Regular Expressions and Context Free Grammars

COMP7607 — Lecture 5

Lingpeng Kong

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

Regular Expressions

A formal language for specifying text strings

How can we search for any of these?

woodchuck
woodchucks
Woodchuck
Woodchucks

Slides from Dan Jurafsky https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html

Regular Expressions: Disjunctions

Letters inside square brackets []

Pattern	Matches
[wW]oodchuck	Woodchuck, woodchuck
[1234567890]	Any digit

Ranges [A-Z]

Pattern	Matches	
[A-Z]	An upper case letter	Drenched Blossoms
[a-z]	A lower case letter	my beans were impatient
[0-9]	A single digit	Chapter 1: Down the Rabbit Hole

Regular Expressions: Negation in Disjunction

Negations [^Ss]
  Caret means negation only when first in []

Pattern	Matches	
[^A-Z]	Not an upper case letter	Oyfn pripetchik
[^Ss]	Neither ‘S’ nor ‘s’	I have no exquisite reason”
[^e^]	Neither e nor ^	Look here
a^b	The pattern a caret b	Look up a^b now

Regular Expressions: More Disjunction

Woodchucks is another name for groundhog!
The pipe | for disjunction

Pattern	Matches
groundhog|woodchuck	
yours|mine	yours mine
a|b|c	= [abc]
[gG]roundhog|[Ww]oodchuck	

Regular Expressions: ? * + .

Pattern	Matches	
colou?r	Optional previous char	color colour
oo*h!	0 or more of previous char	oh! ooh! oooh! ooooh!
o+h!	1 or more of previous char	oh! ooh! oooh! ooooh!
baa+		baa baaa baaaa baaaaa
beg.n		begin begun begun beg3n

Regular Expressions: Anchors ^ $

Pattern	Matches
^[A-Z]	Palo Alto
^[^A-Za-z]	1 "Hello"
\.$	The end.
.$	The end? The end!

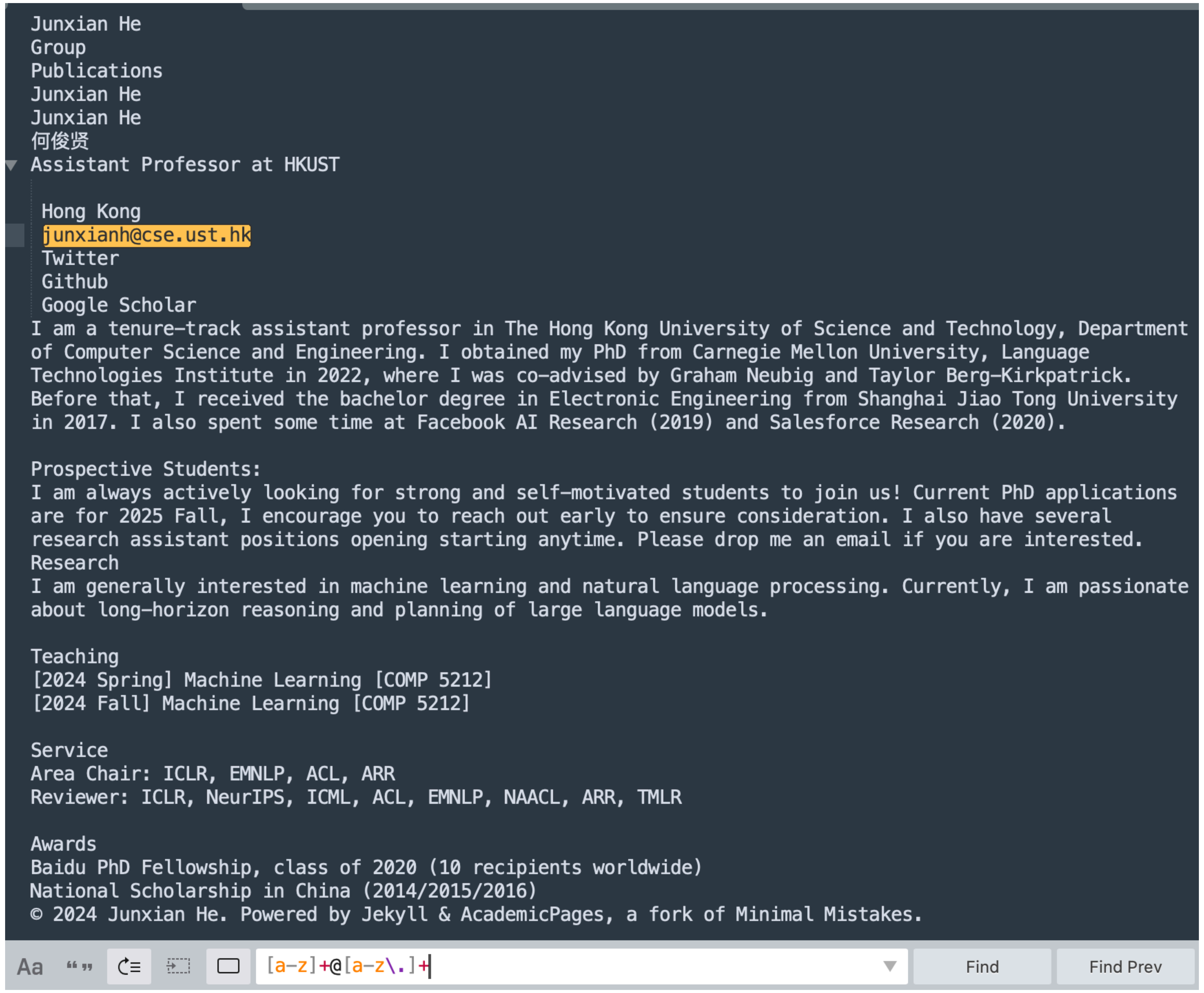
Example

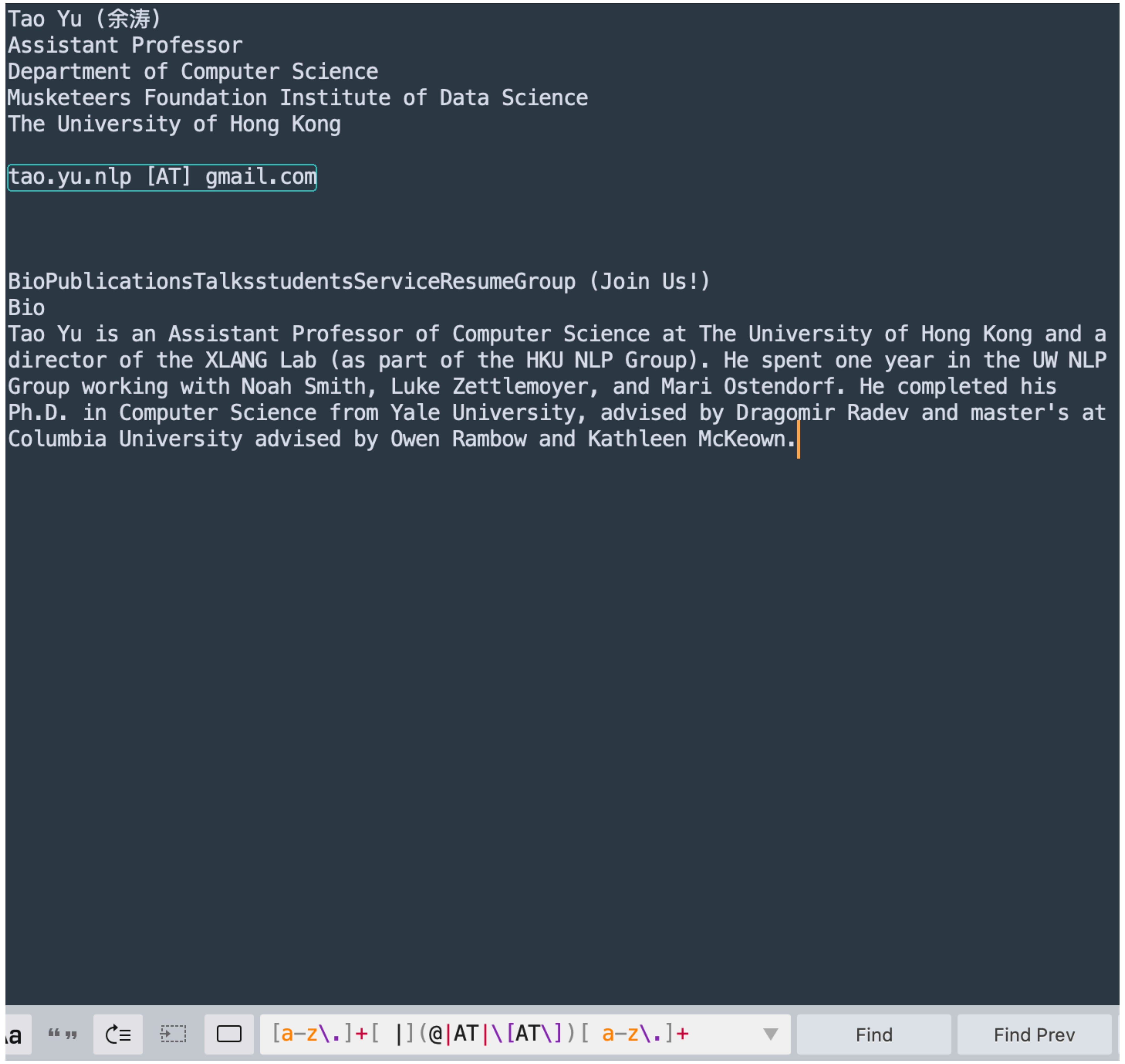
Find me all instances of the word “the” in a text.

the	Misses capitalized examples
[tT]he	Incorrectly returns other or theology
[^a-zA-Z][tT]he[^a-zA-Z]	

Find me the email address on a webpage.

Example


Junxian He
Group
Publications
Junxian He
Junxian He
何俊贤
Assistant Professor at HKUST
Hong Kong
junxianh@cse.ust.hk
Twitter
Github
Google Scholar
I am a tenure-track assistant professor in The Hong Kong University of Science and Technology, Department of Computer Science and Engineering. I obtained my PhD from Carnegie Mellon University, Language Technologies Institute in 2022, where I was co-advised by Graham Neubig and Taylor Berg-Kirkpatrick. Before that, I received the bachelor degree in Electronic Engineering from Shanghai Jiao Tong University in 2017. I also spent some time at Facebook AI Research (2019) and Salesforce Research (2020).
Prospective Students:
I am always actively looking for strong and self-motivated students to join us! Current PhD applications are for 2025 Fall, I encourage you to reach out early to ensure consideration. I also have several research assistant positions opening starting anytime. Please drop me an email if you are interested.
Research
I am generally interested in machine learning and natural language processing. Currently, I am passionate about long-horizon reasoning and planning of large language models.
Teaching
[2024 Spring] Machine Learning [COMP 5212]
[2024 Fall] Machine Learning [COMP 5212]
Service
Area Chair: ICLR, EMNLP, ACL, ARR
Reviewer: ICLR, NeurIPS, ICML, ACL, EMNLP, NAACL, ARR, TMLR
Awards
Baidu PhD Fellowship, class of 2020 (10 recipients worldwide)
National Scholarship in China (2014/2015/2016)
© 2024 Junxian He. Powered by Jekyll & AcademicPages, a fork of Minimal Mistakes.
Aa
[a-z]+@[a-z\.]+
Find
Find Prev

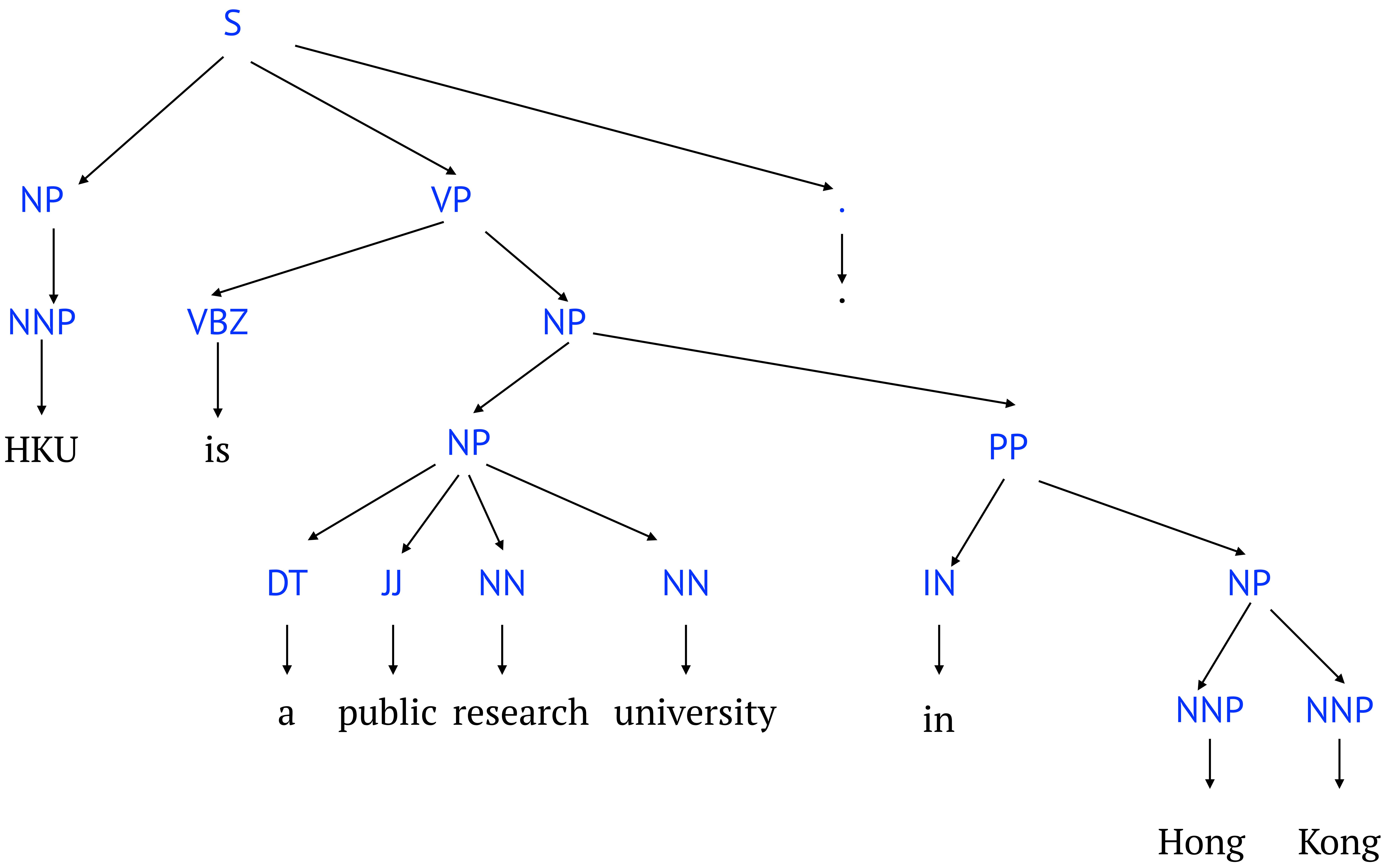

Tao Yu (余涛)
Assistant Professor
Department of Computer Science
Musketeers Foundation Institute of Data Science
The University of Hong Kong
tao.yu.nlp [AT] gmail.com
BioPublicationsTalksstudentsServiceResumeGroup (Join Us!)
Bio
Tao Yu is an Assistant Professor of Computer Science at The University of Hong Kong and a director of the XLANG Lab (as part of the HKU NLP Group). He spent one year in the UW NLP Group working with Noah Smith, Luke Zettlemoyer, and Mari Ostendorf. He completed his Ph.D. in Computer Science from Yale University, advised by Dragomir Radev and master's at Columbia University advised by Owen Rambow and Kathleen McKeown.
[a-z\.]+[|](@|AT|\[AT\])[a-z\.]+
Find
Find Prev

Example

Match strings with equal number of a and b.

e.g., aaabbb, aabb, aaaabbbb...

Linguistic Structures



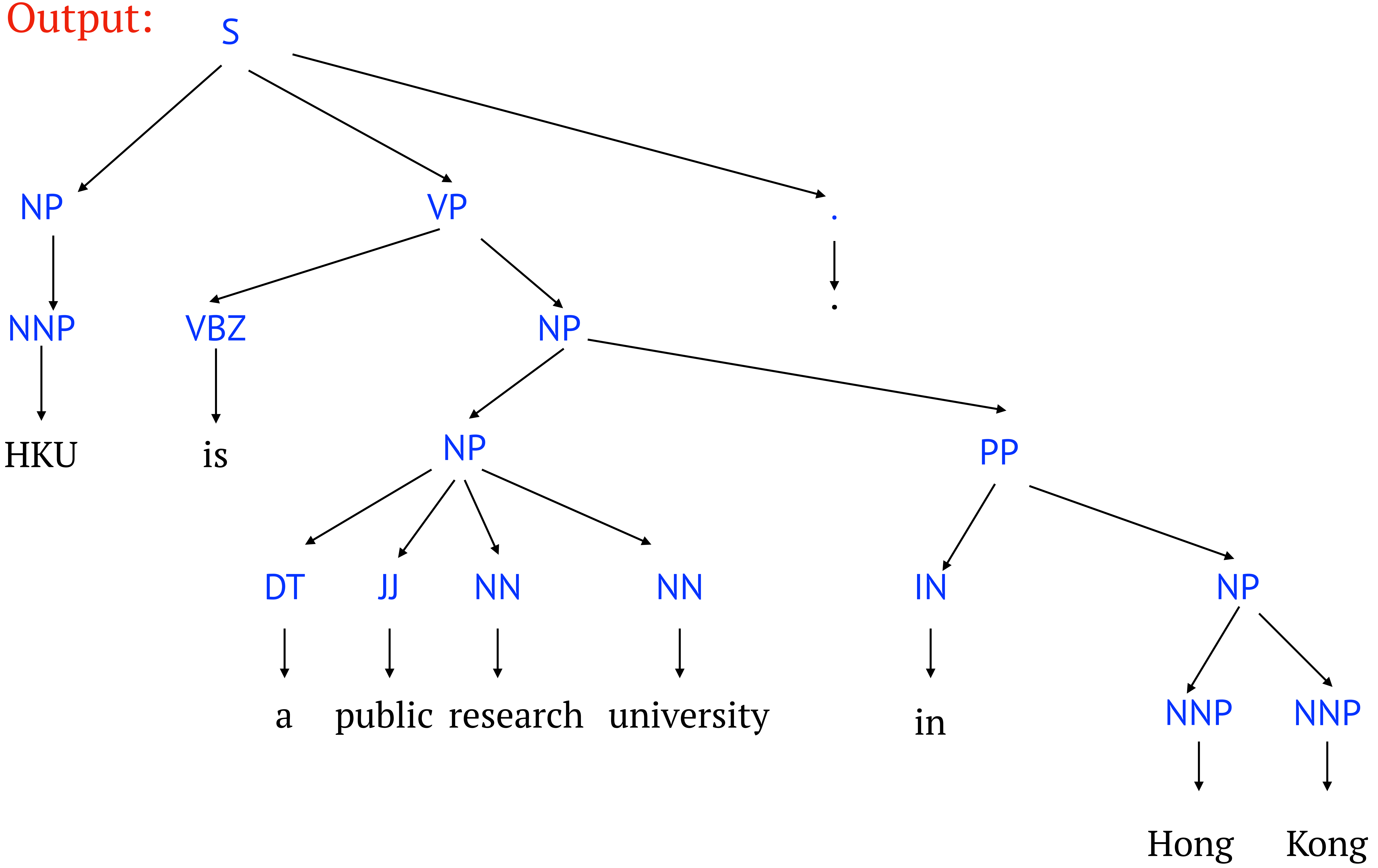
phrase-Structure tree,
constituency tree

Parsing (Phrase-structure Parsing)

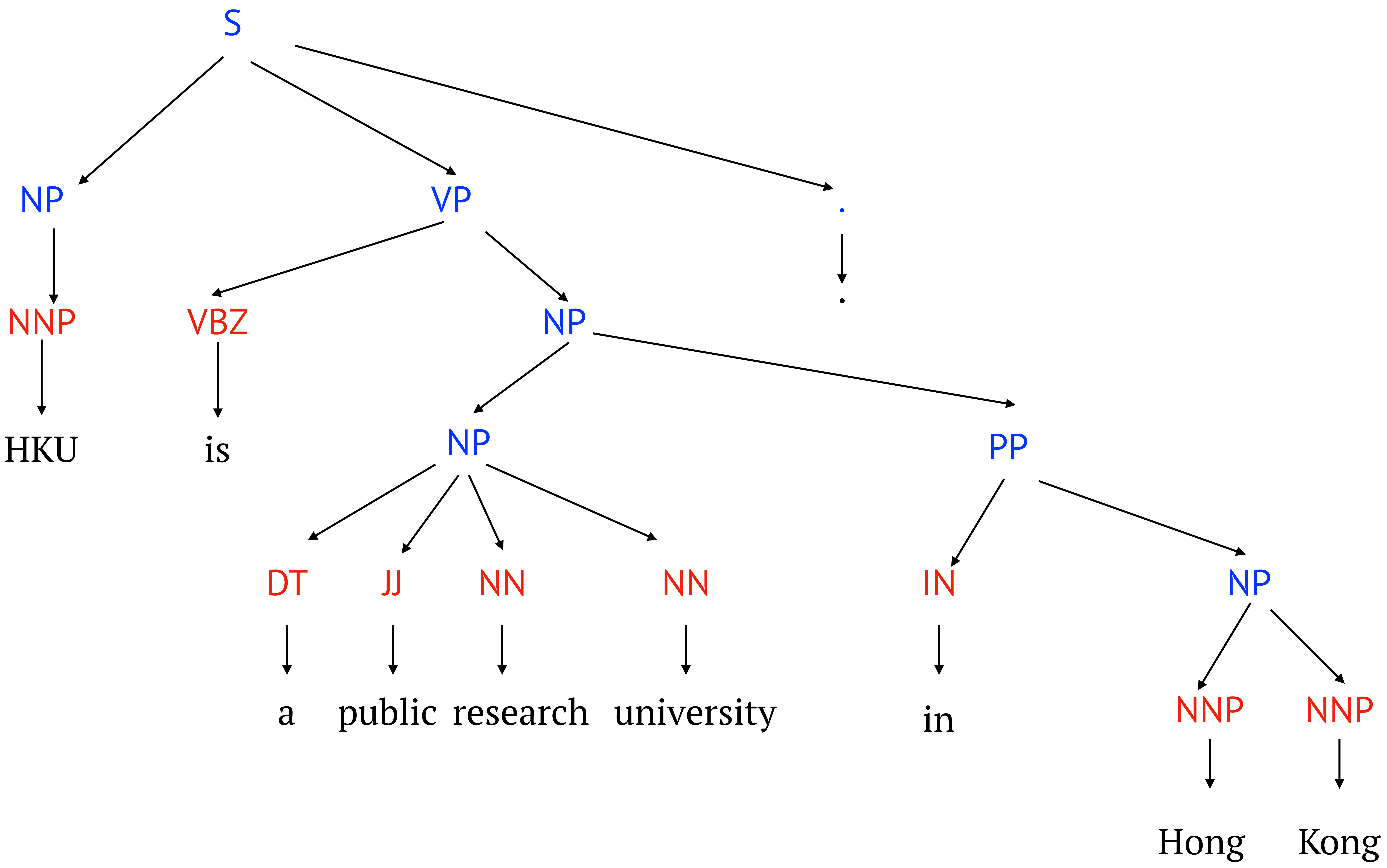
Input:

HKU is a public research university in Hong Kong.

