

Parsing, and Context Free Grammars

COMP3361 – Week 6

Lingpeng Kong

Department of Computer Science, The University of Hong Kong
Many materials from Columbia CS4705 with special thanks!

Linguistic Structures

HKU is a public research university in Hong Kong.



Part-of-speech tagging
(word class)

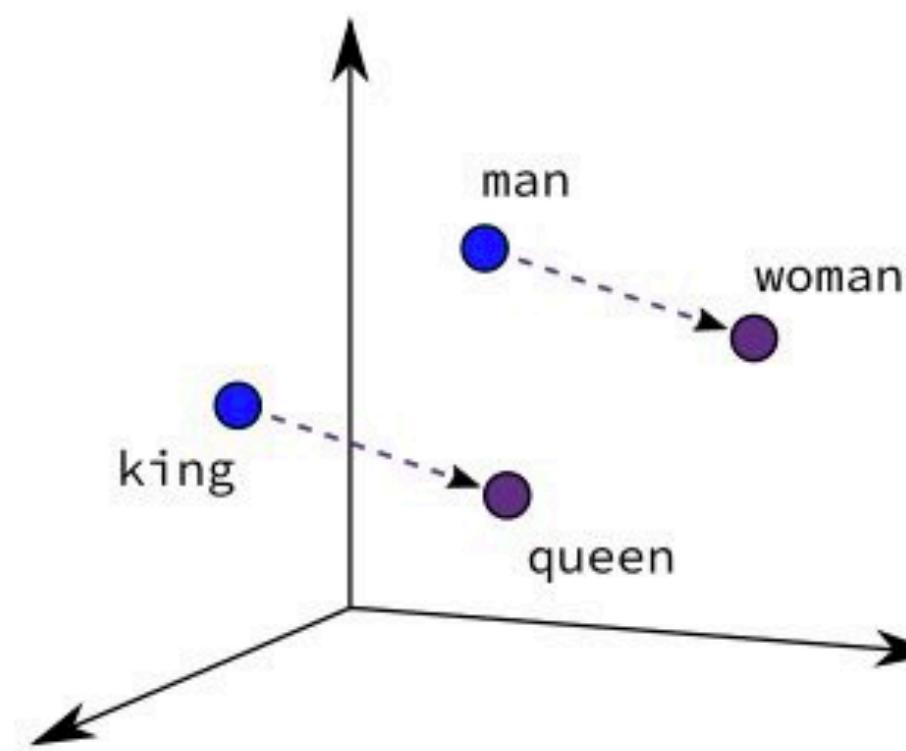


Named entity recognition

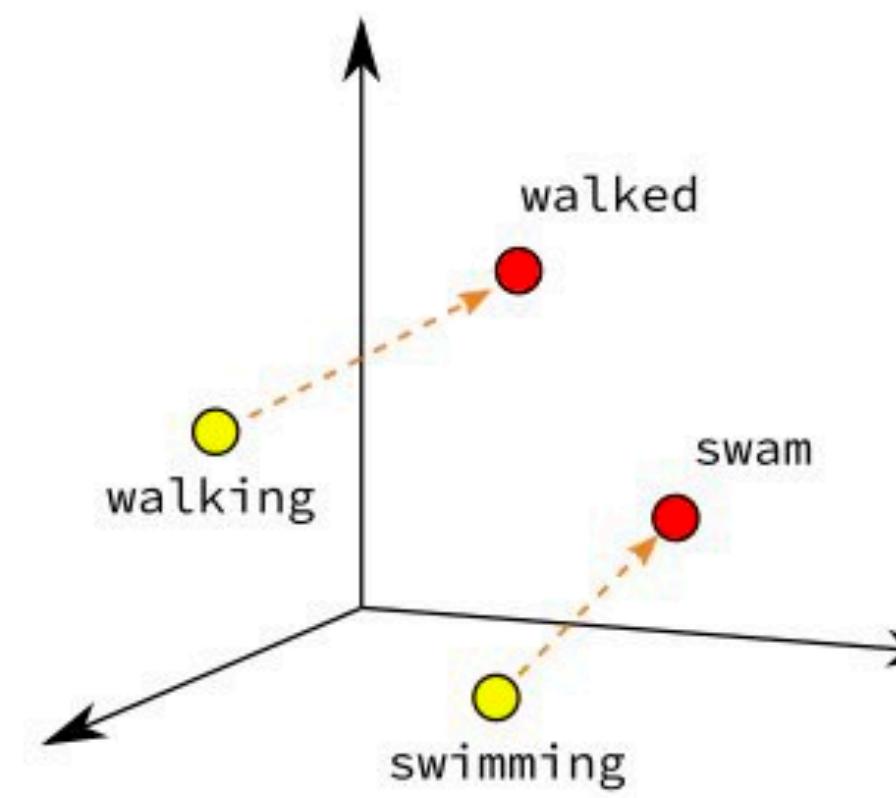
Linguistic Structures



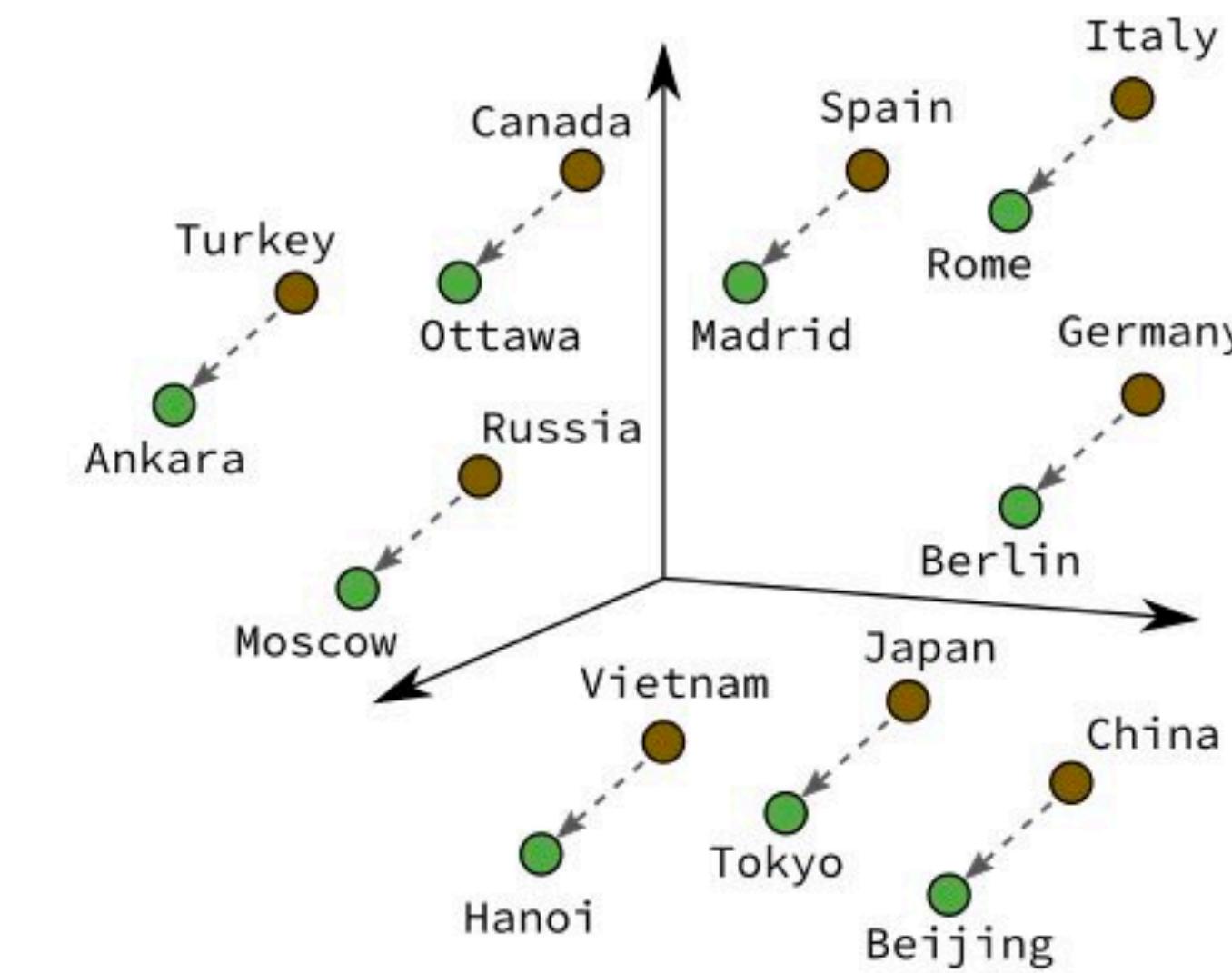
Part-of-speech tagging
(word class)



Male-Female

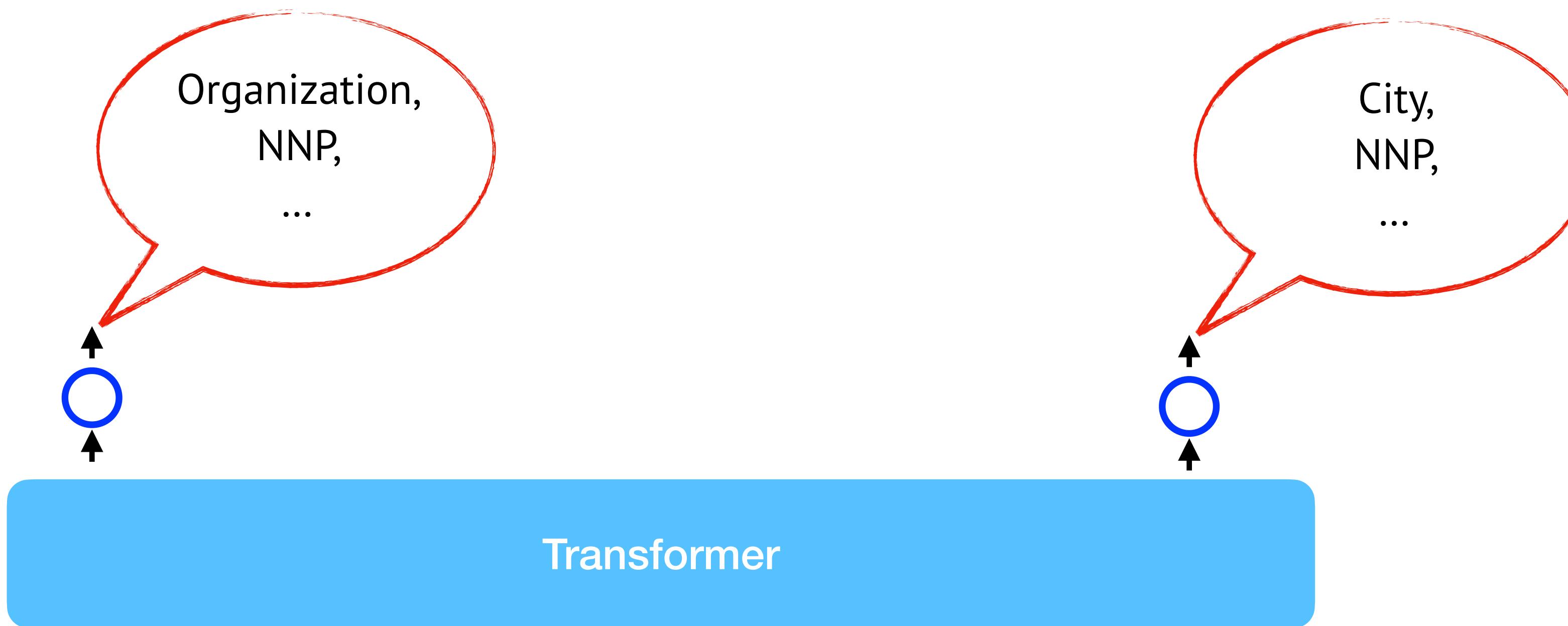


Verb Tense



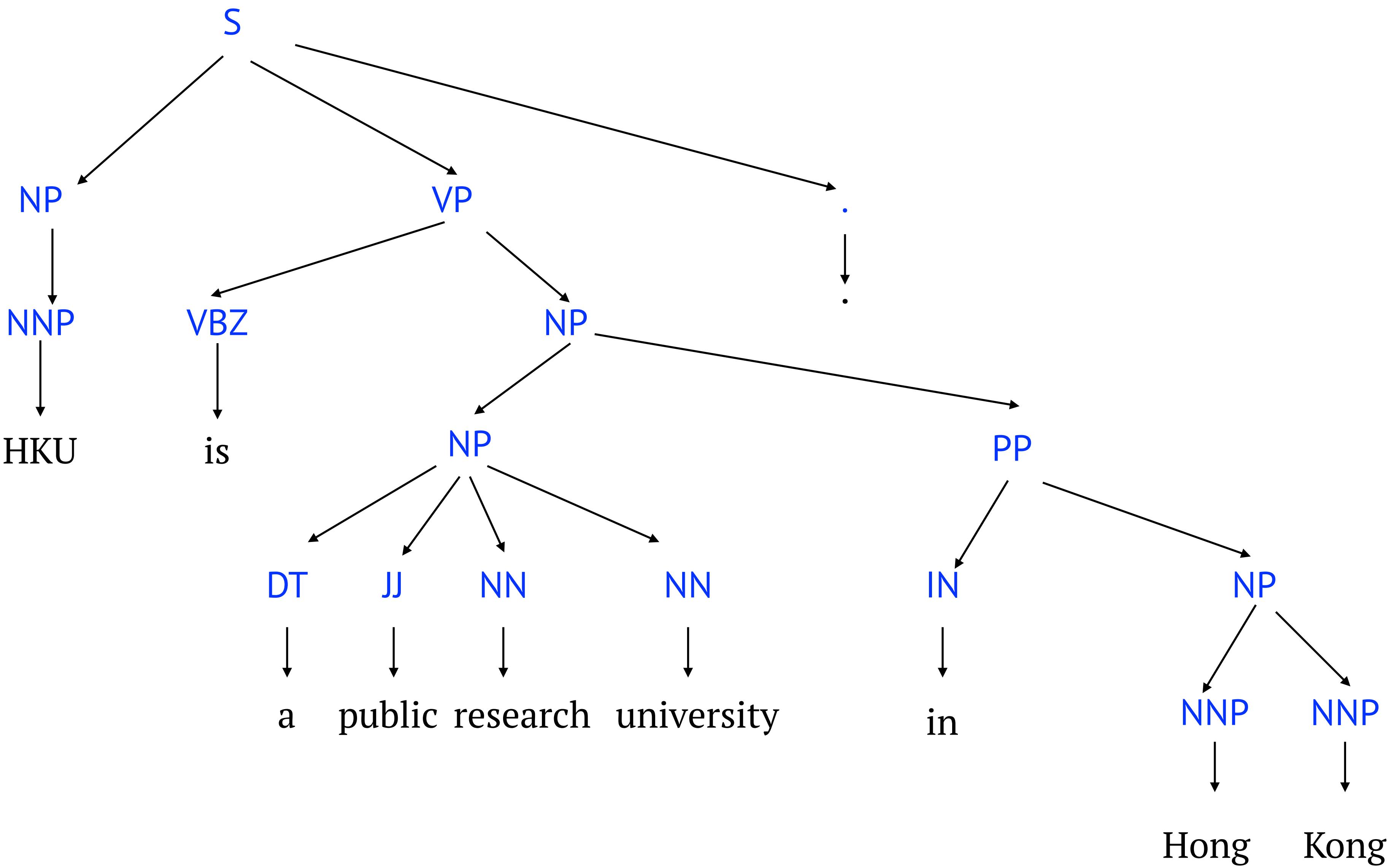
Country-Capital

Linguistic Structures



HKU is a public research university in Hong Kong.

