

COMP90042

Web search and text analysis

Workshop Week 10

xudong.han@unimelb.edu.au

https://github.com/HanXudong/COMP90042_Workshops

regular grammar

- For example: a simple regular grammar
- Rules: $S \rightarrow A$, $A \rightarrow aA$, $A \rightarrow \epsilon$
- S is the start symbol
- It will generate words such as a , aa , aaa , aaa
- The set of words generated by this regular grammar is a regular language
- This regular language can also be expressed in regular expression a^*

Q1 What is chart parsing? Why is it important?

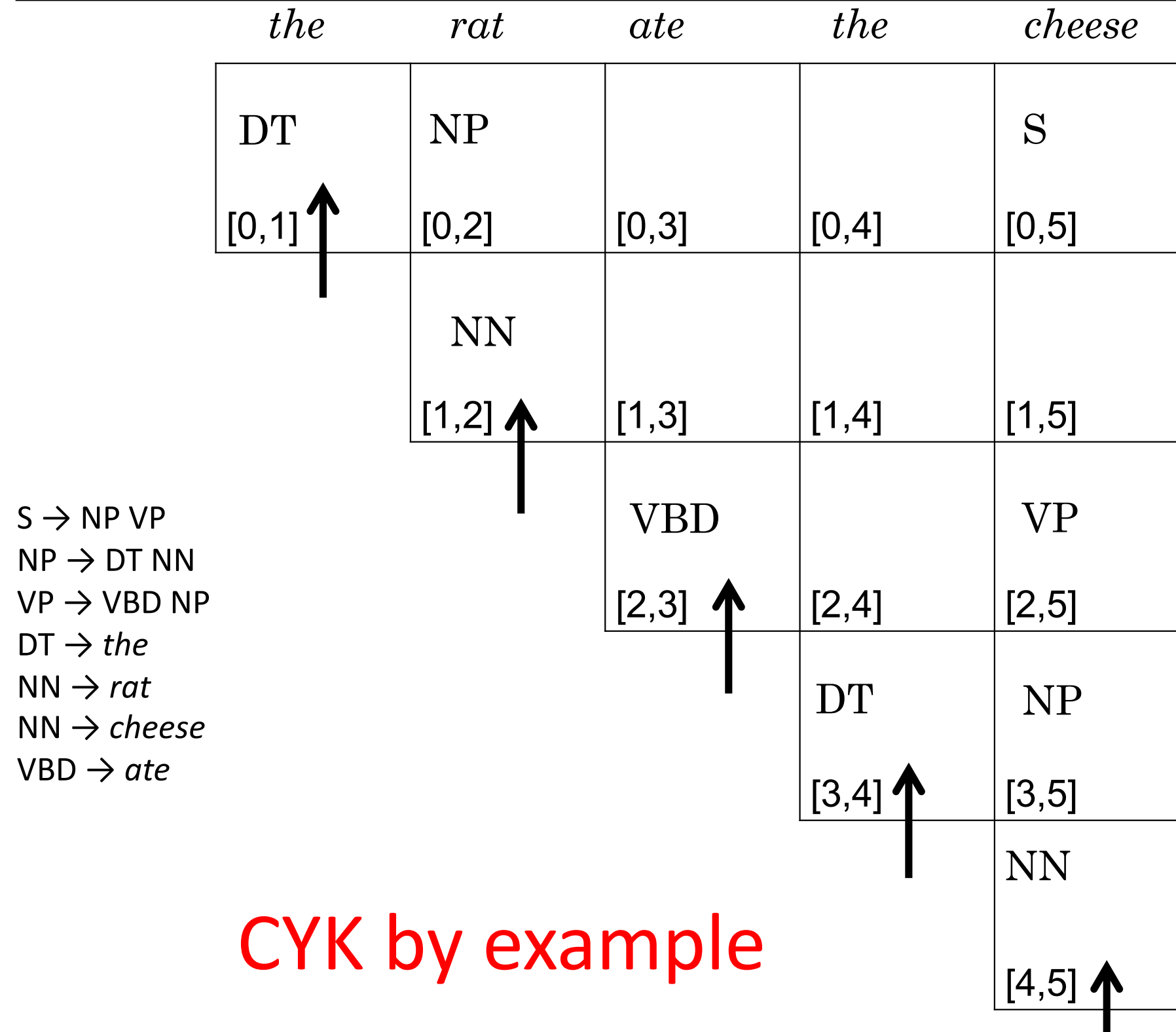
Parsing CFGs

- Parsing: given string, identify possible structures
- Brute force search is intractable for non-trivial grammars
 - * Good solutions use dynamic programming
- Two general strategies
 - * Bottom-up
 - Start with words, work up towards S
 - CYK parsing

Q1 What is chart parsing? Why is it important?

COMP90042 W.S.T.A. (S1 2019)

L17



Q2. What changes need to be made to grammar?