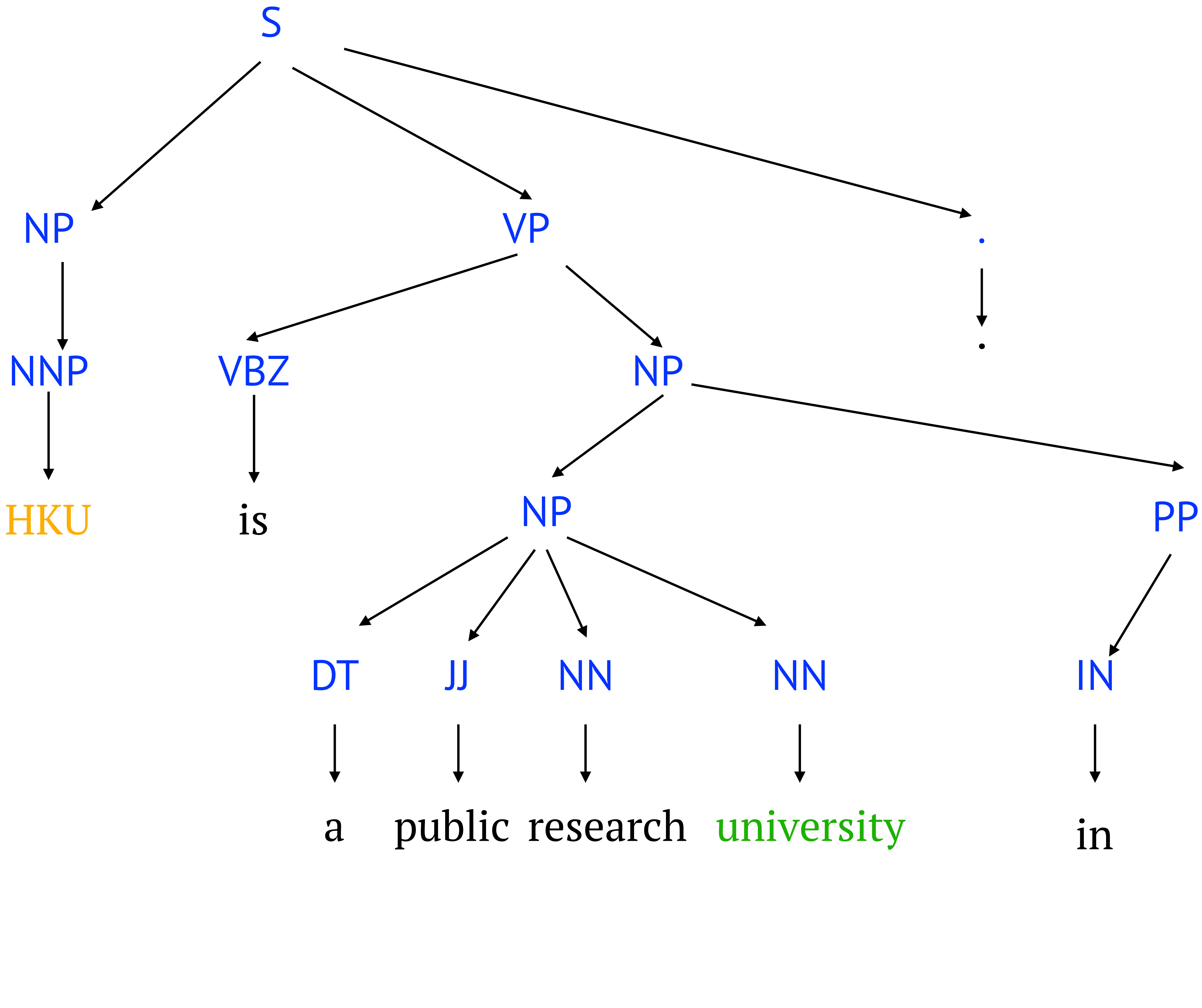
Output:

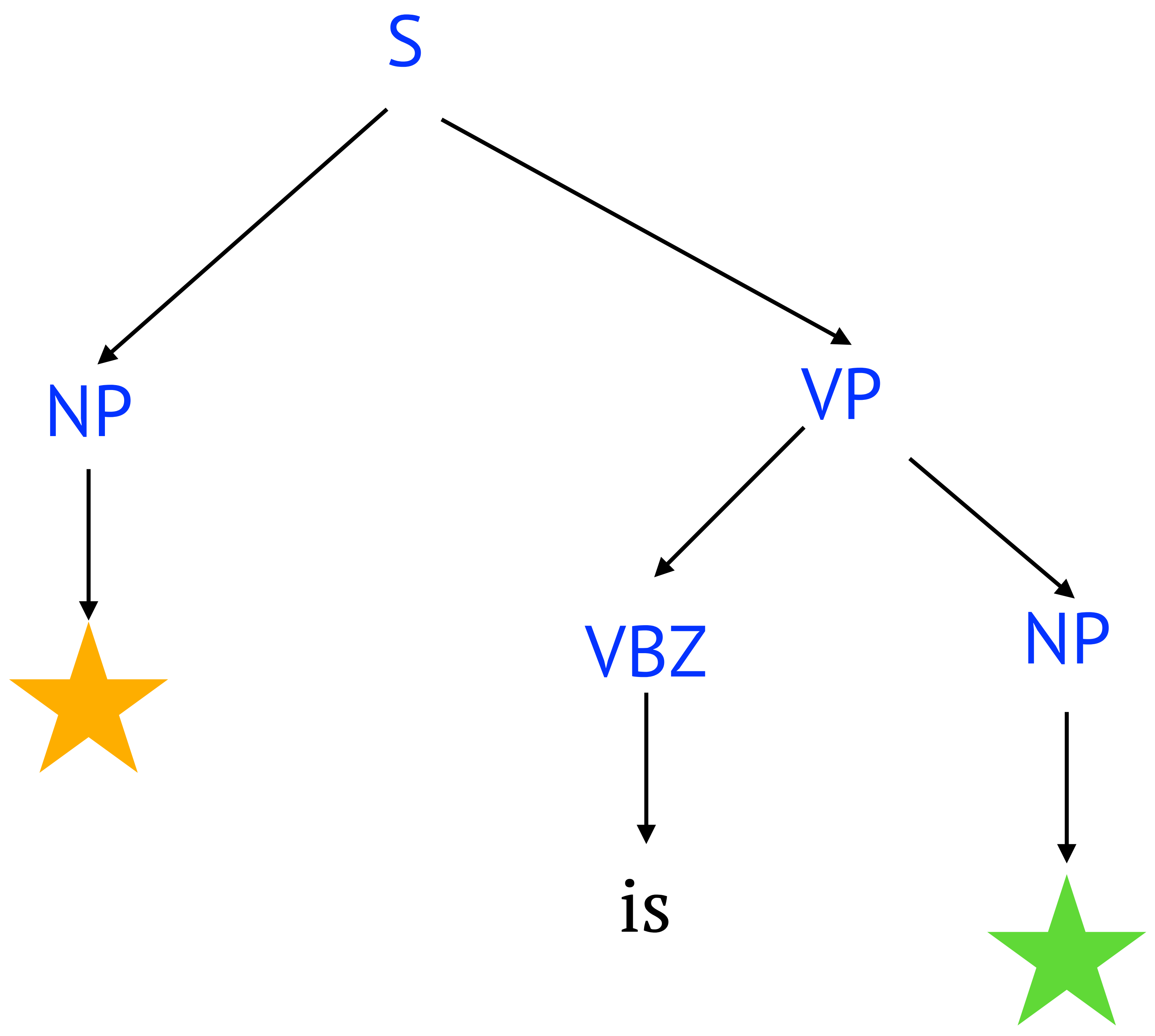


Parse Trees

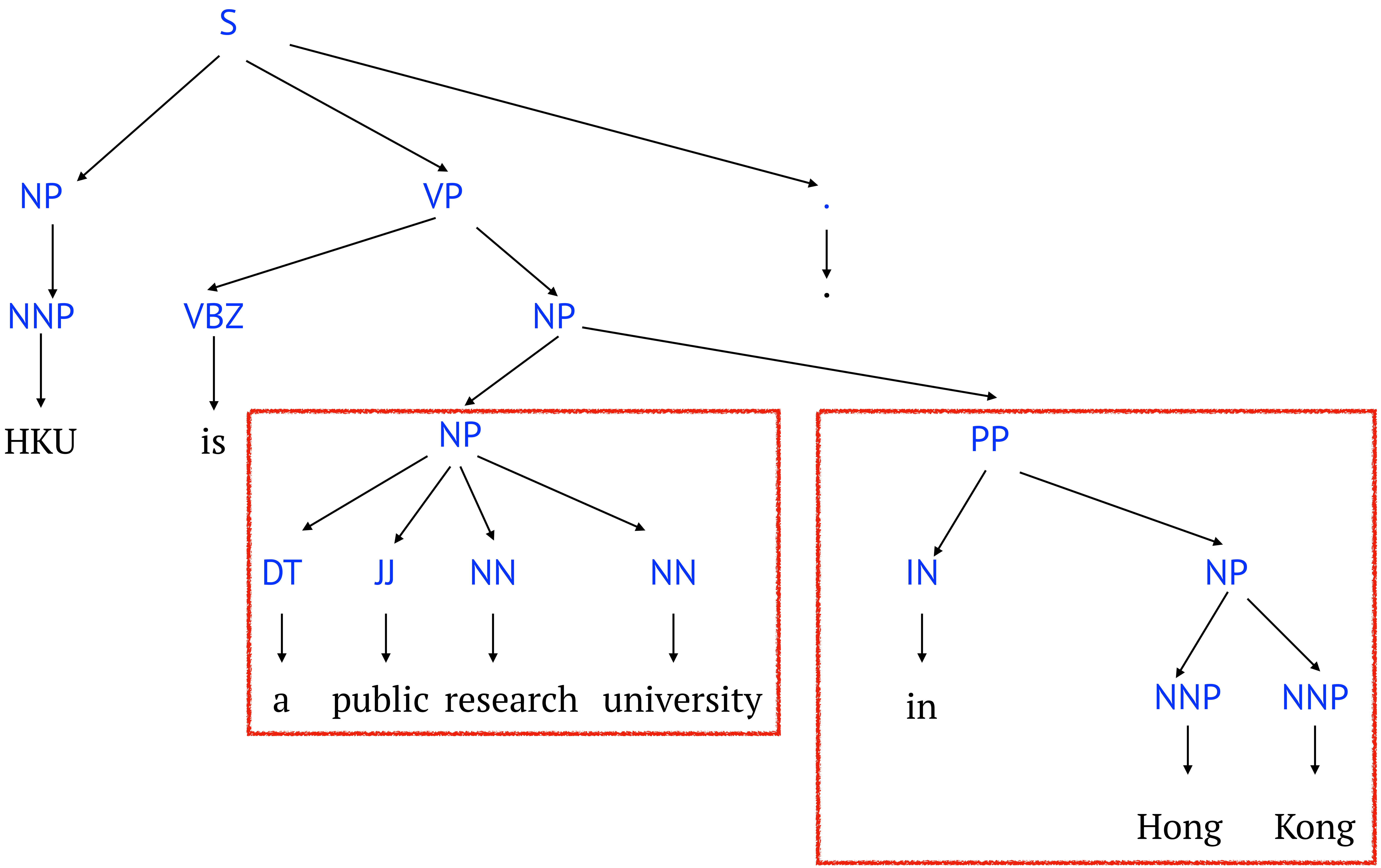

S
NP
VP
.
NNP
VBZ
NP
.
HKU
is
NP
PP
DT
JJ
NN
NN
IN
NP
a
public
research
university
in
NNP
NNP
Hong
Kong

Parse Trees

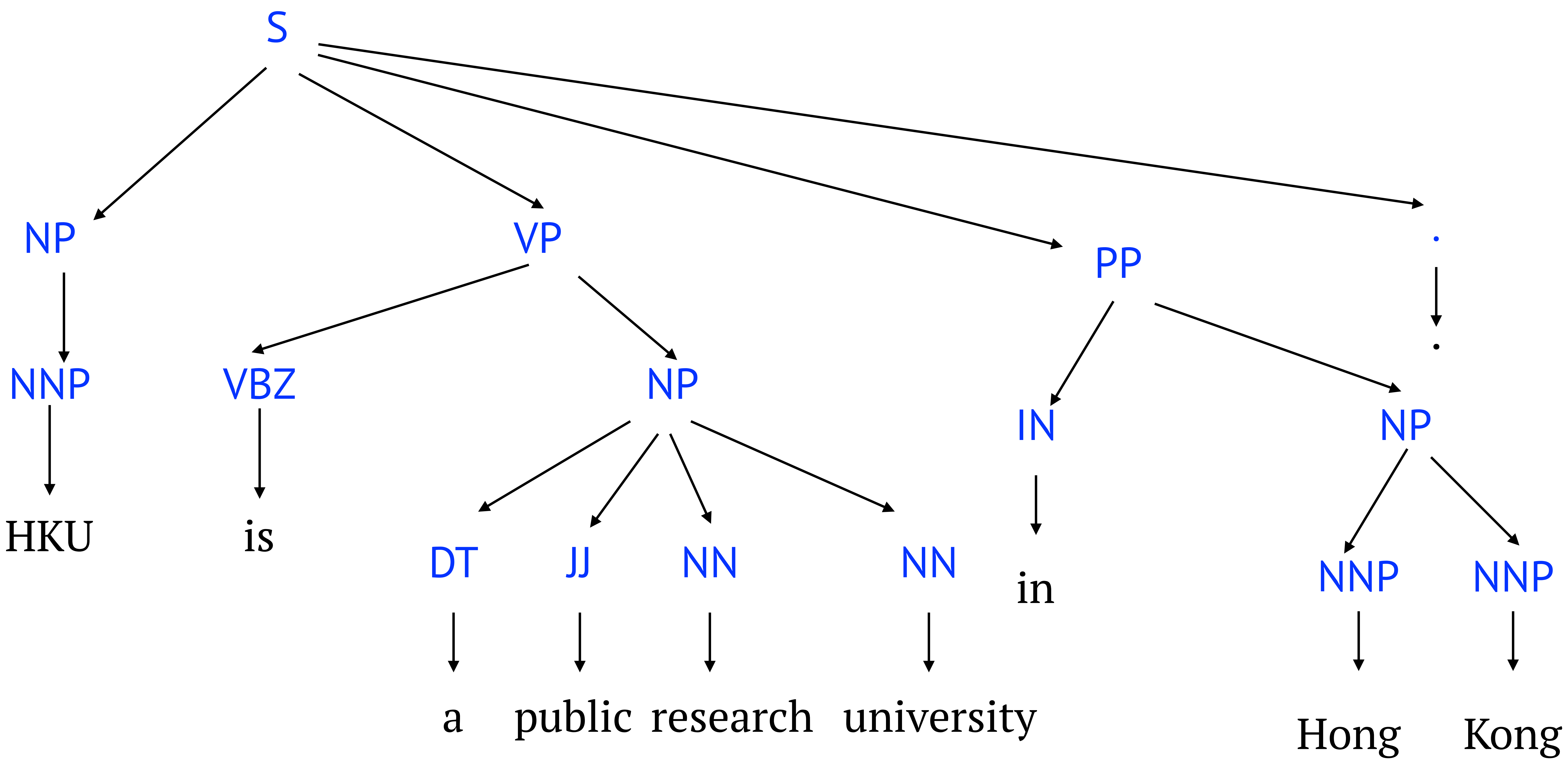

S
NP
VP
.
NNP
VBZ
NP
.
HKU
is
NP
PP
DT
JJ
NN
NN
IN
NP
a
public
research
university
in
NNP
NNP
Hong
Kong


S
NP
VP
VBZ
NP
is

Parse Trees

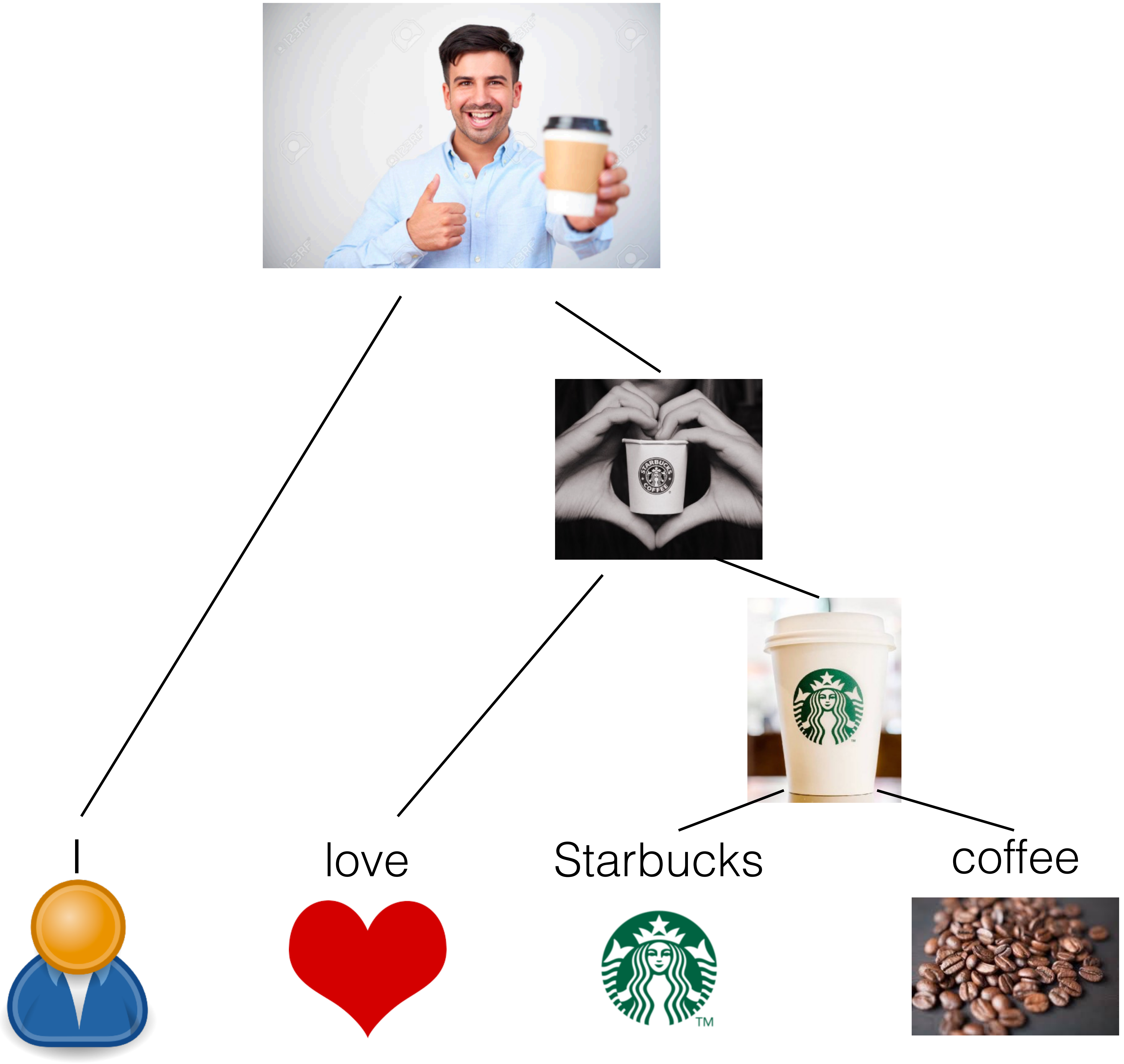

S
NP
VP
.
NNP
VBZ
NP
.
HKU
is
NP
PP
DT
JJ
NN
NN
IN
NP
a
public
research
university
in
NNP
NNP
Hong
Kong

Parse Trees



S
NP
VP
PP
.
NNP
VBZ
NP
IN
NP
HKU
is
DT
JJ
NN
NN
in
NNP
NNP
.
a
public
research
university
Hong
Kong

Compositional Meaning



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 $X \rightarrow Y_1 Y_2 \ldots Y_n$

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

$S \in N$ is a distinguished start symbol

A Context-Free Grammar for English

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

S = S

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

R =

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

Left-Most Derivations

R =

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

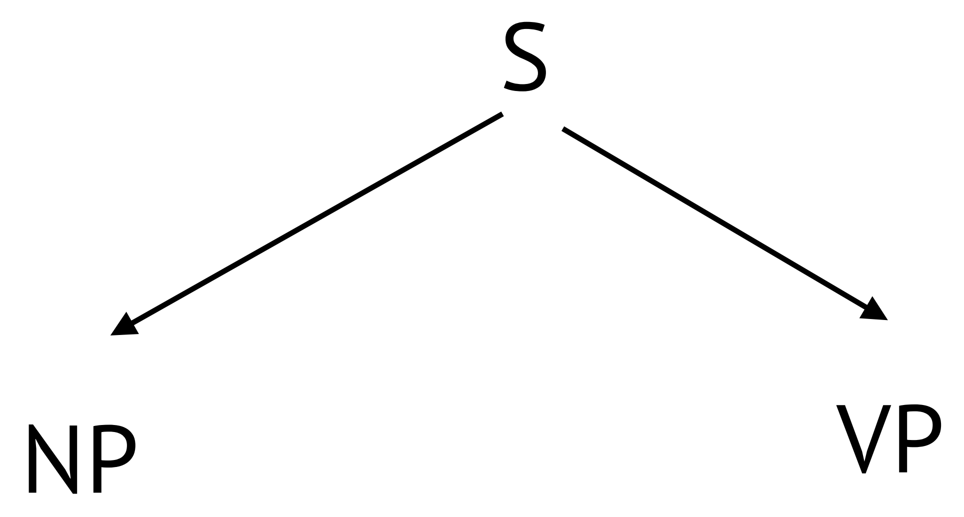
S

(S)

Left-Most Derivations

R =

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

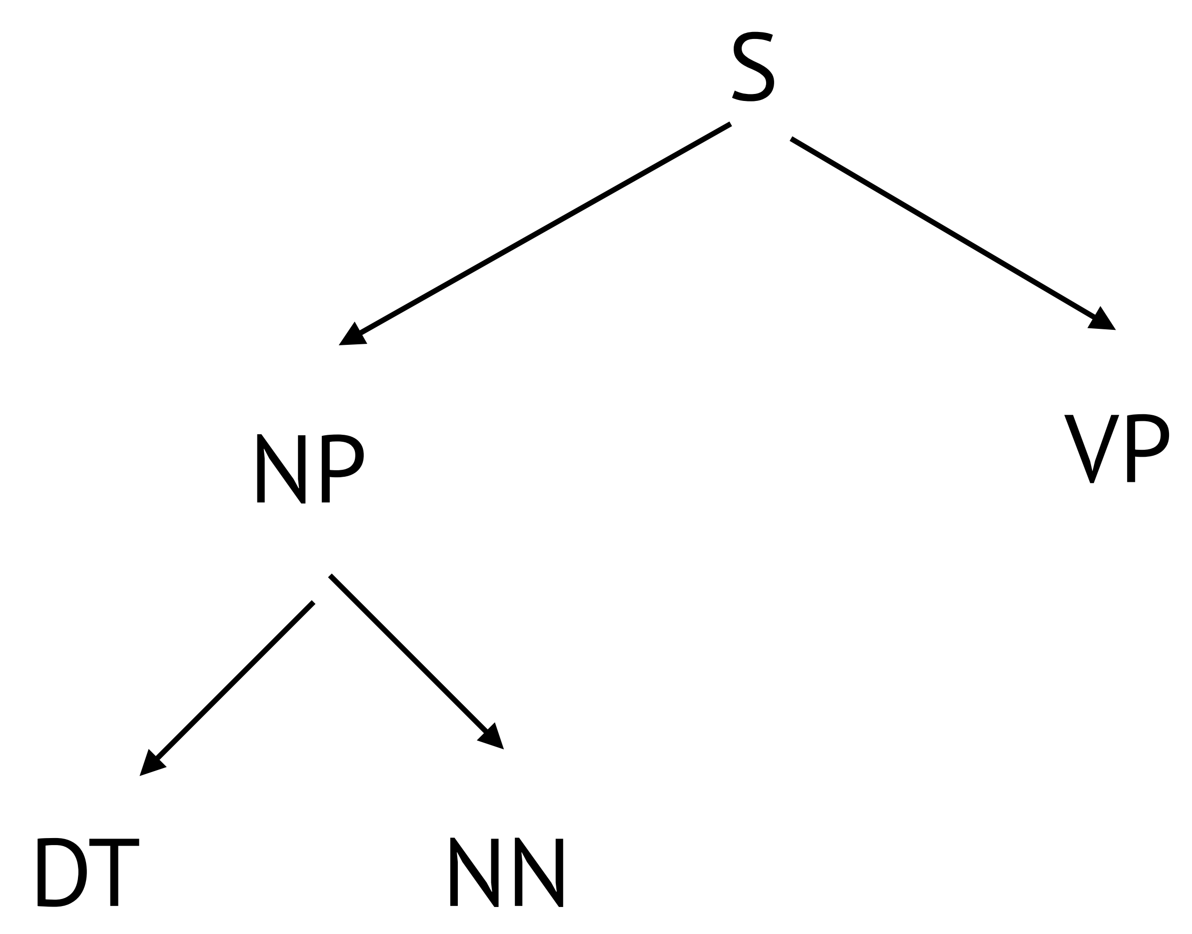


(NP VP)

Left-Most Derivations

R =

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

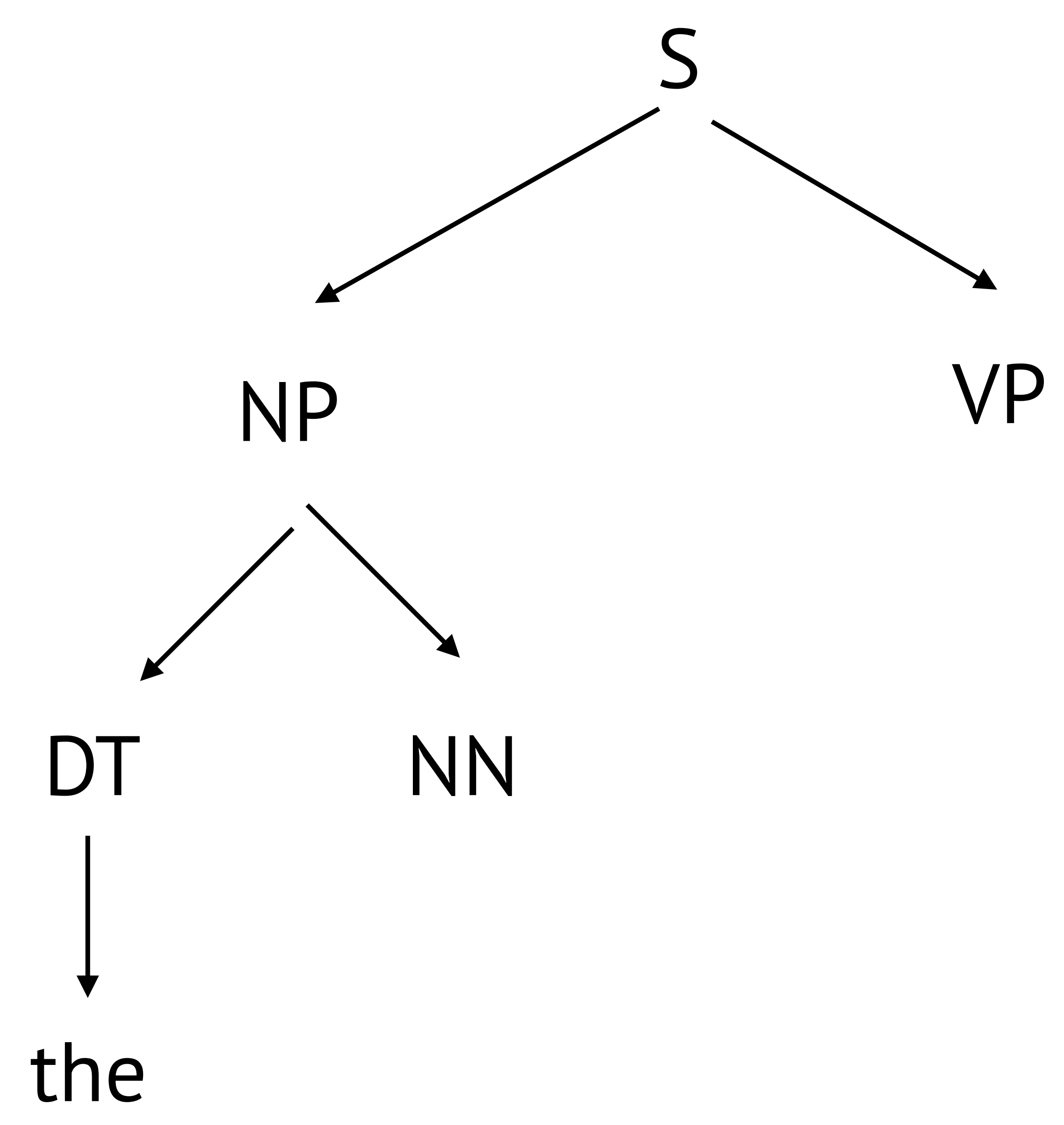


(DT NN VP)

