MODULE 3

SYNTAX ANALYSIS

WORLD STAR
Date :
Page :

* ~~OPEN CLASS WORDS~~:

Two broad categories of words are:

1) Open class words (Content words):
eg) nouns, verbs, adverbs, adjectives

2) Closed class words (functional words)
eg) Pronouns, preposition, conjunctions, auxilary verbs

- When words are grouped into similar classes, it is called as Parts of speech (POS).

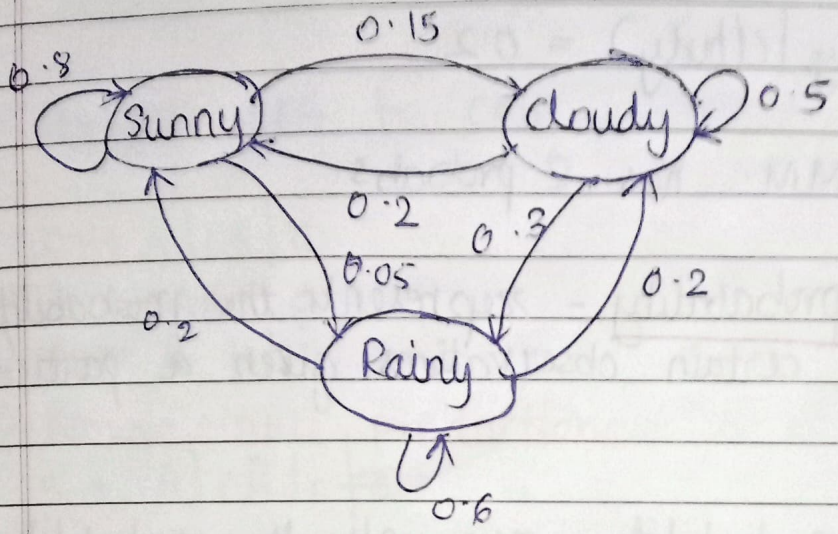
- Penn Treebank tags:

"The grand jury commented on a number of other topics".

The/DT grand/JJ jury/~~NOUN~~ NN commented/VBD on/IN a/DT number/NN of/IN other/JJ topics/NNP

* Hidden Markov Model

- It is a generative model which works on observation and hidden states
- Depends only on ~~current~~ and previous state.



P(sunny/sunny) = 0.8	P(sunny/cloudy) = 0.2
P(rainy/sunny) = 0.05	P(cloudy/cloudy) = 0.5
P(cloudy/sunny) = 0.15	P(rainy/cloudy) = 0.3

P(sunny/rainy) = 0.2
P(cloudy/rainy) = 0.2
P(rainy/rainy) = 0.6

Q - Given that today is sunny, what is the probability that tomorrow is sunny and the next day is rainy.

P(sunny/sunny) × P(rainy/sunny)
= 0.8 × 0.05
= 0.04

Q- Assume that yesterday was rainy, today is cloudy. What is the probability that tomorrow is sunny.

~~P(cloudy|rainy) =~~

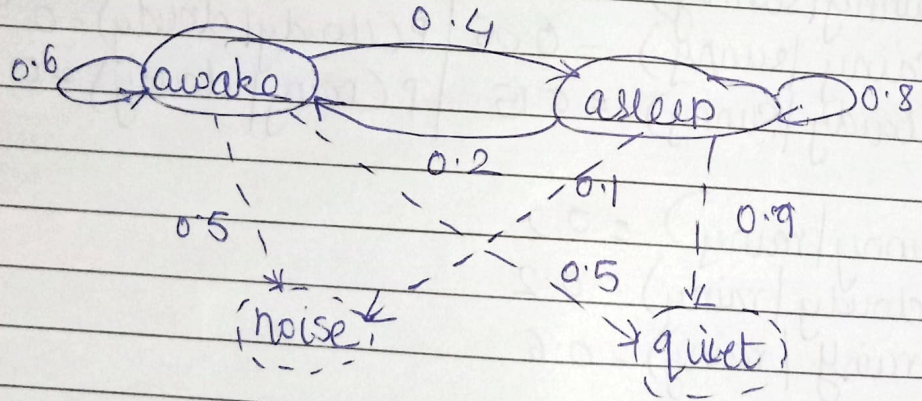
P(sunny | cloudy) = 0.2

- Hidden MM has 2 probabs.

