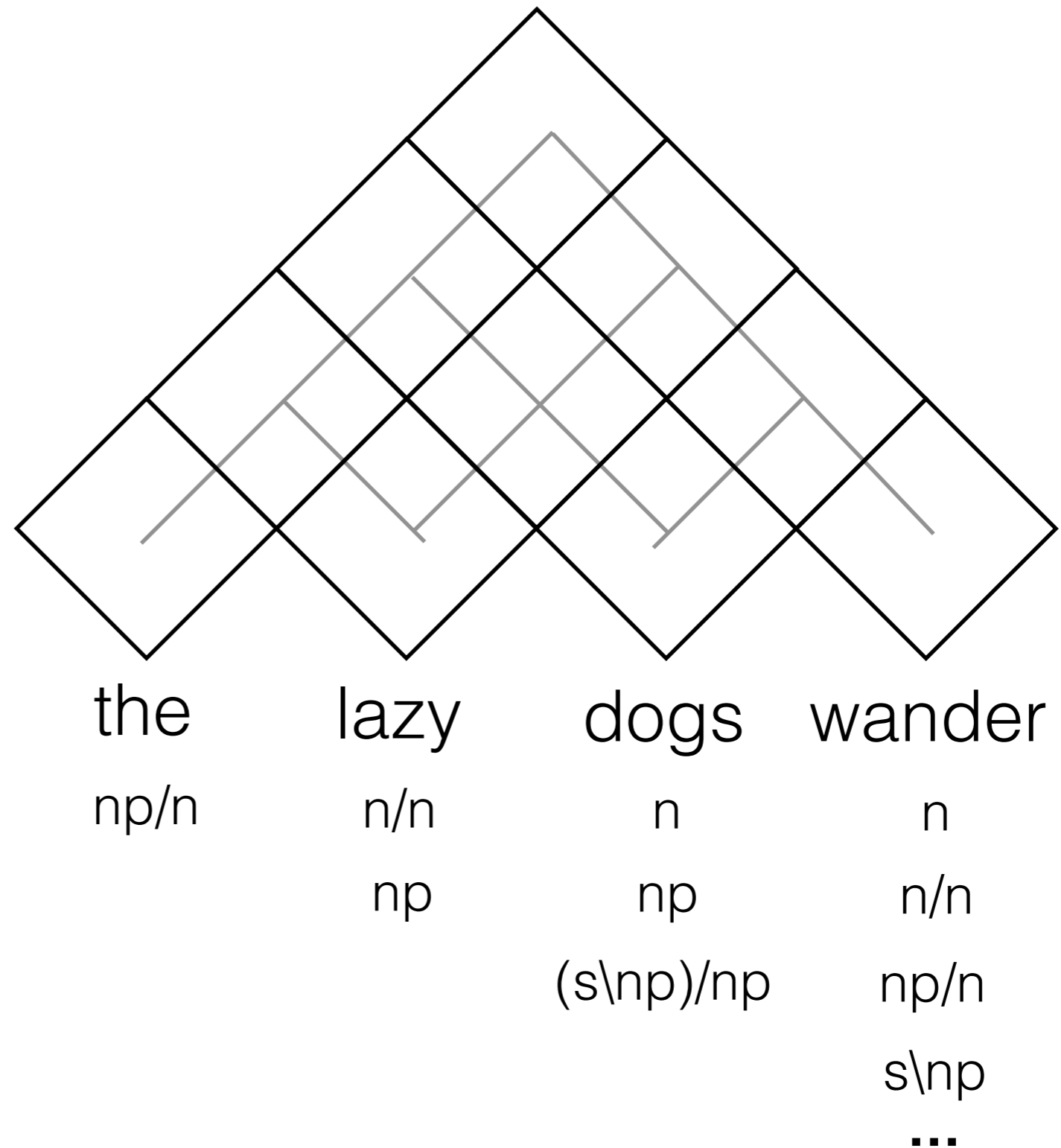
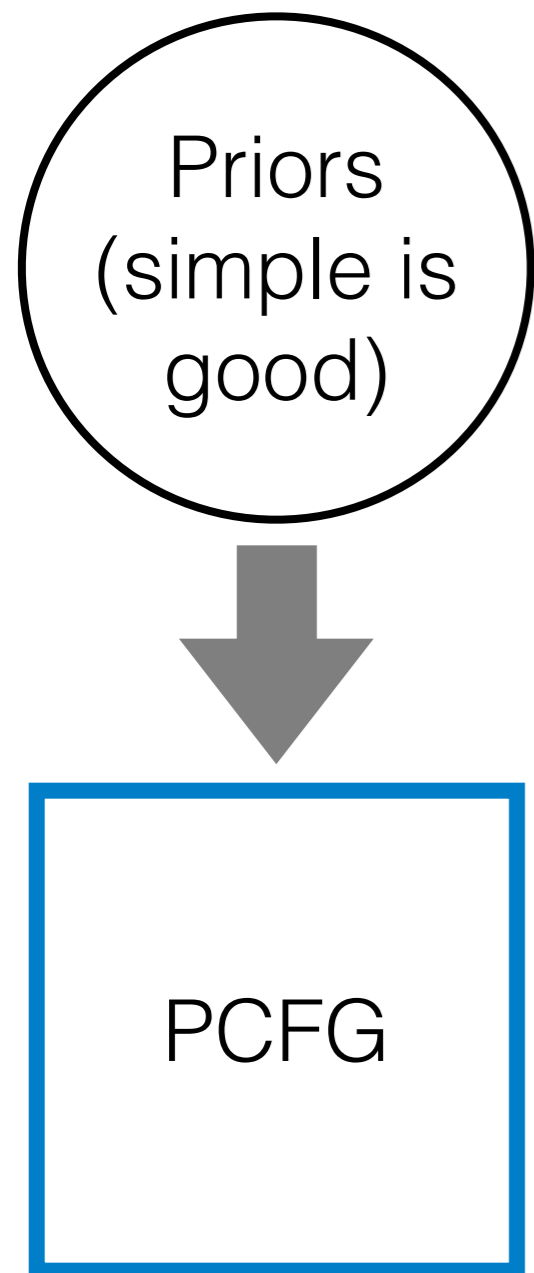
Prefer Plausible Categories