Left-Most Derivations

R =

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

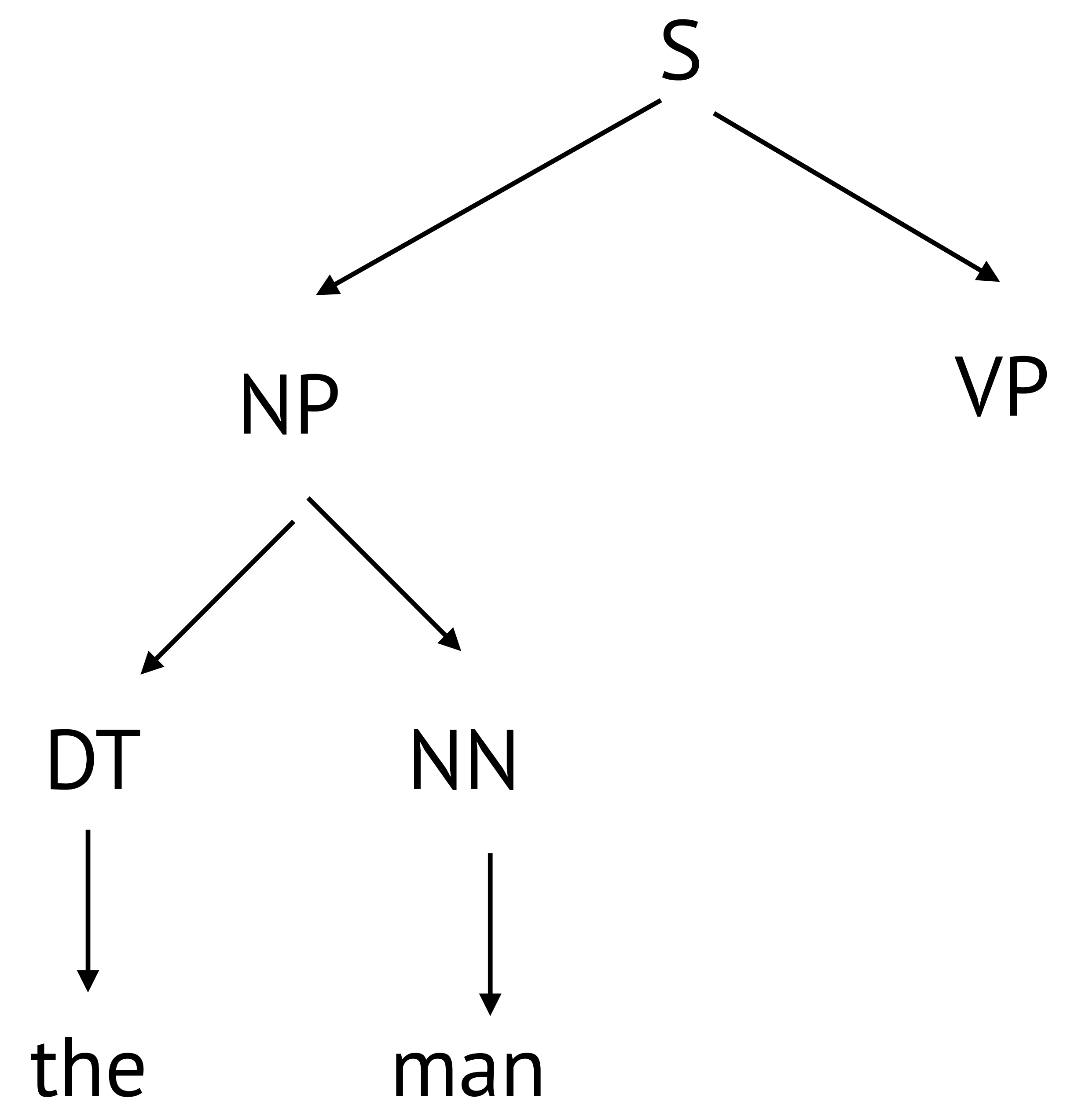


(the NN VP)

Left-Most Derivations

R =

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

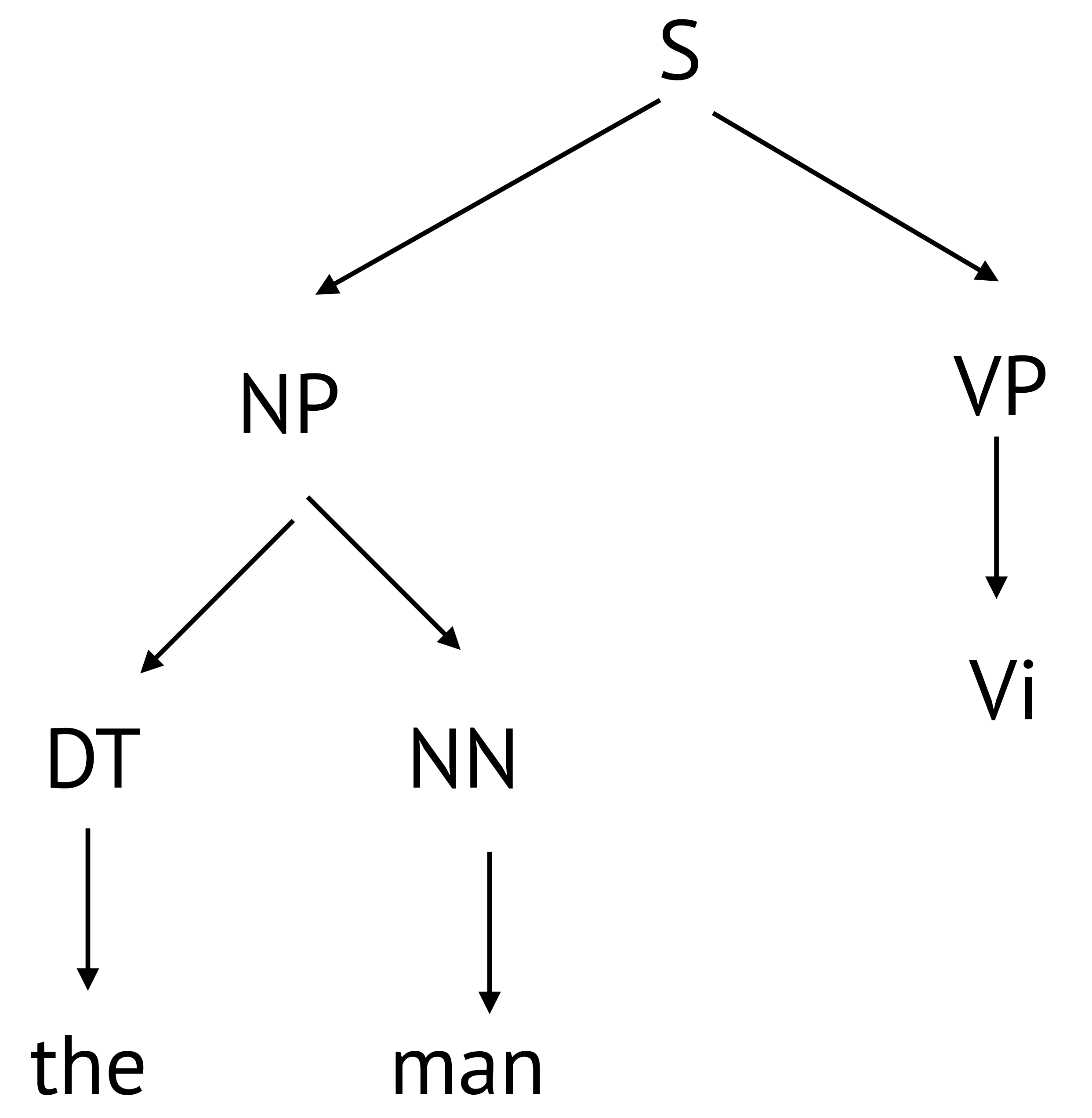


(the man VP)

Left-Most Derivations

R =

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

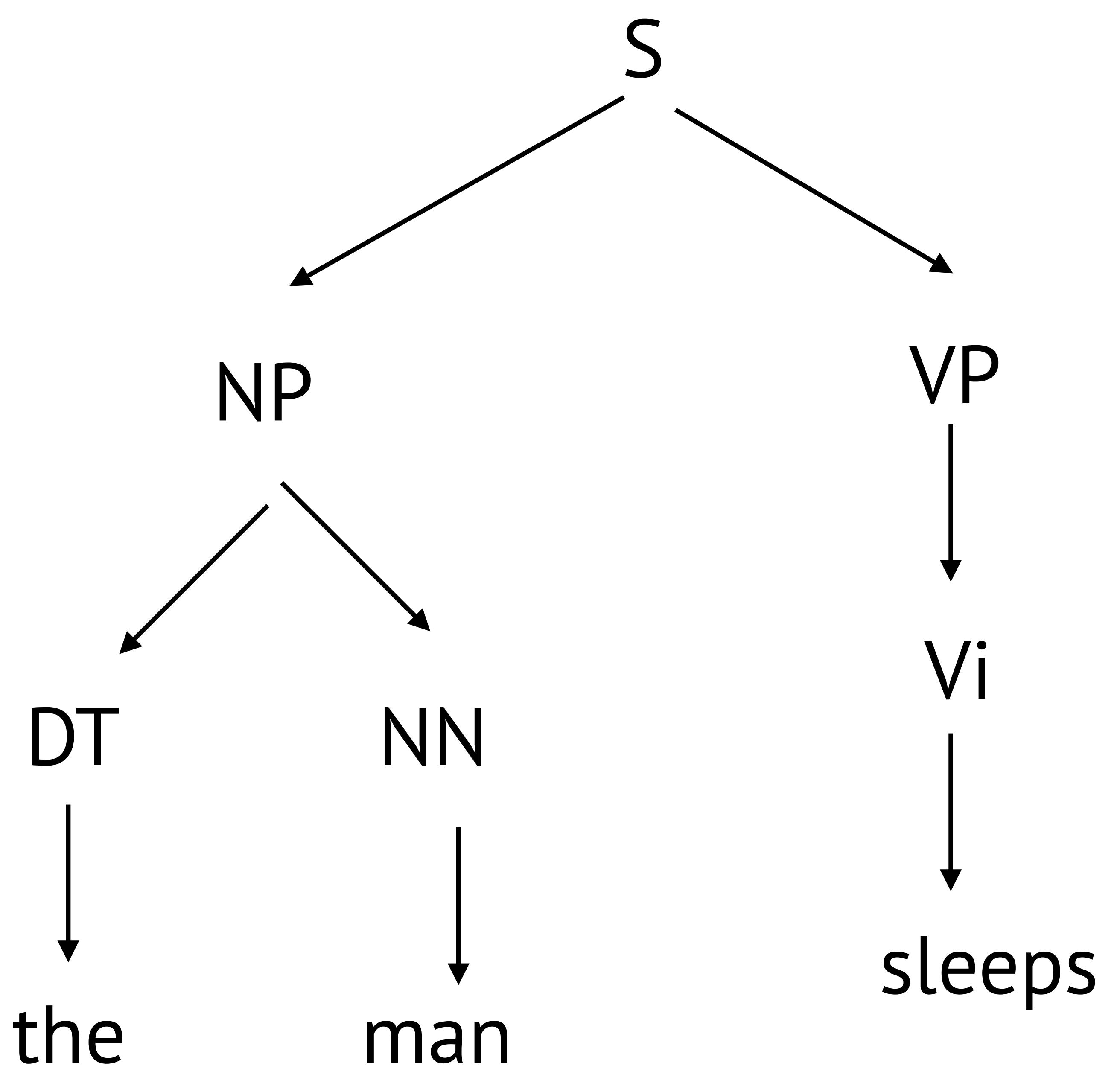


(the man Vi)

Left-Most Derivations

R =

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

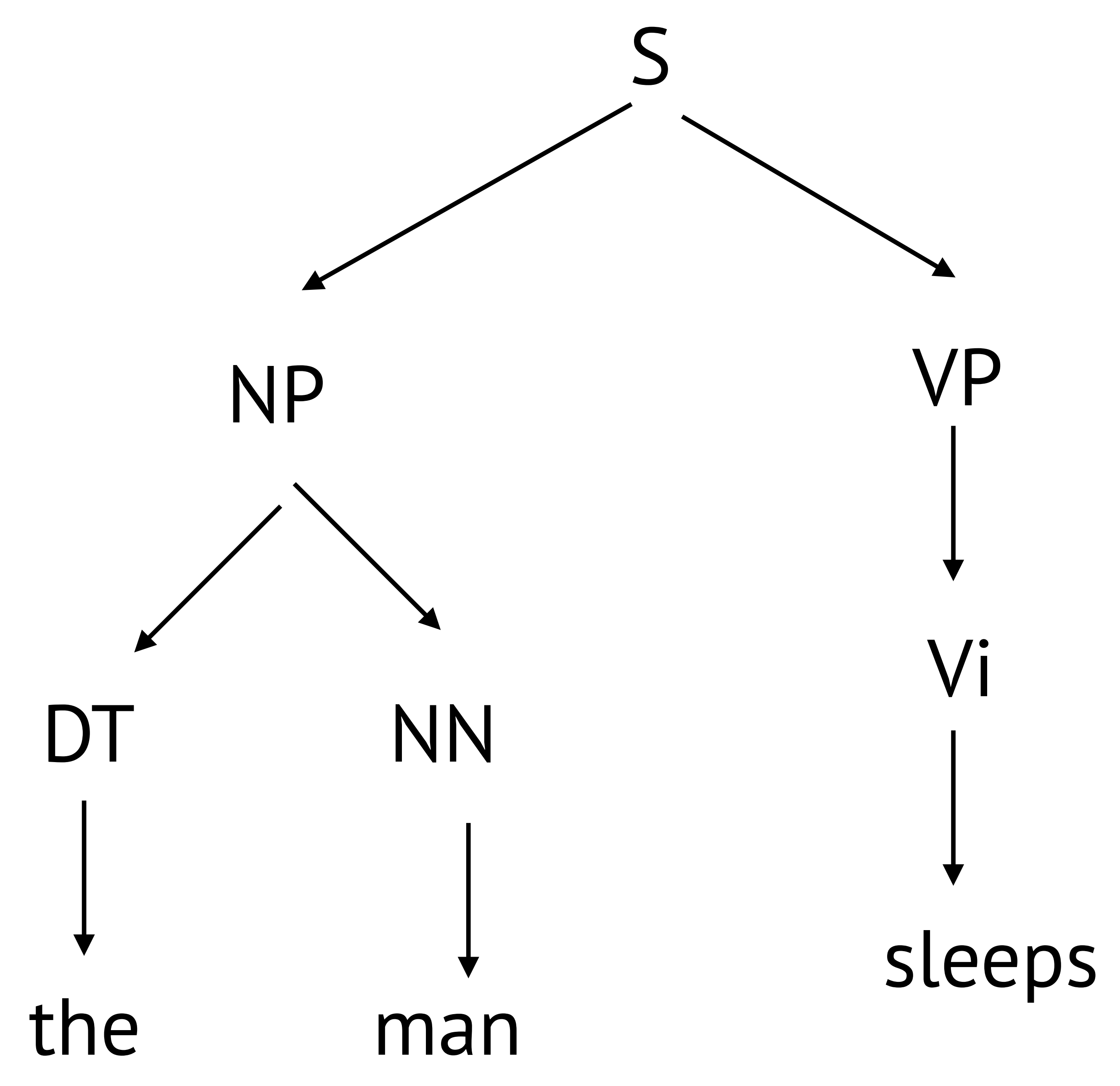


(the man sleeps)

Left-Most Derivations

R =

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

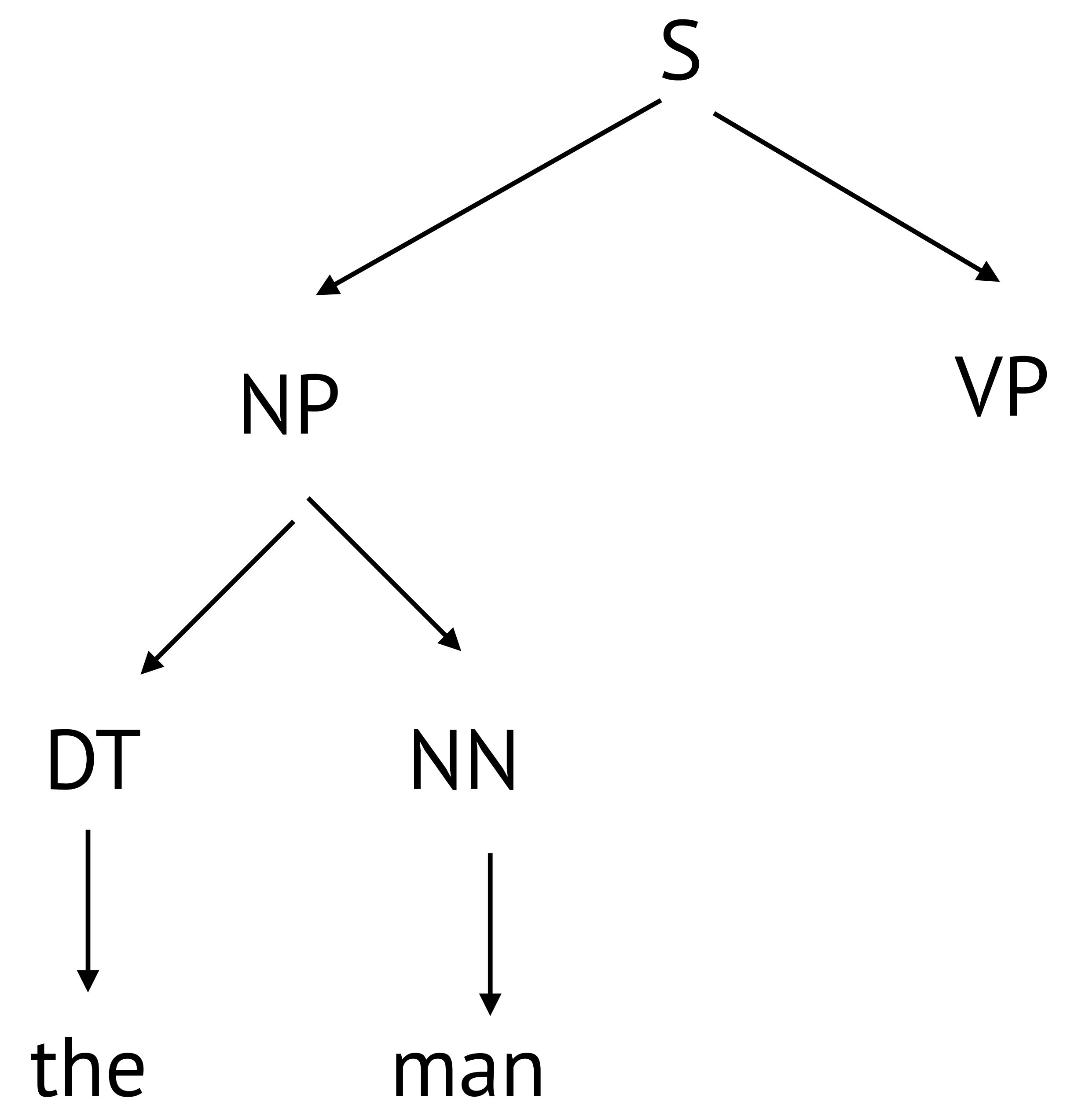


(the man sleeps)

Left-Most Derivations

R =

S → NP VP	Vi → sleeps
	Vt → saw
VP → Vi	
VP → Vt NP	NN → man
VP → VP PP	NN → woman
	NN → telescope
NP → DT NN	DT → the
NP → NP PP	
	IN → with
PP → IN NP	IN → in

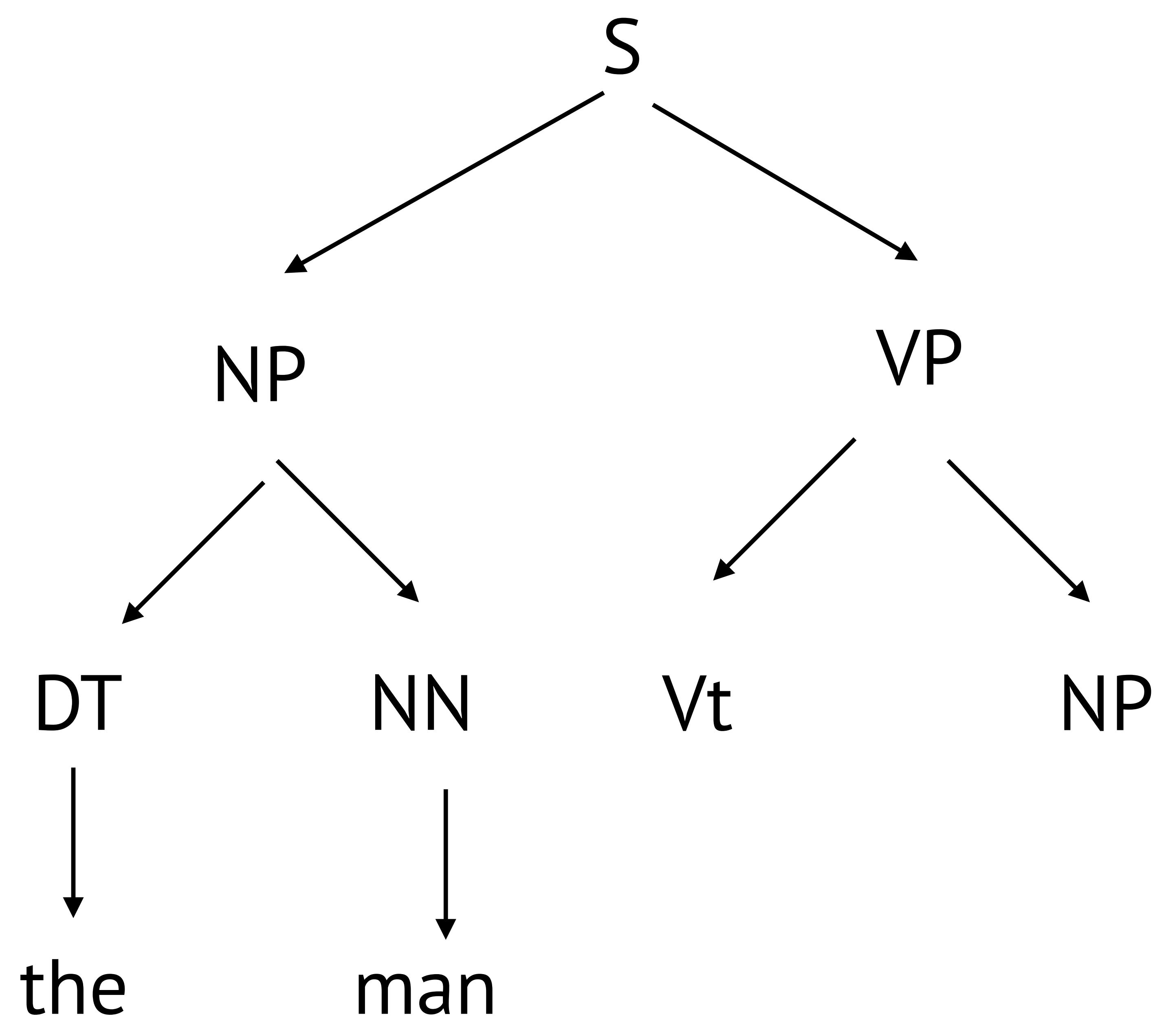


(the man VP)

Left-Most Derivations

R =

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

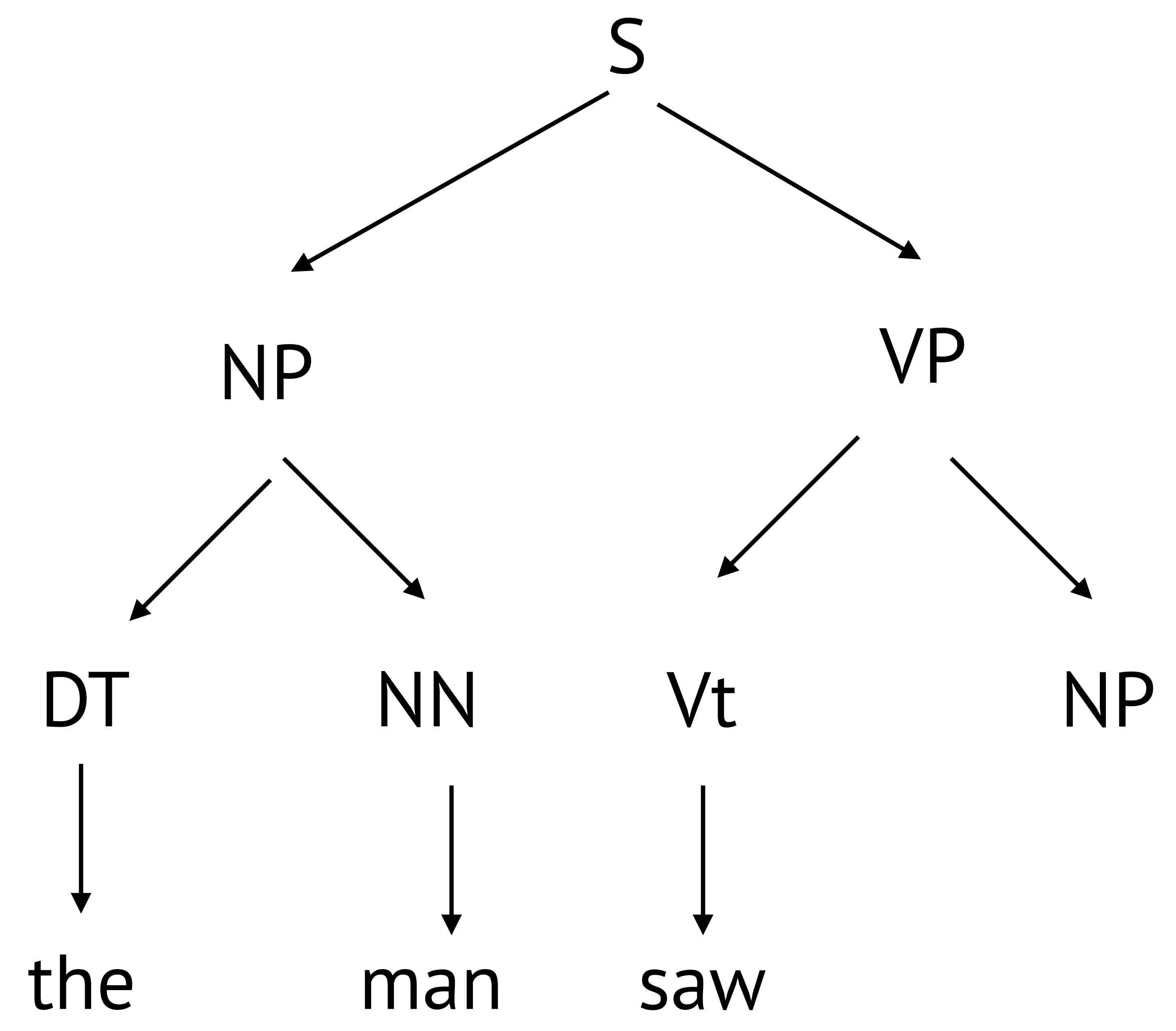


(the man Vt NP)