1) Emission probability - represents the probability of making certain observations given a particular state

2) Transition probability - represents the probability of transition from a particular state to another state.

eg)



Transition matrix:

	awake	asleep	
awake	0.6	0.4	→ 1
asleep	0.2	0.8	→ 1

	noise	quiet
awake	0.5	0.5
asleep	0.1	0.9

Emission Probability Matrix

Q Convert CFG to CNF:

$S \to A \mid cB \mid c$
$B \to cAA \mid \epsilon$
$A \to Bc \mid b$

i) Remove null productions:

$S \to A \mid cB \mid c \mid \cancel{\epsilon}$
$B \to cAA$
$A \to Bc \mid b \mid c$

ii) Remove unit productions:

$S \to Bc \mid b \mid cB \mid c$
$B \to cAA$
$A \to Bc \mid b \mid c$

iii) Remove useless ~~pro~~ production
(none in this case)

iv) ~~Let~~ extra variables:

Let $X \to c$, $Y \to AA$
$S \to Bx \mid b \mid c \mid XB$
$B \to XY$
$A \to BX \mid b \mid c$

Q LMD for: $E \rightarrow E+E \mid E*E \mid id$

derive: id + id * id

$E \rightarrow E+E$

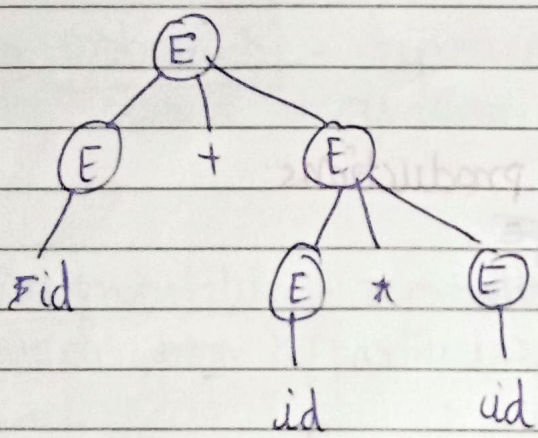
$\Rightarrow E+E*E$

$\Rightarrow id+E*E$

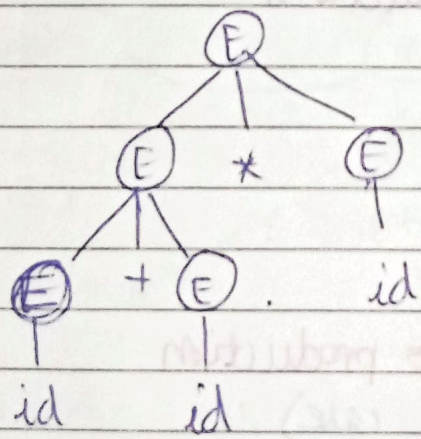
$\Rightarrow id+id*E$

$\Rightarrow id+id*id$

LMD1:



LMD2:



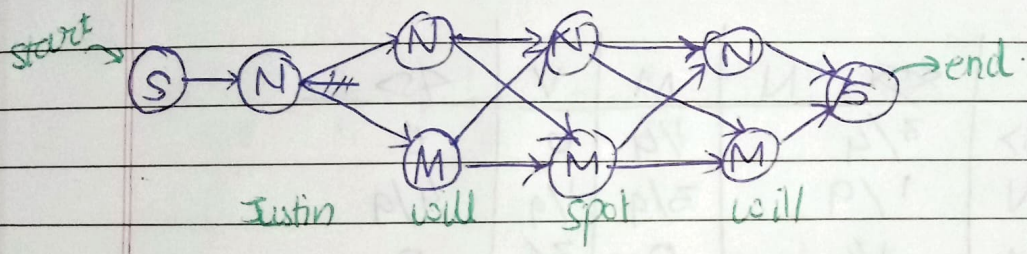
Q) Using HMM part of speech tagging, tag the following:

test data: <s> Justin will spot Will </s>

corpus:
<s> Martin, Justin can watch Will </s>
<s> Spop spot will watch Martin </s>
<s> Will Justin spot Martin </s>
<s> Martin will spot pat spot </s>

Step 1:

"Justin will spot Will".