Linguistic Structures



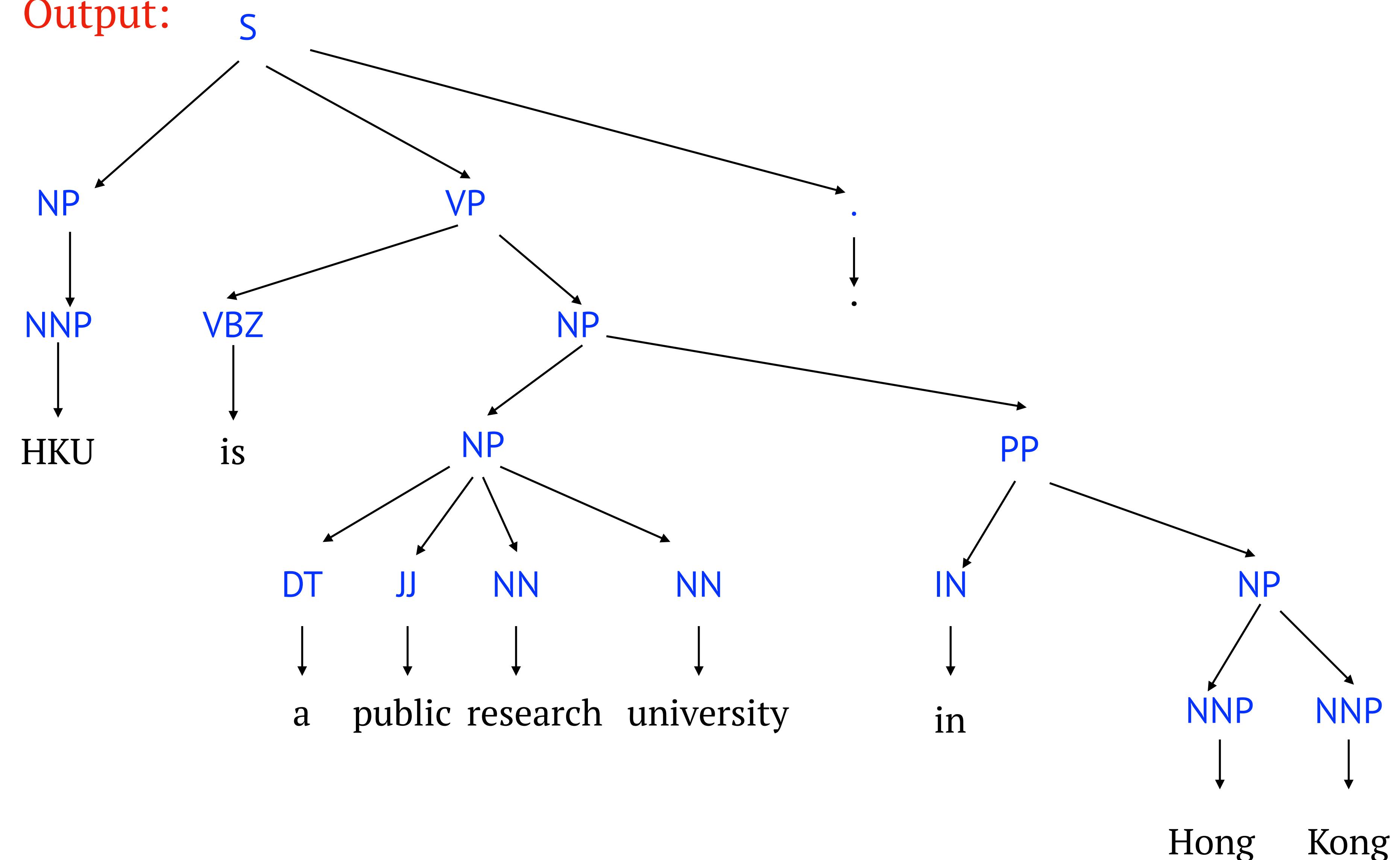
phrase-Structure tree,
constituency tree

Parsing (Phrase-structure Parsing)

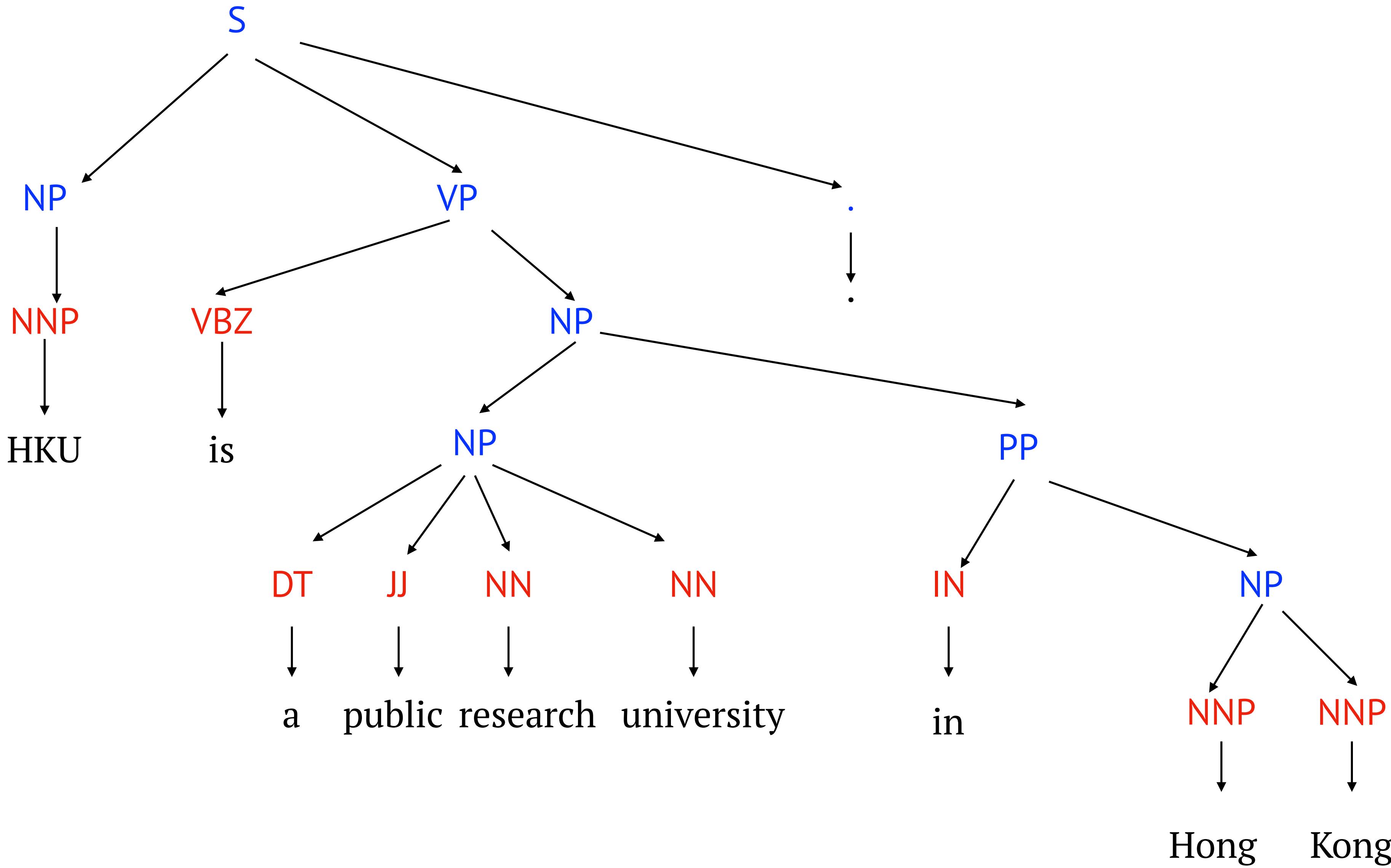
Input:

HKU is a public
research university
in Hong Kong.

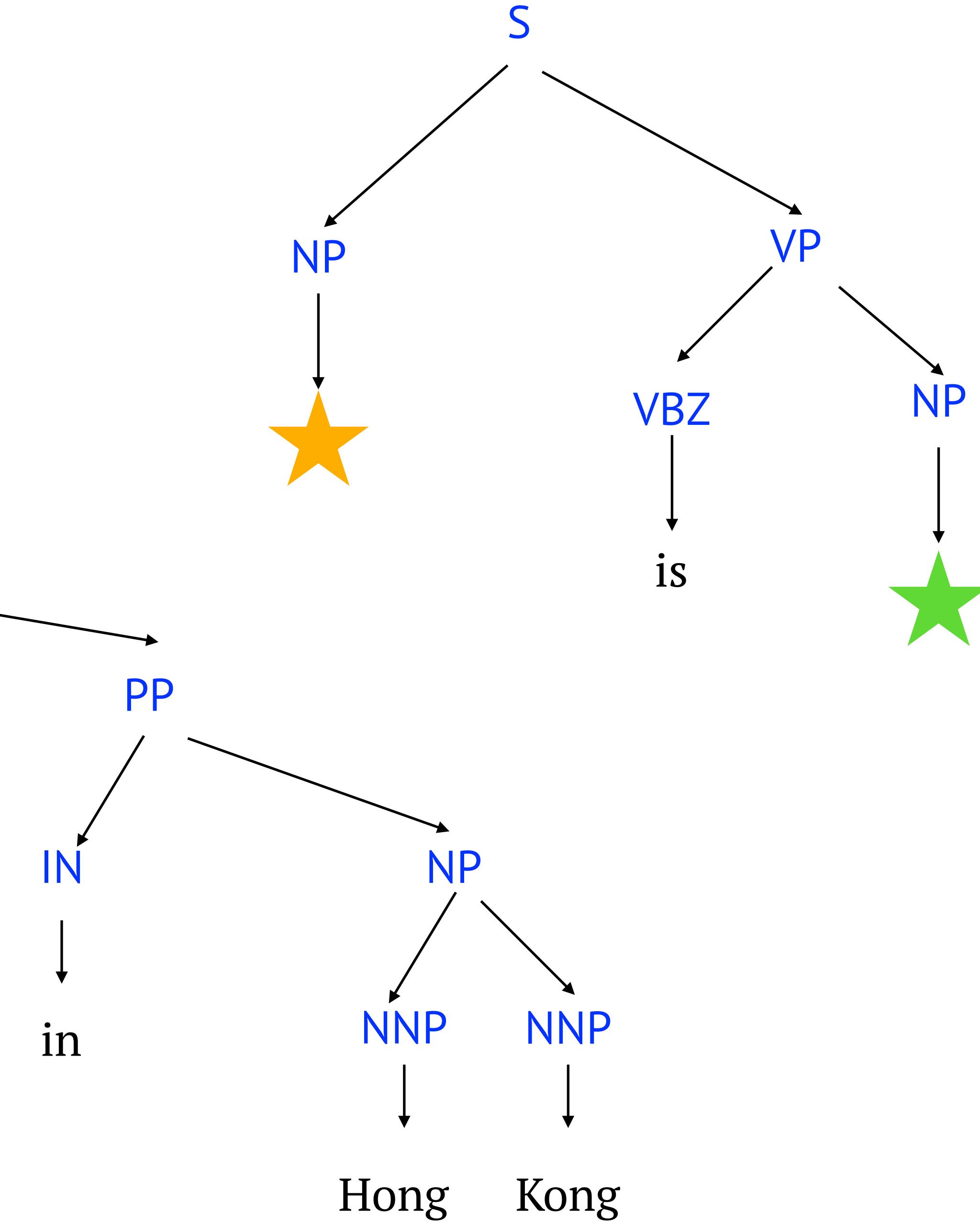
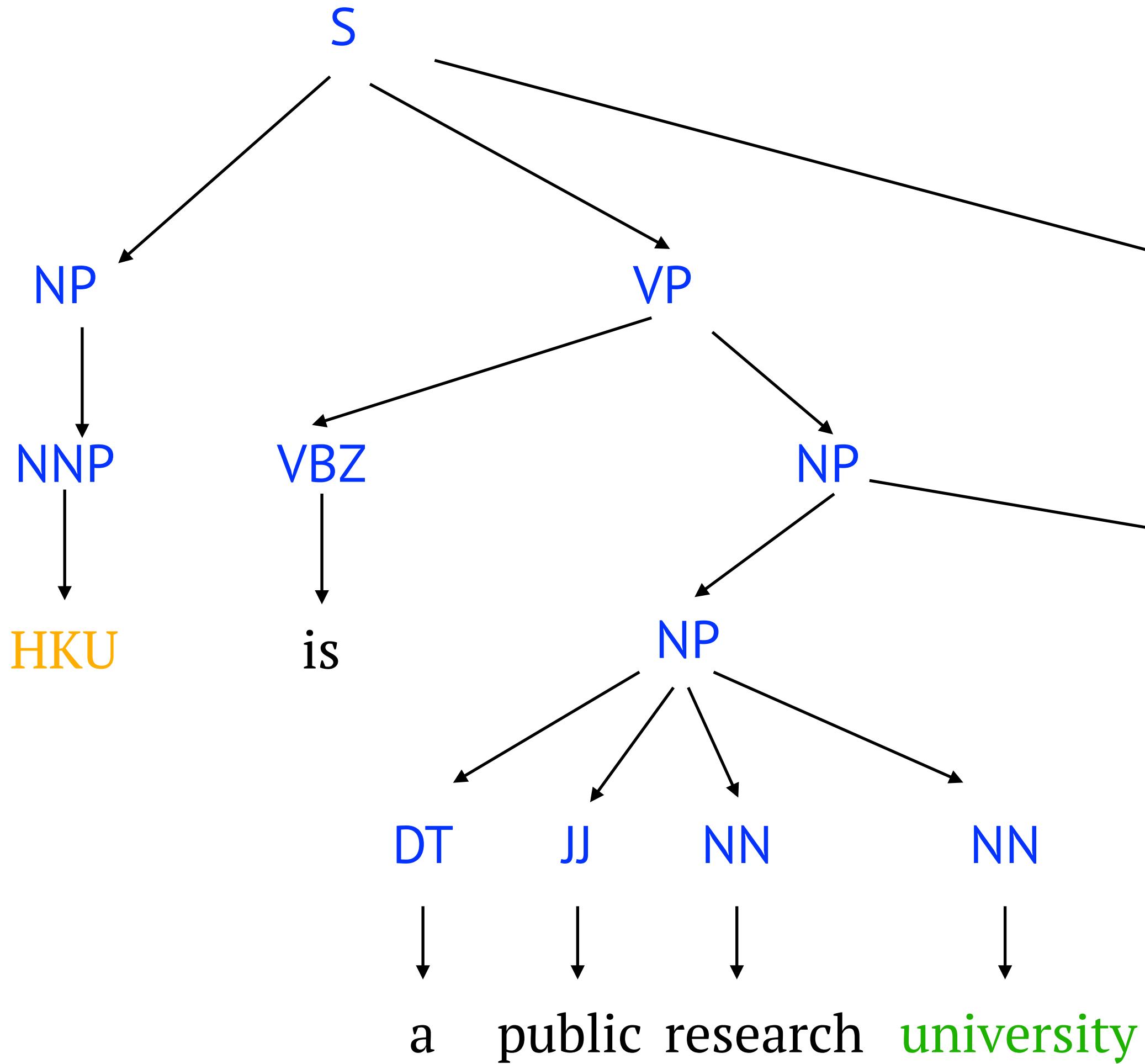
Output:



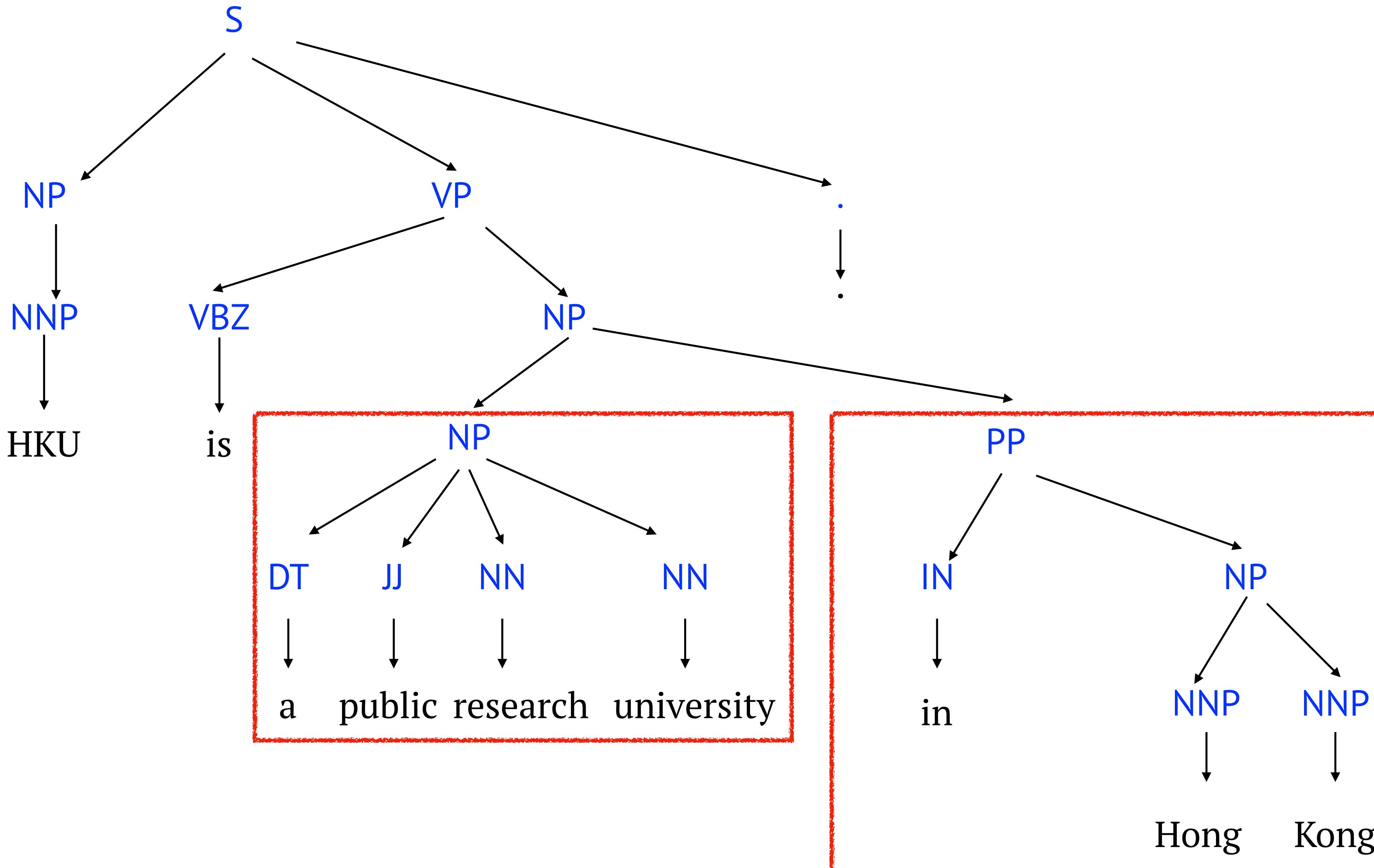
Parse Trees



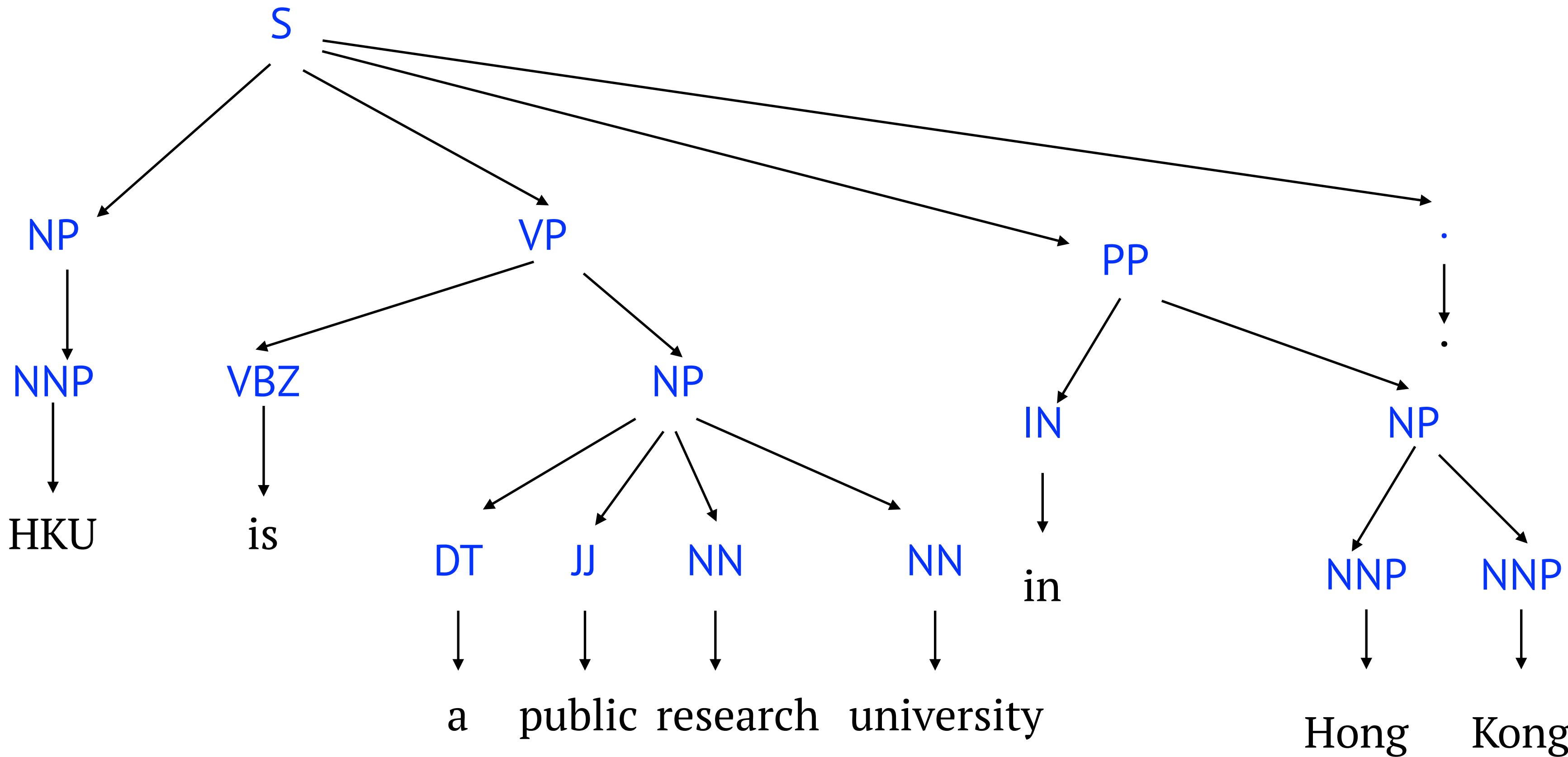
Parse Trees



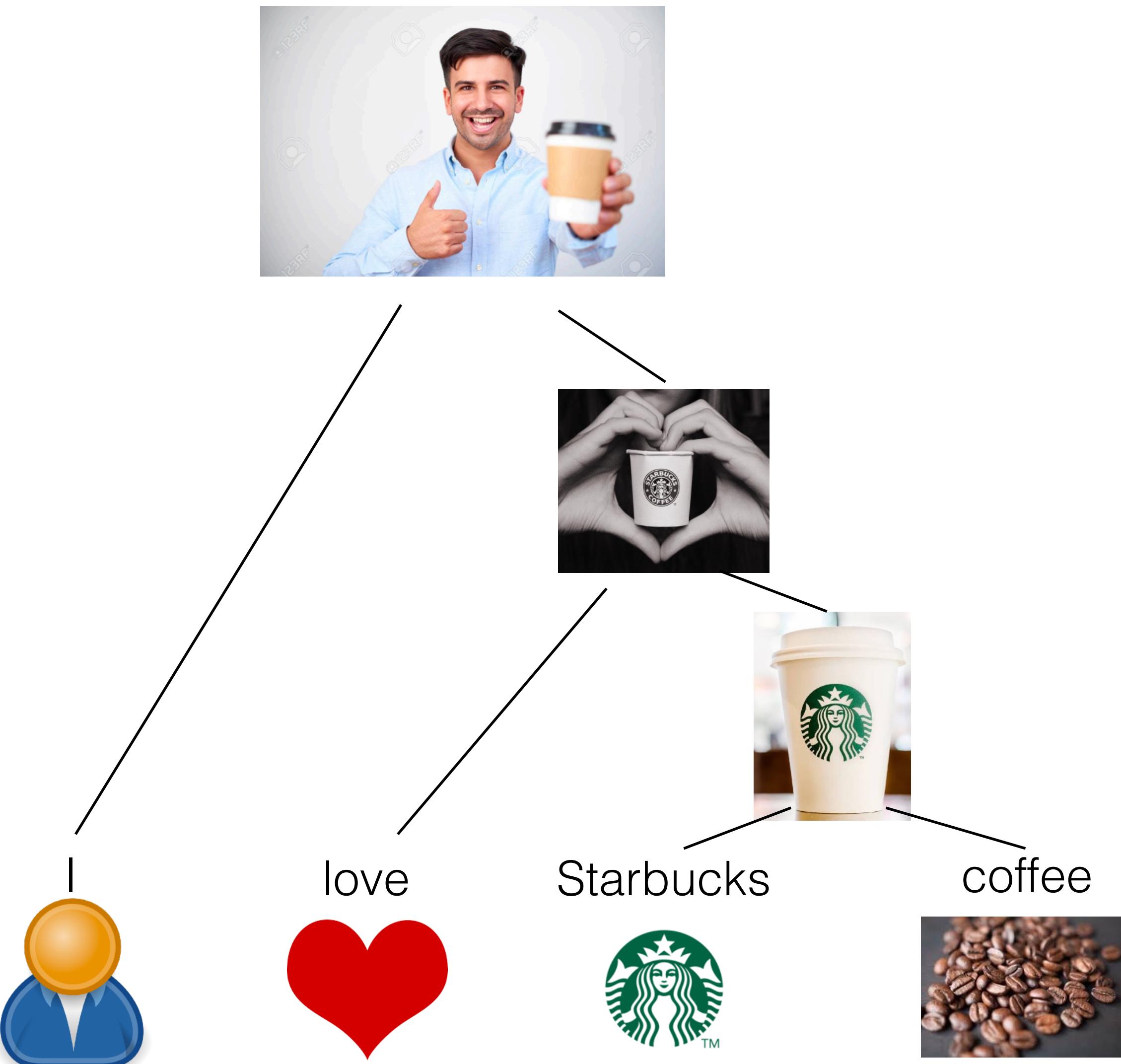
Parse Trees



Parse Trees



Linguistic Structures



Context-Free Grammars

A context free grammar $G = (N, \Sigma, R, S)$ where:

N is a set of non-terminal symbols

Σ is a set of terminal symbols

R is a set of rules of the form $X \rightarrow Y_1 Y_2 \dots Y_n$

for $n \geq 0$ $X \in N$ $Y_i \in (N \cup \Sigma)$

$S \in N$ is a distinguished start symbol

A Context-Free Grammar for English

$N = \{S, NP, VP, PP, DT, Vi, Vt, NN, IN\}$

$S = S$

$\Sigma = \{\text{sleeps}, \text{saw}, \text{man}, \text{woman}, \text{telescope}, \text{the}, \text{with}, \text{in}\}$

$R =$	$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
	$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$
	$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
	$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
	$NP \longrightarrow DT \ NN$	$NN \longrightarrow \text{telescope}$
	$NP \longrightarrow NP \ PP$	$DT \longrightarrow \text{the}$
	$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{with}$
		$IN \longrightarrow \text{in}$

Left-Most Derivations

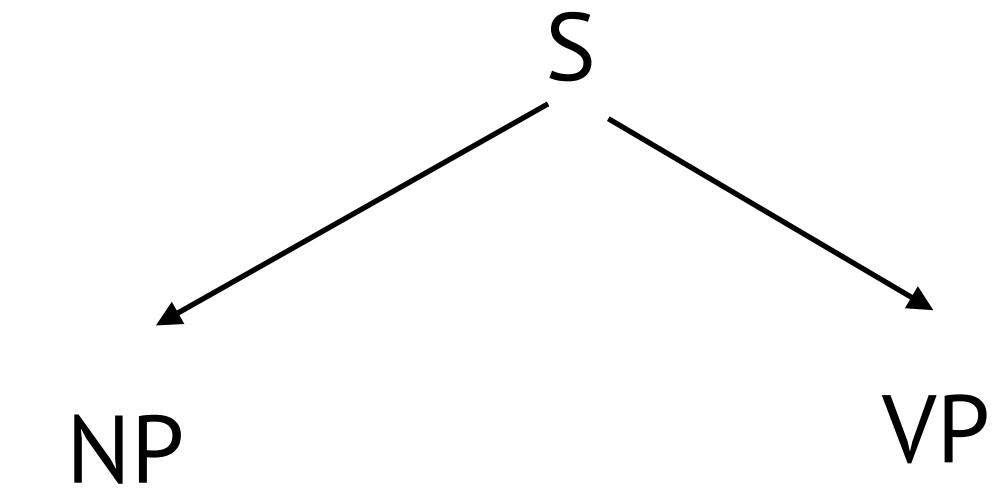
$R =$	$S \longrightarrow NP\ VP$	$Vi \longrightarrow \text{sleeps}$	S
	$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$	
	$VP \longrightarrow Vt\ NP$	$NN \longrightarrow \text{man}$	
	$VP \longrightarrow VP\ PP$	$NN \longrightarrow \text{woman}$	
		$NN \longrightarrow \text{telescope}$	
	$NP \longrightarrow DT\ NN$	$DT \longrightarrow \text{the}$	
	$NP \longrightarrow NP\ PP$		
	$PP \longrightarrow IN\ NP$	$IN \longrightarrow \text{with}$	
		$IN \longrightarrow \text{in}$	

(S)

Left-Most Derivations

$R =$	$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
	$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$
	$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
	$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
	$NP \longrightarrow DT \ NN$	$NN \longrightarrow \text{telescope}$
	$NP \longrightarrow NP \ PP$	$DT \longrightarrow \text{the}$
	$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{with}$
		$IN \longrightarrow \text{in}$

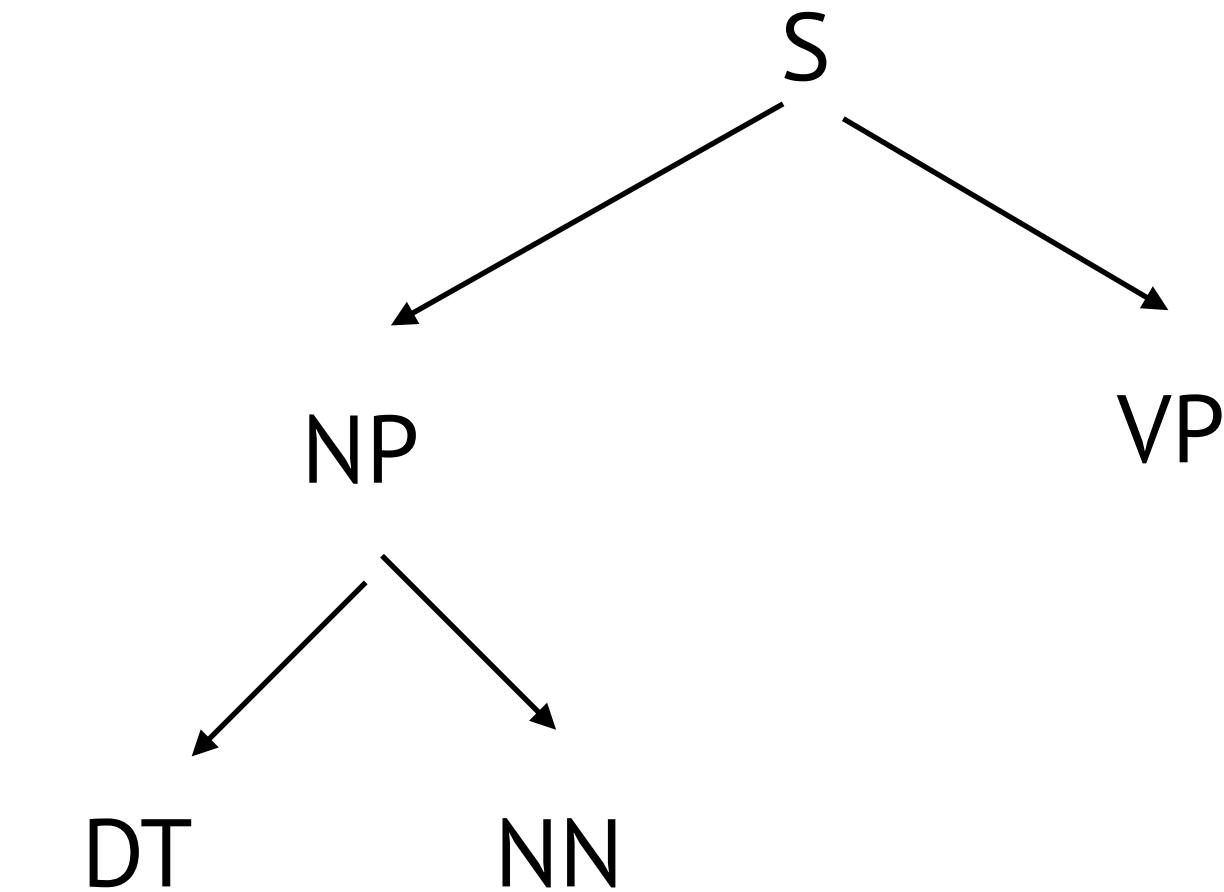
(NP VP)



Left-Most Derivations

$R =$	$S \rightarrow NP\ VP$	$Vi \rightarrow \text{sleeps}$
	$VP \rightarrow Vi$	$Vt \rightarrow \text{saw}$
	$VP \rightarrow Vt\ NP$	$NN \rightarrow \text{man}$
	$VP \rightarrow VP\ PP$	$NN \rightarrow \text{woman}$
	$NP \rightarrow DT\ NN$	$NN \rightarrow \text{telescope}$
	$NP \rightarrow NP\ PP$	$DT \rightarrow \text{the}$
	$PP \rightarrow IN\ NP$	$IN \rightarrow \text{with}$
		$IN \rightarrow \text{in}$

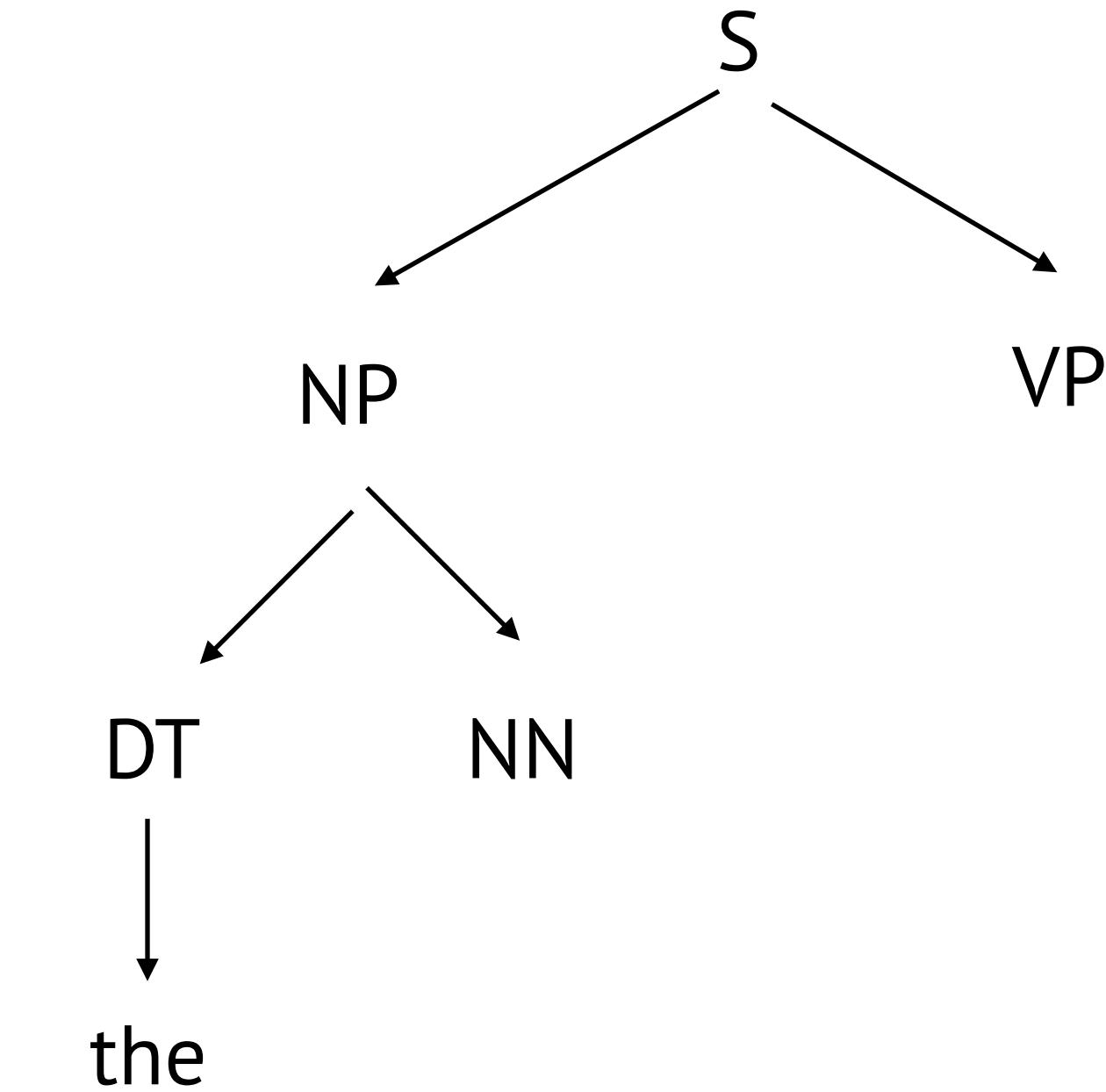
(DT NN VP)