Left-Most Derivations

R =

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

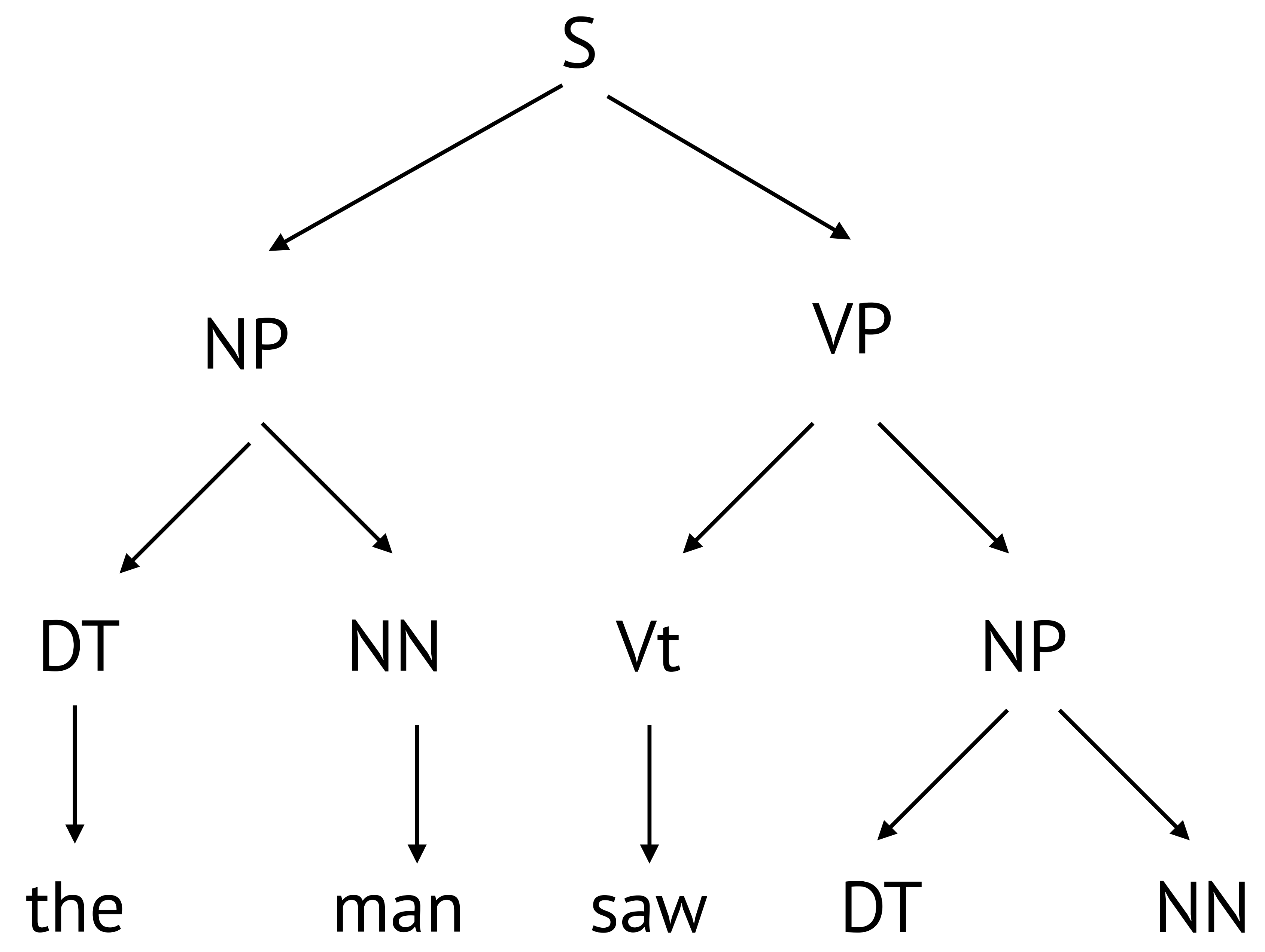


(the man saw NP)

Left-Most Derivations

R =

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



(the man saw DT NN)

Left-Most Derivations

R =

$S \longrightarrow NP\ VP$

$VP \longrightarrow Vi$
$VP \longrightarrow Vt\ NP$
$VP \longrightarrow VP\ PP$

$NP \longrightarrow DT\ NN$
$NP \longrightarrow NP\ PP$

$PP \longrightarrow IN\ NP$

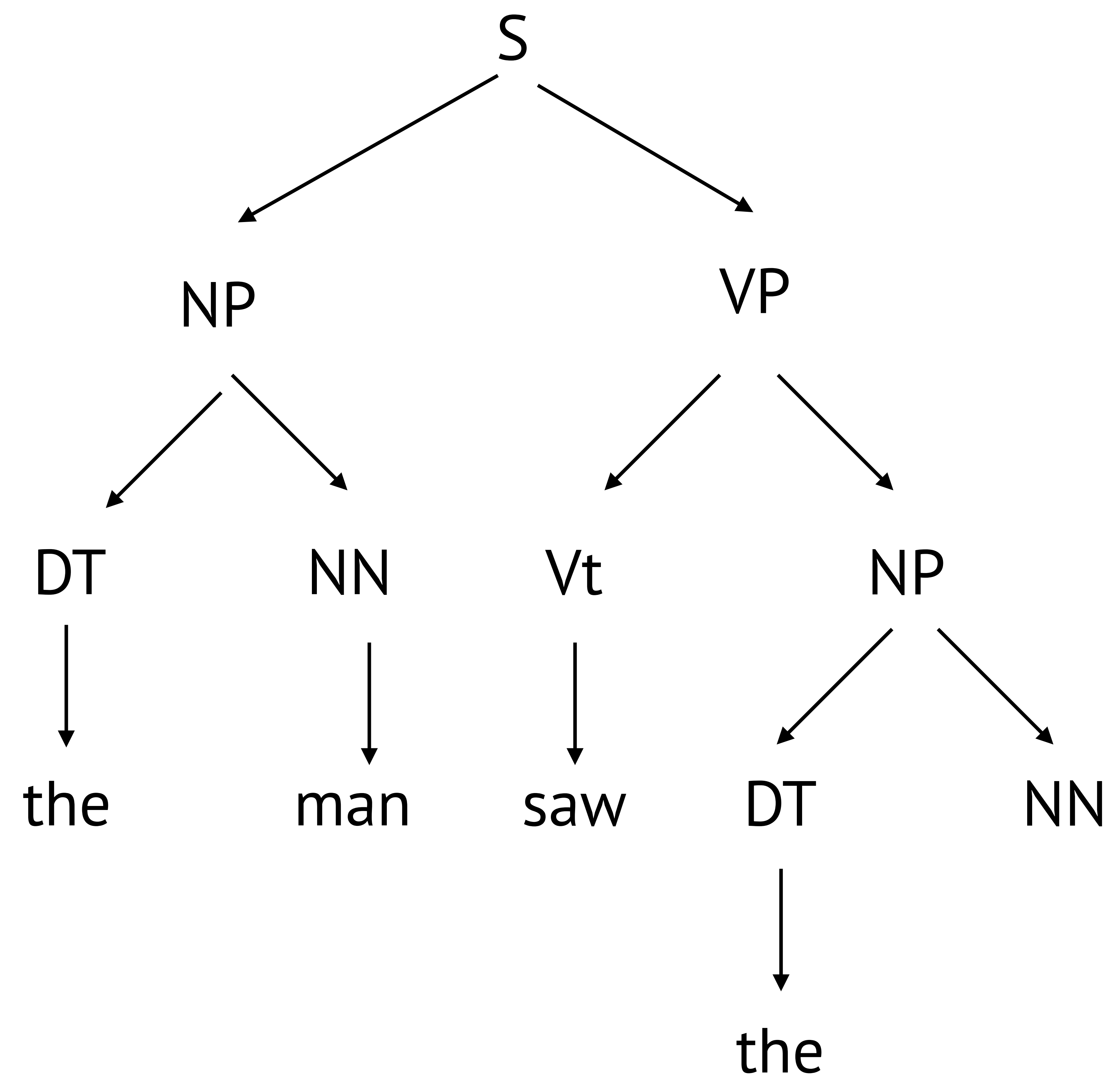
$Vi \longrightarrow$ sleeps
$Vt \longrightarrow$ saw

$NN \longrightarrow$ man
$NN \longrightarrow$ woman
$NN \longrightarrow$ telescope

$DT \longrightarrow$ the

$IN \longrightarrow$ with
$IN \longrightarrow$ in

(the man saw the NN)

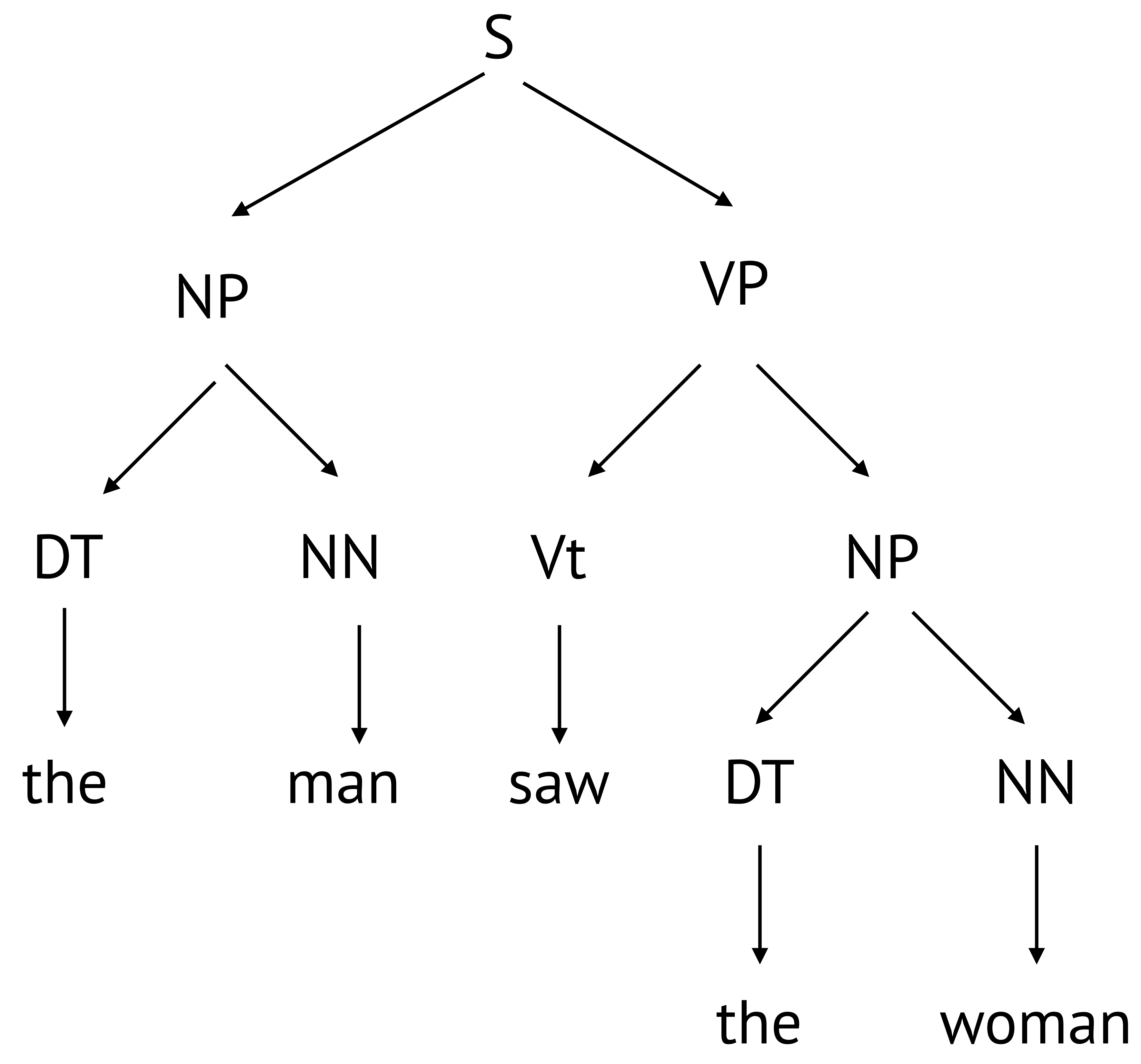


Left-Most Derivations

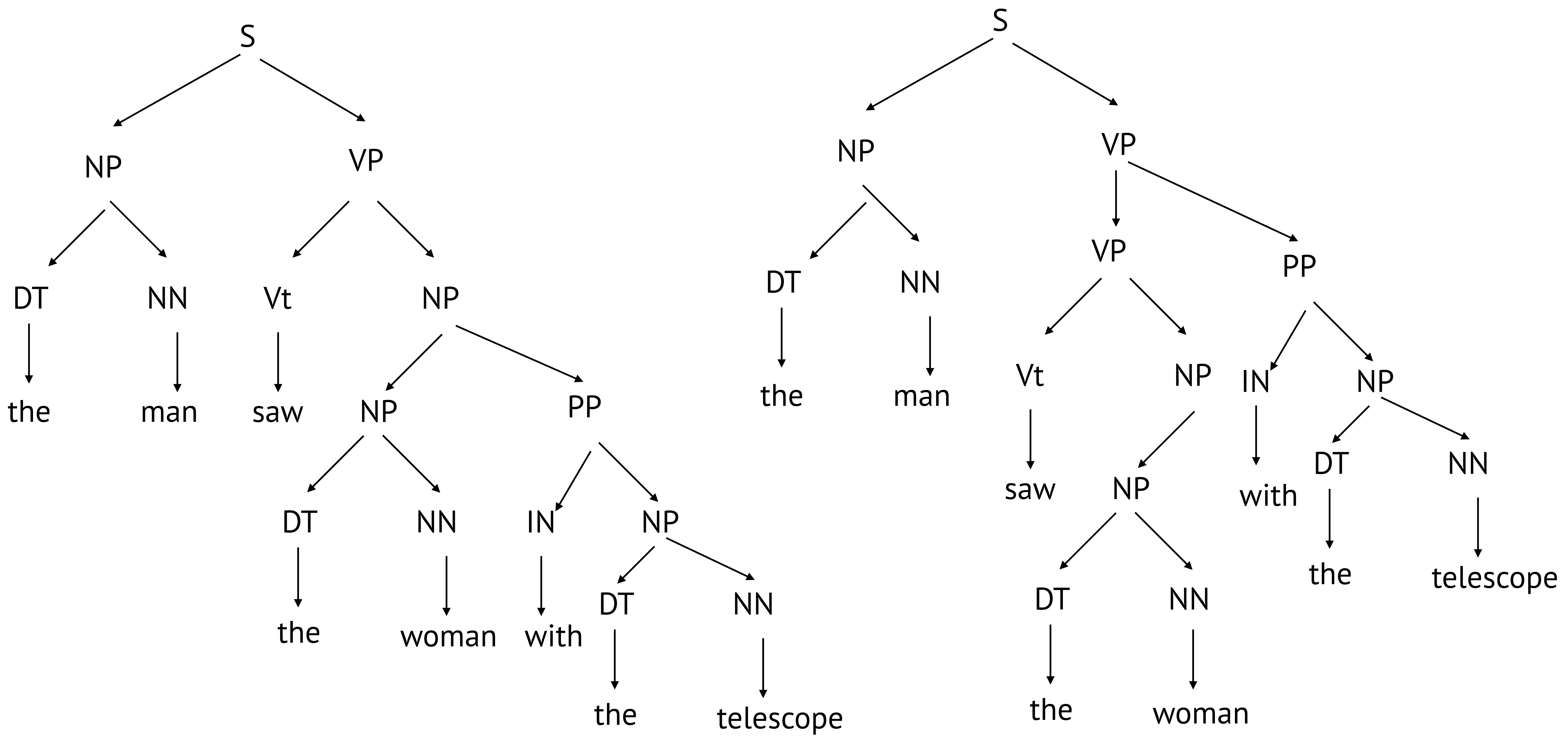
R =

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

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

R =

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

Vt NP $\longrightarrow$ Vt NP PP

"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		Vi $\longrightarrow$ sleeps	
		Vt $\longrightarrow$ saw	
VP $\longrightarrow$ Vi			
VP $\longrightarrow$ Vt NP		NN $\longrightarrow$ man	
VP $\longrightarrow$ VP PP		NN $\longrightarrow$ woman	
		NN $\longrightarrow$ telescope	
NP $\longrightarrow$ DT NN			
Vt NP $\longrightarrow$ Vt NP PP		DT $\longrightarrow$ the	non-context-free
PP $\longrightarrow$ IN NP		IN $\longrightarrow$ with	
		IN $\longrightarrow$ in	

Probabilistic Context-Free Grammars

R =

S	$\longrightarrow$	NP VP	1.0
VP	$\longrightarrow$	Vi	0.4
VP	$\longrightarrow$	Vt NP	0.4
VP	$\longrightarrow$	VP PP	0.2
NP	$\longrightarrow$	DT NN	0.3
NP	$\longrightarrow$	NP PP	0.7
PP	$\longrightarrow$	IN NP	1.0

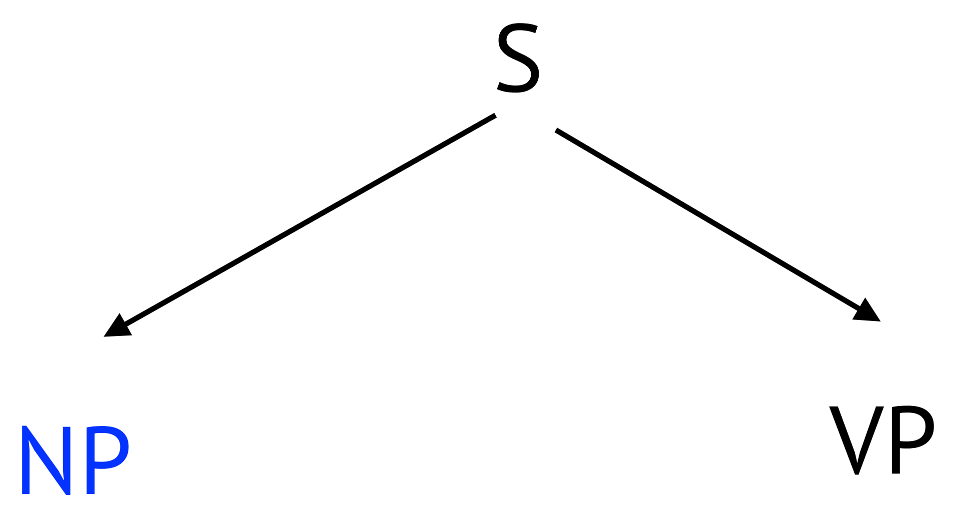
Vi	$\longrightarrow$	sleeps	1.0
Vt	$\longrightarrow$	saw	1.0
NN	$\longrightarrow$	man	0.7
NN	$\longrightarrow$	woman	0.2
NN	$\longrightarrow$	telescope	0.1
DT	$\longrightarrow$	the	1.0
IN	$\longrightarrow$	with	0.5
IN	$\longrightarrow$	in	0.5

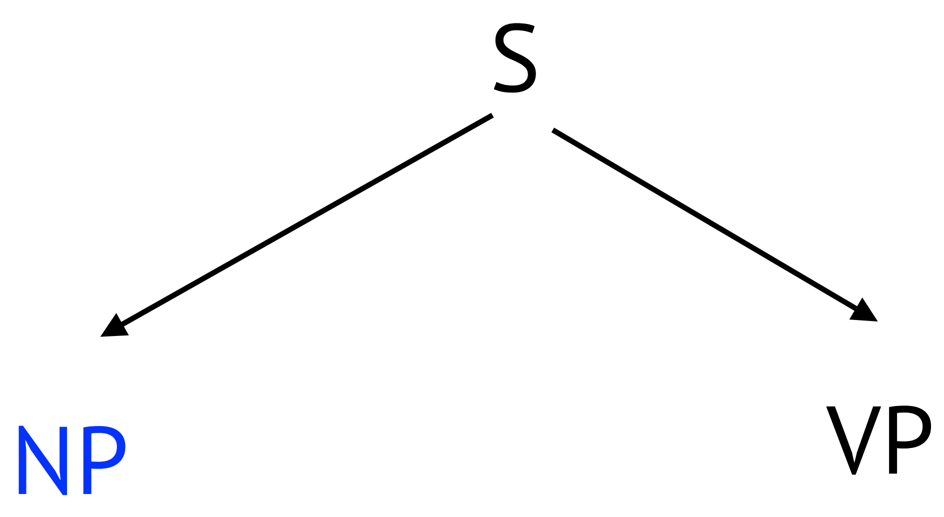
Probabilistic Context-Free Grammars

R =

S	$\longrightarrow$	NP VP	1.0
VP	$\longrightarrow$	Vi	0.4
VP	$\longrightarrow$	Vt NP	0.4
VP	$\longrightarrow$	VP PP	0.2
NP	$\longrightarrow$	DT NN	0.3
NP	$\longrightarrow$	NP PP	0.7
PP	$\longrightarrow$	IN NP	1.0

Vi	$\longrightarrow$	sleeps	1.0
Vt	$\longrightarrow$	saw	1.0
NN	$\longrightarrow$	man	0.7
NN	$\longrightarrow$	woman	0.2
NN	$\longrightarrow$	telescope	0.1
DT	$\longrightarrow$	the	1.0
IN	$\longrightarrow$	with	0.5
IN	$\longrightarrow$	in	0.5

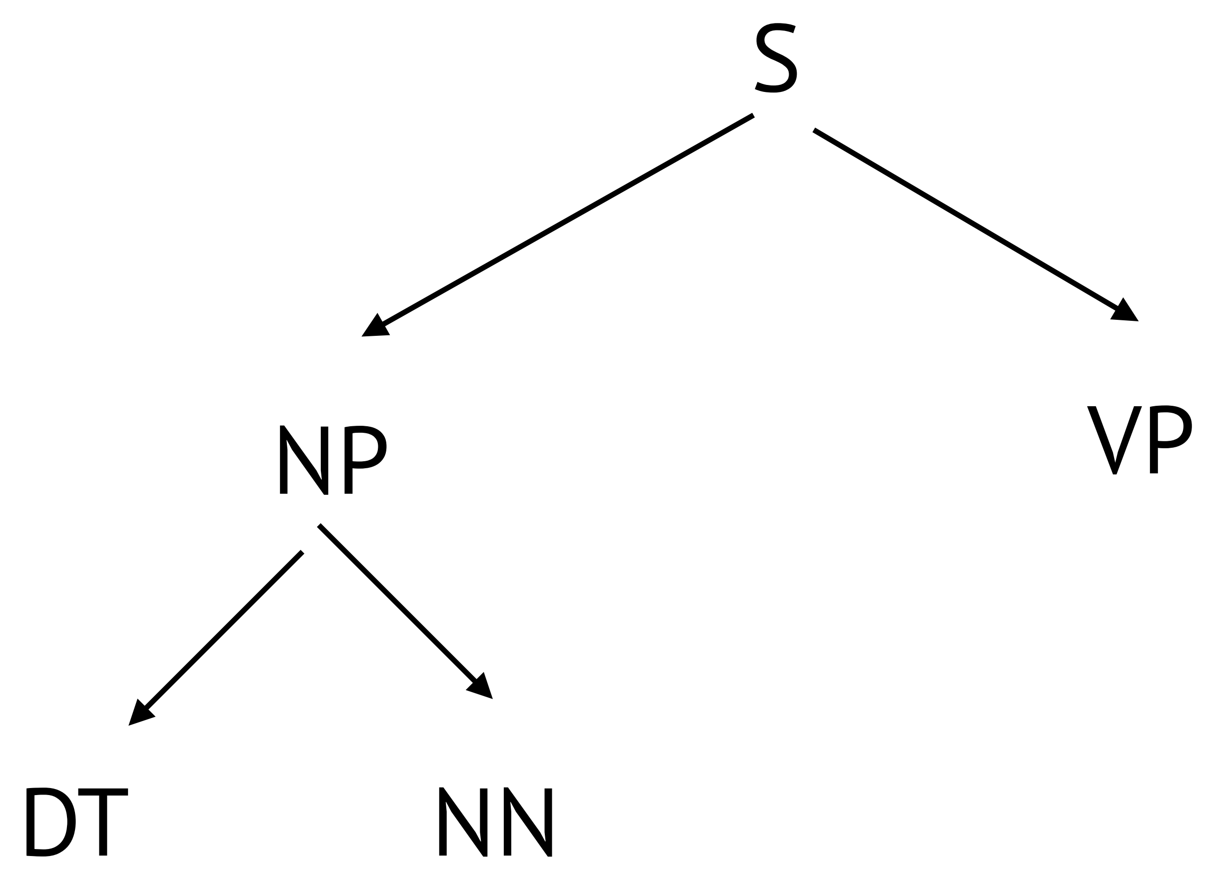


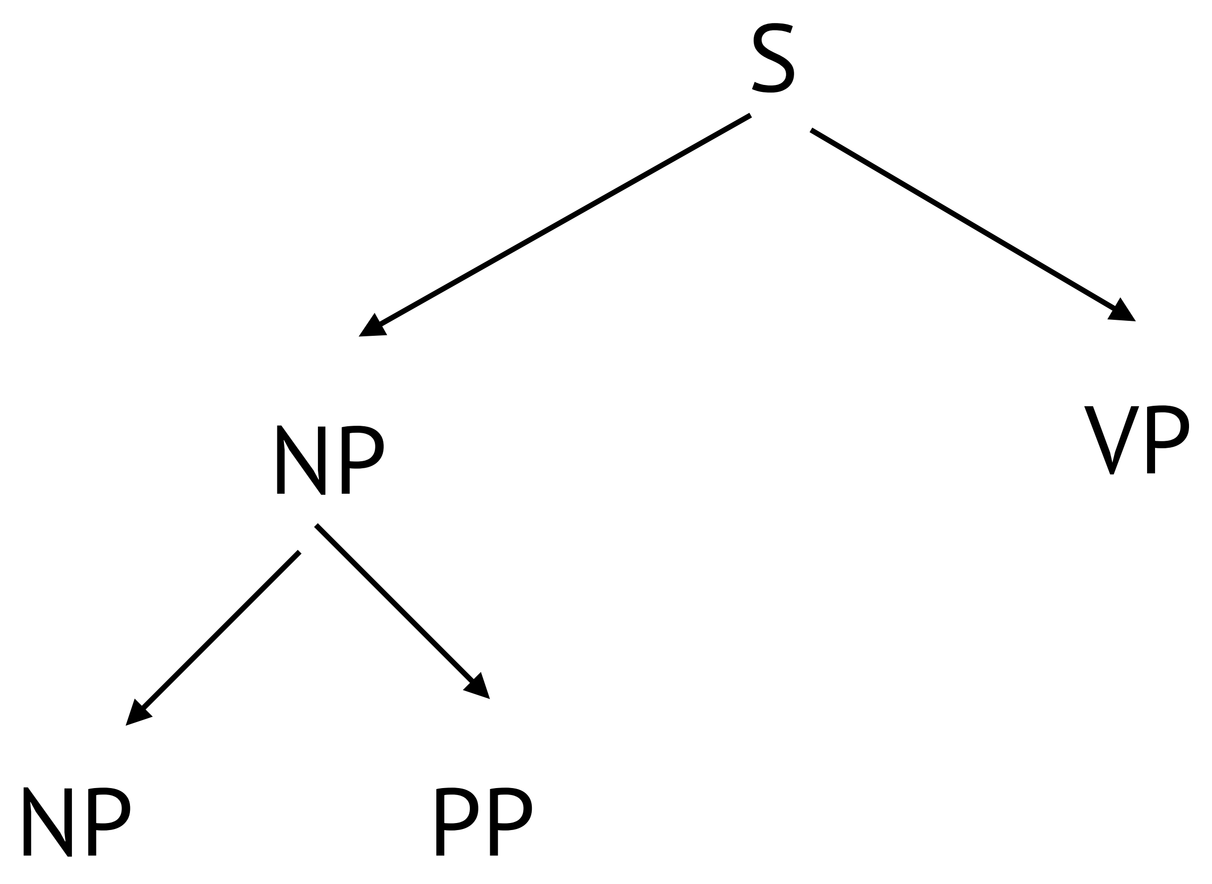


Probabilistic Context-Free Grammars

R =

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





Probabilistic Context-Free Grammars

R =

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

Derivation Example

R =

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

S

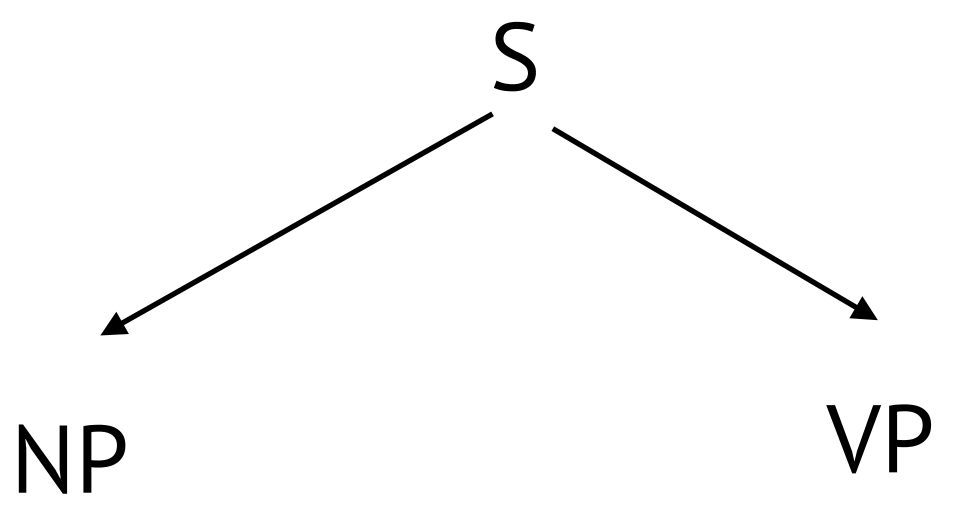
(S)