Step 2:

<s>	Martin	Justin	can	watch	Will	</s>
	N	N	M	V	N	
<s>	spot	will	watch	Martin	</s>	
	N	M	V	N		
<s>	Will	Justin	spot	martin	</s>	
	M	N	V	N		
<s>	martin	will	pat	Spot	</s>	
	N	M	V	N		

Step 3: Emission Probability matrix.

$$\frac{P(word, class)}{P(class)}$$

CPT.O

	N	M	V
Martin	4/9	0	0
Justin	2/9	0	0
can	0	1/4	0
watch	0	0	2/4
will	1/9	3/4	0
spot	2/9	0	1/4
pat	0	0	1/4
Total	9	4	4

Step 4: Transition matrix.

word after class
class

	<S>	N	M	V	</S>
<S>	3/4	1/4	0	0	
N	1/9	3/9	1/9	4/9	
M	1/4	0	3/4	0	
V	4/4	0	0	0	

i) let Justin be Noun

$$P(N \mid \text{Justin}, <S>)$$
$$= P(\text{Justin} \mid N) \times P(N \mid <S>)$$
$$= \frac{2}{9} \times \frac{3}{4} = \frac{1}{6}$$

ii) let will be Noun
$P(\text{will} \mid N, <S> N)$
$= P(\text{will} \mid N) \times P(N \mid N)$

$= \frac{1}{9} \times \frac{1}{9} = \frac{1}{81} \times \frac{1}{6} = \frac{1}{486}$

Let will be ~~con~~ model verb.

P(will / model, noun)
$= P(will/m) \times P(M/N)$
$= \frac{3}{4} \times \frac{3}{9} = \frac{9}{36} =$

$= \frac{9}{36} \times \frac{1}{6} = \frac{1}{4} \quad \frac{9}{144}$
$\frac{1}{32}$

$\frac{1}{32} > \frac{1}{486}$ ∴ ~~will~~ = M

$= \frac{9}{36} = \frac{1}{4} \times \frac{1}{6} = \frac{1}{24}$

$\frac{1}{24} > \frac{1}{486}$ ∴ will is ~~mod~~ M.

iii) Let spot be Noun.

P(Spot / N, M)
$= P(Spot/N) \times P(N/M)$
$= \frac{2}{9} \times \frac{1}{4} = \frac{2}{36} = \frac{1}{18}$

$\frac{1}{18} \times \frac{1}{24} = \frac{1}{432}$

Spot as verb:

$P(\text{spot}|V, M)$

$= P(\text{spot}|V) \times P(V|M)$

$= \frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$

$\frac{3}{16} \times \frac{1}{24} = \frac{3}{384}$

$= \frac{1}{128}$

$\frac{1}{128} > \frac{1}{432}$ ∴ Spot is verb.

iv) Let will be Noun:

$P(\text{will}|\text{Noun}, V)$

$P(\text{will}|\text{Noun}) \times P(\text{Noun}|V)$

$= \frac{1}{9} \times 1 = \frac{1}{9}$

$\frac{1}{9} \times \frac{1}{128} = \frac{1}{1152}$

Let will be modelt verb (M):

$P(\text{will}|M, V)$

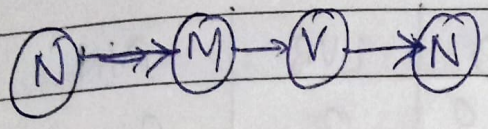
$= P(\text{will}|M) * P(M, V)$

$= \frac{3}{4} \times 0$

$= 0$

∴ will is a Noun

: Justin will Spot Will

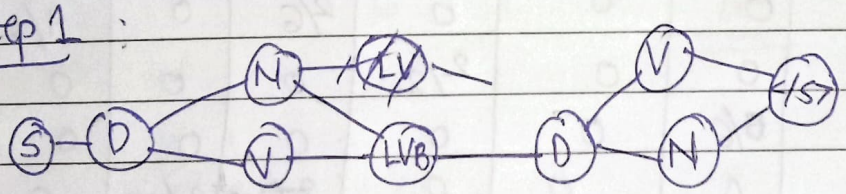


2) Generate the state transition and emission probability matrix by applying the Pos tagging using HMM:

i) Book a car.
ii) Park the car.
iii) The book is in the car.
iv) The car is in a park.

