Vansh Ahuja C33 04

THADOMAL SHAHANI TSEC ENGINEERING COLLEGE

NLP WRITTEN ASSIGNMENT

Q.1] Using the following CFG, parse the sentence "Book the flight".

S → NP/VP
S → VP
NP → Det / N
NP → Det / Adj / N
VP → V
VP → V / NP

1. Top-down parsing

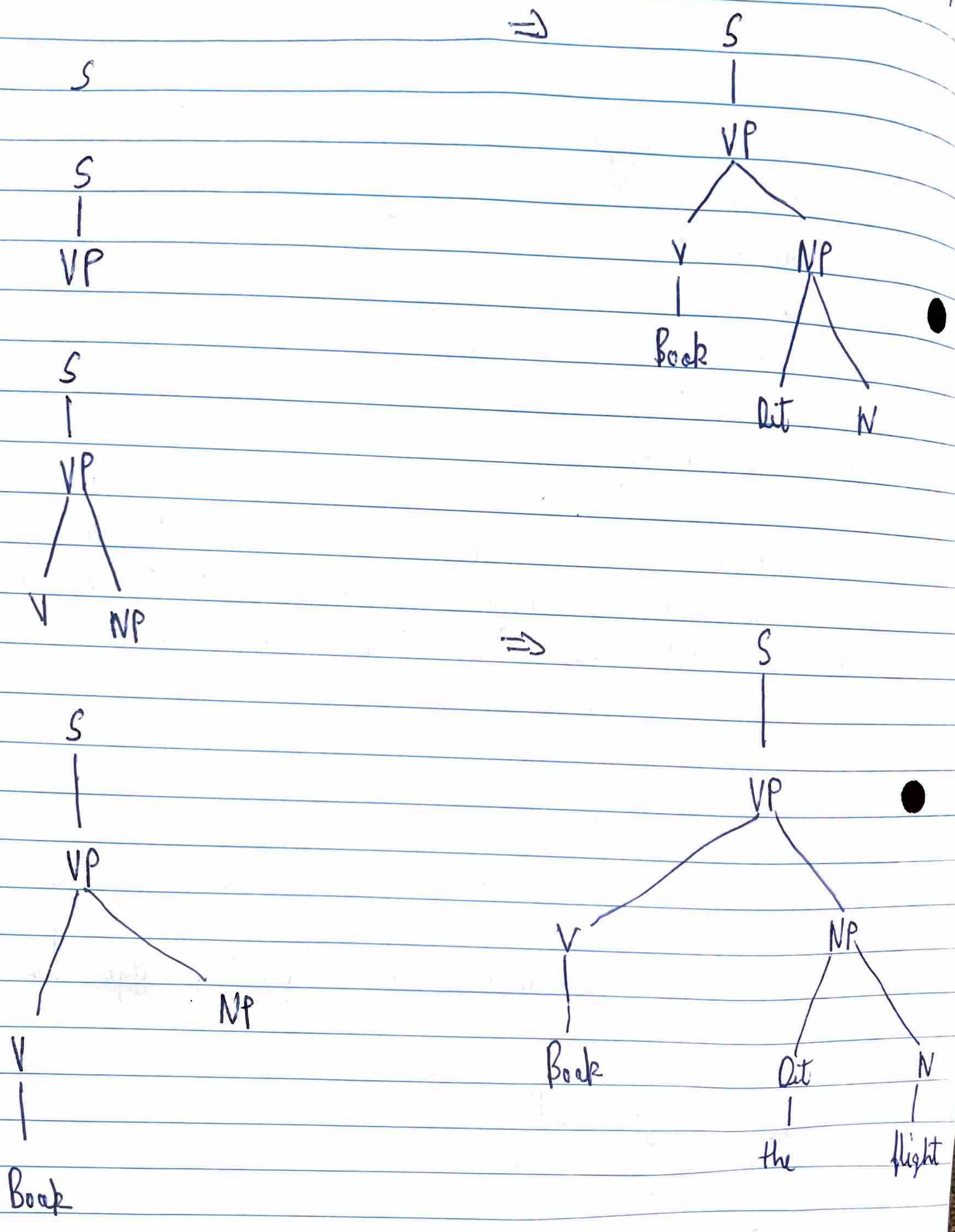
A top down is a goal-oriented, parse start root and then with a list of constituents to be built. It rewrites the goal in goal list by matching one against the LHS of the grammar rules.