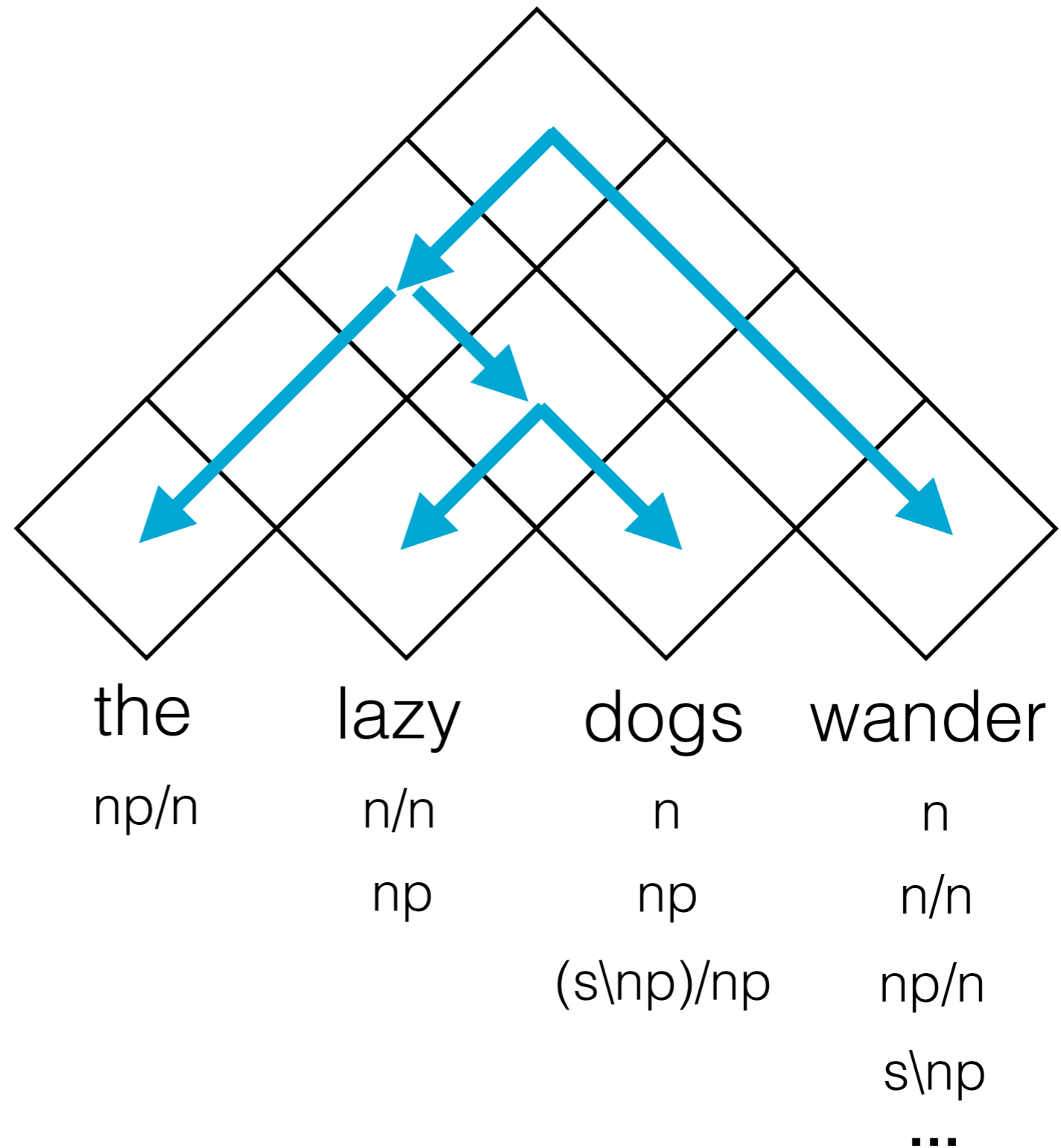
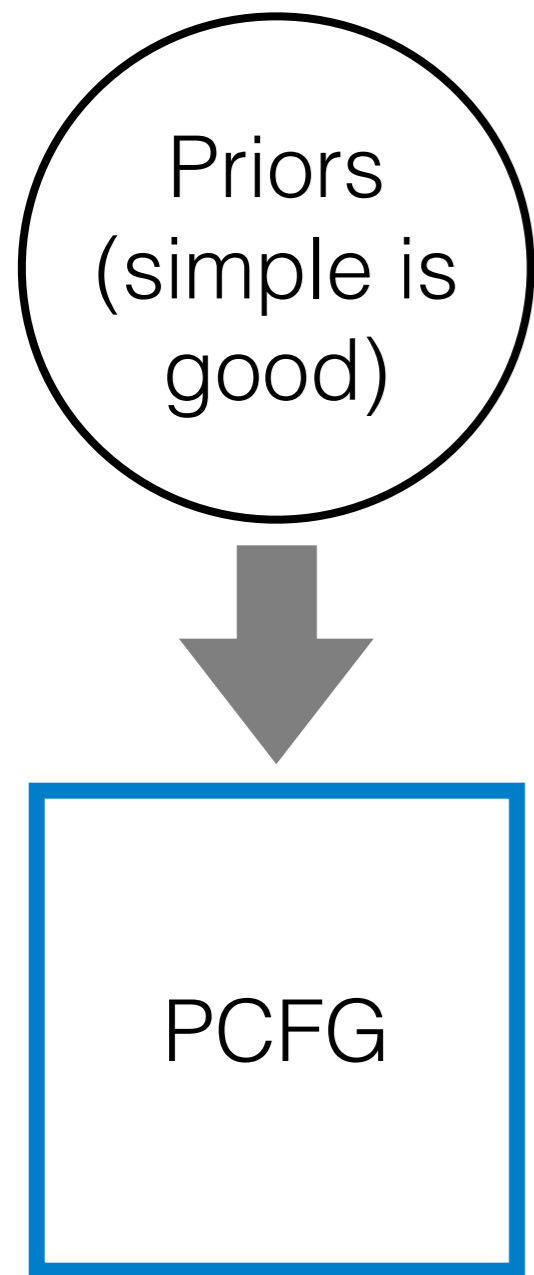
Posterior Inference