test data => The park is a book </s>

Step 1:



step2:

<s>	Book	a	car	</s>	
	V	D	N		
<s>	Park	the	car	</s>	
	V	D	N		
<s>	The	book	is	in	the car </s>
	D	N	LVB	IN	D N
<s>	The	car	is	in	a park </s>
	D	N	LVB	IN	D N

Step 3: Emission probability matrix:

B	N	V	D	LVB	IN
Book	1/6	1/2	0	0	0
a	0	0	2/6	0	0
car	4/6	0	0	0	0
Park	1/6	1/2	0	0	0
the	0	0	4/6	0	0
is	0	0	0	2/2	0
in	0	0	0	0	2/2

Step 4: Transition matrix

	N	V	D	LVB	IN	</S>
<S>	0	2/4	2/4	0	0	0
N	0	0	0	2/6	0	4/6
V	0	0	2/2	0	0	0
D	5/6	0	0	0	0	0
LVB	0	0	0	0	2/2	0
IN	0	0	2/2	0	0	0

Step 5:

i) Let The be a determinant:

$$P(D|The, <S>) = P(The|D) \times P(D|<S>)$$

$$= \frac{4}{6} \times \frac{2}{4}$$

$$= \frac{1}{3}$$

ii) Assume Park as verb.

$$P(\text{Park}|V, D) = P(\text{Park}|V) * P(V|D)$$
$$= \frac{1}{2} \times \frac{1}{1000}$$
$$= \frac{1}{2000}$$

$$\frac{1}{2000} \times \frac{1}{3} = \frac{1}{6000}$$

Assume Park as Noun:

$$P(\text{Park}|N, D) = P(\text{Park}|N) \times P(N|D)$$
$$= \frac{1}{6} \times \frac{1}{1} \times \frac{1}{3} = \frac{1}{18}$$

$\because \frac{1}{18} > \frac{1}{6000}$

$\therefore$ Park is a Noun.

iii) Let is be a linking verb.

$$P(\text{is}|\text{LVB}, N) = P(\text{is}|\text{LVB}) * P(\text{LVB}|N)$$
$$= \frac{2}{2} \times \frac{2}{6} \times \frac{1}{18}$$
$$= \frac{1}{54}$$

iv) Let 'a' be a determinant:

$$P(a|D, LVB) = P(a|D) \times P(D, LVB)$$
$$= \frac{2}{6} \times \frac{1}{1000} \times \frac{1}{54}$$
$$= \frac{1}{162000}$$

v) Let 'book' be a verb:

$$P(book|V, D) = P(book|V) \times P(V, D)$$
$$= \frac{1}{2} \times \frac{1}{1000} \times \frac{1}{162000}$$
$$= 3.086 \times 10^{-9}$$

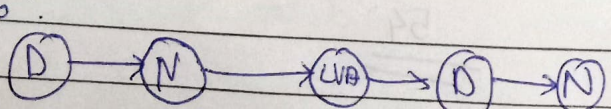
Let 'book' be a noun

$$P(book|N, D) = P(book|N) \times P(N|D)$$
$$= \frac{1}{6} \times \frac{6}{6} \times \frac{1}{162000}$$
$$= \frac{1}{972000}$$

$$\therefore \frac{1}{972000} > 3.086 \times 10^{-6}$$

∴ book is a noun

Step 6:



* Probabilistic Context free Grammar:

1)

$S \to AB$	0.5
$S \to BC$	0.5
$A \to BA$	0.3
$A \to a$	0.7
$B \to CC$	0.4
$B \to b$	0.6
$C \to AB$	0.2
$C \to a$	0.8

Calculate the probability of deriving the string "aaab" from the given grammar.