S	->	NP	VP
VP	->	V NP	V NP PP
PP	->	P NP	
V	->	saw	walked
NP	->	Joh	Bob Det N Det N PP
Det	->	a	an the my
N	->	ma	cat telescope park
P	->	on	by with

- **Productions (rules)**

$$W \rightarrow X Y Z$$

- * Exactly one non-terminal on left-hand side (LHS)
- * An ordered list of symbols on right-hand side (RHS)
 - can be **Terminals** or **Non-terminals**

Q2. What changes need to be made to grammar?

S	->	NP	VP						
VP	->	V	NP		V	NP	PP		
PP	->	P	NP						
V	->	saw		walked					
NP	->	Joh		Bob		Det	N		Det N PP
Det	->	a		an		the		my	
N	->	ma		cat		telescope		park	
P	->	on		by		with			

CYK parsing:

Convert grammar to Chomsky Normal Form (CNF)

- Change grammar so all rules of form
 $A \rightarrow BC$ or $A \rightarrow a$

Q2-a-(i) “a man saw John”

0	1	2	3	4
a	man	saw	John	
[0,1]	[0,2]	[0,3]	[0,4]	
	[1,2]	[1,3]	[1,4]	
		[2,3]	[2,4]	
			[3,4]	

S	->	NP	VP						
VP	->	V	NP		V	X			
PP	->	P	NP						
V	->	saw		walked					
NP	->	John		Bob		Det	N		Det Y
Det	->	a		an		the		my	
N	->	man		cat		telescope		park	
P	->	on		by		with			
X	->	NP	PP						
Y	->	N	PP						

Q2-a-(i) “a man saw John”

0	1	2	3	4
a	man	saw	John	
[0,1] Det	[0,2] NP	[0,3]	[0,4] S	
	[1,2] N	[1,3]	[1,4]	
		[2,3] V	[2,4] VP	
			[3,4] NP	

S	->	NP	VP						
VP	->	V	NP		V	X			
PP	->	P	NP						
V	->	saw		walked					
NP	->	John		Bob		Det	N		Det Y
Det	->	a		an		the		my	
N	->	man		cat		telescope		park	
P	->	on		by		with			
X	->	NP	PP						
Y	->	N	PP						

Q2-a-(iii) “park by the cat with my telescope”

S	->	NP	VP						
VP	->	V	NP		V	X			
PP	->	P	NP						
V	->	saw		walked					
NP	->	John		Bob		Det	N		Det Y
Det	->	a		an		the		my	
N	->	man		cat		telescope		park	
P	->	on		by		with			
X	->	NP	PP						
Y	->	N	PP						

<i>park</i>	<i>by</i>	<i>the</i>	<i>cat</i>	<i>with</i>	<i>my</i>	<i>telescope</i>
[0,1] N	[0,2] -	[0,3] -	[0,4] Y	[0,5] -	[0,6] -	[0,7] Y
	[1,2] P	[1,3] -	[1,4] PP	[1,5] -	[1,6] -	[1,7] PP
		[2,3] Det	[2,4] NP	[2,5] -	[2,6] -	[2,7] NP, X
			[3,4] N	[3,5] -	[3,6] -	[3,7] Y
				[4,5] P	[4,6] -	[4,7] PP
					[5,6] Det	[5,7] NP
						[6,7] N

Q2-a-(ii) “an park by Bob walked an park with Bob”

[illegible]

Q3 and Q4

- What differentiates probabilistic parsing from chart parsing? Why is this important? How does this affect the algorithms used for parsing?
- What is a probabilistic grammar and what problem does it attempt to solve?

Q3 and Q4

- Given a tree, we can compute its probability
 - Decomposes into probability of each production