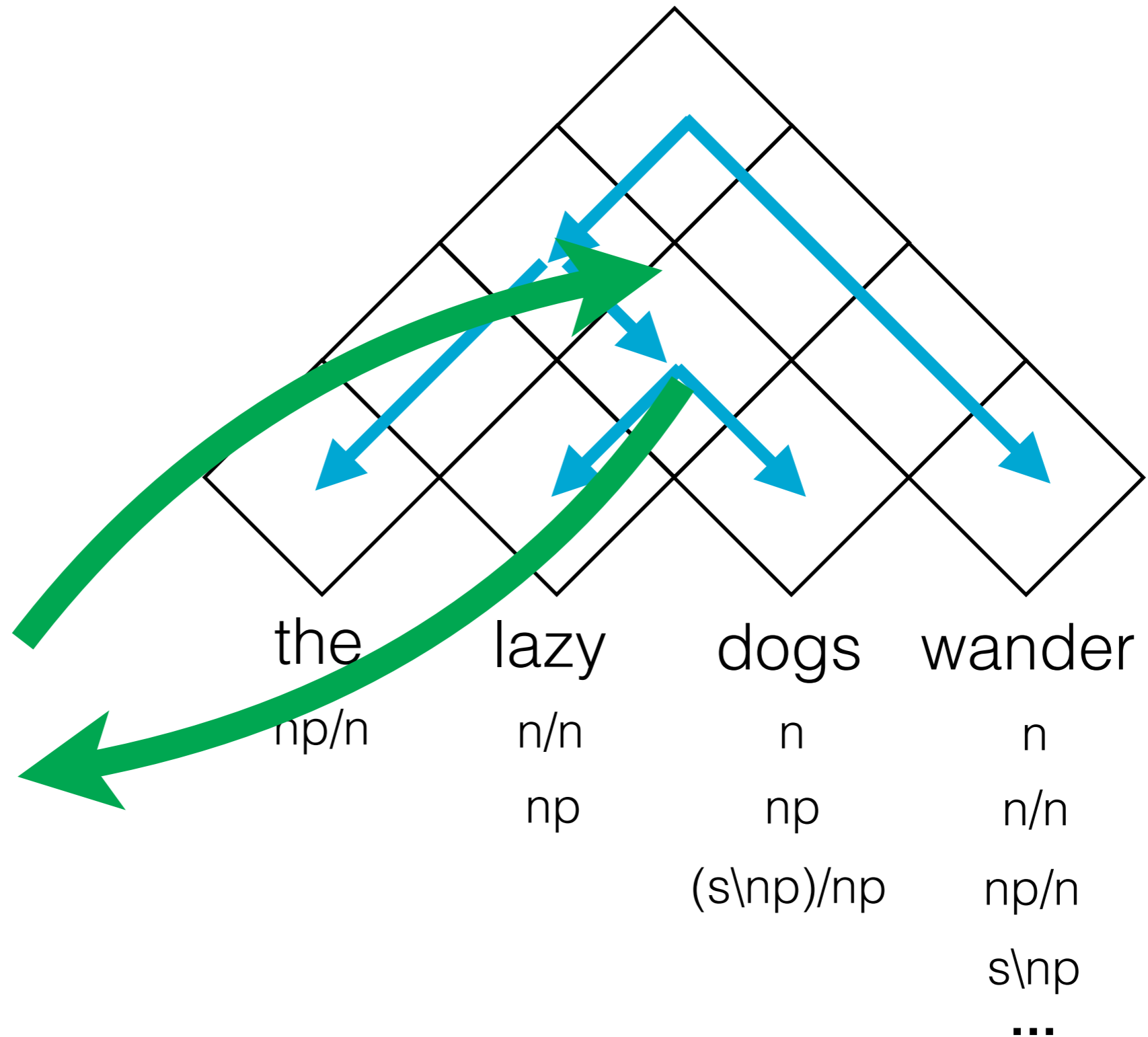
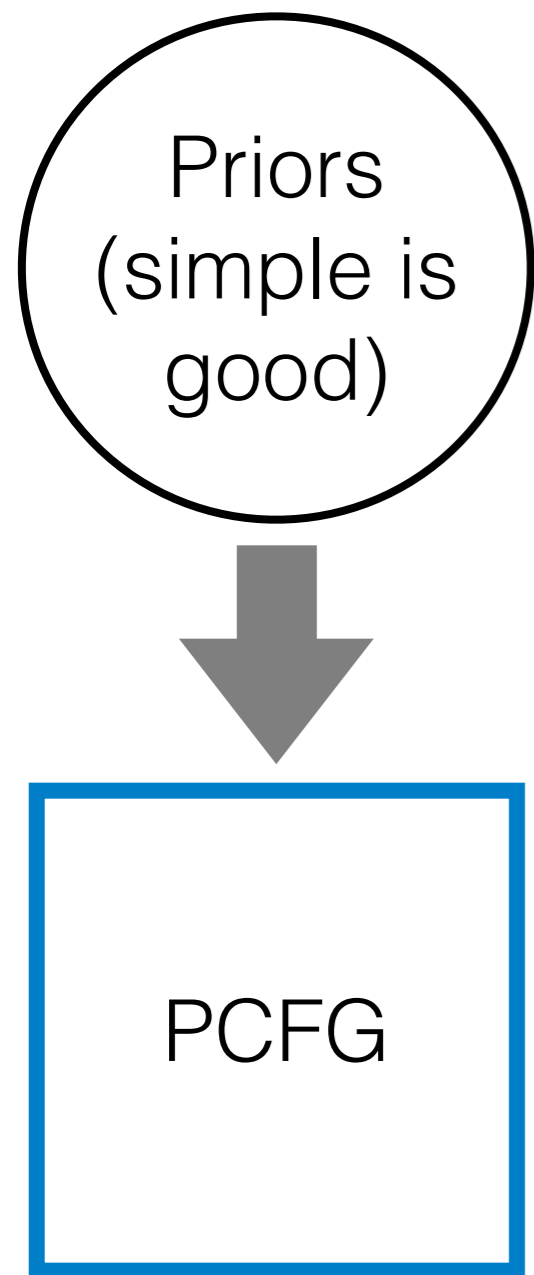
Posterior Inference