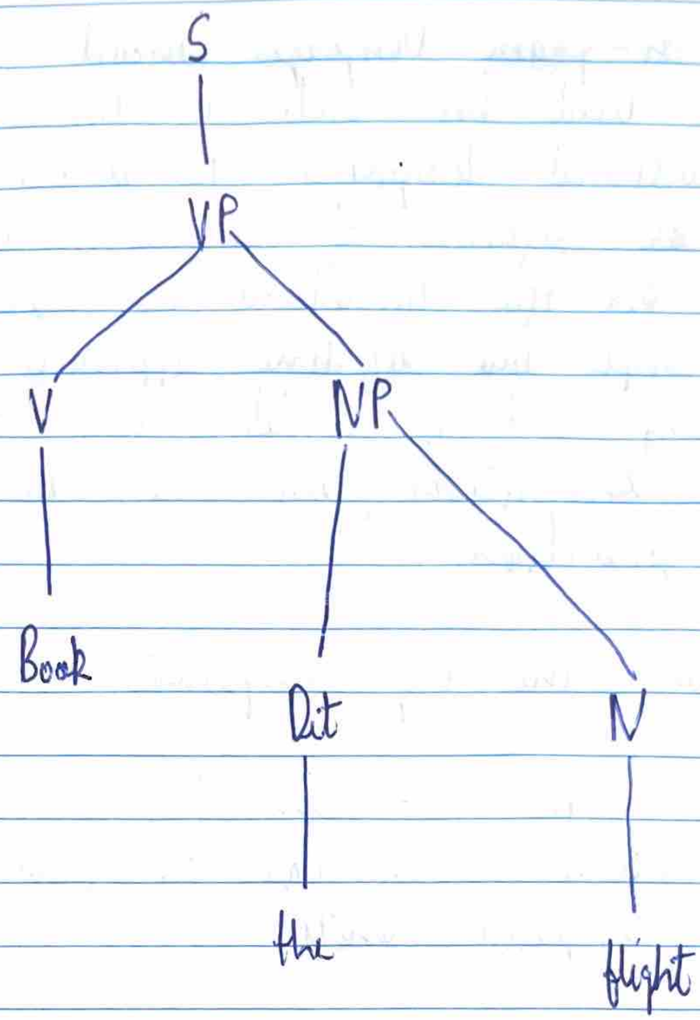
2. Bottom-up parsing

Bottom up is data-directed. The initial goal list of a bottom up parse is the string to be parse. Bottom-up parsing is earliest known parsing algorithm.

A

Lets parse "Book the flight" using top-down parsing





Q.2] Explain the n-gram language model. Given following corpus

<s> I am henry </s>
<s> I like college </s>
<s> Do henry like college </s>
<s> henry I am </s>
<s> Do i like college </s>

Ans An n-gram language model is a statistical model used in natural language processing and computational linguistics to predict the probability of word or sequence of words occurring in a text based on the conditional probabilities of n-1 words. Its simple but effective approach for language modelling and is widely used in tasks such as speech recognition, machine translation and text generation

Here are the key components

1. N-gram : An n-gram is a contiguous sequence of n-items (which can be words, characters or other linguistic units

2. N- gram probability : The n- gram calculates the probability of an n-gram based on the conditional probability of last n-1 words

3. Laplace smoothing : Laplace smoothing in diagram involves adding a small constant to the frequency of each ~~&~~ unique biagram to address problem of zero probability for unseen combinations

<s> I am Henry </s>
<s> I like college </s>
<s> Do henry like college </s>
<s> Henry I am </s>
<s> Do I like henry </s>
<s> Do I like college </s>
<s> I do like henry </s>