$P(\boldsymbol{x}, \boldsymbol{y}) =$

Derivation Example

R =

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



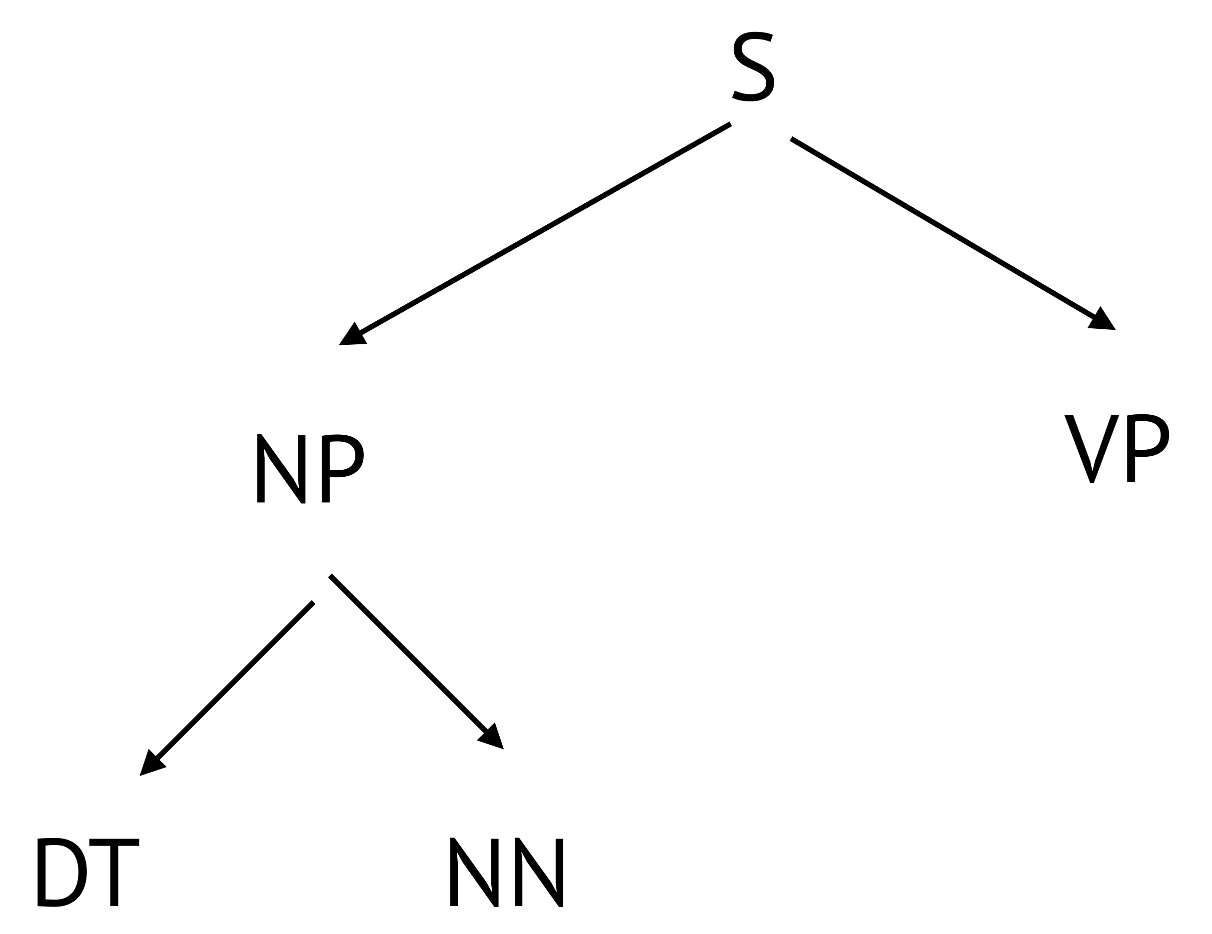
(S (NP VP))

$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times$

Derivation Example

R =

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



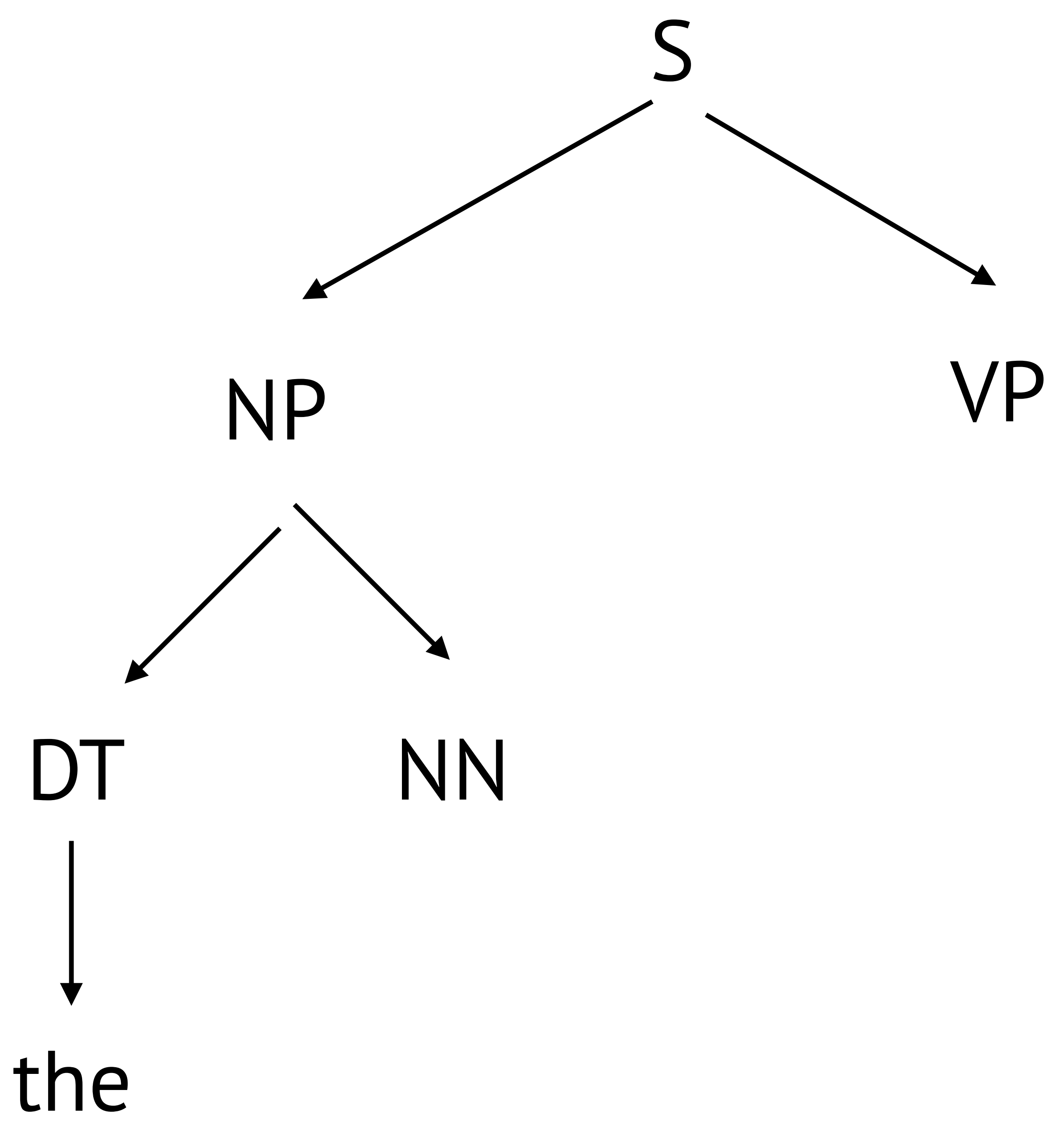
(S (NP (DT NN) VP))

$$P(\boldsymbol{x}, \boldsymbol{y}) = \ 1 \times \ 0.3 \times$$

Derivation Example

R =

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



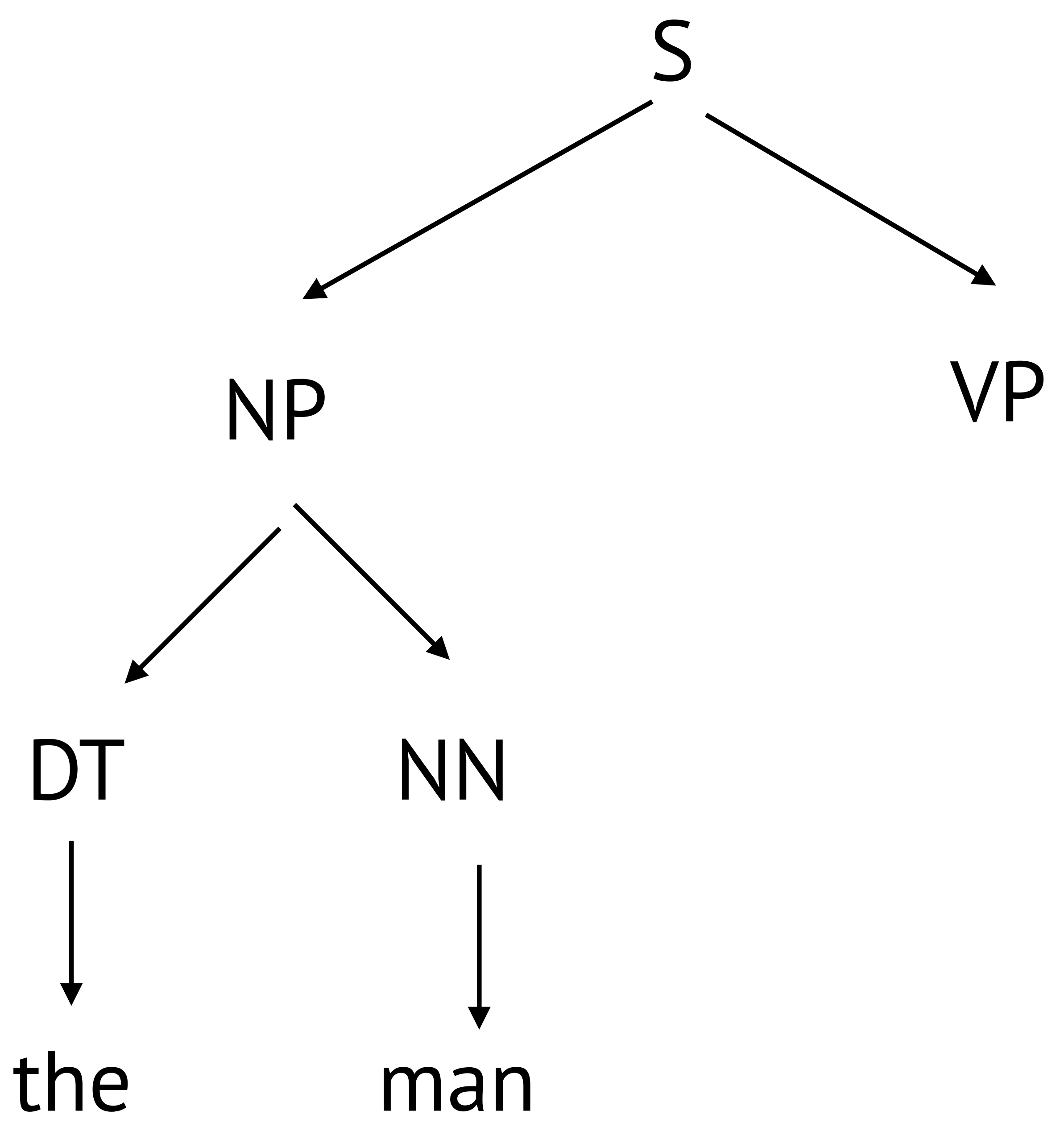
(S (NP ((DT the) NN) VP))

$$P(\boldsymbol{x}, \boldsymbol{y}) = \ 1\times \ 0.3\times \ 1\times$$

Derivation Example

R =

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



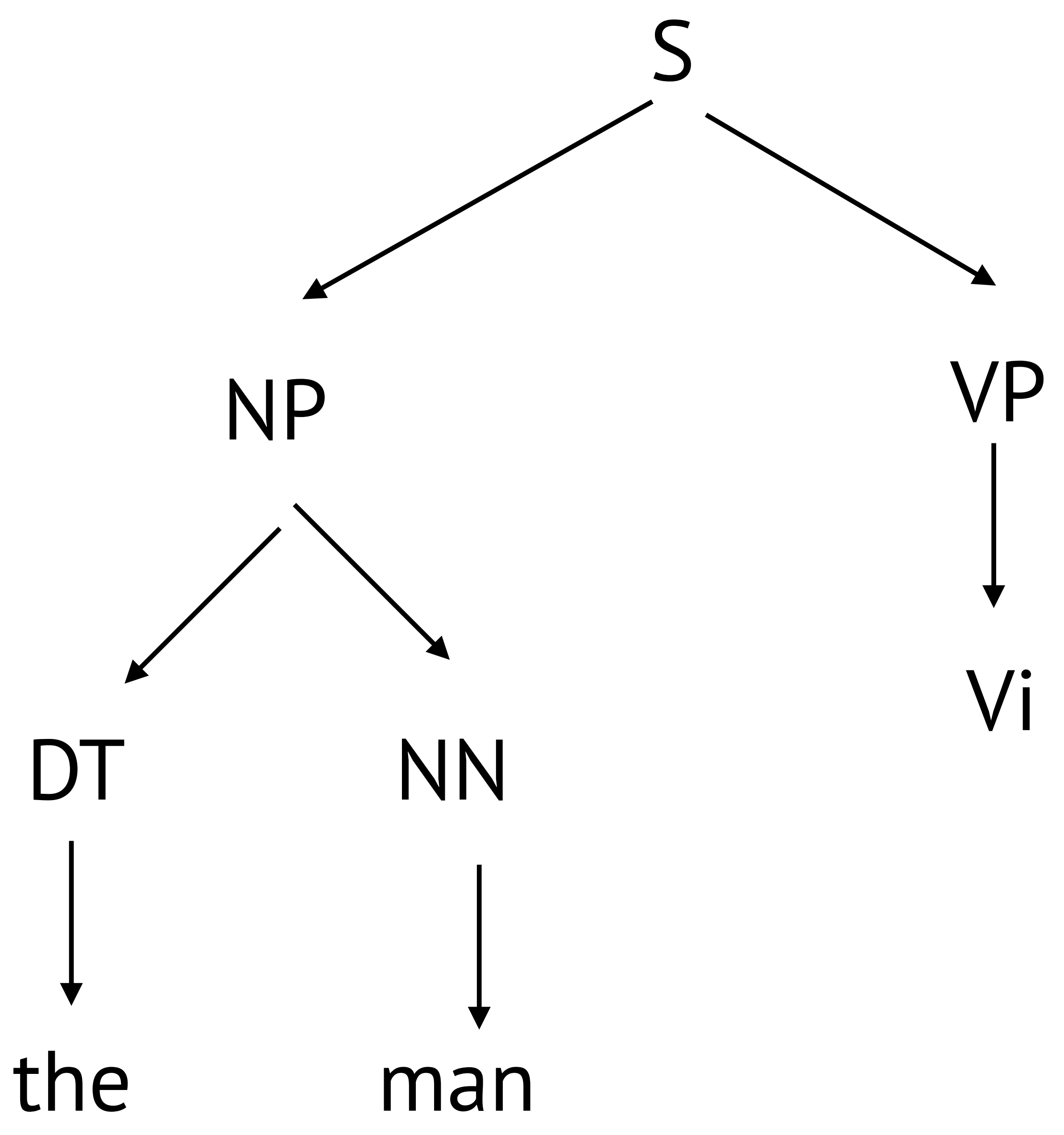
(S (NP ((DT the) (NN man)) VP))

$$P(\boldsymbol{x}, \boldsymbol{y}) = 1\times\ 0.3\times\ 1\times\ 0.7\times$$

Derivation Example

R =

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



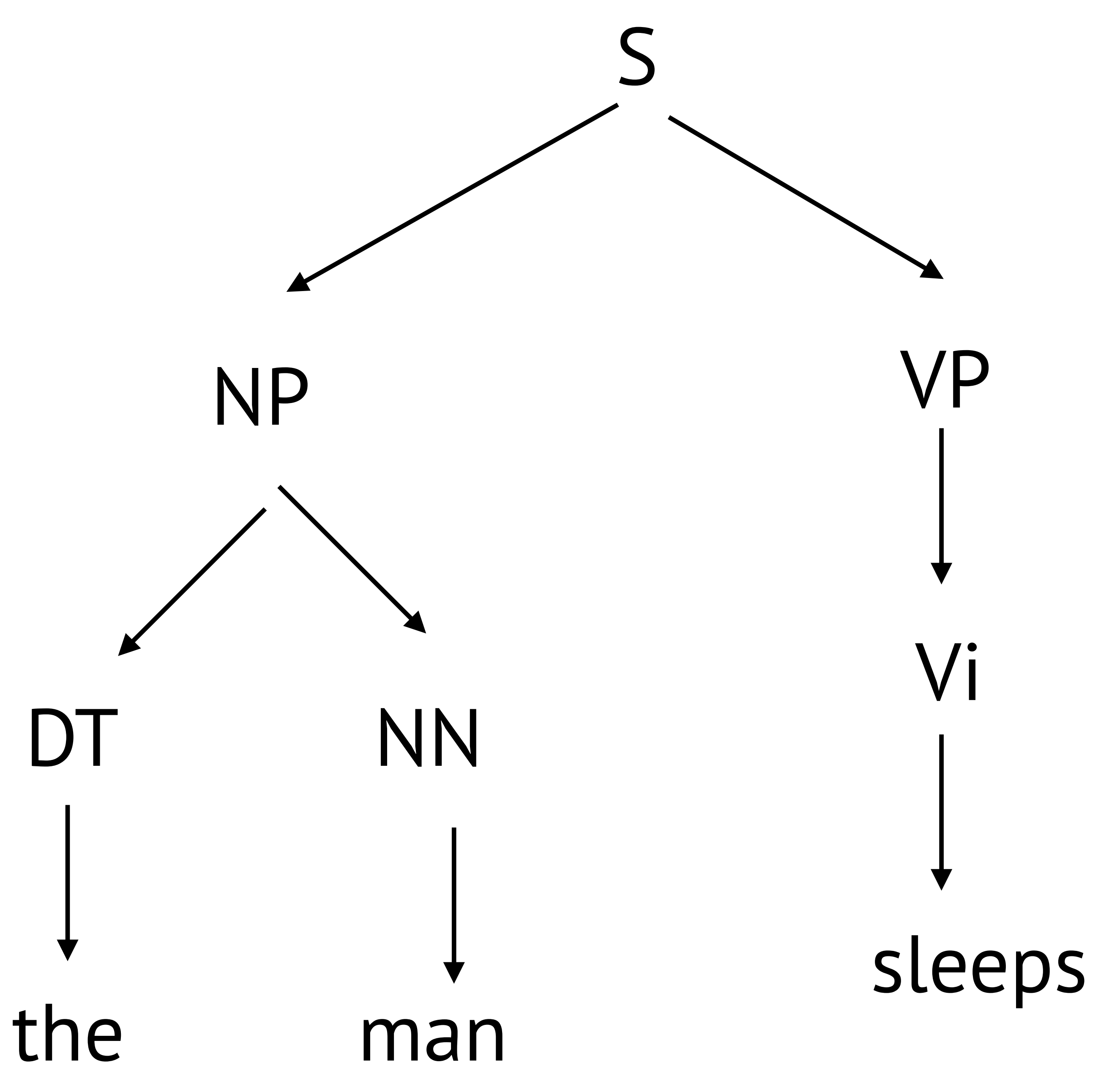
(S (NP ((DT the) (NN man)) (VP (Vi))))

$P(\boldsymbol{x}, \boldsymbol{y}) = \; 1\times \; 0.3\times \; 1\times \; 0.7\times \; 0.4\times$

Derivation Example

R =

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



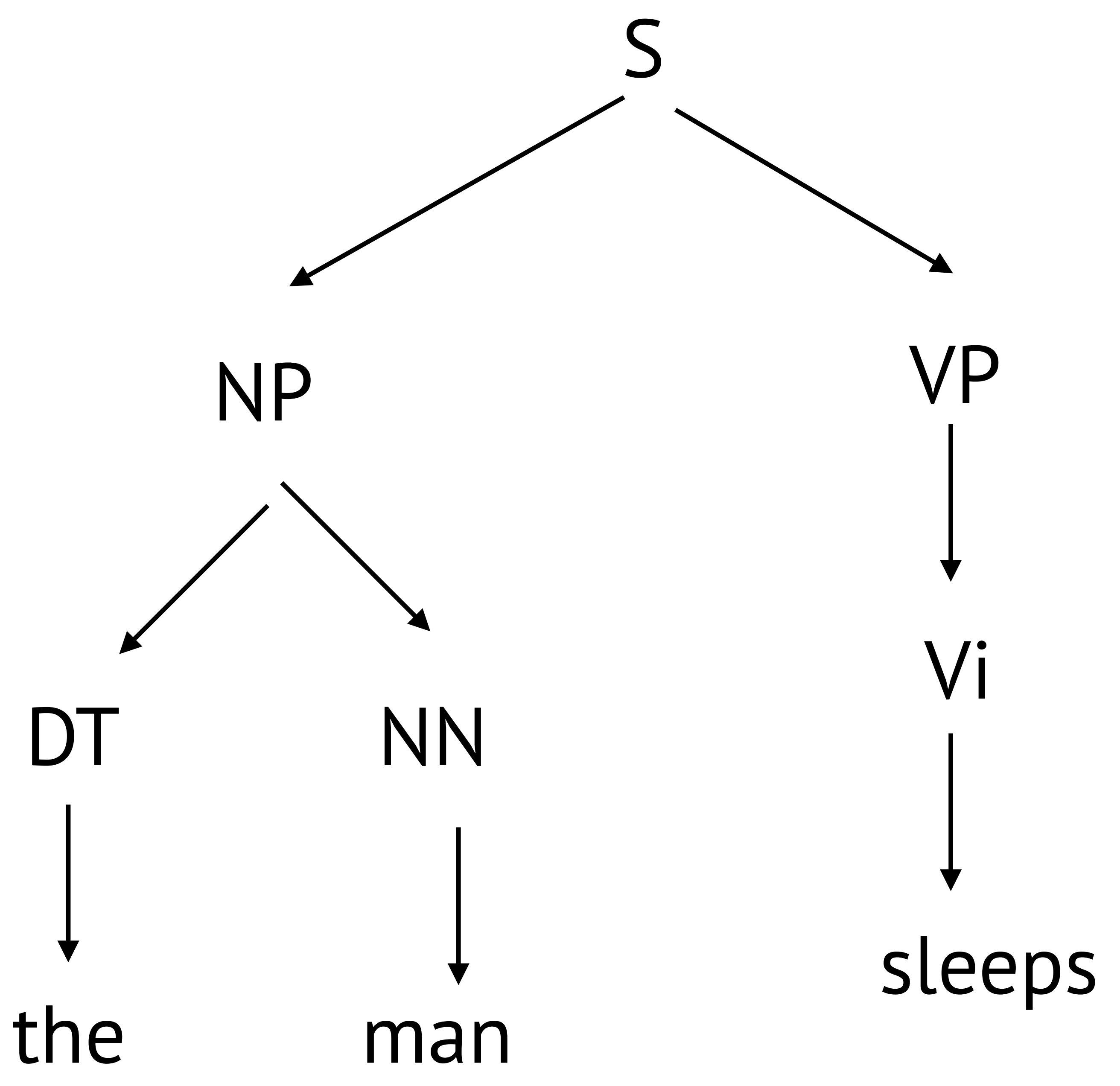
(S (NP ((DT the) (NN man)) (VP (Vi sleeps))))

$$P(\boldsymbol{x}, \boldsymbol{y}) = 1\times \ 0.3\times \ 1\times \ 0.7\times \ 0.4\times 1$$

Derivation Example

R =

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



(S (NP ((DT the) (NN man)) (VP (Vi sleeps))))

Probabilistic Context-Free Grammars

R =

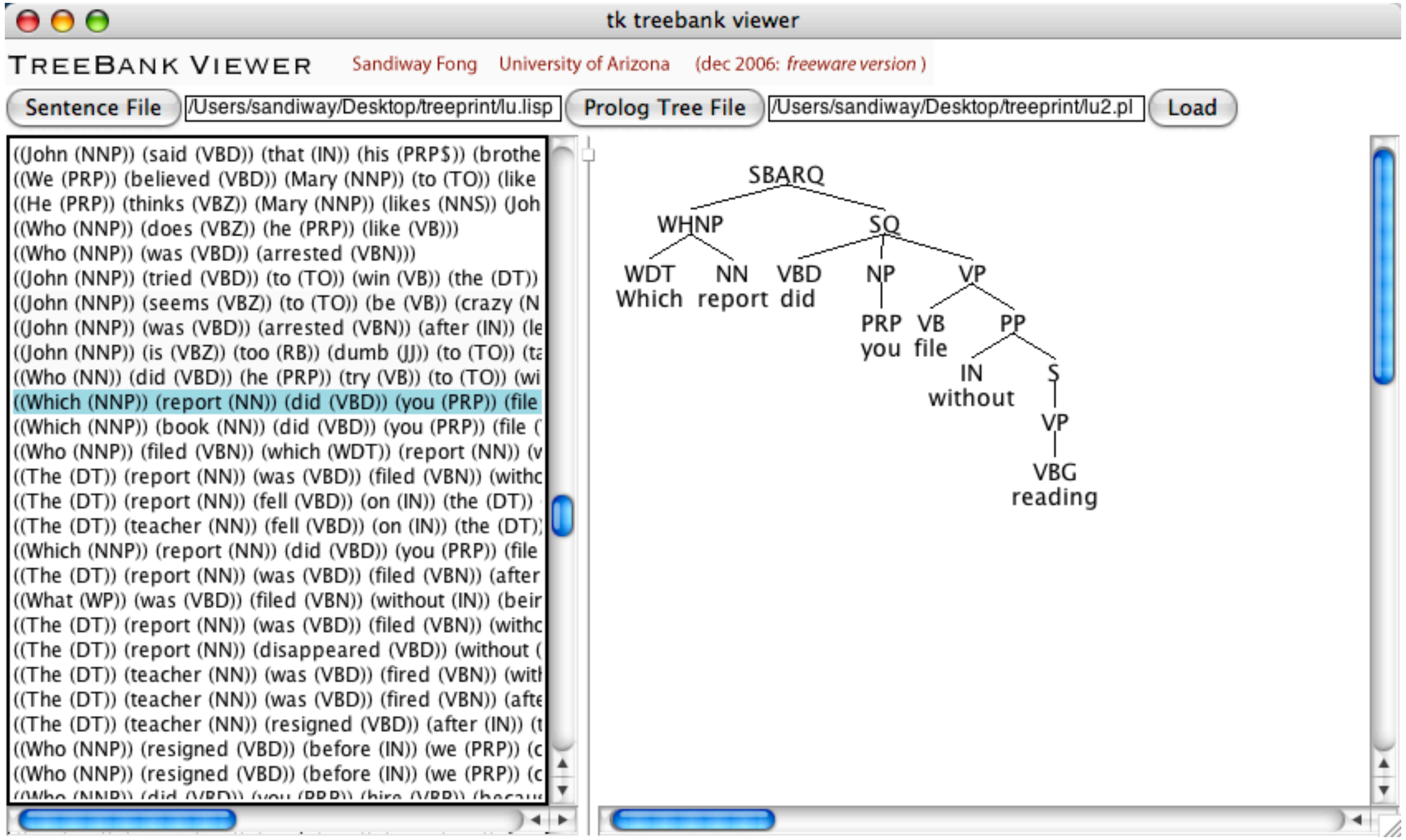
S	→ NP VP	1.0		Vi	→ sleeps	1.0
				Vt	→ saw	1.0
VP	→ Vi	0.4				
VP	→ Vt NP	0.4		NN	→ man	0.7
VP	→ VP PP	0.2		NN	→ woman	0.2
				NN	→ telescope	0.1
NP	→ DT NN	0.3		DT	→ the	1.0
NP	→ NP PP	0.7				
PP	→ IN NP	1.0		IN	→ with	0.5
				IN	→ in	0.5