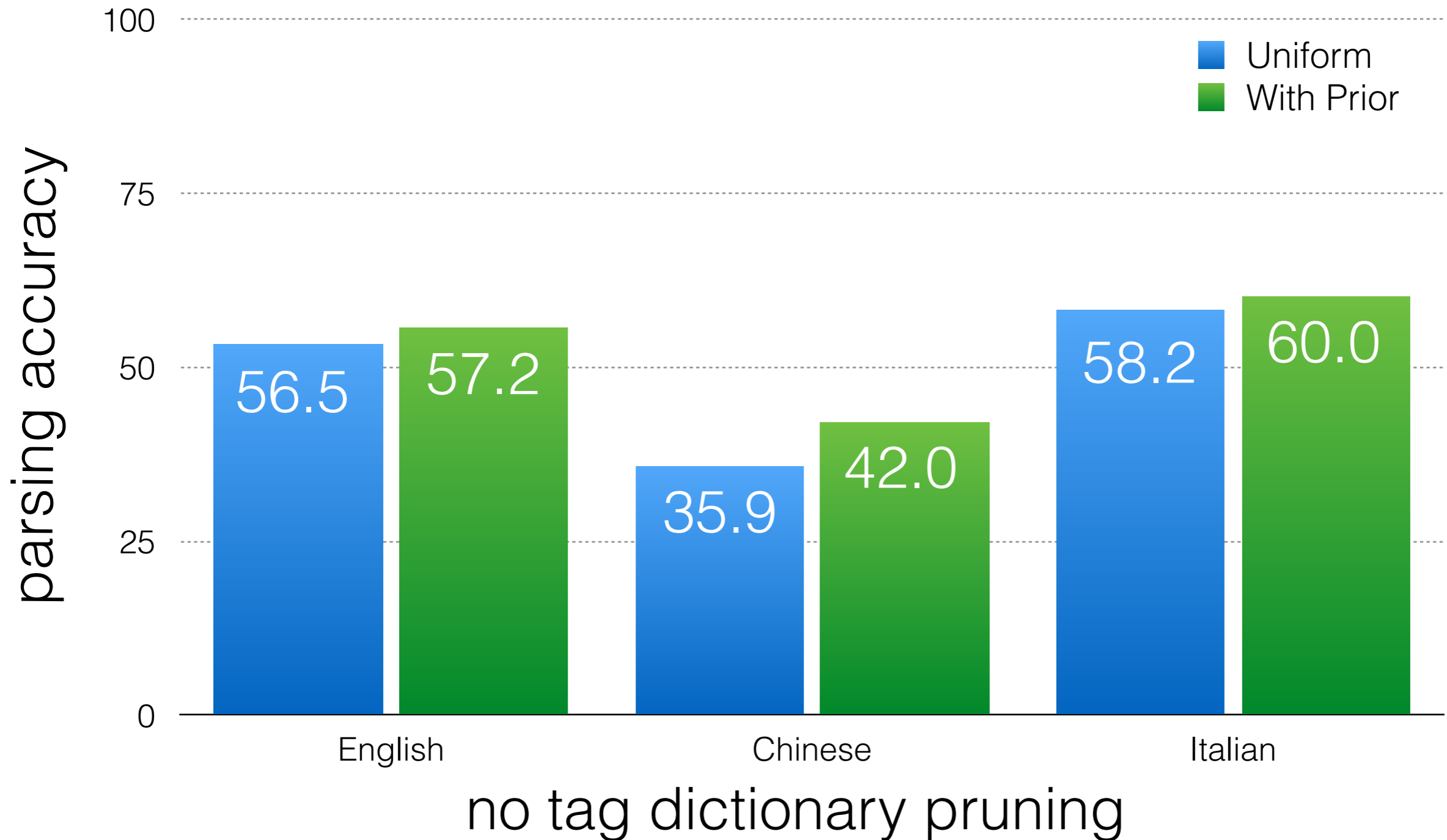
Posterior Inference



Posterior Inference



CCG Parsing Results



Conclusion

Combining annotation exploitation with
universal grammatical knowledge
yields good models
from weak supervision