Bigram probabilities

P(I|<s>) = count (<s>, I) / count (<s>) = 3/7 = 0.4
P(am|I) = ~~H~~ 1/6 = 0.1 P(I|am) = 0.1
P(Henry|am) = 1/5 = 0.2 P(am|henry) = 0.2
P(</s>|Henry) = 3/7 = 0.4 P(I|like|</s>) = 0.4
P(like|I) = 3/5 = 0.6 P(I|like) = 0.6
P(college|like) = 3/5 = 0.6 P(like|college) = 0.1
P(</s>|college) = 3/7 = 0.4 P(college|</s>) = 0.4
P(~~I~~ Henry|<s>) = 1/5 = 0.2 P(<s>|henry) = 0.2
P(I|Henry) = 1/6 = 0.1 P(henry|I) = 0.1
P(</s>|am) = 1/7 = 0.14 ~~P(<s>~~ P(am|</s>) = 0.14
P(Do|<s>) = 3/7 = 0.4 P(<s>|Do) = 0.4
P(henry|like) = 1/4 = 0.25 P(like|henry) = 0.25
P(</s>|henry) = 3/7 = 0.4 P(henry|</s>) = 0.4
P(like|(do)) = 1/4 = 0.25 P(do|like) = 0.25
P(do|I) = 1/6 = 0.1 P(I|do) = 0.1
P(I|do) = 2/4 = 0.5 P(do|I) = 0.5

1) <I like college>

$P(I \mid \langle s \rangle) \cdot P(like \mid I) \cdot P(college \mid like) \cdot P(\langle /s \rangle \mid college)$

$= \frac{3}{7} \cdot \frac{3}{5} \cdot \frac{3}{3} \cdot \frac{3}{7} \cdot \frac{3}{5}$

$= 0.096 \times 0.6 = 0.05$

<s> do I like Henry </s>

$P(do \mid \langle s \rangle) \cdot P(I \mid do) \cdot P(like \mid I) \cdot P(Henry \mid like) \cdot$
$P(\langle /s \rangle \mid Henry)$

$= 0.4 \times 0.5 \times 0.6 \times 0.2 \times 0.4$
$= 0.0096$

2) Next probable word for "I Like" can be founding by finding max probability

$P(college \mid like) = 0.6$

$P(Henry \mid like) = 0.25$

Next probable word is "college"

For applying Laplace smoothing

$$P^*_{laplace}(W_n \mid W_{n-1}) = \frac{C(W_{n-1} W_n) + 1}{C(W_{n-1}) + V}$$

V = unique words

1) <s> I like college </s>

$P(\langle s\rangle | I) \cdot P(I | like) \; P(like | college) \cdot P(college | \langle /s\rangle)$

$\frac{3+1}{7+6} \cdot \frac{3+1}{5+6} \cdot \frac{3+1}{3+6} \cdot \frac{3+1}{7+6}$

$= 0.015$

2) <s> do I like honey </s>

$P(\langle s\rangle | do) \cdot P(do | I) \cdot P(I | like) \cdot P(like | honey)$
$P(honey)(\langle /s\rangle)$

$= \frac{3+1}{7+6} \cdot \frac{2+1}{4+6} \cdot \frac{3+1}{5+6} \cdot \frac{1+1}{4+6} \cdot \frac{3+1}{7+1}$

$= 0.0020$

Q.3] List all bi-grams and compute first 5 bigram probabilities for sentence given below

<s> the student pass the test </s>
<s> the student went for the pass </s>
<s> the ~~student~~ teacher test student </s>

Ans <s> the student pass the test </s>
<s> the student wait for the pass </s>
<s> the teacher test student. </s>

Bigrams are:
(the|<s>, (the, student) (student, pass), (pass, the)
(the, test), (test, </s>) (<s>, the), (the, student),
(student, wait), (wait, for), (for, the), (the, pass),
(pass, </s>), (<s>, teacher), (teacher, test), (test, student),
(student, </s>)

P(the|student) = P(student|the) | P(student)
= 0

P(the|<s>) = P(<s>|the) | P(<s>)
= 2/3
=

P(student|the) = P(the| student) | P(the)
= 2/4

P(</s>|test) = P(test|</s> | P(test)
= 1/2
= 0.5

spelling correction using N-gram is a simple and effective approach to address misspelled words in a text N-gram are used to identify and suggest corrections for misspelled words by comparing them to correctly spelled words in a language model.