Left-Most Derivations

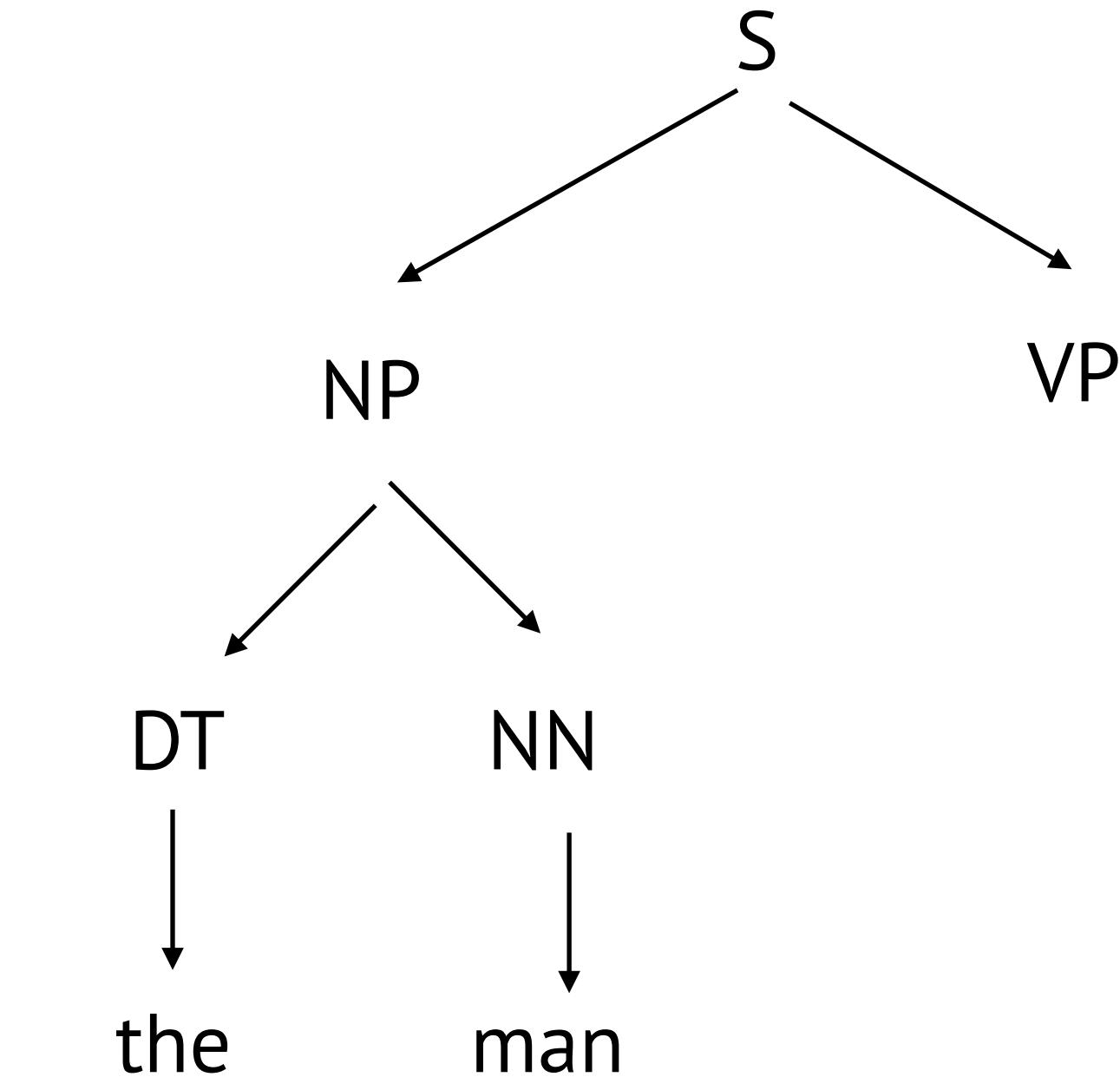
$R =$	$S \rightarrow NP\ VP$	$Vi \rightarrow \text{sleeps}$
	$VP \rightarrow Vi$	$Vt \rightarrow \text{saw}$
	$VP \rightarrow Vt\ NP$	$NN \rightarrow \text{man}$
	$VP \rightarrow VP\ PP$	$NN \rightarrow \text{woman}$
	$NP \rightarrow DT\ NN$	$NN \rightarrow \text{telescope}$
	$NP \rightarrow NP\ PP$	$DT \rightarrow \text{the}$
	$PP \rightarrow IN\ NP$	$IN \rightarrow \text{with}$
		$IN \rightarrow \text{in}$

(the NN VP)



Left-Most Derivations

$R =$	$S \rightarrow NP\ VP$	$Vi \rightarrow \text{sleeps}$
	$VP \rightarrow Vi$	$Vt \rightarrow \text{saw}$
	$VP \rightarrow Vt\ NP$	$NN \rightarrow \text{man}$
	$VP \rightarrow VP\ PP$	$NN \rightarrow \text{woman}$
	$NP \rightarrow DT\ NN$	$NN \rightarrow \text{telescope}$
	$NP \rightarrow NP\ PP$	$DT \rightarrow \text{the}$
	$PP \rightarrow IN\ NP$	$IN \rightarrow \text{with}$
		$IN \rightarrow \text{in}$

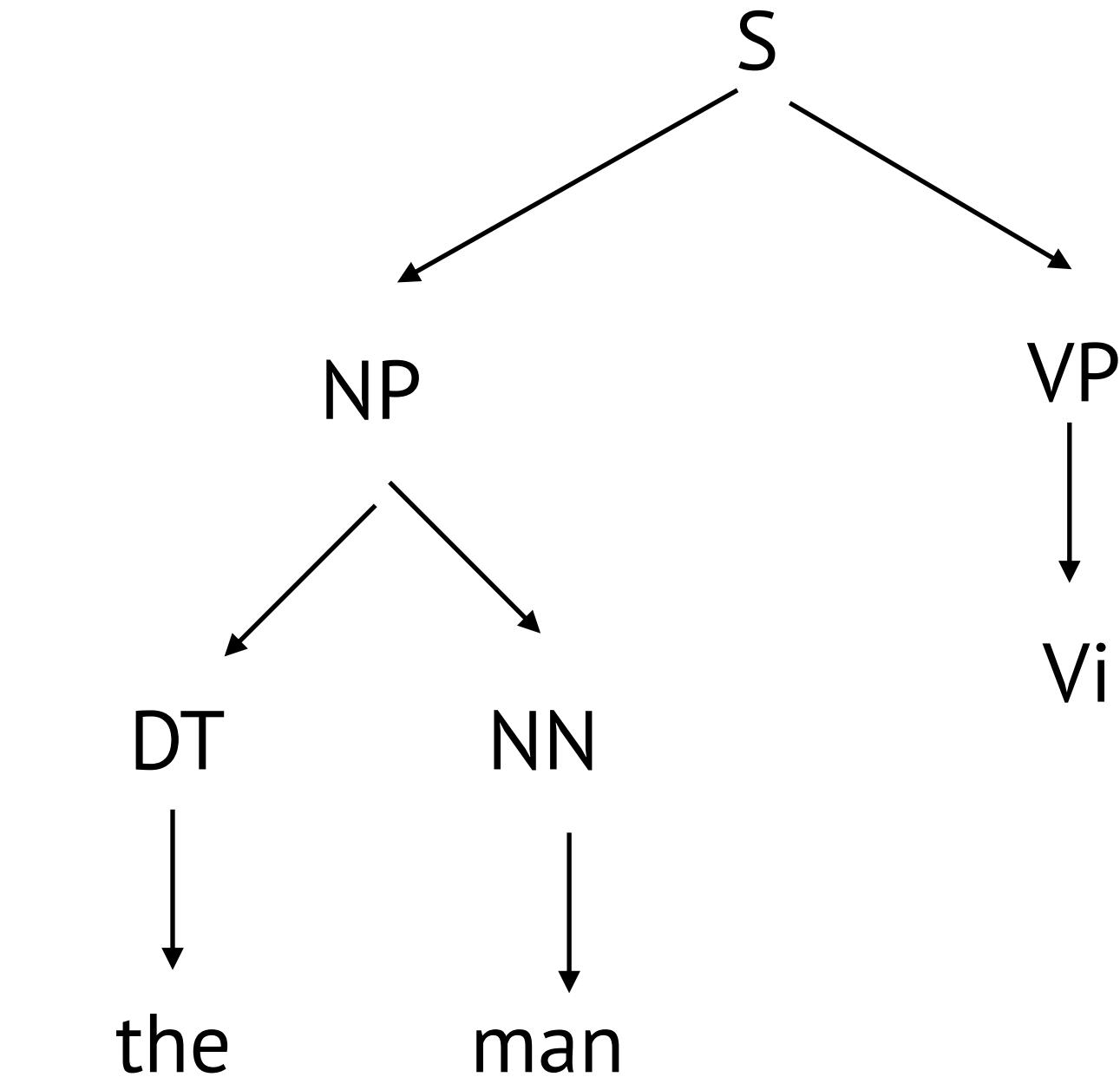


(the man VP)

Left-Most Derivations

$R =$

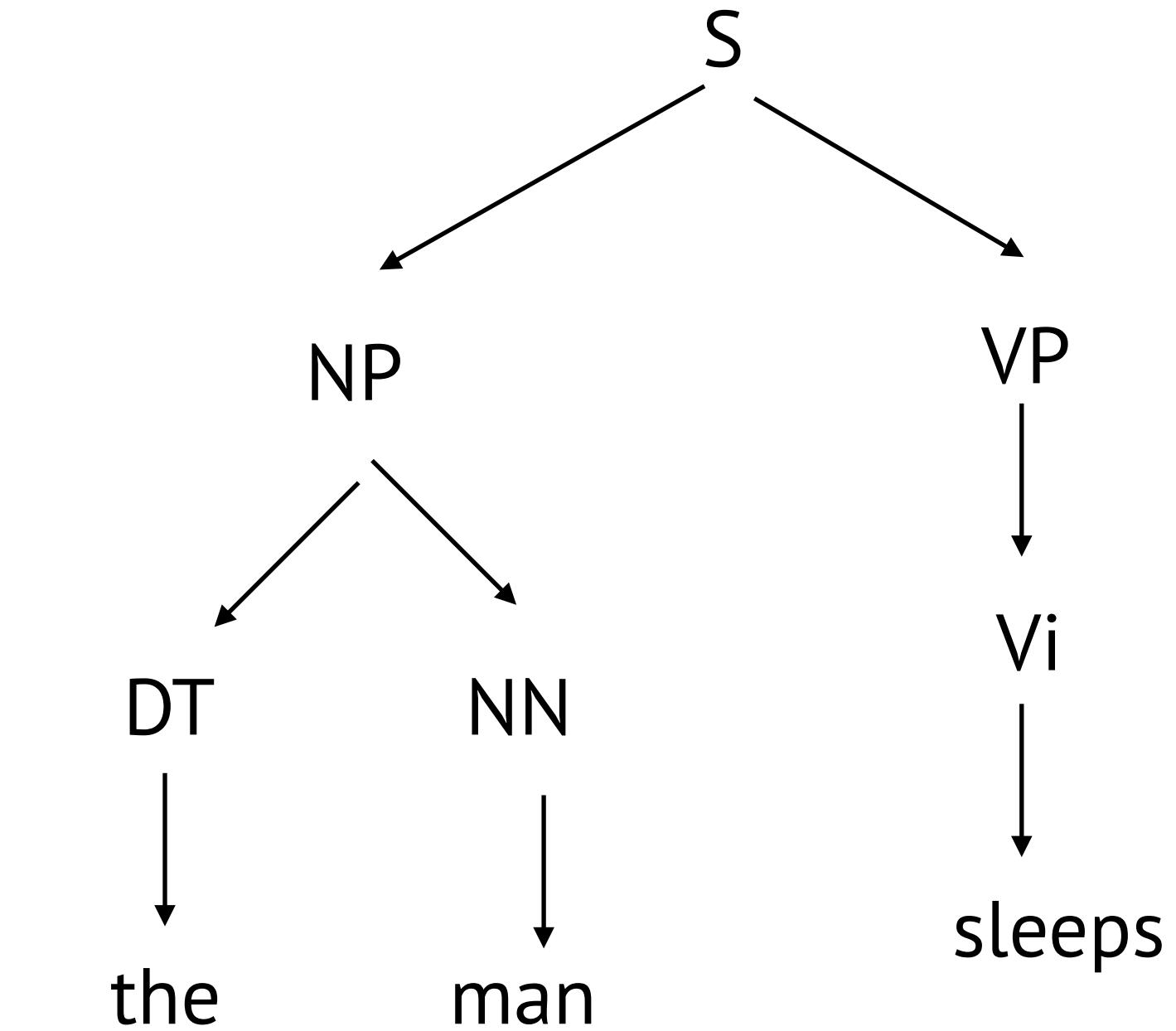
$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$
$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
$NP \longrightarrow DT \ NN$	$NN \longrightarrow \text{telescope}$
$NP \longrightarrow NP \ PP$	$DT \longrightarrow \text{the}$
$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{with}$
	$IN \longrightarrow \text{in}$



(the man Vi)