i.e PCFG (aaab)

LMD

$S \to AB$ 0.5
$\Rightarrow aB$ 0.7
$\Rightarrow aCC$ 0.4
$\Rightarrow aaC$ 0.8
$\Rightarrow aaAB$ 0.2
$\Rightarrow aaaB$ 0.7
$\Rightarrow aaab$ 0.6

$P_1(aaab) = 0.5 \times 0.7 \times 0.4 \times 0.8 \times 0.2 \times 0.7 \times 0.6$
$= 0.0094$

RMD

$S \to AB$ 0.5
$\Rightarrow ACC$ 0.4
$\Rightarrow ACAB$ 0.2
$\Rightarrow ACAb$ 0.6
$\Rightarrow ACaab$ 0.7
$\Rightarrow Aaaab$ 0.8
$\Rightarrow aaab$ 0.7

$S \rightarrow AB \quad 0.5$

$\Rightarrow Ab \quad 0.6$

$\Rightarrow BAb \quad 0.3$

$\Rightarrow Bab \quad 0.7$

$\Rightarrow CCab \quad 0.4$

$\rightarrow Caab \quad 0.8$

$\Rightarrow aaab \quad 0.8$

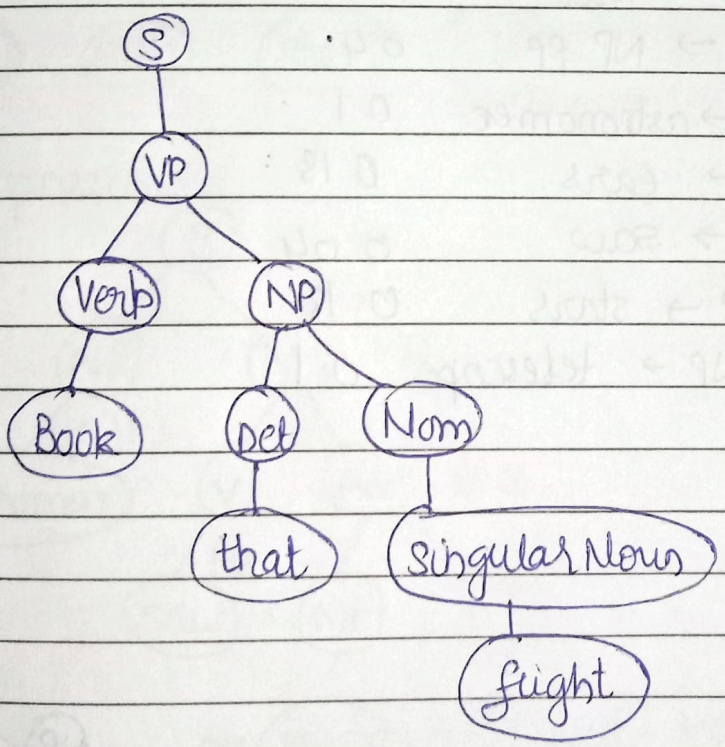
$P_2(aaab) = 0.016$

7. Design the parse tree for:

"Book that flight" using the following:

S → VP
VP → verb NP
NP → Det ~~Noun~~ NOM
ND → Det Noun
Det → that
NOM → singular Noun
verb → ~~Bok~~ Book
singular Noun → flight