1. Building an n-gram model:

First, we need to arrange a corpus of text with correctly spelled words to build an N-gram language model.

2. Tokenization into N-gram

The text in the corpus is tokenized into N-gram typically biagram or trigrams.

3. Counting N-gram: for each N-gram, you count the number of times it appears in corpus

4. Spelling correction: To correct a misspelled word you tokenize the misspelled words into N-grams you compare these N-grams to N-grams in your language model you calculate the likelihood probability of misspelled n-gram for each n-gram from misspelled, you consider possible correct alternatives.

Q.4] The following POS tagged corpus is given

<s> the /DT students /NN pass /V test /NN </s>
<s> the /DT students /NN wait /Y for /P pass /NN </s>
<s> teacher /N test /Y students /NN </s>

Write down transition and emission probabilities

Ans & Emission probabilities

Words	Noun	Verb	Det	Preposition
The	0	0	4	0
Student	3	0	0	0
Pass	0	1	0	0
test	1	1	0	0
wait	0	1	0	0
for	0	0	0	1
student	1	0	0	0

Probabilities

Words	Noun	Verb	Det	Preposition
The	0	0	1	0
Student	3/5 = 0.6	0	0	0
Pass	0	0.3	0	0
Test	0.2	0.3	0	0
wait	0	0.3	0	0
for	0	0	0	1
Student	0.2	0	0	0

Transition Probabilities

	N	V	D	P	</S>
<S>	1	0	2	0	0
N	0	3	0	0	3
V	1	0	1	1	0
D	4	0	0	0	0
P	0	0	1	0	0

Probabilities

	N	V	D	P	</S>
<S>	0.3	0	0.6	0	0
N	0	0.5	0	0	0.5
V	0.3	0	0.3	0.3	0
D	1	0	0	0	0
P	0	0	1	0	0

Q.5] For, each sentence, identify whether the different meaning arise from Structural ambiguity, semantic ambiguity or pragmatic ambigity.

a) Time flies like an arrow

b) He crushed the key to my heart

Ans Here in our example

In the two sentences, different meaning arise from different type of ambiguity.

a) Time flies like arrow

The ambiguity in this sentence is an example of structural ambiguity. It can be represented in two different way based on structural grouping of words.

1) "Time-flies": (i.e insects known as flies that exist over time) are similar to an arrow. In this interpretation, "like an arrow" is a simple indicating a comparison between time flies and arrow.

2) "Time" (the concept of time") passes quickly or moves swiftly in a manner similar to an arrow, like an arrow" is simile indicating comparision between speed of time and arrow.

b. "He crushed the key to my heart"

1) Semantic ambiguity: The word "crushed" is the source of semantic ambiguity. It can b interpreted in two different ways.

- Literally as in physically crushing into piece
- Figuratively, as in having strong emotions impact

2) Pragmatic ambiguity. The phrase "to my heart" introduces some pramatic ambiguity.

- Literally physical location (unlikely)
- Figurative emotional impact: (crushed) it is in emotional sense

Q.6] Identify the morphological type (Noun phrase, Verb phrase, Adjective phrase)
a) important to Bill
b) looked up the tree

Ans. a) Important to Bill

- Morphological type: Adjective phrase

The phrase "important to the bill" functions as an adjective modifying as a noun. It describes the quality of something in relation to "Bill"

b) "looked up there"

- Morphological Type: Verb phrase

The phrase "looked up the tree" functions as a verb in past list and describes an actions. It include the main verb "looked" and a prepositional phrase "up the tree" where "up" is a preposition and tree is the object of the preposition. This structure constitutes a verb phrase

Q.7] For given corpus

<S> Martin Justin can watch will <E>
<S> Pat will watch martin <E>
<S> Will Justin spat martin <E>
<S> Martin will pat spat <E>

N : noun [martin Justin, Will, Spat, Pat]
M : Model [can, will]
V : Verb [watch, spat, pat]

Create transition and emission probability matrix apply HMM and do POS of

"Justin will spat Will"