Left-Most Derivations

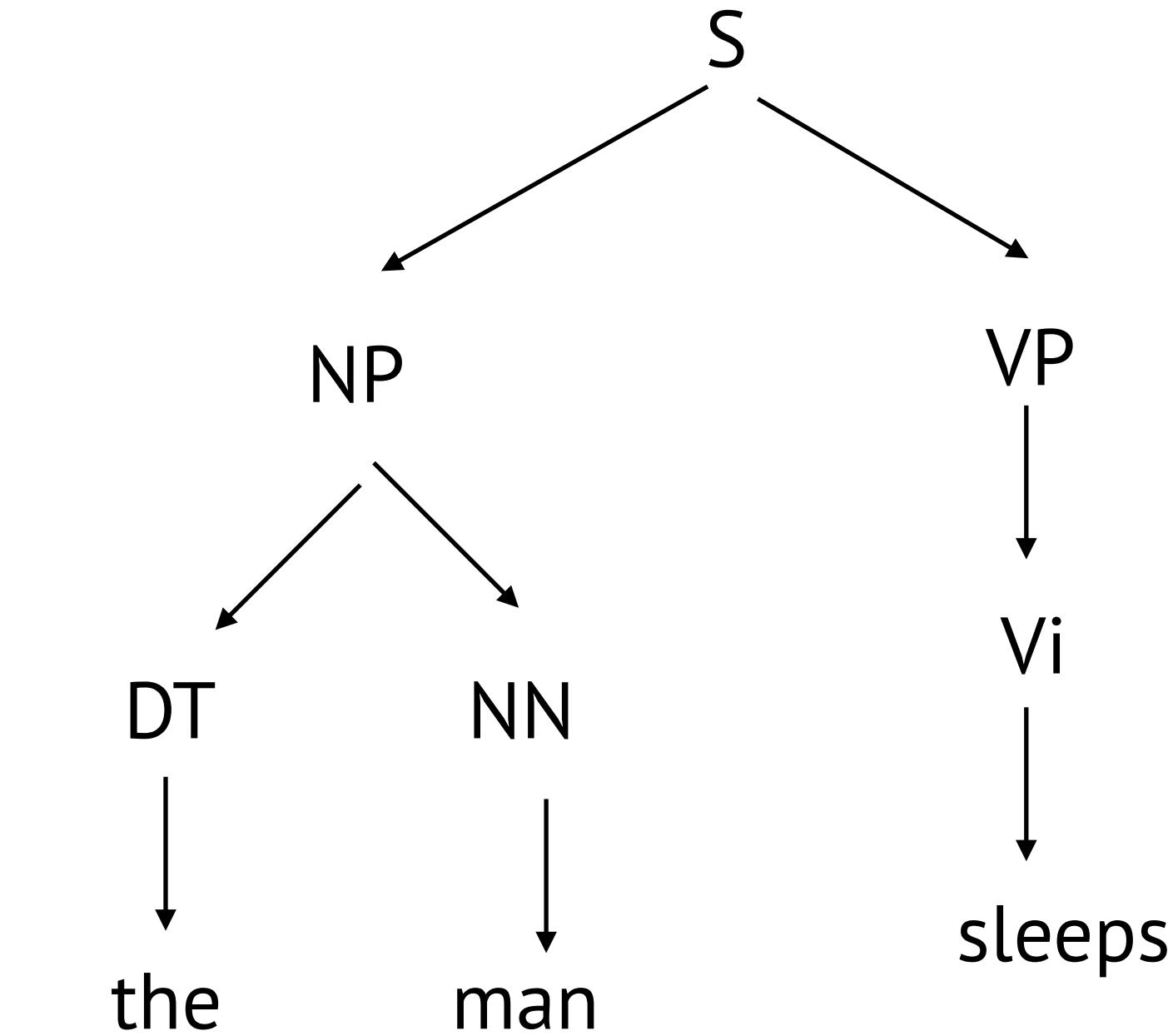
$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$
$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
$NP \longrightarrow DT \ NN$	$NN \longrightarrow \text{telescope}$
$NP \longrightarrow NP \ PP$	$DT \longrightarrow \text{the}$
$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{with}$
	$IN \longrightarrow \text{in}$



(the man sleeps)

Left-Most Derivations

$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
	$Vt \longrightarrow \text{saw}$
$VP \longrightarrow Vi$	
$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
	$NN \longrightarrow \text{telescope}$
$NP \longrightarrow DT \ NN$	$DT \longrightarrow \text{the}$
$NP \longrightarrow NP \ PP$	$IN \longrightarrow \text{with}$
$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{in}$

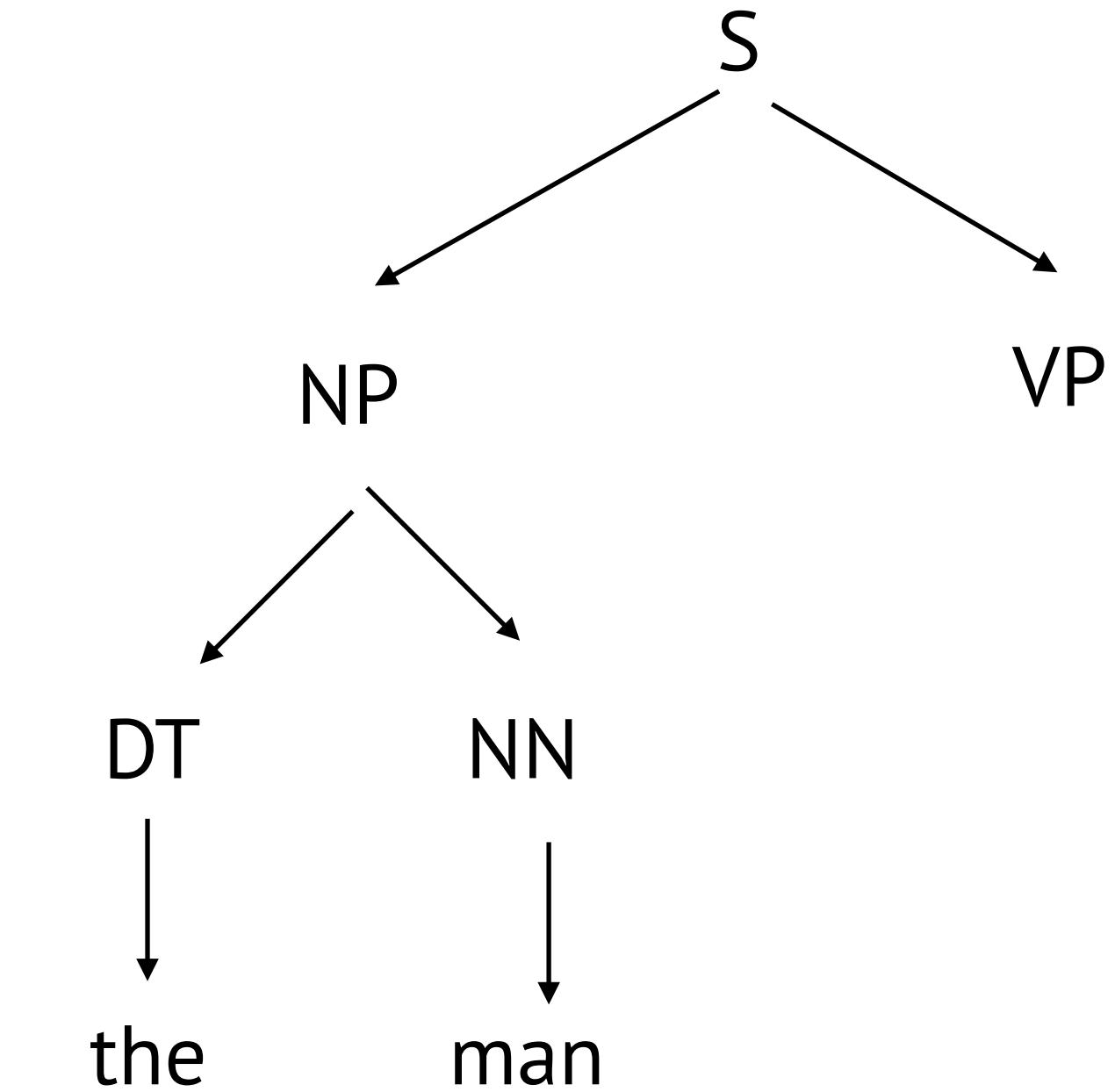


(the man sleeps)

Left-Most Derivations

$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$
$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
$NP \longrightarrow DT \ NN$	$NN \longrightarrow \text{telescope}$
$NP \longrightarrow NP \ PP$	$DT \longrightarrow \text{the}$
$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{with}$
	$IN \longrightarrow \text{in}$

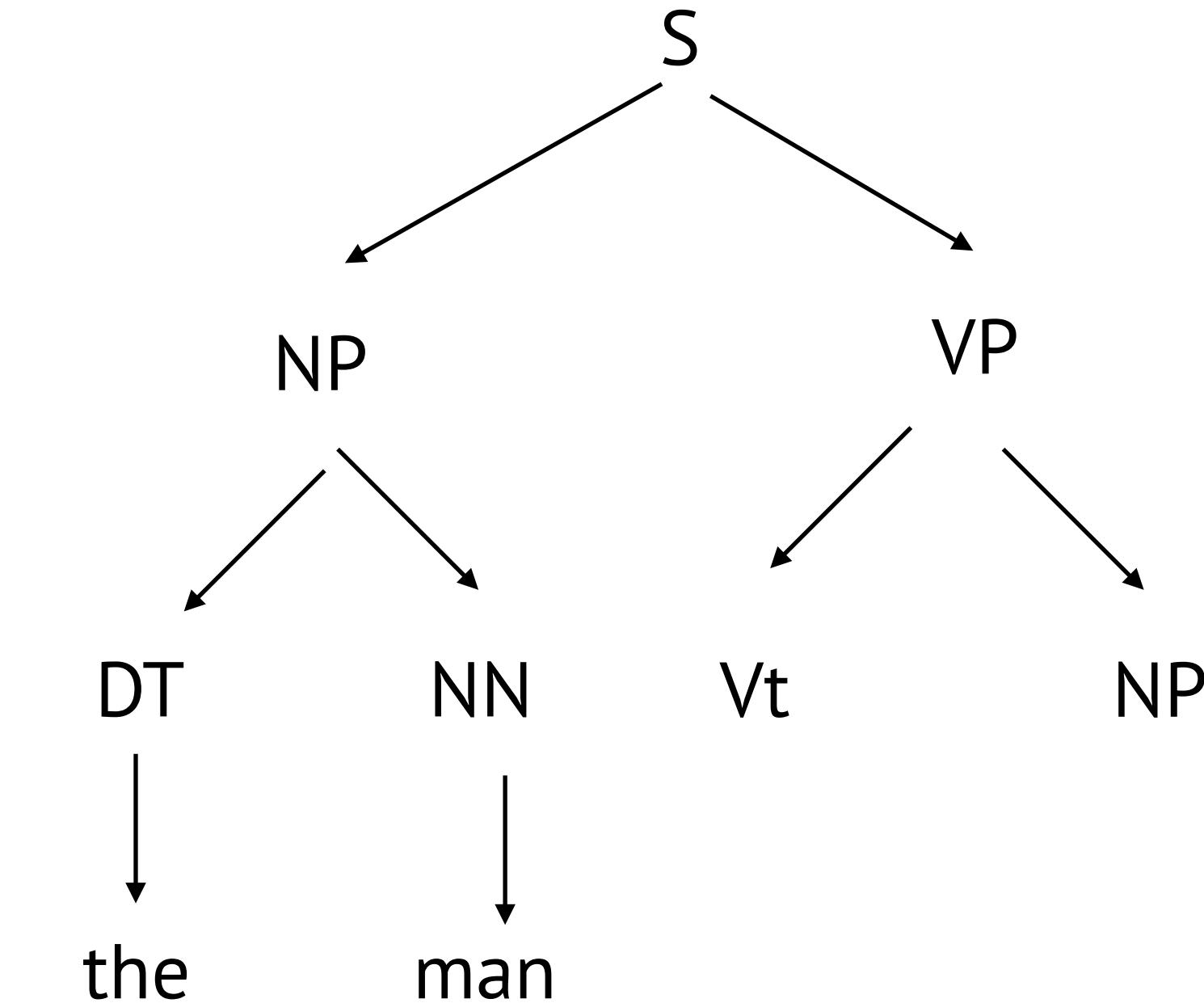
(the man VP)



Left-Most Derivations

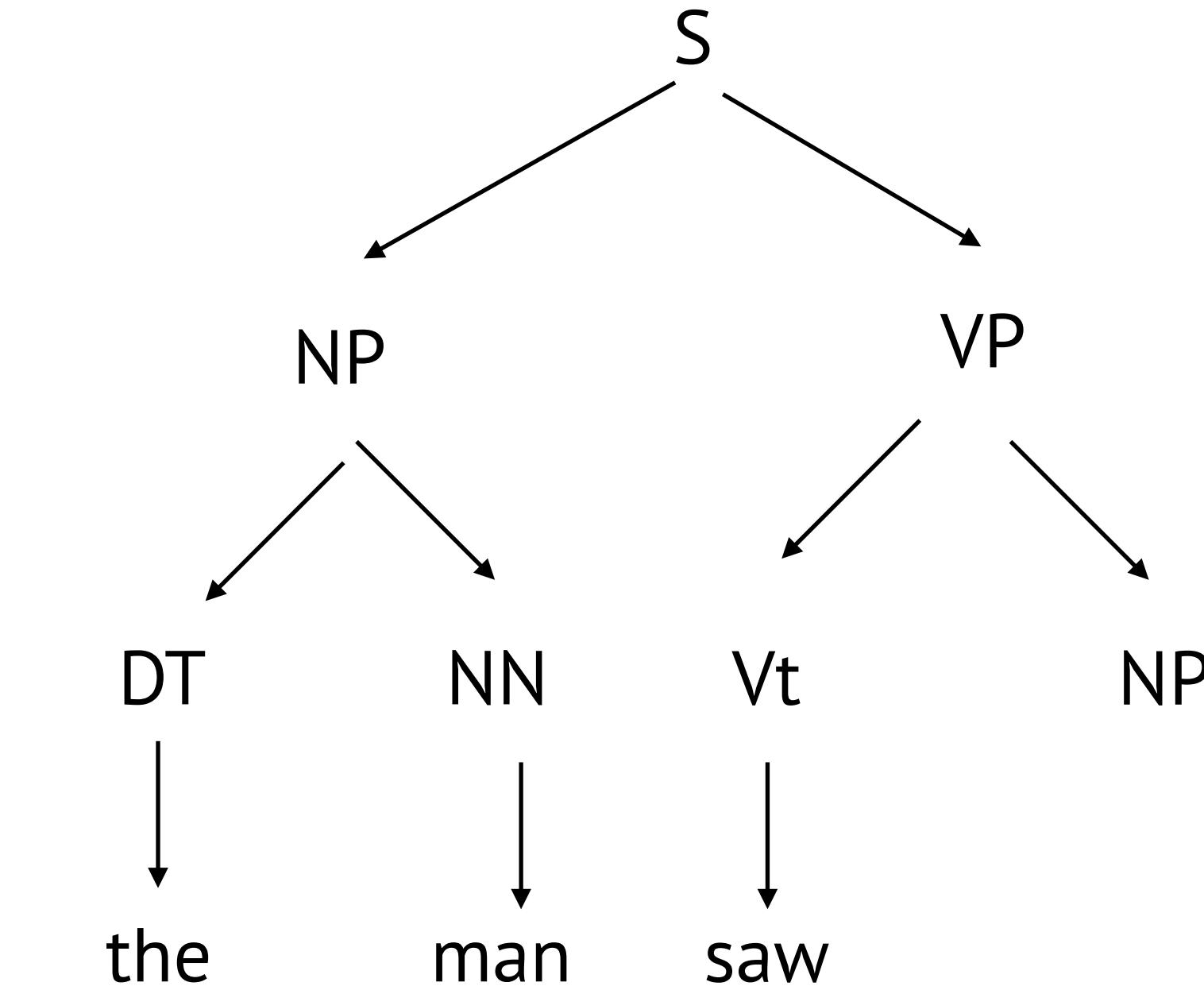
$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$
$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
$NP \longrightarrow DT \ NN$	$NN \longrightarrow \text{telescope}$
$NP \longrightarrow NP \ PP$	$DT \longrightarrow \text{the}$
$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{with}$
	$IN \longrightarrow \text{in}$