- E.g., for tree on right,

* $P(\text{tree}) =$

$P(S \rightarrow VP) \times$

$P(VP \rightarrow \text{Verb NP}) \times$

$P(\text{Verb} \rightarrow \textit{Book}) \times$

$P(NP \rightarrow \text{Det Nominal}) \times$

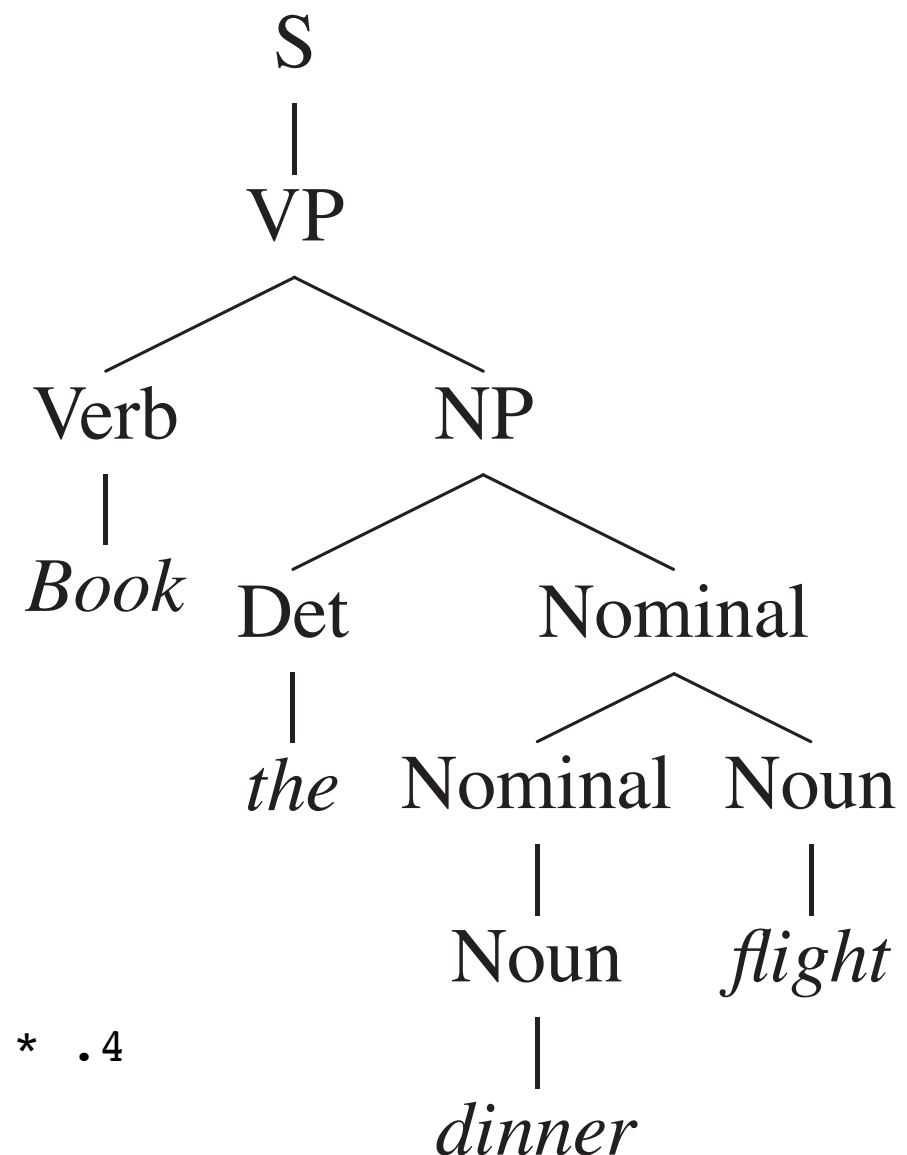
$P(\text{Det} \rightarrow \textit{the}) \times$

$P(\text{Nominal} \rightarrow \text{Nominal Noun}) \times$

$P(\text{Nominal} \rightarrow \text{Noun}) \times$

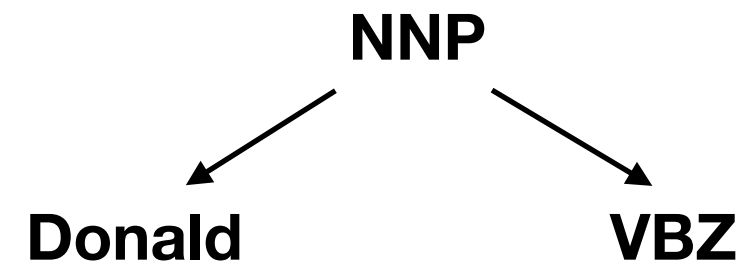
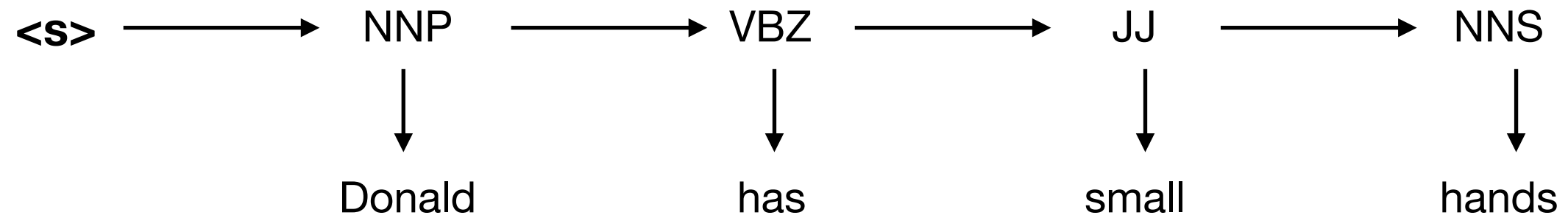
$P(\text{Noun} \rightarrow \textit{dinner}) \times$

$P(\text{Noun} \rightarrow \textit{flight}) = 2.16 \times 10^{-6}$



I.e., $.05 * .2 * .3 * .2 * .6 * .2 * .75 * .1 * .4$

Q5



Transition	Emission
<s> -> NNP NNP'	NNP -> Donald
NNP' -> VBZ VBZ'	VBZ -> has
VBZ' -> JJ JJ'	JJ -> small
JJ' -> NNS	NNS -> hands