Ans <S> Martin/N , Justin/N, Can/M , Watch/V
Will/N <E>

<S> Spat/N will/M , watch/V , martin/N
<E>

<S> Will/N Justin/N Spat/V martin/N
</S>

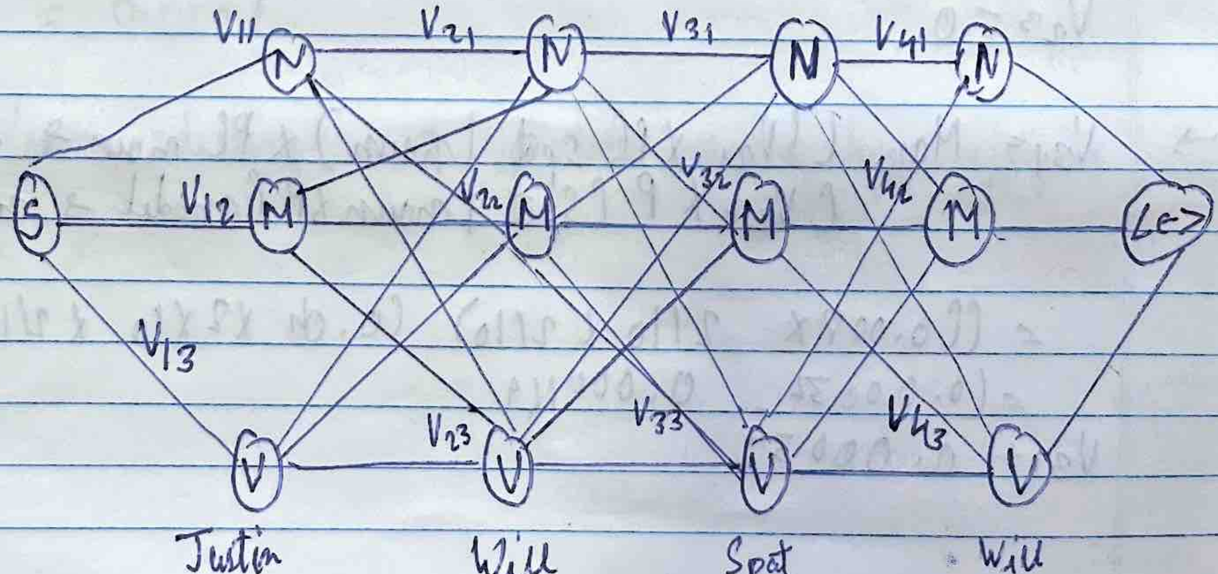
<S> Martin/W Will/M pat W spat/W <E>

Emission probability

Words	Noun	Model	Verb
Martin	4/10	0	0
Justin	2/10	0	0
can	0	1/3	0
watch	0	0	2/4
will	2/10	2/3	0
spat	2/10	0	1/4
pat	0	0	~~1~~ 1/4

Transition probability

	N	M	V	<S>
<S>	3/4	0/4	0	0
N	2/10	3/10	1/10	4/10
M	0	0	3/3	0
V	4/4	0	0	0



⇒ $V_{11} = P(\text{Tuition} \mid N) \times P(S \to N)$

$= \frac{2}{10} \times 1$

$= \frac{2}{10}$

$V_{12} = P(\text{Tuition} \mid M) \times P(S \to M)$

$= 0$

$V_{13} = 0$

⇒ $V_{21} = V_{11} \times P(\text{Will} \mid \text{noun}) \times P(\text{noun} \to \text{noun})$

$= \frac{2}{10} \times \frac{2}{10} \times \frac{2}{10}$

$= 0.008$

$V_{22} = V_{11} \times P(\text{Will} \mid \text{model}) \times P(N \to M)$

$= \frac{2}{10} \times \frac{2}{3} \times \frac{3}{10}$

$= 0.04$

$V_{23} = 0$

→ $V_{31} = \text{Max}\,((V_{21} \times P(\text{spot} \mid \text{noun}) \times P(\text{noun} \to \text{noun})$

$(V_{22} \times P(\text{Spot} \mid \text{noun} \times P(\text{model} \to \text{noun}))$