(the man Vt NP)



Left-Most Derivations

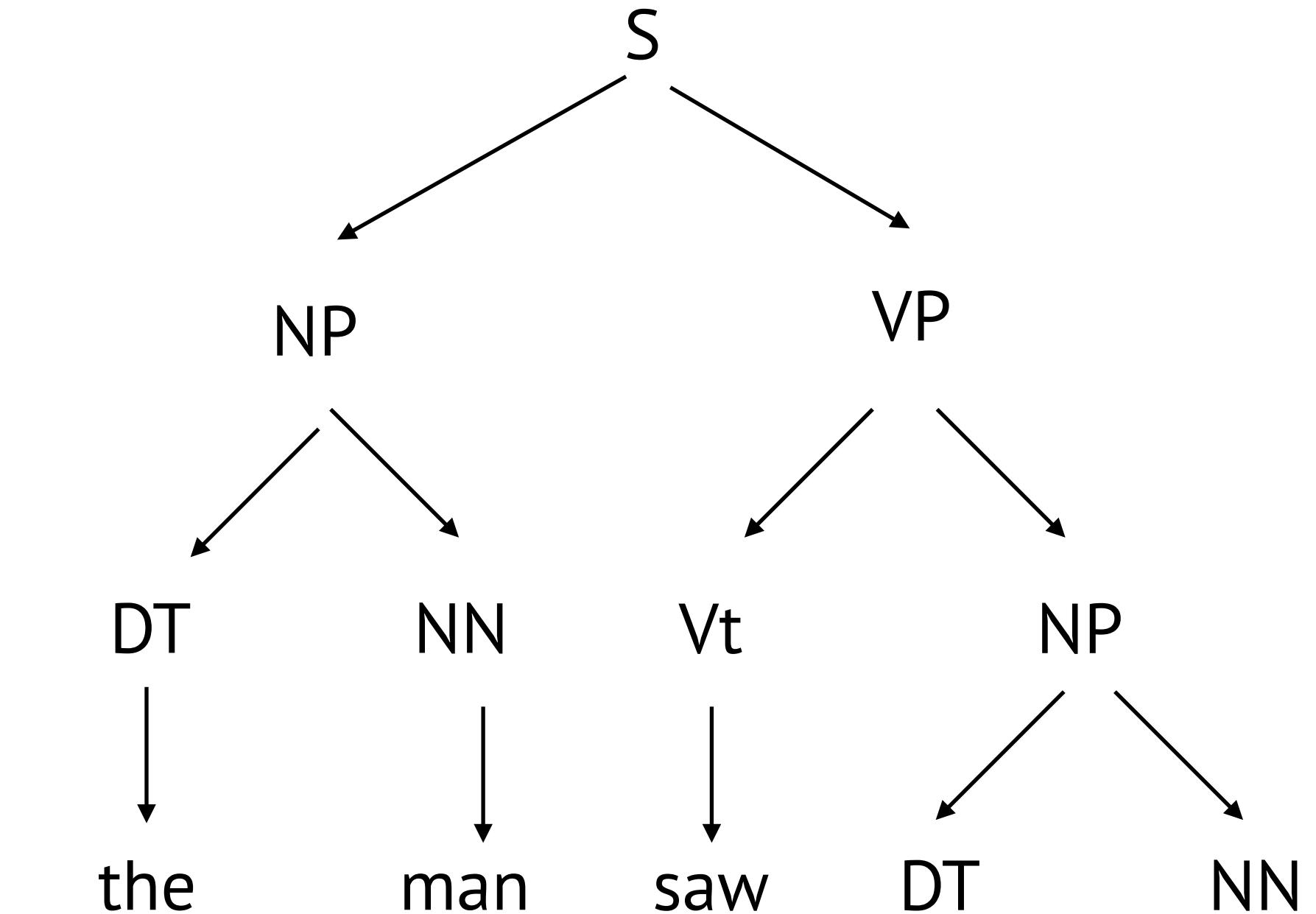
$S \longrightarrow NP \ VP$	$Vi \longrightarrow \text{sleeps}$
$VP \longrightarrow Vi$	$Vt \longrightarrow \text{saw}$
$VP \longrightarrow Vt \ NP$	$NN \longrightarrow \text{man}$
$VP \longrightarrow VP \ PP$	$NN \longrightarrow \text{woman}$
$NP \longrightarrow DT \ NN$	$NN \longrightarrow \text{telescope}$
$NP \longrightarrow NP \ PP$	$DT \longrightarrow \text{the}$
$PP \longrightarrow IN \ NP$	$IN \longrightarrow \text{with}$
	$IN \longrightarrow \text{in}$



(the man saw NP)

Left-Most Derivations

$R =$	$S \rightarrow NP\ VP$	$Vi \rightarrow \text{sleeps}$
	$VP \rightarrow Vi$	$Vt \rightarrow \text{saw}$
	$VP \rightarrow Vt\ NP$	$NN \rightarrow \text{man}$
	$VP \rightarrow VP\ PP$	$NN \rightarrow \text{woman}$
	$NP \rightarrow DT\ NN$	$NN \rightarrow \text{telescope}$
	$NP \rightarrow NP\ PP$	$DT \rightarrow \text{the}$
	$PP \rightarrow IN\ NP$	$IN \rightarrow \text{with}$
		$IN \rightarrow \text{in}$

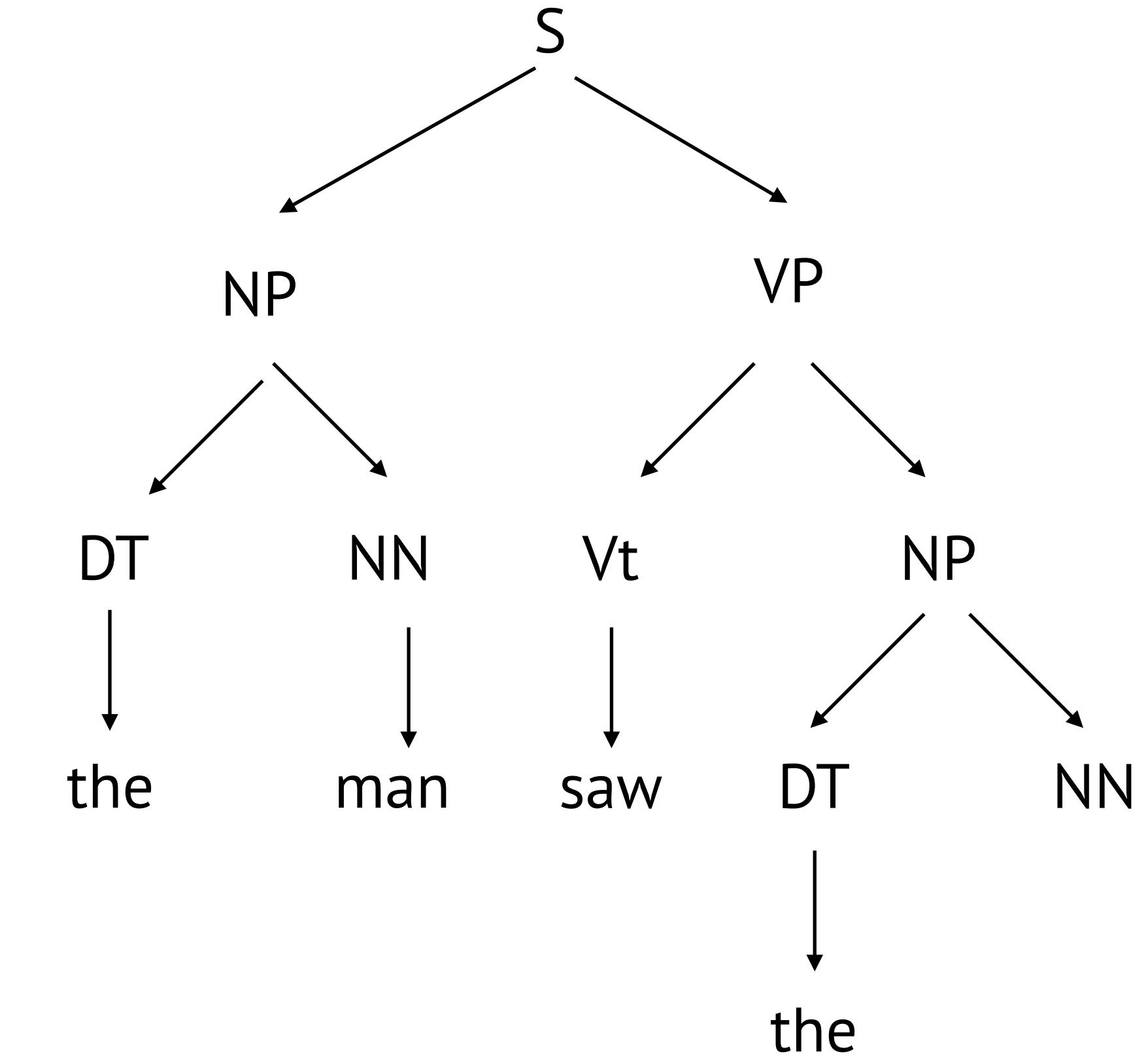


(the man saw DT NN)

Left-Most Derivations

$R =$	$S \rightarrow NP\ VP$	$Vi \rightarrow \text{sleeps}$
	$VP \rightarrow Vi$	$Vt \rightarrow \text{saw}$
	$VP \rightarrow Vt\ NP$	$NN \rightarrow \text{man}$
	$VP \rightarrow VP\ PP$	$NN \rightarrow \text{woman}$
	$NP \rightarrow DT\ NN$	$NN \rightarrow \text{telescope}$
	$NP \rightarrow NP\ PP$	$DT \rightarrow \text{the}$
	$PP \rightarrow IN\ NP$	$IN \rightarrow \text{with}$
		$IN \rightarrow \text{in}$

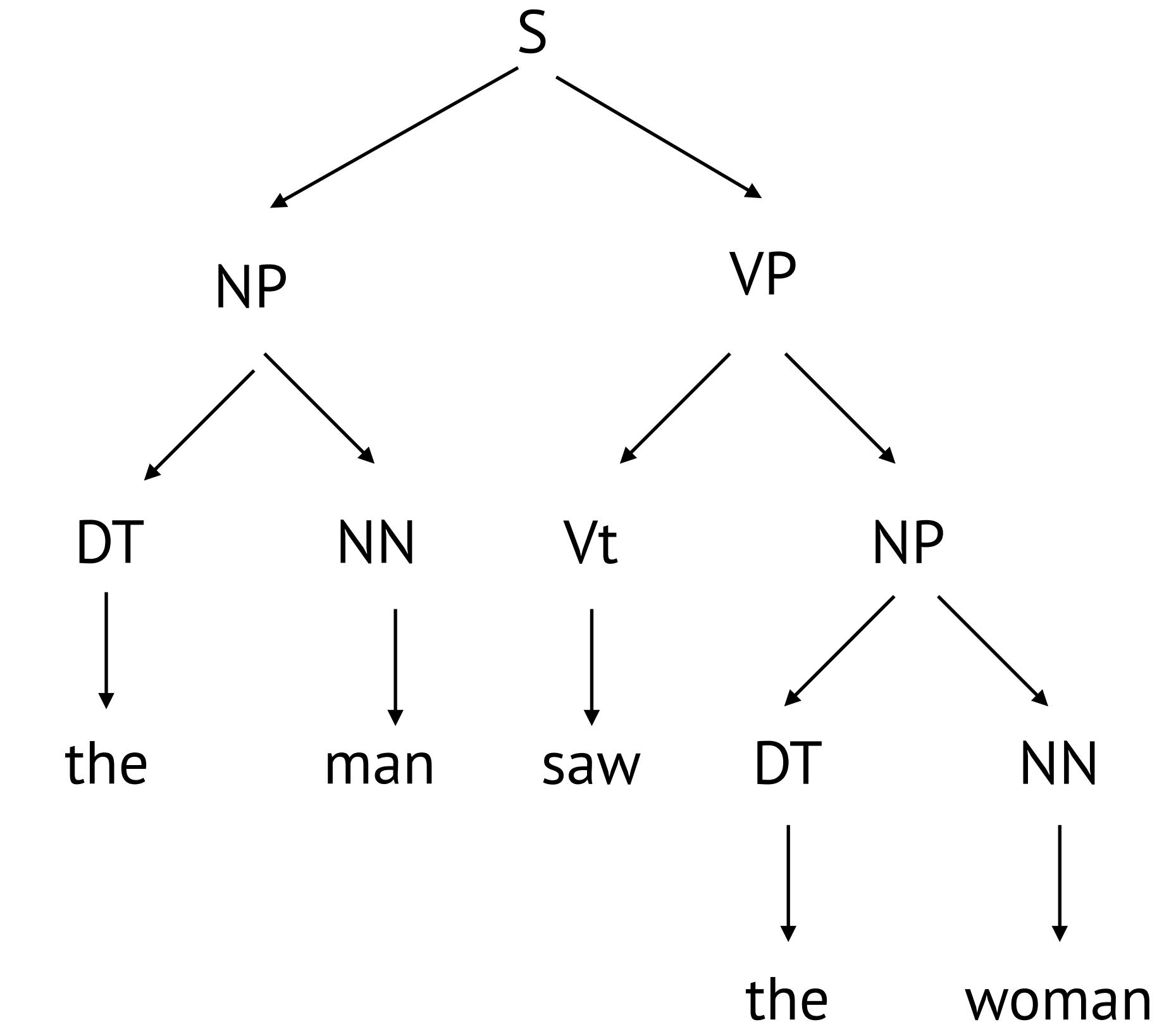
(the man saw the NN)