1. How do we get those weights?

2. How to get the (marginal) probability of this sentence and the most likely parse?

the man saw the woman with a telescope

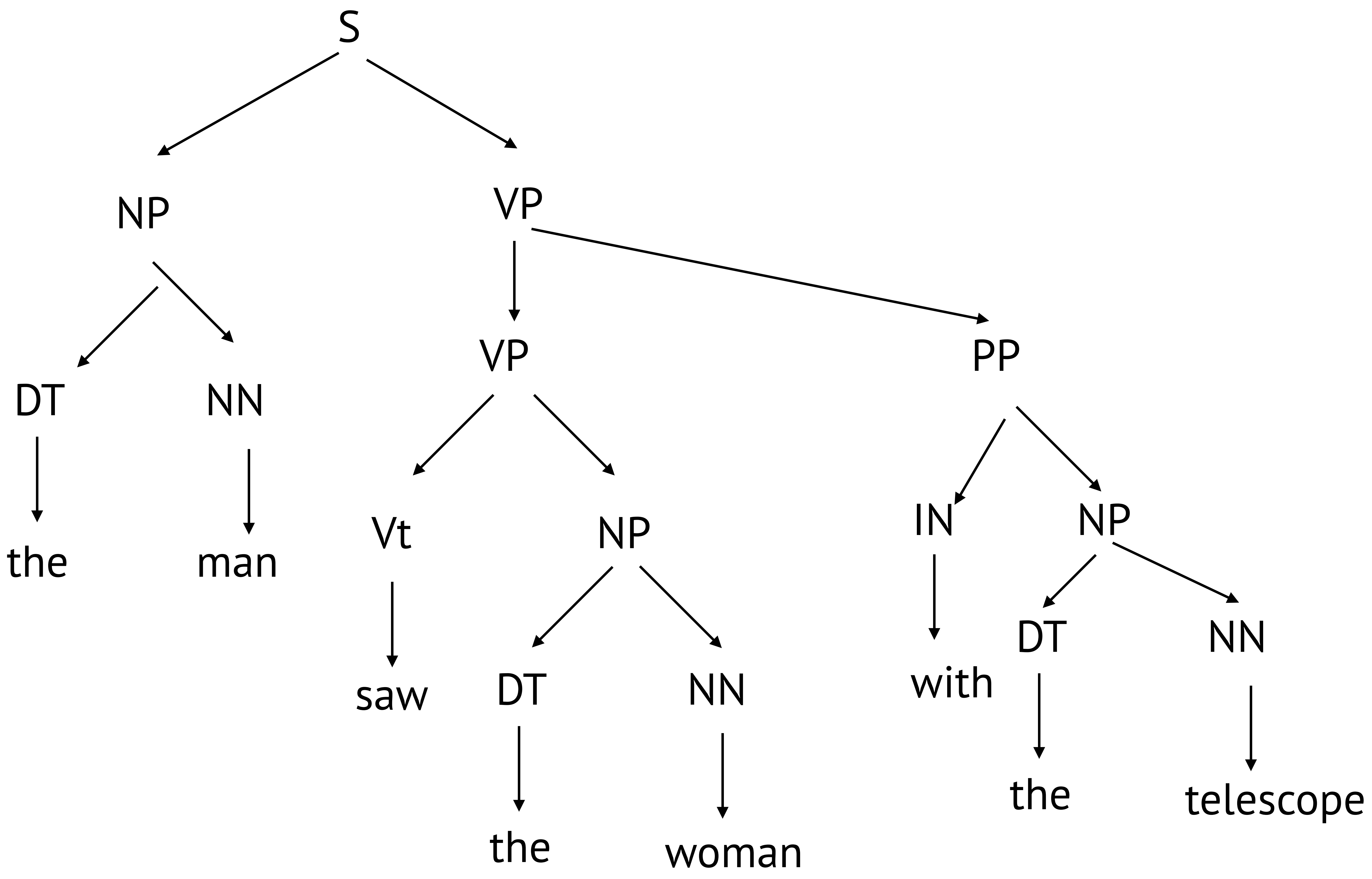
Treebanks (Penn Treebank)



4.5 million words of American English

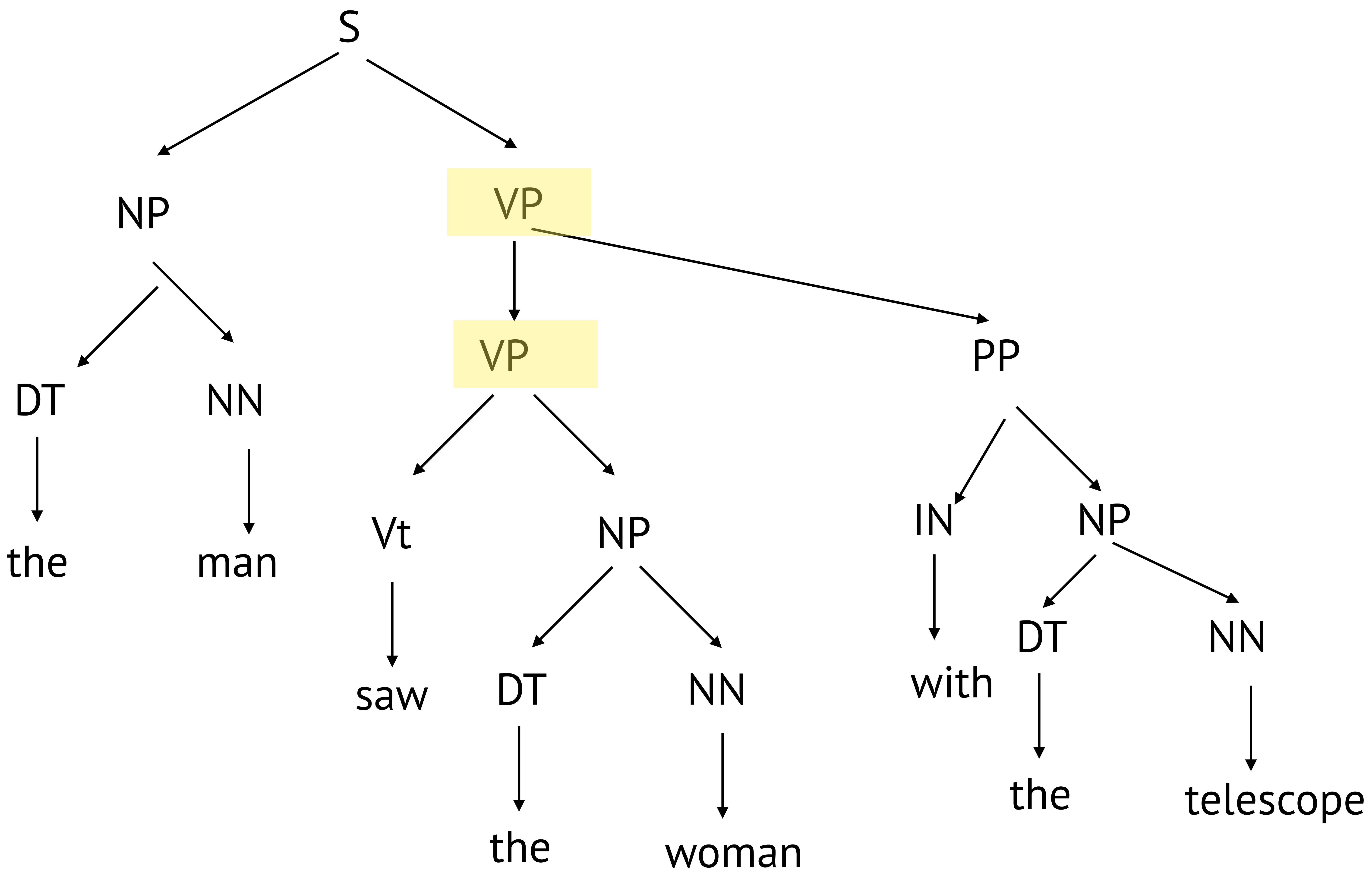
Mitch Marcus

Estimating the Probabilities from the Treebank



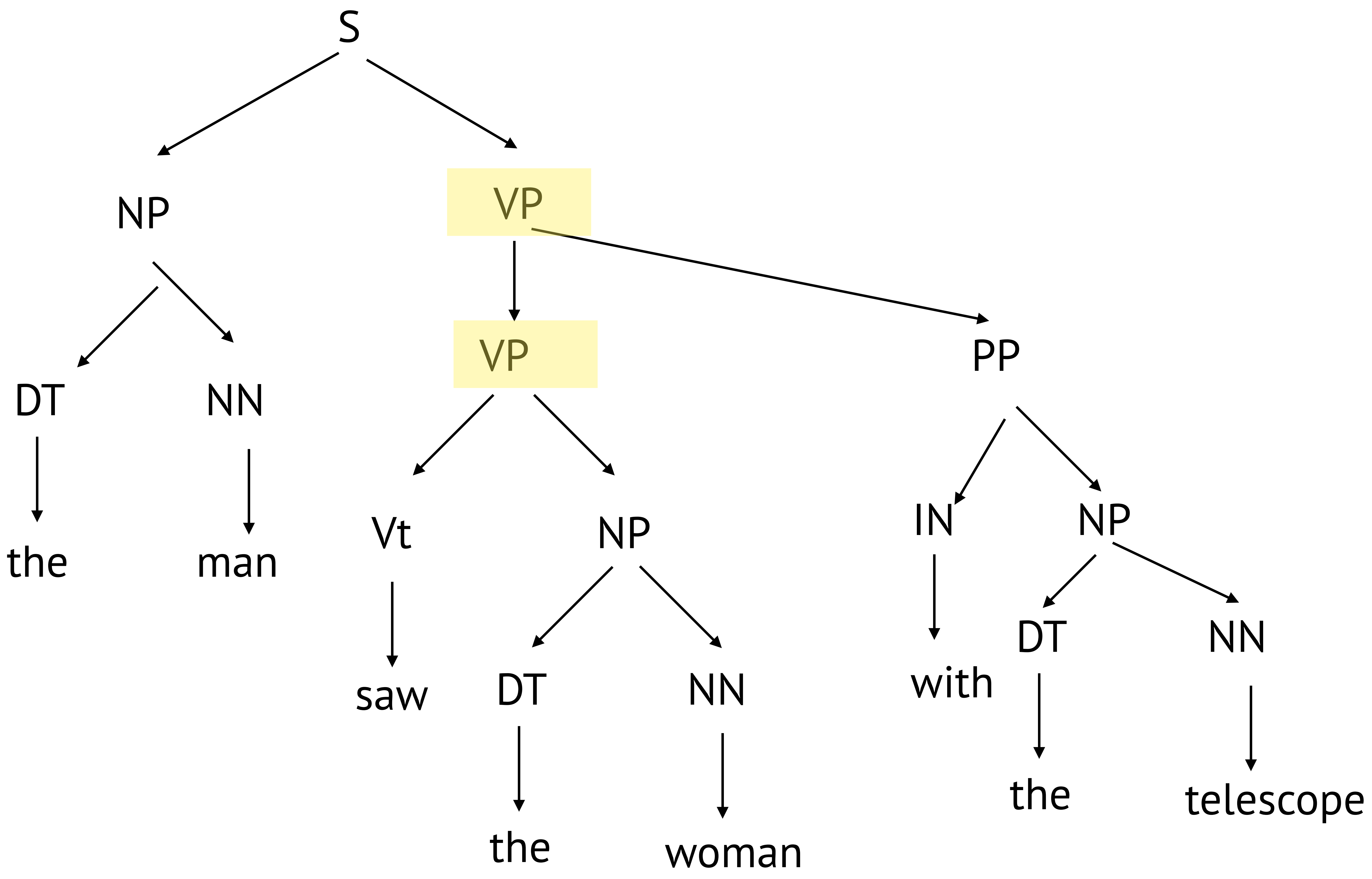
$$p(\mathrm{VP} \rightarrow \mathrm{VP\ PP}) = ?$$

Estimating the Probabilities from the Treebank



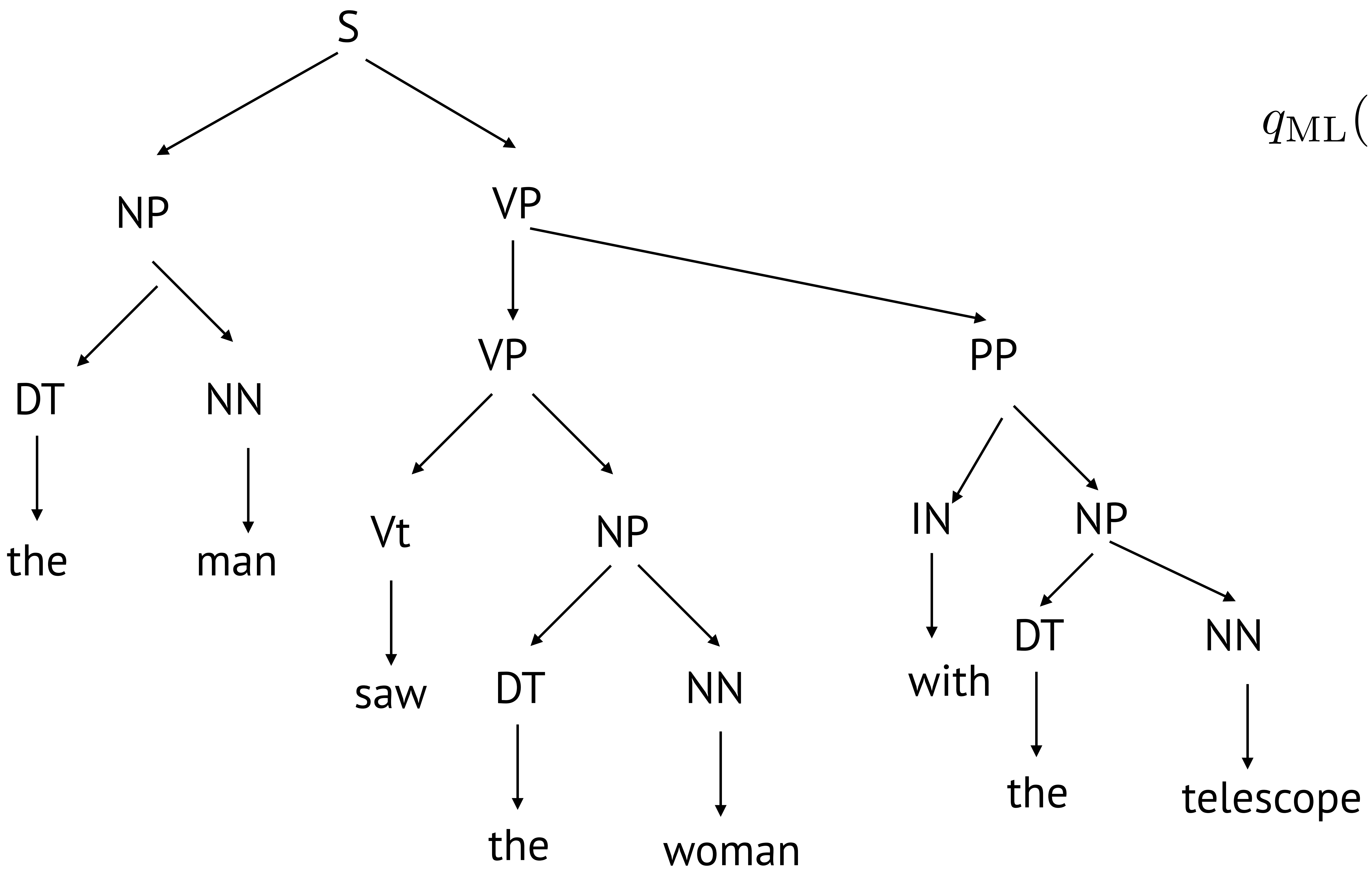
$$p(\text{VP} \rightarrow \text{VP PP}) =?$$

Estimating the Probabilities from the Treebank



$$p(\text{VP} \rightarrow \text{VP PP}) = \frac{1}{2}$$

Estimating the Probabilities from the Treebank

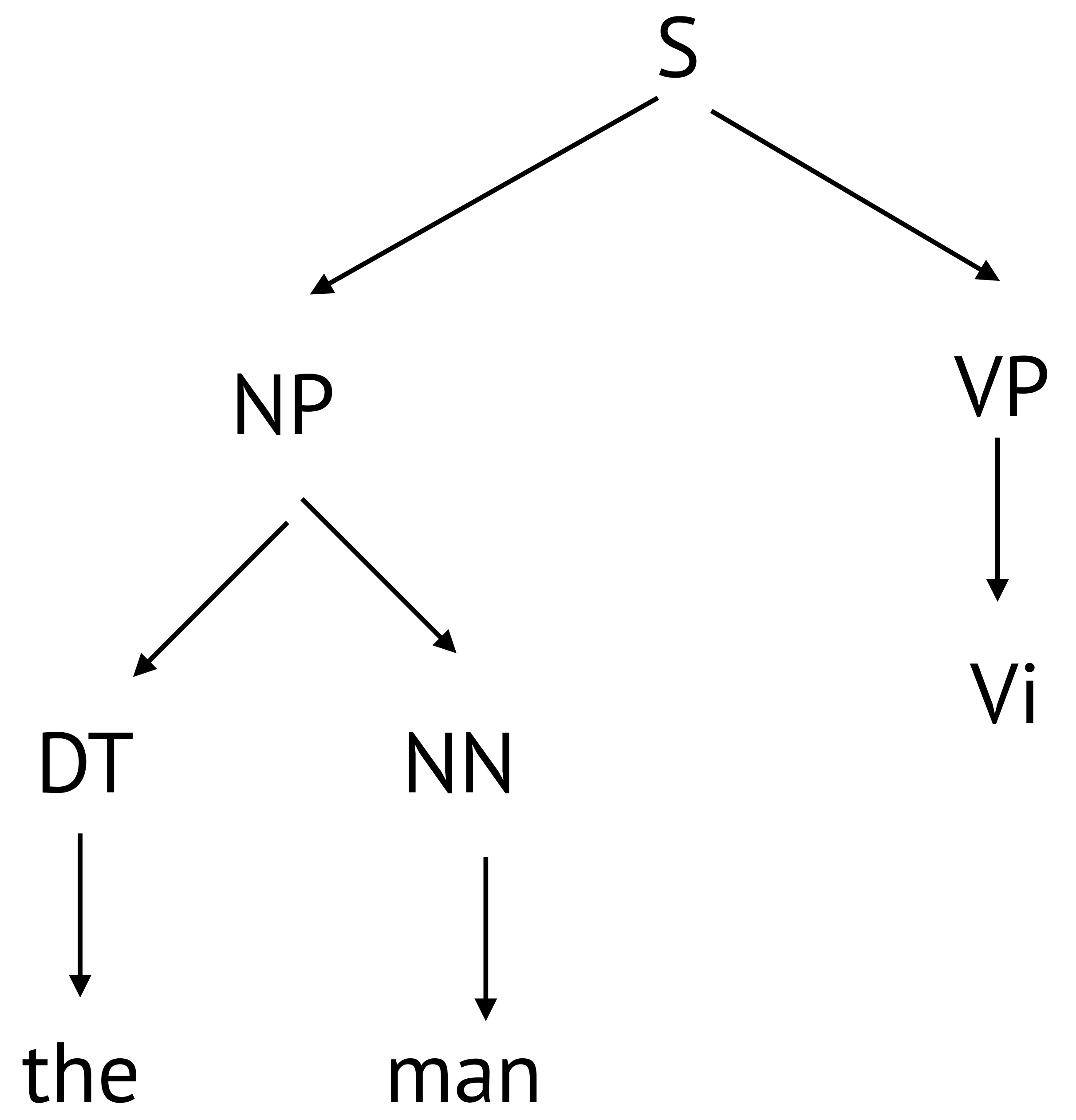


$$q_{\mathrm{ML}}(\alpha \rightarrow \beta) = \frac{\mathrm{Count}(\alpha \rightarrow \beta)}{\mathrm{Count}(\alpha)}$$

MLE estimation!

Parsing with PCFGs

Parsing: $$\arg\max_{\boldsymbol{y} \in \boldsymbol{Y}} p(\boldsymbol{x}, \boldsymbol{y})$$



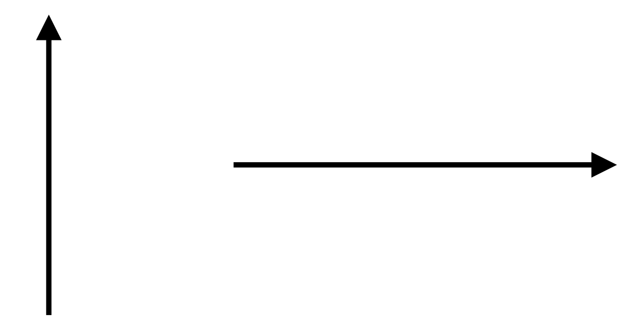
(S (NP ((DT the) (NN man)) (VP (Vi sleeps))))

$$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times 0.3 \times 1 \times 0.7 \times 0.4 \times 1$$

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

Find the lowest cost path from bottom left corner to the upper right corner



Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

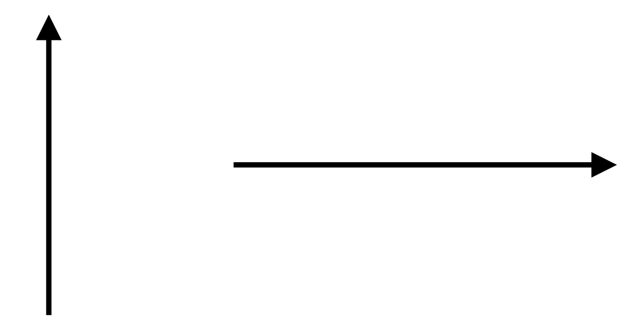
Find the lowest cost path from bottom left corner to the upper right corner

$2 + 2 + 5 + 4 + 2 + 9 + 7 = 31$

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

Find the lowest cost path from bottom left corner to the upper right corner



2 + 2 + 6 + 1 + 1 + 1 + 7 = 20

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

2			

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)			
4 (d)			
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)			
4 (d)			
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)			
4 (d)	10 (l)		
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)	13 (l)		
4 (d)	10 (l)	11 (l)	
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)	15 (d)		
9 (d)	13 (l)	14 (d)	
4 (d)	10 (l)	11 (l)	12 (l)
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)	15 (d)	23 (d)	
9 (d)	13 (l)	14 (d)	13 (d)
4 (d)	10 (l)	11 (l)	12 (l)
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)	15 (d)	23 (d)	20 (d)
9 (d)	13 (l)	14 (d)	13 (d)
4 (d)	10 (l)	11 (l)	12 (l)
2	7 (l)	11 (l)	13 (l)

Parsing with PCFGs

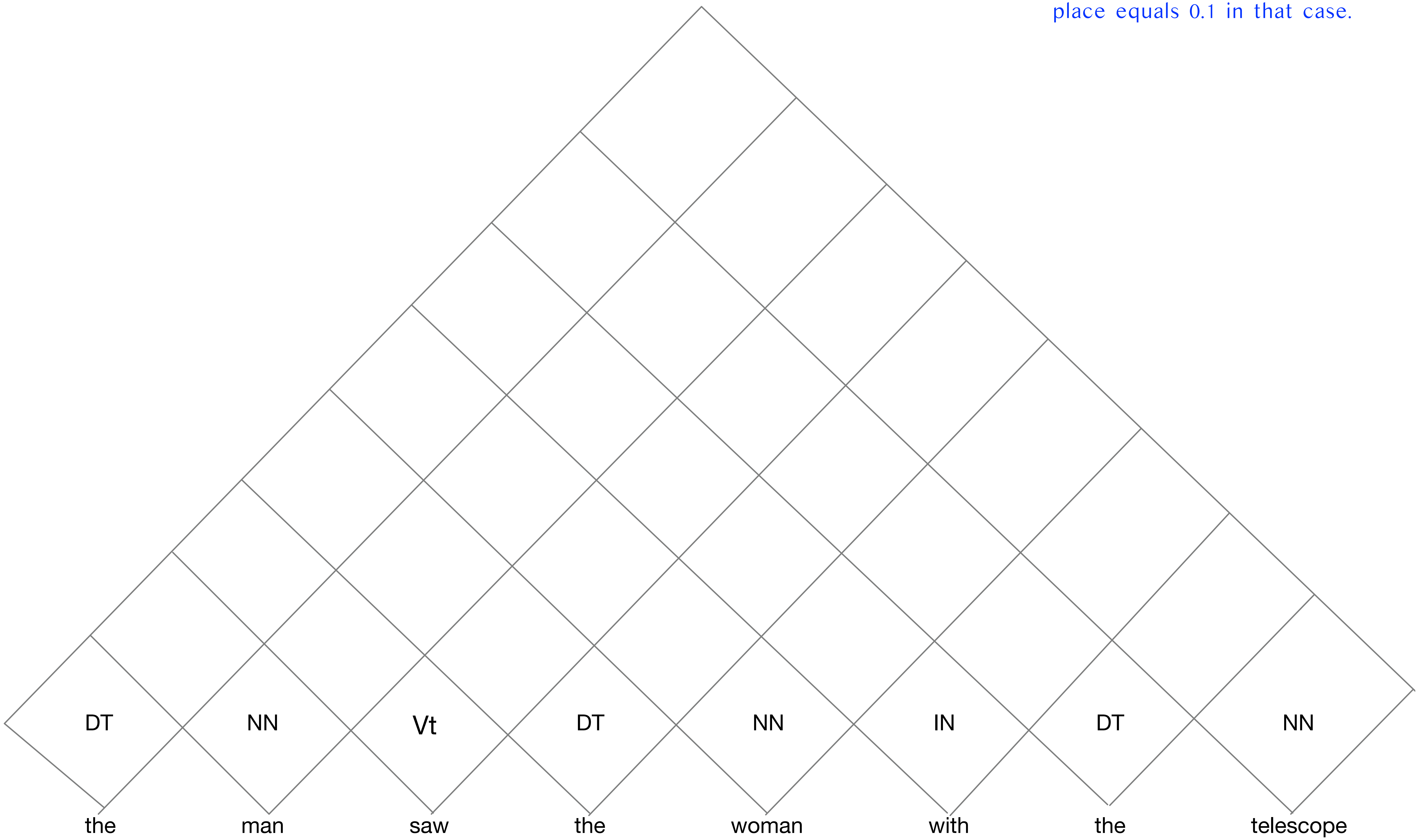
R =

S	$\longrightarrow$	NP VP	1.0
VP	$\longrightarrow$	Vt NP	0.8
VP	$\longrightarrow$	VP PP	0.2
NP	$\longrightarrow$	DT NN	0.8
NP	$\longrightarrow$	NP PP	0.2
PP	$\longrightarrow$	IN NP	1.0