Parse tree



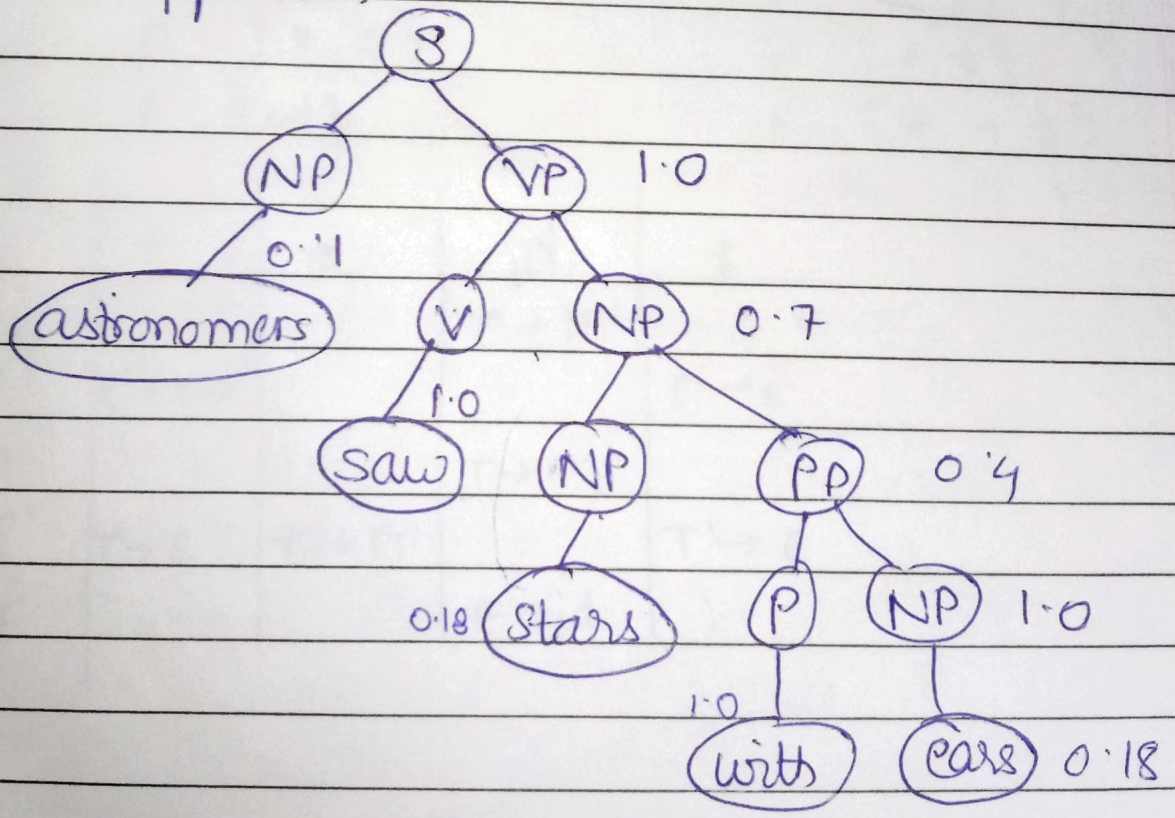
3) Design the parse tree for foll.

"Astronomers saw stars with ears."

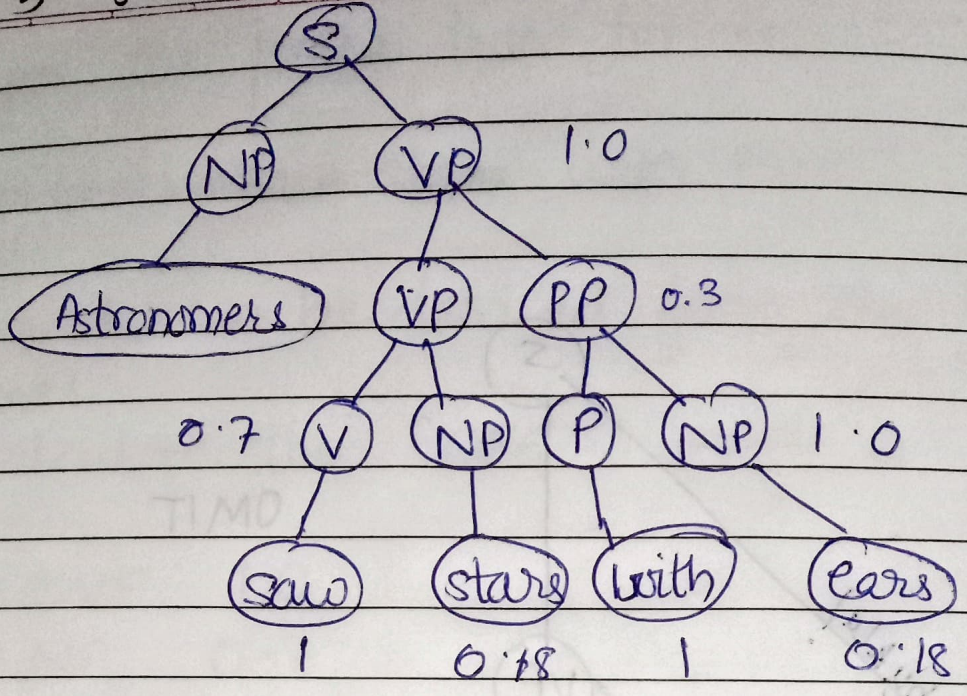
Provided it can be derived from the foll grammer.
Also calculate the PCFG.