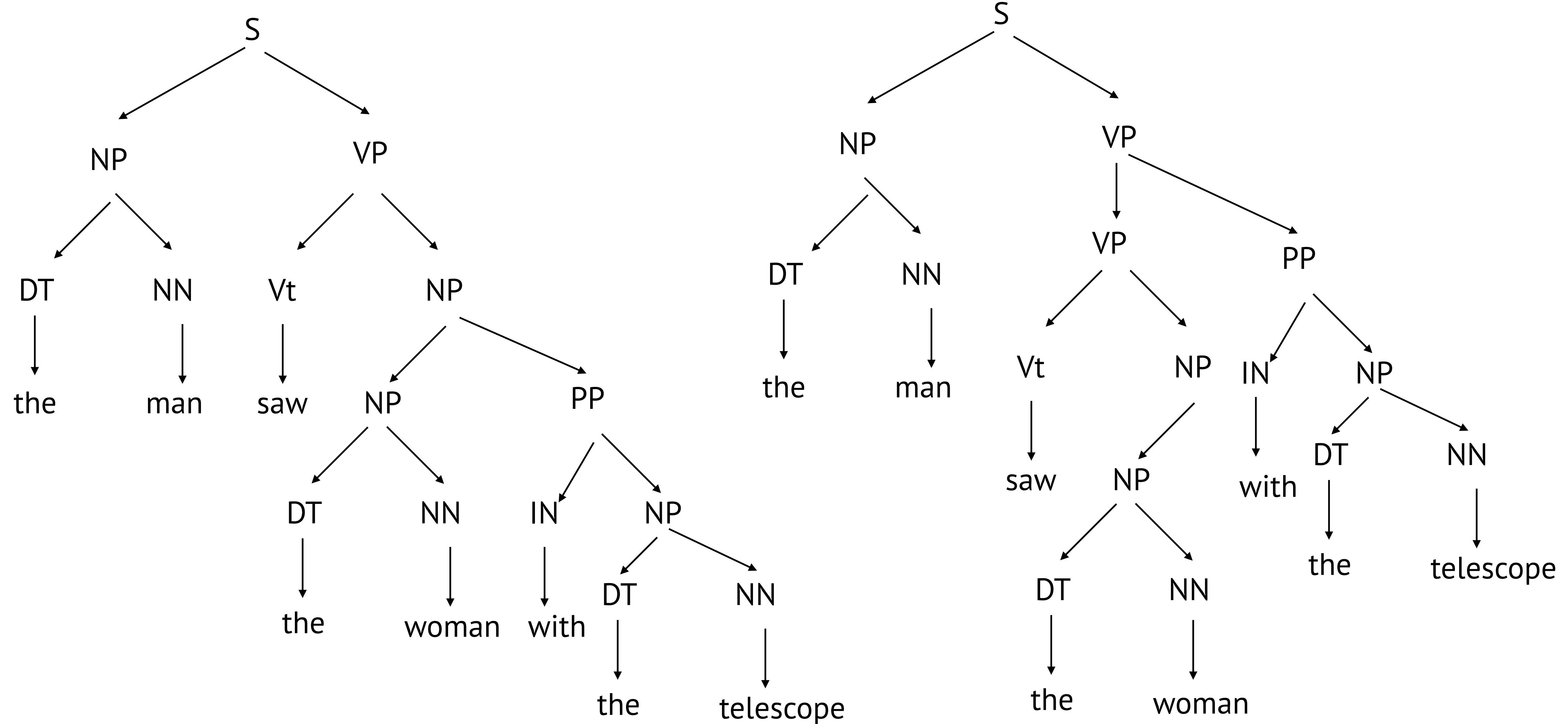
Left-Most Derivations

$R =$	$S \rightarrow NP\ VP$	$Vi \rightarrow \text{sleeps}$
	$VP \rightarrow Vi$	$Vt \rightarrow \text{saw}$
	$VP \rightarrow Vt\ NP$	$NN \rightarrow \text{man}$
	$VP \rightarrow VP\ PP$	$NN \rightarrow \text{woman}$
	$NP \rightarrow DT\ NN$	$NN \rightarrow \text{telescope}$
	$NP \rightarrow NP\ PP$	$DT \rightarrow \text{the}$
	$PP \rightarrow IN\ NP$	$IN \rightarrow \text{with}$
		$IN \rightarrow \text{in}$

(the man saw the woman)



Different Derivations Can Lead to the Same String

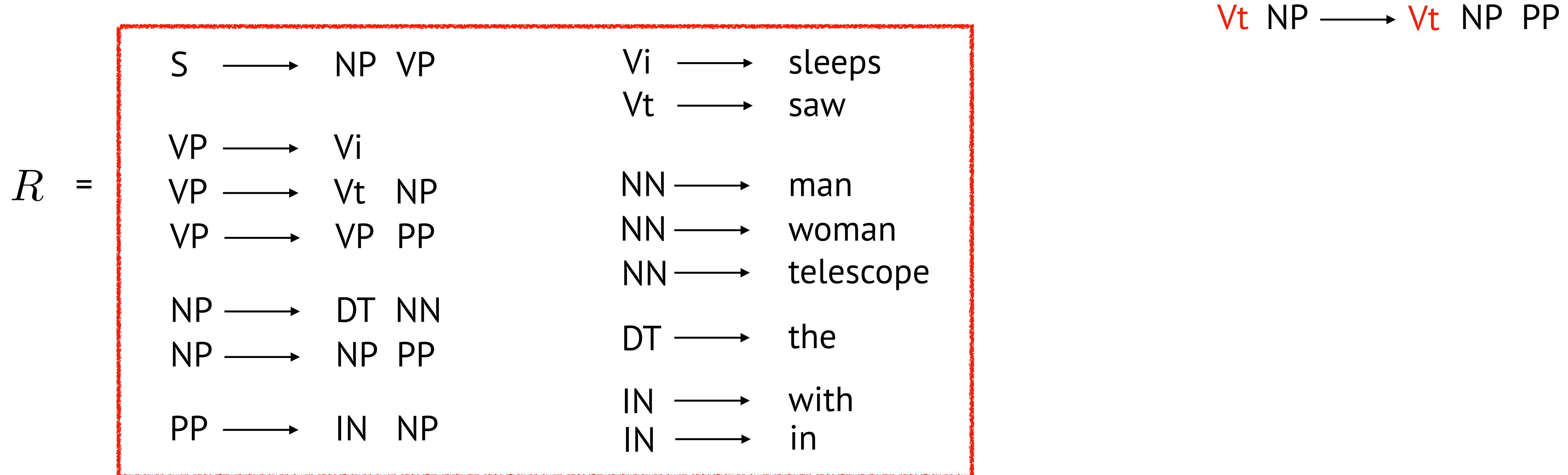


“Context-Free” – What does it mean?

$N = \{S, NP, VP, PP, DT, Vi, Vt, NN, IN\}$

$S = S$

$\Sigma = \{\text{sleeps, saw, man, woman, telescope, the, with, in}\}$



“Context-Free” – What does it mean?

N = {S, NP, VP, PP, DT, Vi, Vt, NN, IN}

S = S

Σ = {sleeps, saw, man, woman, telescope, the, with, in}

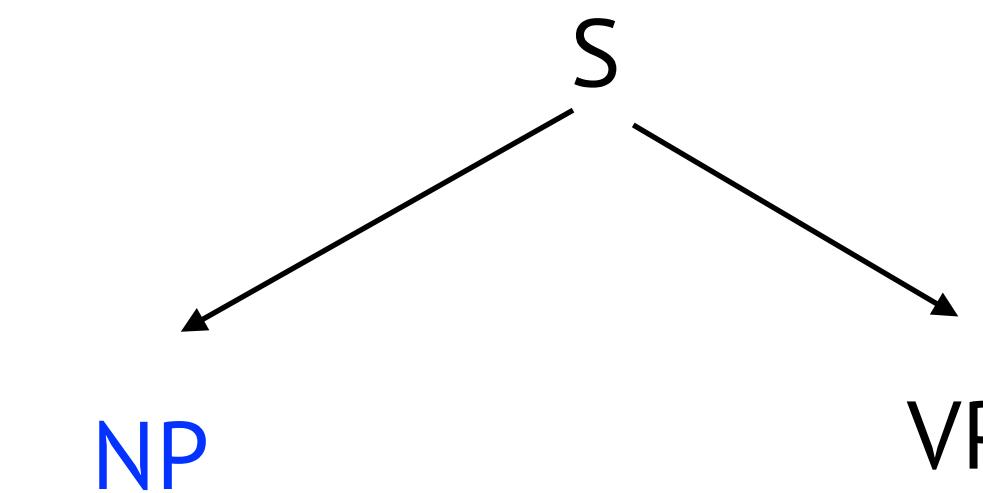
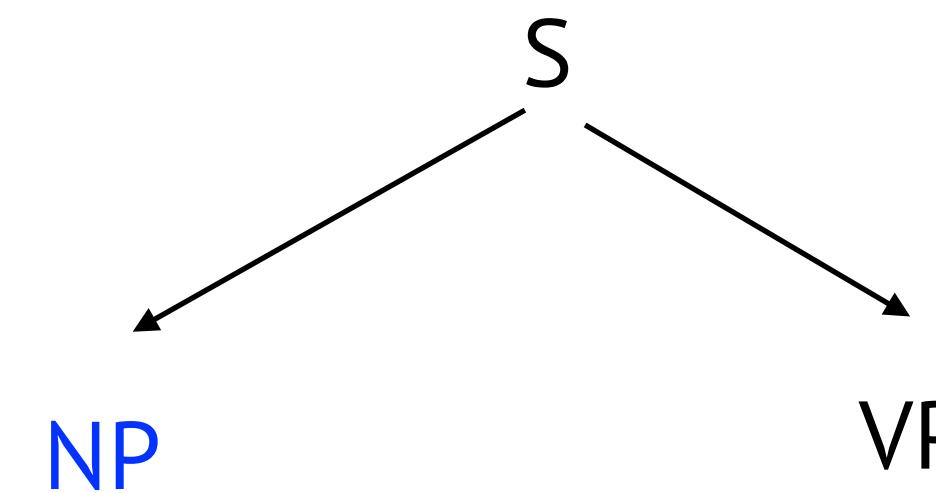
$R =$	$S \longrightarrow NP\ VP$ $VP \longrightarrow Vi$ $VP \longrightarrow Vt\ NP$ $VP \longrightarrow VP\ PP$ $NP \longrightarrow DT\ NN$ $Vt\ NP \longrightarrow Vt\ NP\ PP$ $PP \longrightarrow IN\ NP$	$Vi \longrightarrow \text{sleeps}$ $Vt \longrightarrow \text{saw}$ $NN \longrightarrow \text{man}$ $NN \longrightarrow \text{woman}$ $NN \longrightarrow \text{telescope}$ $DT \longrightarrow \text{the}$ $IN \longrightarrow \text{with}$ $IN \longrightarrow \text{in}$	non-context-free
-------	--	---	------------------

Probabilistic Context-Free Grammars

$R =$	$S \rightarrow NP\ VP \quad 1.0$	$Vi \rightarrow \text{sleeps} \quad 1.0$
	$VP \rightarrow Vi \quad 0.4$	$Vt \rightarrow \text{saw} \quad 1.0$
	$VP \rightarrow Vt\ NP \quad 0.4$	$NN \rightarrow \text{man} \quad 0.7$
	$VP \rightarrow VP\ PP \quad 0.2$	$NN \rightarrow \text{woman} \quad 0.2$
	$NP \rightarrow DT\ NN \quad 0.3$	$NN \rightarrow \text{telescope} \quad 0.1$
	$NP \rightarrow NP\ PP \quad 0.7$	$DT \rightarrow \text{the} \quad 1.0$
	$PP \rightarrow IN\ NP \quad 1.0$	$IN \rightarrow \text{with} \quad 0.5$
		$IN \rightarrow \text{in} \quad 0.5$

Probabilistic Context-Free Grammars

$R =$	$S \longrightarrow NP\ VP$	1.0	$Vi \longrightarrow sleeps$	1.0
	$VP \longrightarrow Vi$	0.4	$Vt \longrightarrow saw$	1.0
	$VP \longrightarrow Vt\ NP$	0.4	$NN \longrightarrow man$	0.7
	$VP \longrightarrow VP\ PP$	0.2	$NN \longrightarrow woman$	0.2
	$NP \longrightarrow DT\ NN$	0.3	$NN \longrightarrow telescope$	0.1
	$NP \longrightarrow NP\ PP$	0.7	$DT \longrightarrow the$	1.0
	$PP \longrightarrow IN\ NP$	1.0	$IN \longrightarrow with$	0.5
			$IN \longrightarrow in$	0.5



Probabilistic Context-Free Grammars

$R =$	$S \longrightarrow NP\ VP$	1.0	$Vi \longrightarrow sleeps$	1.0
	$VP \longrightarrow Vi$	0.4	$Vt \longrightarrow saw$	1.0
	$VP \longrightarrow Vt\ NP$	0.4	$NN \longrightarrow man$	0.7
	$VP \longrightarrow VP\ PP$	0.2	$NN \longrightarrow woman$	0.2
	$NP \longrightarrow DT\ NN$	0.3	$NN \longrightarrow telescope$	0.1
	$NP \longrightarrow NP\ PP$	0.7	$DT \longrightarrow the$	1.0
	$PP \longrightarrow IN\ NP$	1.0	$IN \longrightarrow with$	0.5
			$IN \longrightarrow in$	0.5