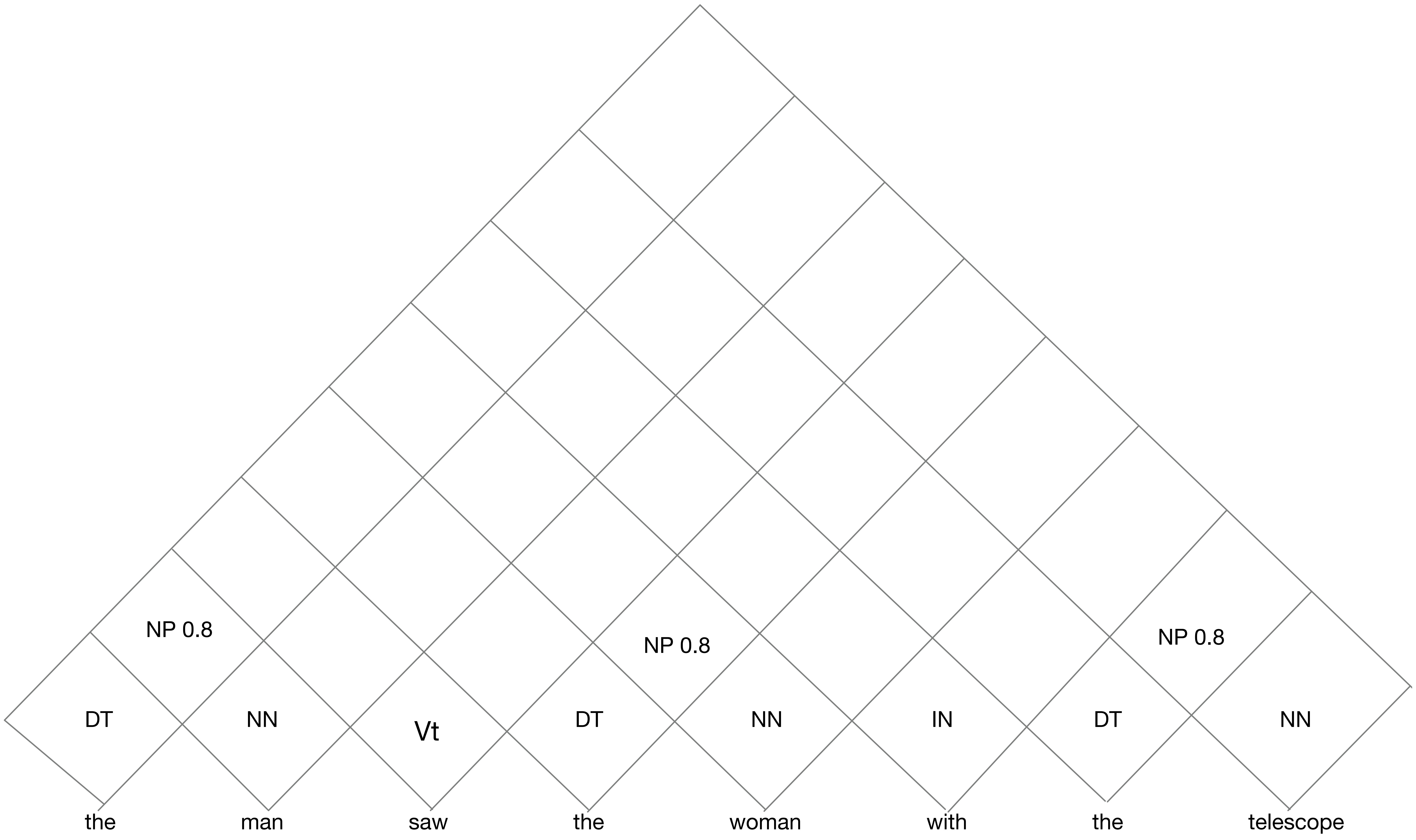
Vt	$\longrightarrow$	saw	1.0
NN	$\longrightarrow$	man	0.1
NN	$\longrightarrow$	woman	0.1
NN	$\longrightarrow$	telescope	0.3
NN	$\longrightarrow$	dog	0.5
DT	$\longrightarrow$	the	1.0
IN	$\longrightarrow$	with	0.6
IN	$\longrightarrow$	in	0.4

Parsing with PCFGs

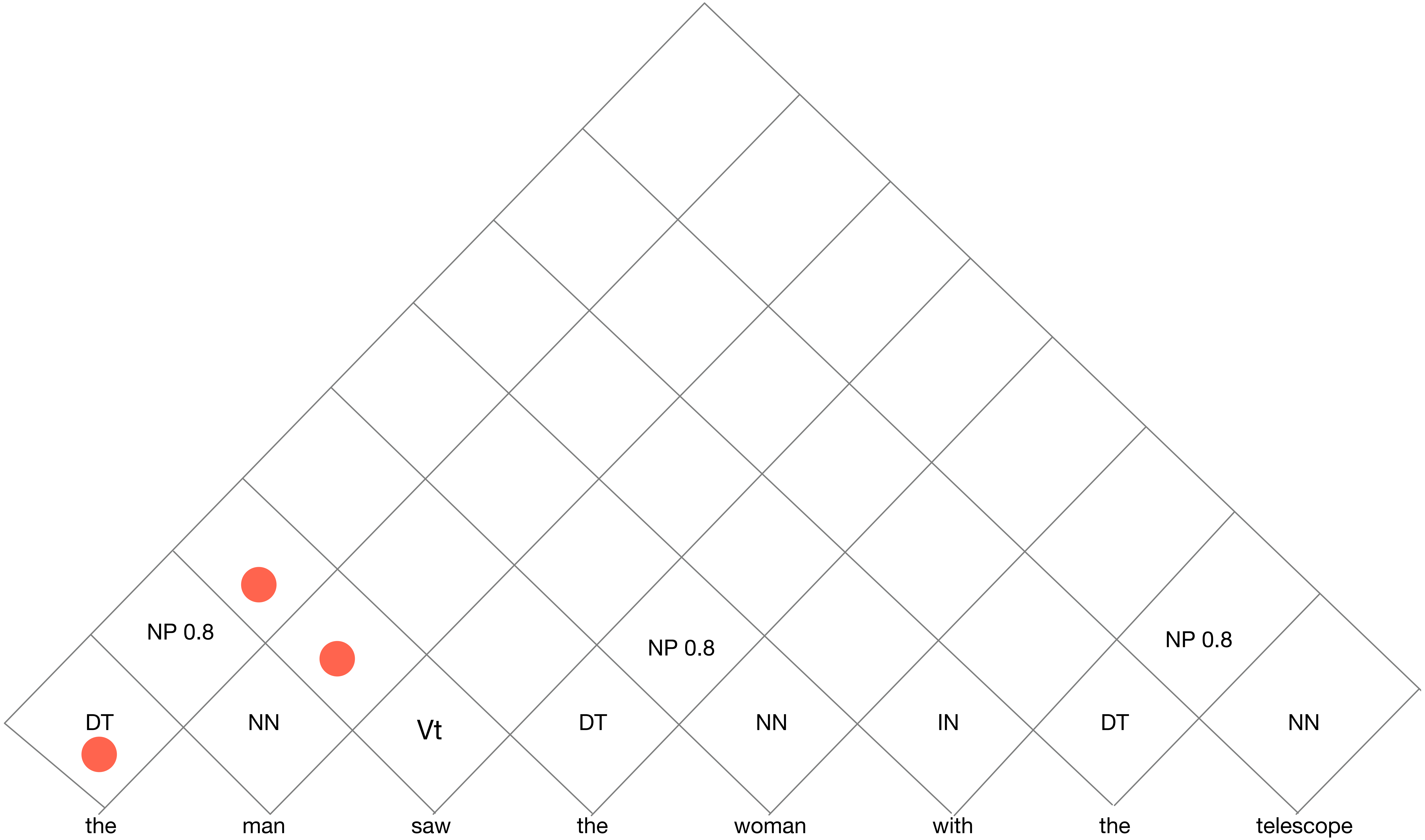
** For simplicity — Let's assume all these pre-terminals are built with probability 1 (this is obviously wrong) — remember that in reality you need to use rules like NN -> man 0.1 to build it. The probability of NN in that place equals 0.1 in that case.