S → NP VP	1.0
VP → V NP	0.7
VP → VP PP	0.3
PP → P NP	1.0
P → with	1.0
V → saw	1.0
NP → NP PP	0.4
NP → astronomers	0.1
NP → ears	0.18
NP → saw	0.04
NP → stars	0.18
NP → telescope	0.1

Top down approach



P(~~Astronomers~~ = 1.0 x 0.1 x 0.7 x 1 x 0.4 x 0.18 x
1.0 x 1.0 x 0.18
= 0.0009072



$P_2 = 1.0 \times 0.1 \times 0.7 \times 1 \times 1 \times 1 \times 0.18 \times 0.18$

$= 0.0006804$

PCFG $= 0.00090072 + 0.0006804$

$= 0.00158876$

* Predictive Parser:

$E \rightarrow E+T \mid T$
$T \rightarrow T*F \mid F$
$F \rightarrow id$

→ Remove left recursion

i) $E \rightarrow TE'$
$E' \rightarrow +TE' \mid \epsilon$
$T \rightarrow FT'$
$T' \rightarrow *FT' \mid \epsilon$
$F \rightarrow id$

ii) First and follow:

First: $E : \{id\}$
$E' : \{+, \epsilon\}$
$T : \{id\}$
$T' : \{*, \epsilon\}$
$F : \{id\}$

Follow: $E : \{\$\}$
$E' : \{\$\}$
$T : \{+, \$\}$
$T' : \{+, \$\}$
$F : \{*, +, \$\}$

	+	*	id	$
E			$E \rightarrow TE'$	
E'	$E' \rightarrow +TE'$			$E' \rightarrow \epsilon$
T			$T \rightarrow FT'$	
T'	$T' \rightarrow \epsilon$	$T' \rightarrow *FT'$		$T' \rightarrow \epsilon$
F			$F \rightarrow id$	

Stack	Input	Production
$E	id + id * id $	
$E'T	id + id * id $	
$E'T'F	id + id * id $	
$E'T'id	id + id * id $	
$E'T'	+ id * id $	
$E'	+ id * id $	
$E'T+	+ id * id $	
$E'T	id * id $	
$E'T'F	id * id $	
$E'T'id	id * id $	
$E'T'	* id $	
$E'T'F*	* id $	
$E'T'F	id $	
$E'T'id	id $	
$E'T'	$	
$E'	$	
$	$	

* CKY Algorithm:

Design a CKY parser for accepting the string "A pilot likes flying planes"

S → NP VP
VP → VBG NNS
VP → VBZ VP
NP → DT NN
NP → JJ NNS
DT → a
NN → pilot

VBZ → likes
VBG → flying
JJ → flying
NNS → planes

a	pilot	likes	flying	planes.
DT	NP	-		S
x_{01}	x_{02}	x_{03}	x_{04}	x_{05}
	NN	-	-	-
	x_{12}	x_{13}	x_{14}	x_{15}
		VBZ	-	VP
		x_{23}	x_{24}	x_{25}
			VBG, JJ	VP, NP
			x_{34}	x_{35}
				NNS
			x_{44}	x_{45}

02:

x_{00} x_{01}

01 11

02 21

03 31

04 41

05 51

13:

x_{10} x_{03}

x_{11} x_{13}

x_{12} x_{23}

x_{13} x_{33}

x_{14} x_{43}

x_{15} x_{53}

03:

00 03

01 13

02 23

03 33

04 43

05 53

24:

20 04

21 14

22 24