$= ((0.008 \times 2/10 \times 2/10)\ (0.04 \times 2 \times 10 \times 2/10)$

$= (0.00032,\ 0.000169$

$V_{31} = 0.00032$

$V_{32} = 0$

$V_{33} = \text{Max}((V_{21} \times P(\text{spot} \mid \text{Verb}) P(n \to V)) \; (V_{22} \times P(\text{spot} \mid \text{Verb}) \times P(\text{Model} \to \text{verb}))$

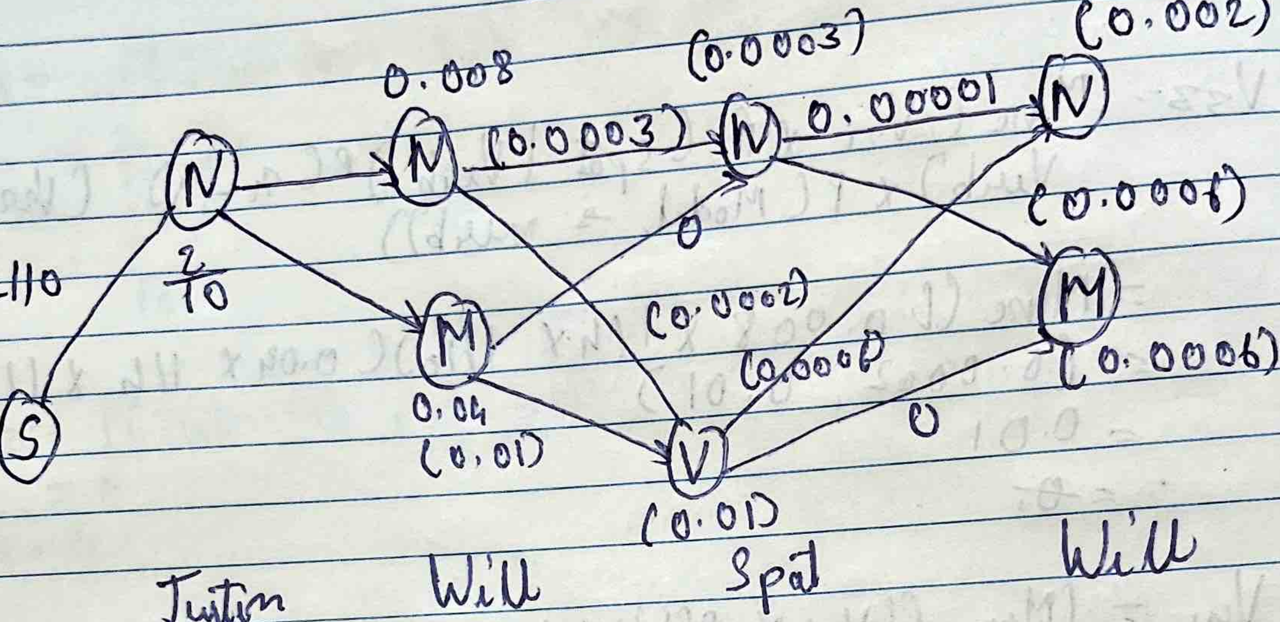
$= \text{Max}((0.008 \times 1/4 \times 1/10)(0.04 \times 1/4 \times 1/10))$
$= (0.0002, 0.01)$
$= 0.01$
$= \cancel{0.}$

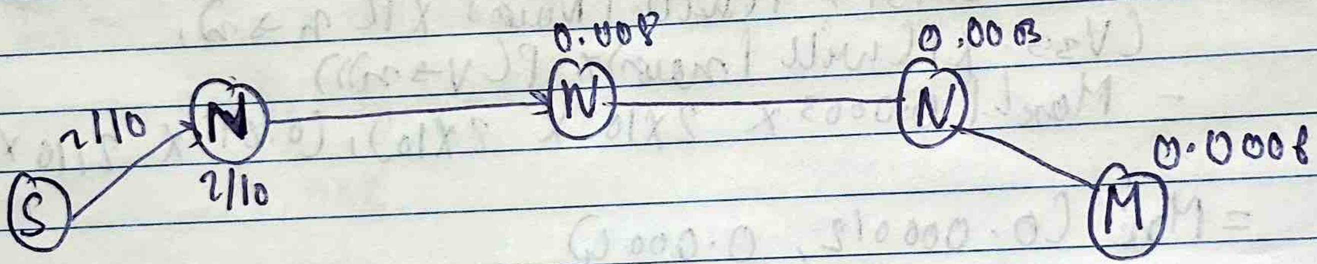
$\Rightarrow V_{41} = (\text{Max}((V_{31} \times P(\text{Will} \mid \text{Noun}) \times P(n \to n),$
$(V_{33} \times P(\text{will} \mid \text{noun}) \times P(V \to n)))$
$= \text{Max}((0.0003 \times 2 \times 10 \times 2 \times 10), (0.01 \times 2/10 \times 1/2.5))$