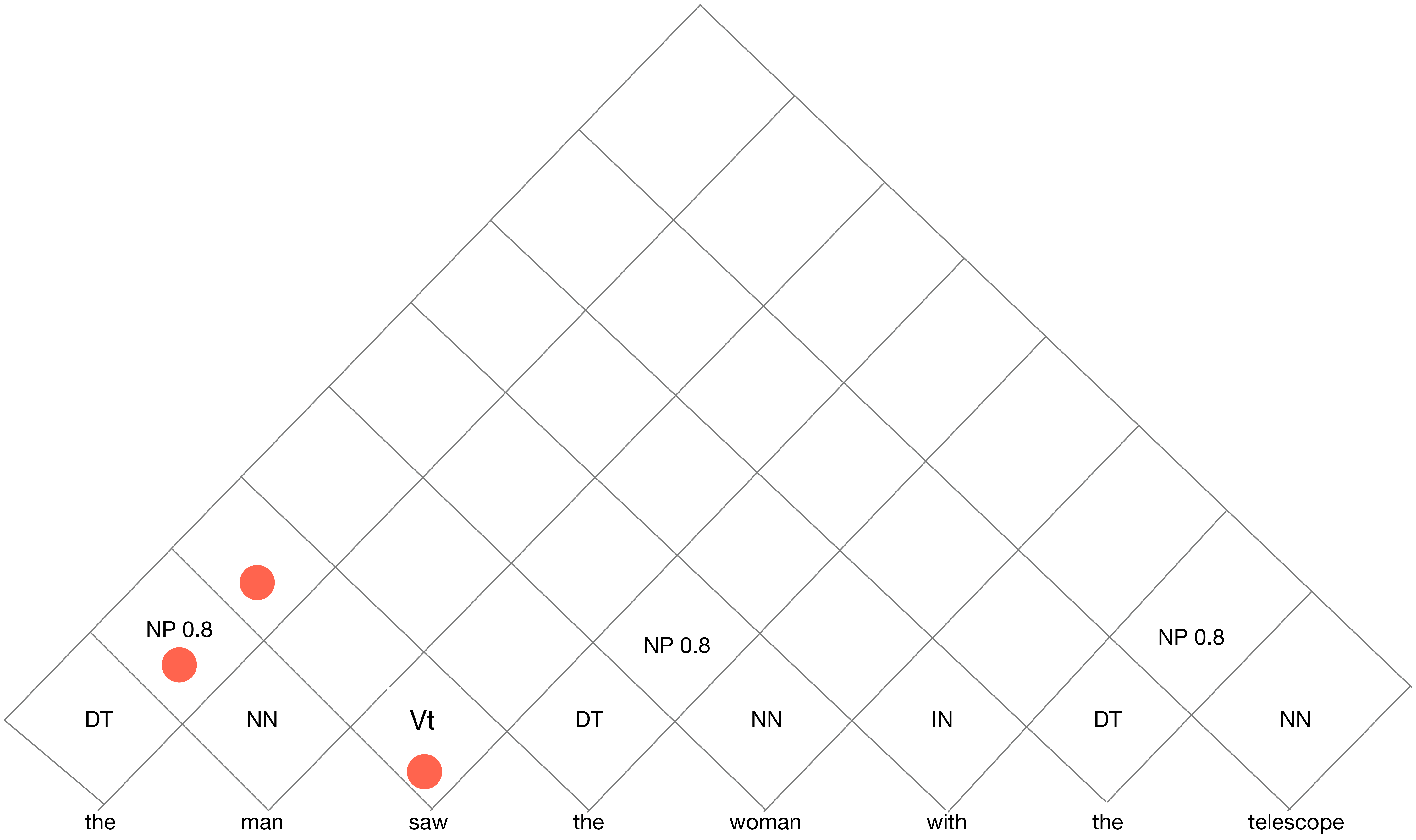
Parsing with PCFGs



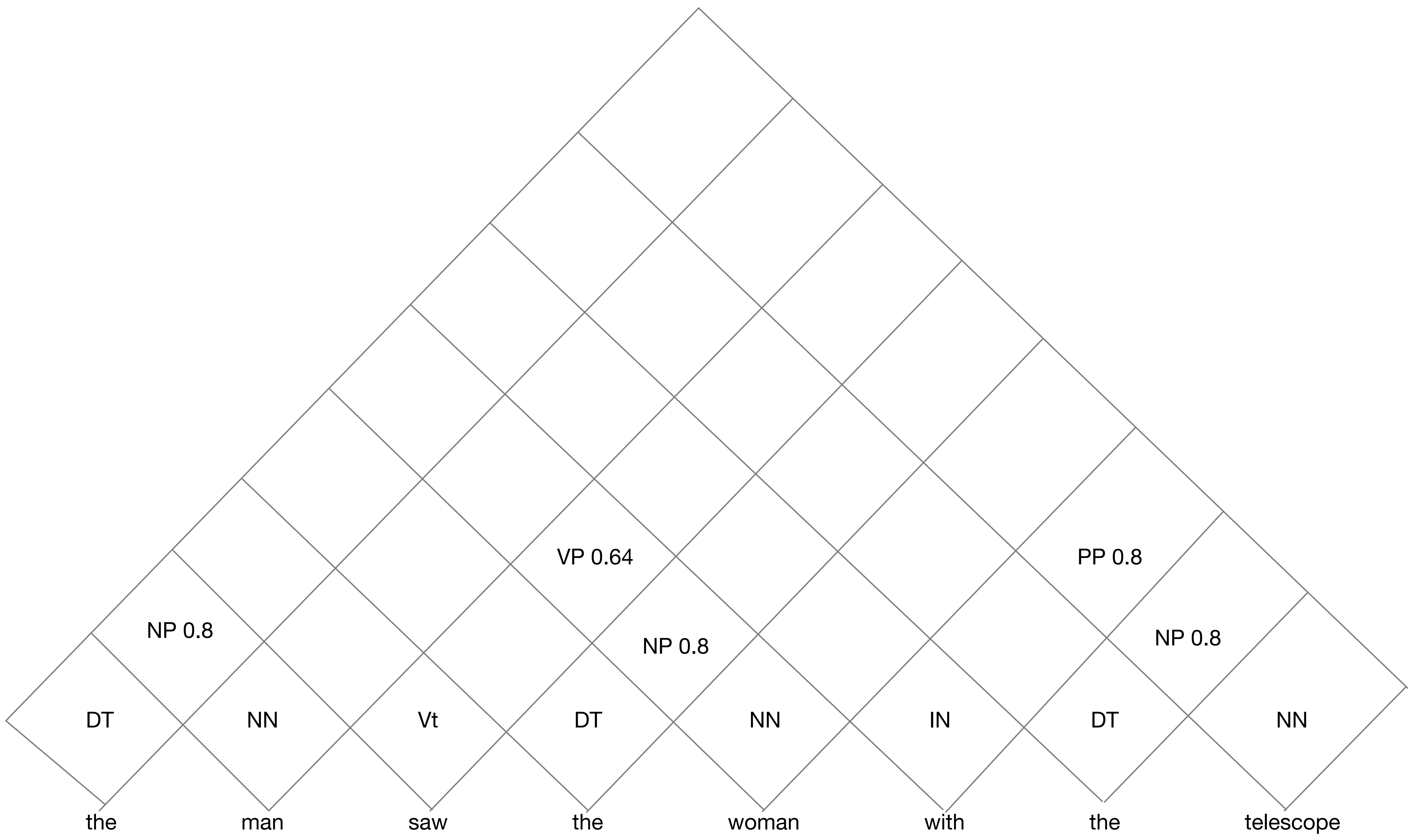
Parsing with PCFGs



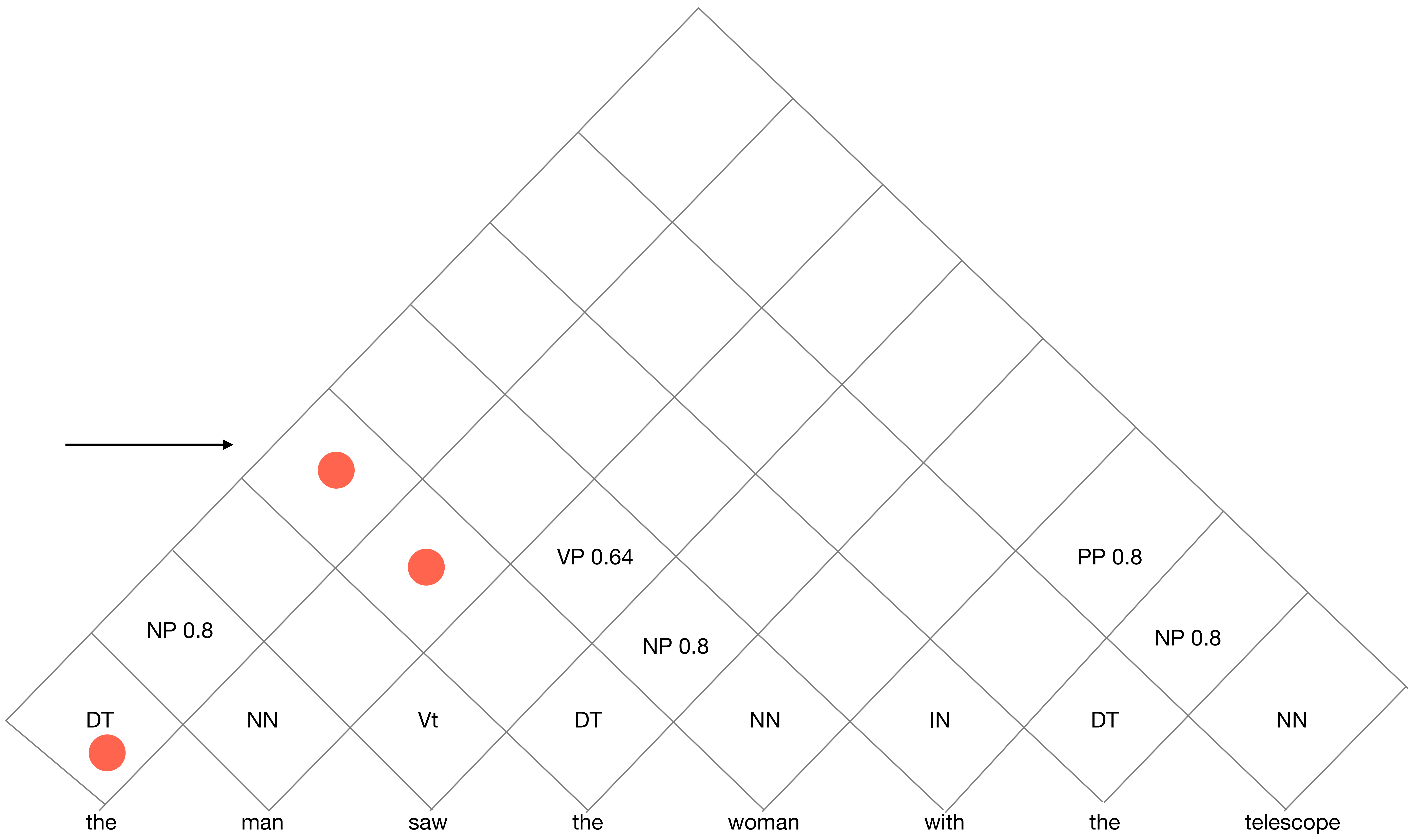
Parsing with PCFGs



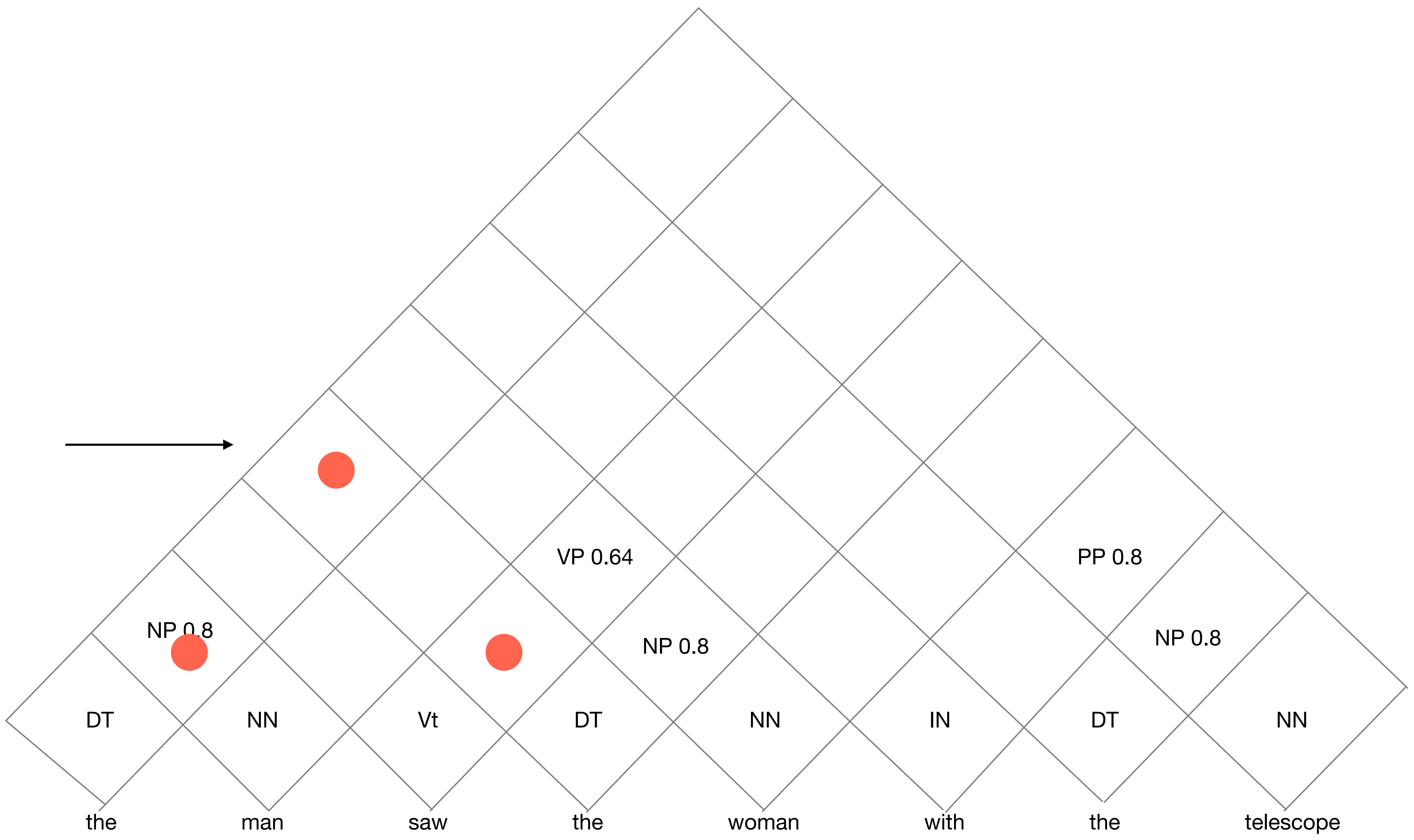
Parsing with PCFGs



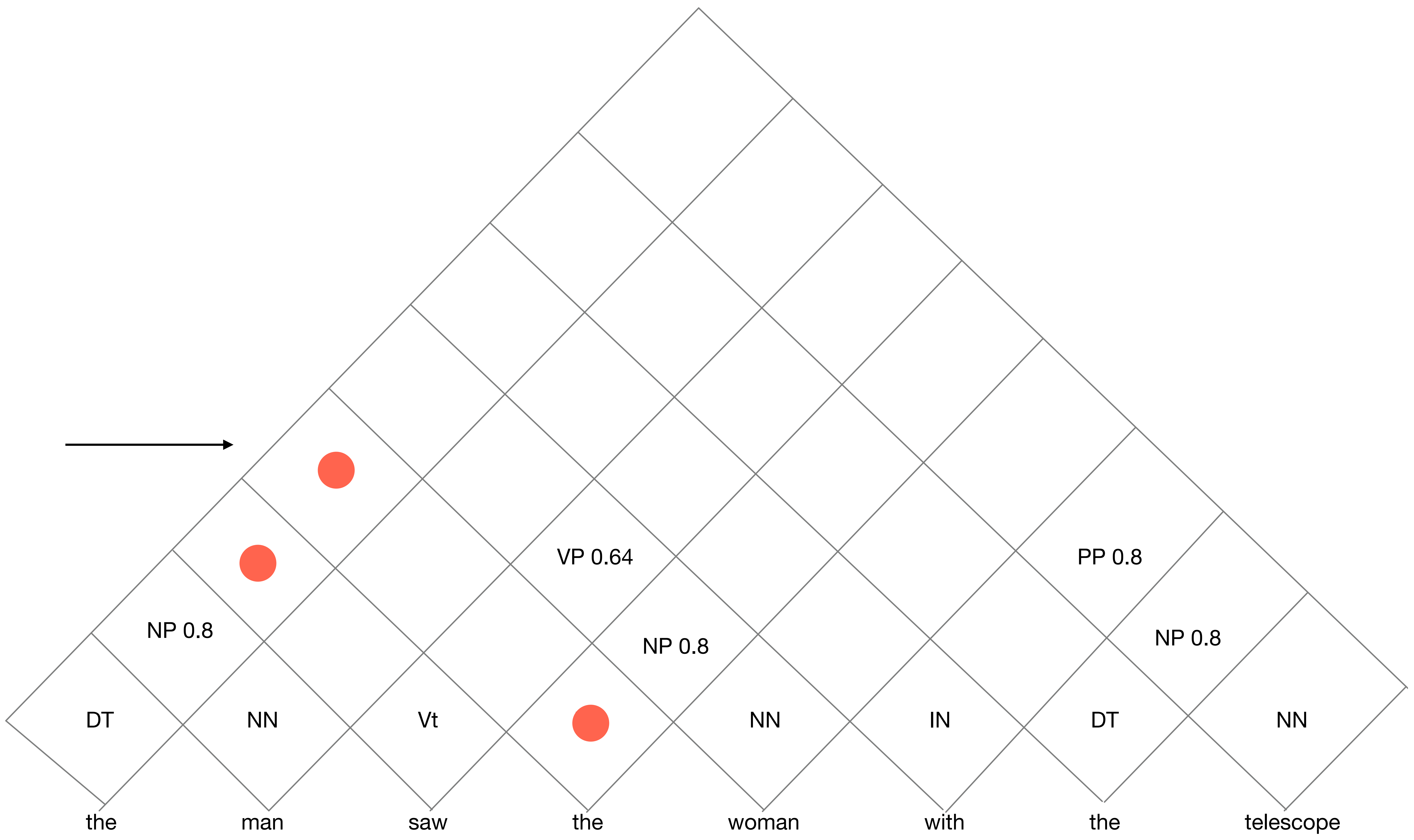
Parsing with PCFGs



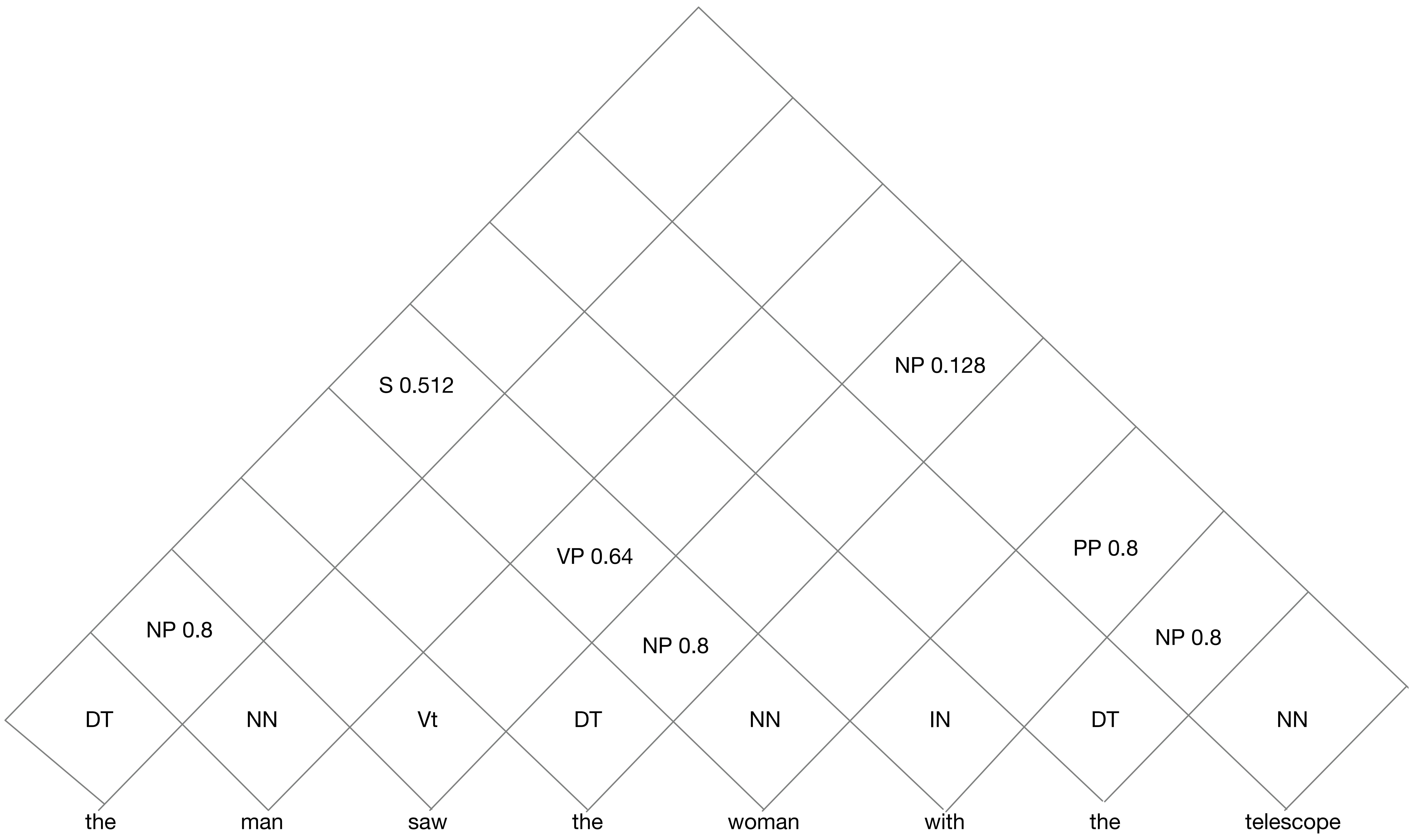
Parsing with PCFGs



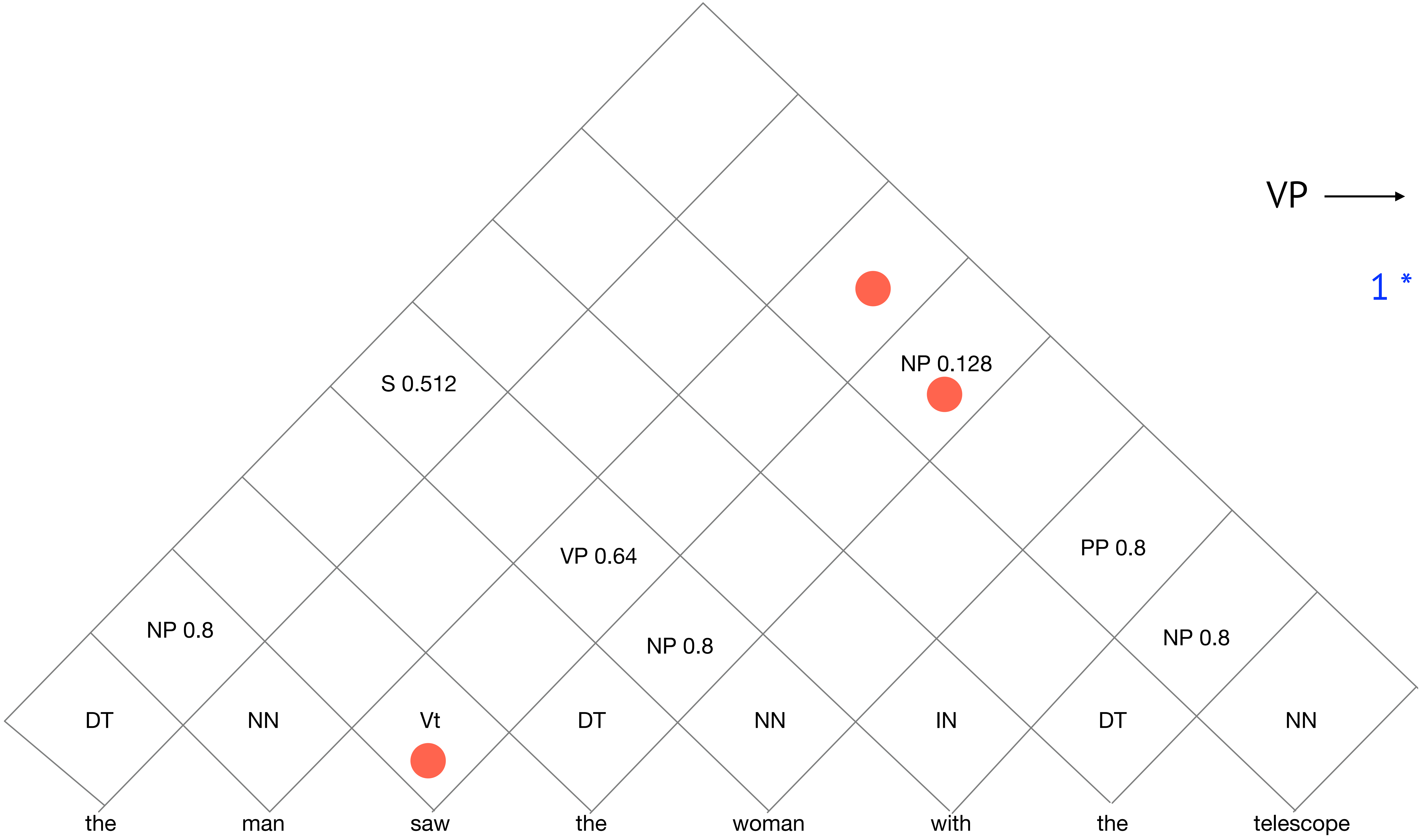
Parsing with PCFGs



Parsing with PCFGs



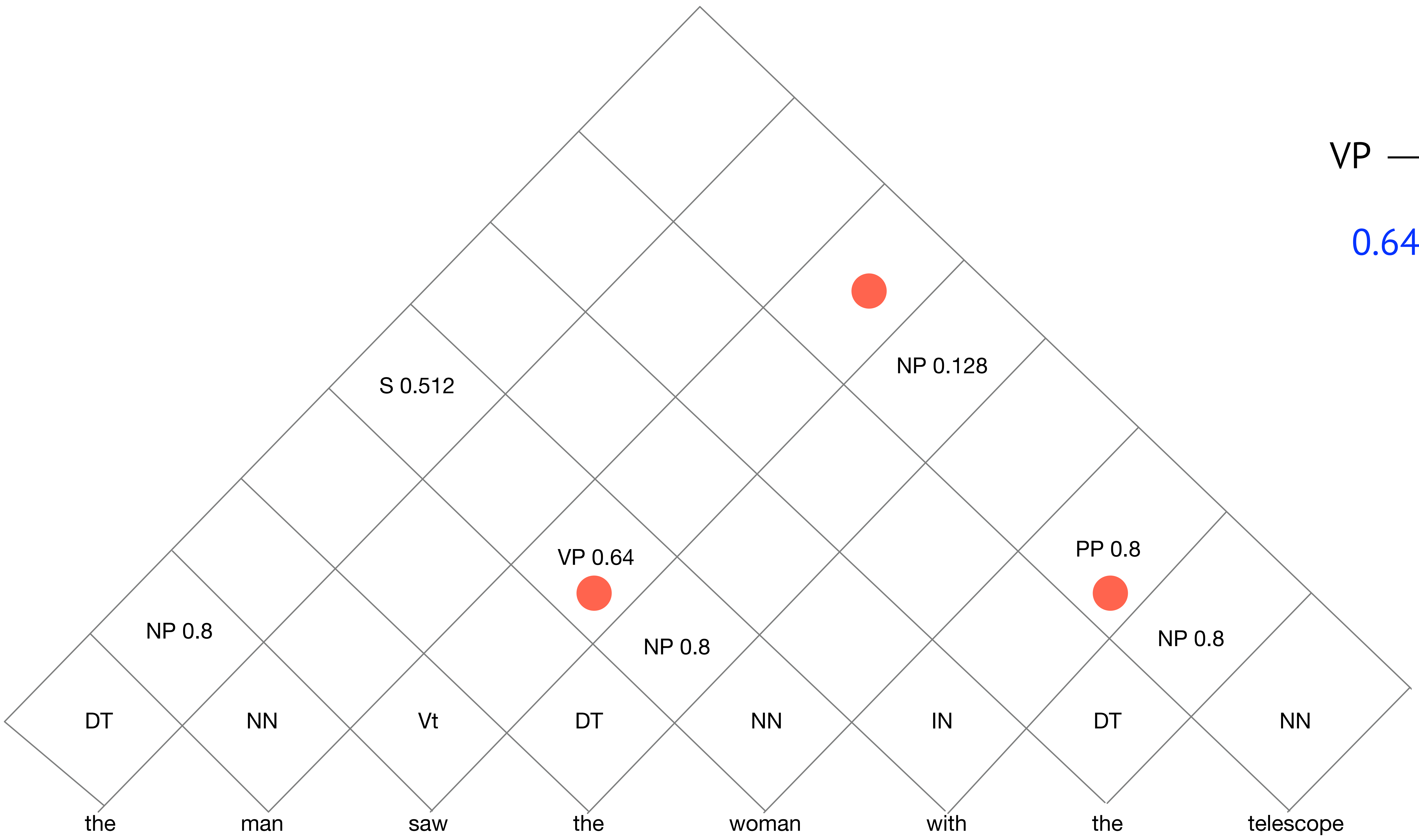
Parsing with PCFGs



VP ⟶ Vt NP 0.8

1 * 0.128 * 0.8 = 0.1024

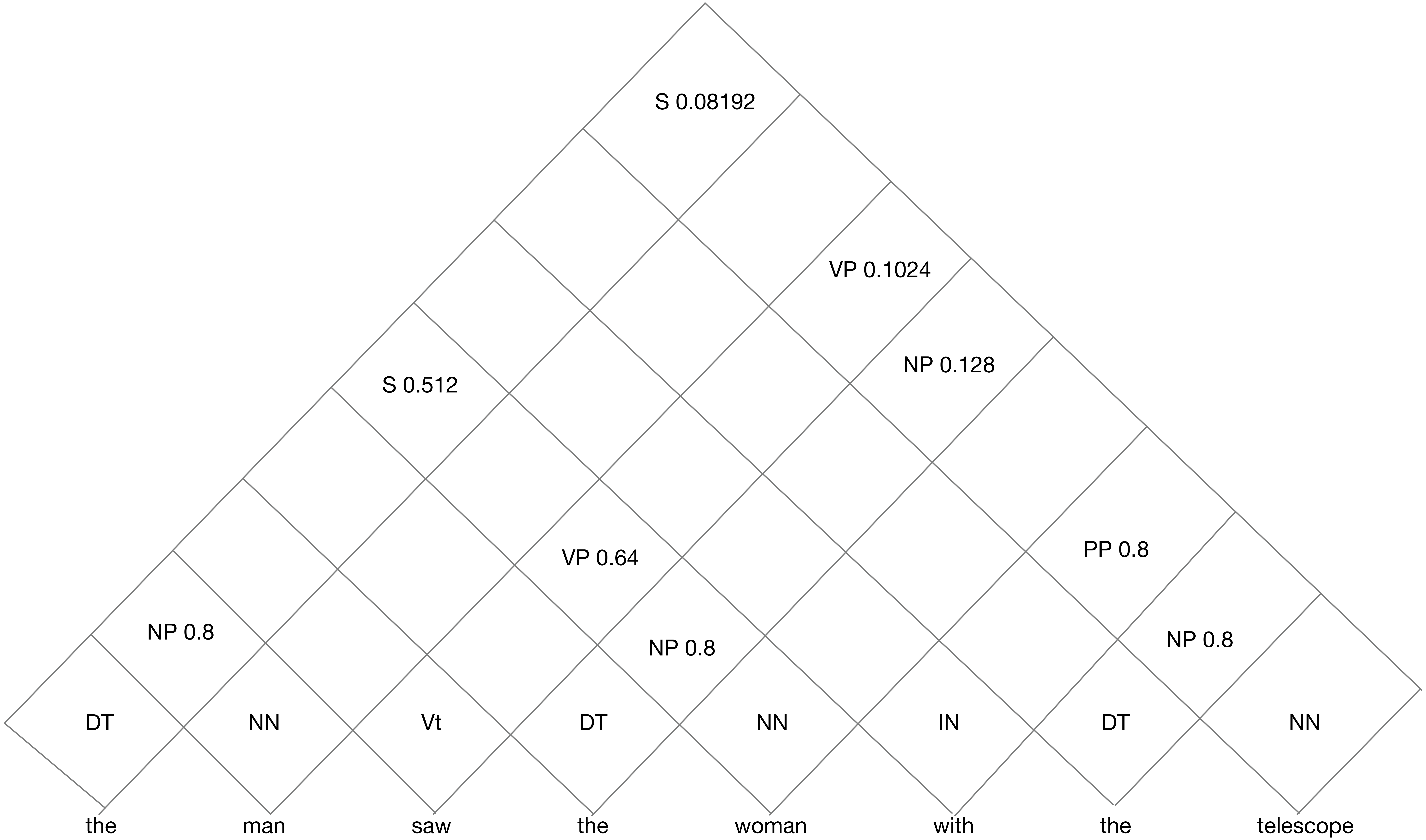
Parsing with PCFGs



$VP \longrightarrow VP\ \ PP \quad 0.2$

$0.64 * 0.8 * 0.2 = 0.1024$

Parsing with PCFGs (CYK Algorithm)

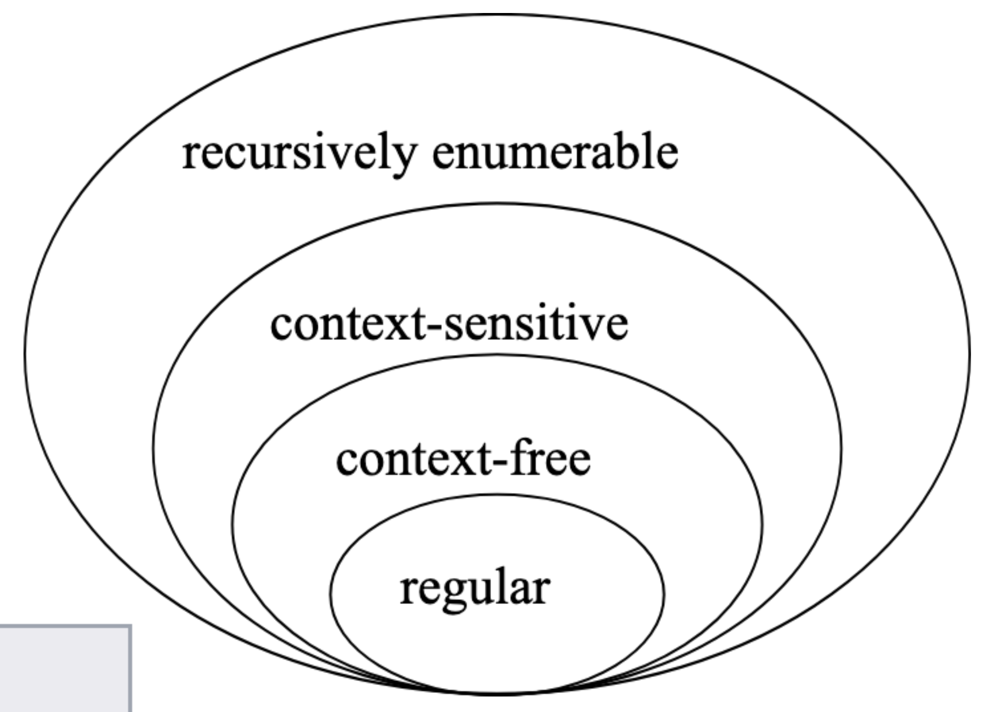


Revisit Question

Match strings with equal number of a and b.

e.g., aaabbb, aabb, aaaabbbb...

Chomsky Hierarchy



Grammar	Languages	Recognizing Automaton	Production rules (constraints)[a]	Examples[5][6]
Type-3	Regular	Finite-state automaton	$A \to \mathrm{a}$ $A \to \mathrm{a}B$ (right regular) or $A \to \mathrm{a}$ $A \to B\mathrm{a}$ (left regular)	$L = \{a^n \| n > 0\}$
Type-2	Context-free	Non-deterministic pushdown automaton	$A \to \alpha$	$L = \{a^n b^n \| n > 0\}$
Type-1	Context-sensitive	Linear-bounded non-deterministic Turing machine	$\alpha A \beta \to \alpha \gamma \beta$	$L = \{a^n b^n c^n \| n > 0\}$
Type-0	Recursively enumerable	Turing machine	$\gamma \to \alpha$ (γ non-empty)	$L = \{w \| w$ describes a terminating Turing machine$\}$

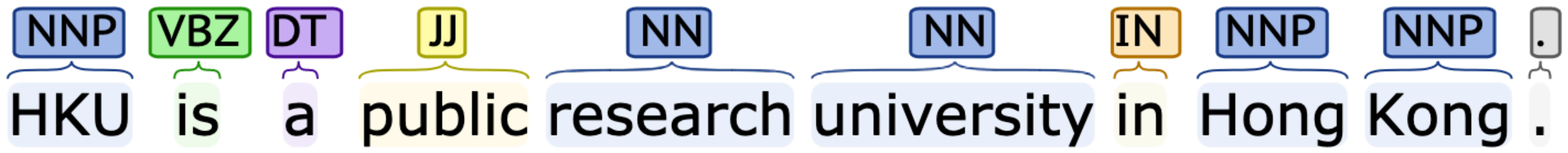
a. ^ Meaning of symbols:

- a = terminal
- A, B = non-terminal
- α, β, γ = string of terminals and/or non-terminals

https://en.wikipedia.org/wiki/Chomsky_hierarchy

Linguistic Structures

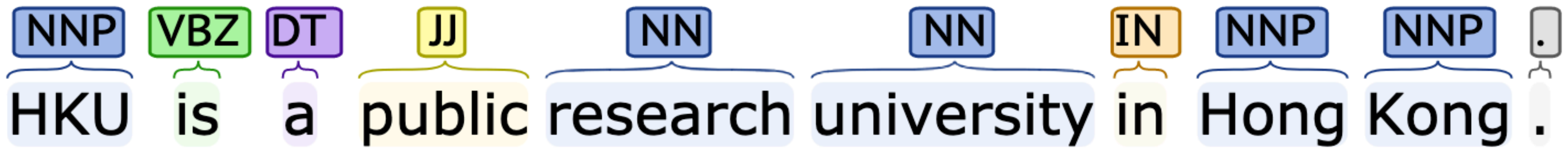
HKU is a public research university in Hong Kong.



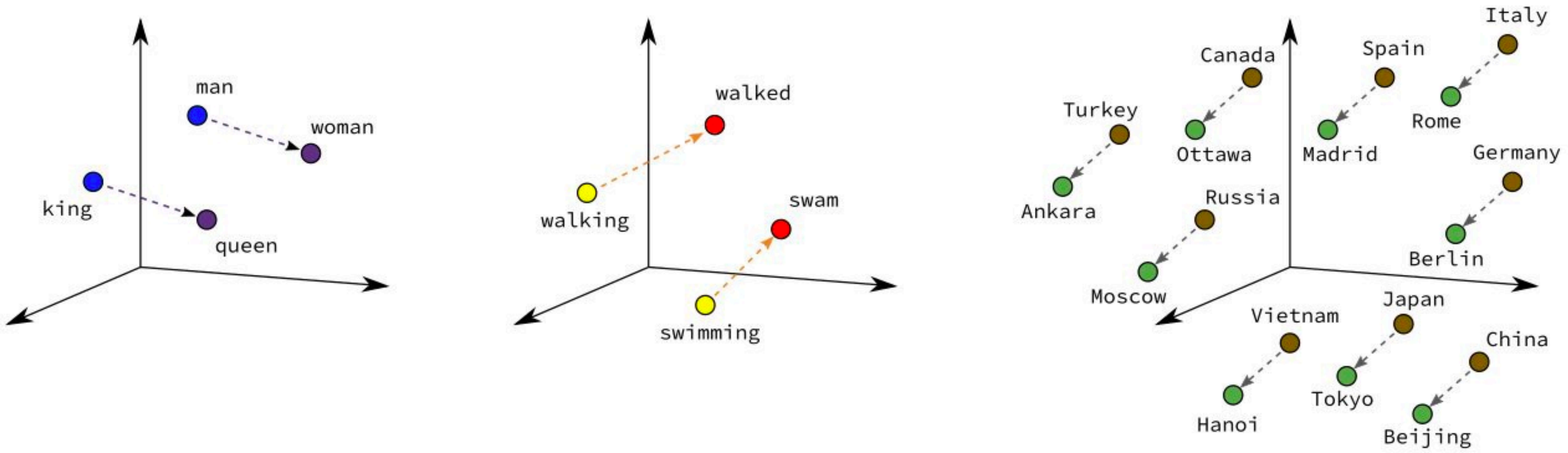
Part-of-speech tagging (word class)

Named entity recognition

Linguistic Structures



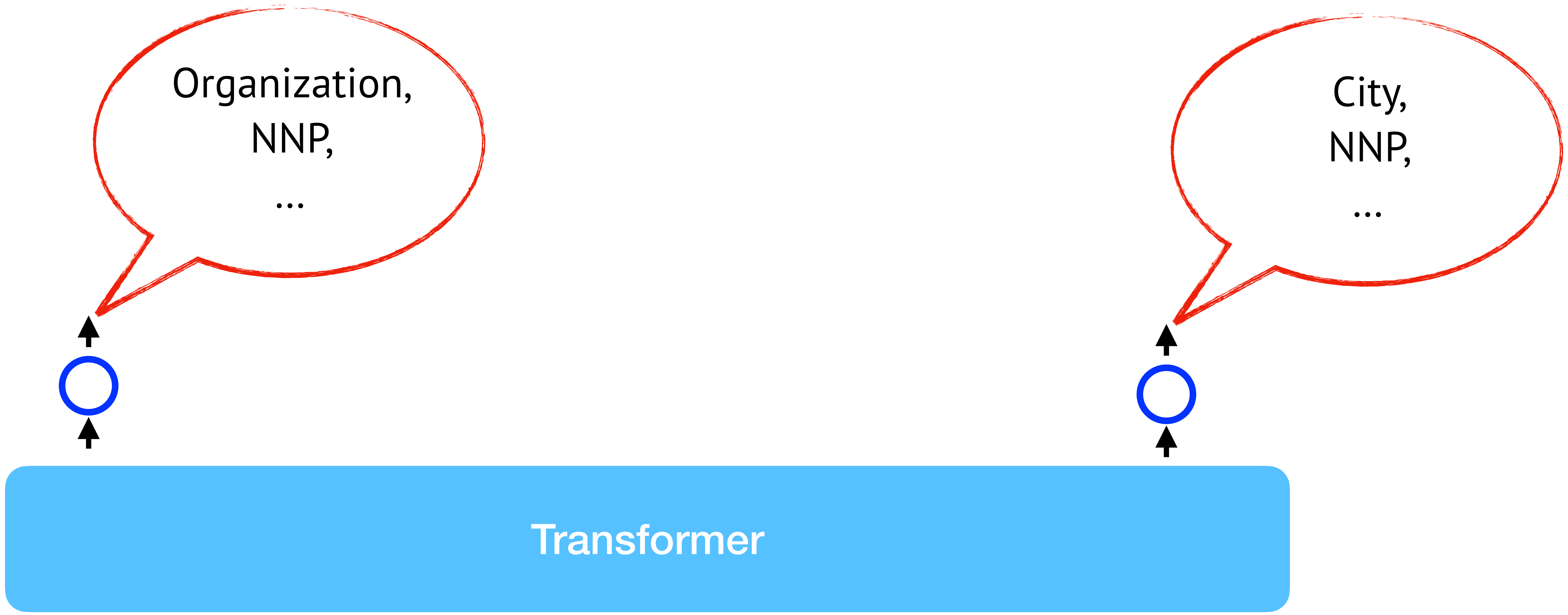
Part-of-speech tagging
(word class)



Male-Female   Verb Tense   Country-Capital

https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space

Linguistic Structures



HKU is a public research university in Hong Kong.