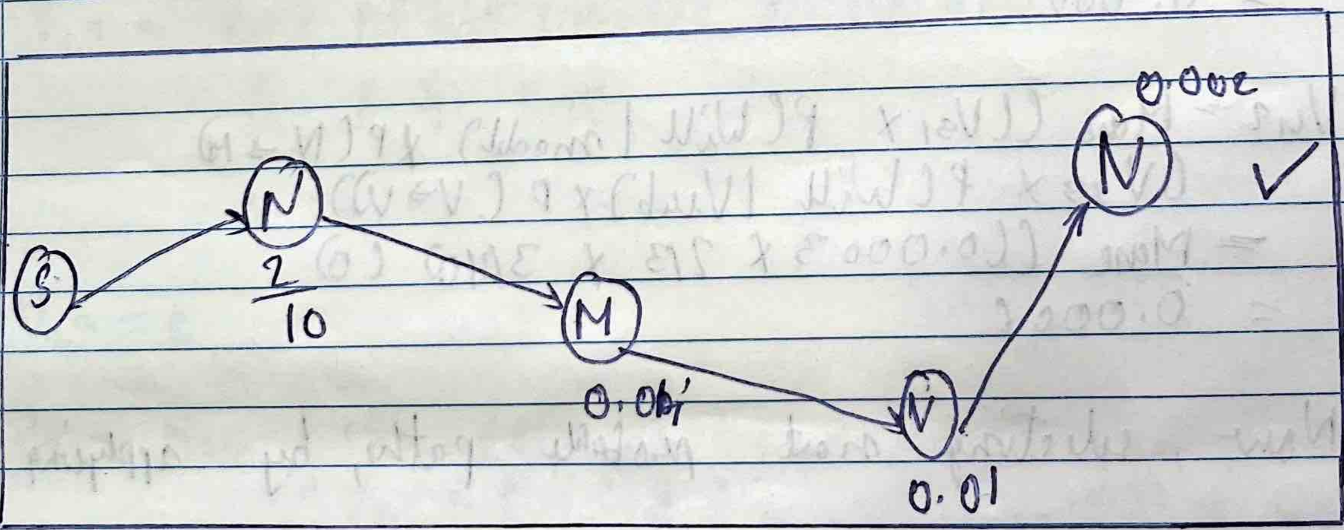
$= \text{Max}(0.00012, 0.0004)$
$= 0.002$

$V_{42} = \text{Max}((V_{31} \times P(\text{Will} \mid \text{model}) \times P(N \to M)$
$(V_{33} \times P(\text{Will} \mid \text{Verb}) \times P(V \to V))$
$= \text{Max}((0.0003 \times 2/3 \times 3/10)(0))$
$= 0.0006$

Now, selecting most probable paths, by applying viterbi







<S> Tuition / N Will / Model spot / V marry / N <E>
final probability = 0.0008

Q.8] Consider the following corpus

<s> I tell you to sit and sleep </s>

<s> I would like to sleep for an hour </s>

<s> sleep helps one to relax </s>

Ans P(<s>|I) = count(<s>|I) P(<s>) = 2/3 =

P(I|tell) = 0.5

P(tell|you) = 1

P(you|to) = 1

P(to|sleep) = 0.66

P(sleep|and) = 0.33

P(and|sit) = 1

P(sit|</s>) = 1

P(I|would) = 0.5

P(would|like) = 1

P(like|to) = 1

P(sleep|for) = 0.33

P(for|an) = 1

P(an|hour) = 1

P(hour|</s>) = 1

P(<s>|sleep) = 0.33

P(sleep|helps) = 0.33

P(one|helps) = 1/1 = 1

P(to|one) = 1/1 = 1

P(relax|to) = 1/3 = 0.33

P(</s>| relax = 1/1 = 1

Next ~~round~~ word for to can be chosen from

P(sleep | to) = 0.66
P(relax | to) = 0.33

The word is 'sleep' as probability is 0.66

Q.9] Using the given training corpus, Identify tags for sentence "Can justin watch Martin" using HMM

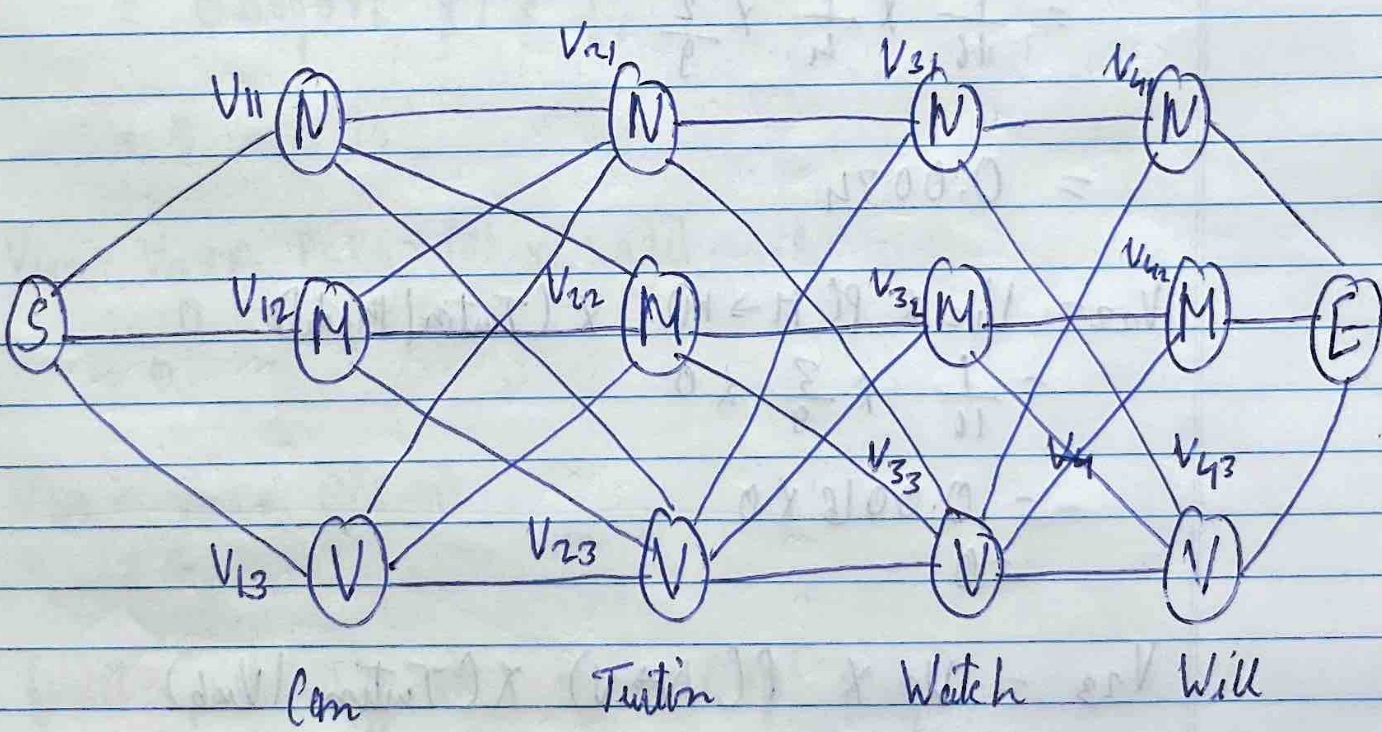
<S> Martin justin can watch will <E>
<S> Spot will watch Martin <E>
<S> Will justin Spot Martin <E>
<S> Martin will pat spot <E>

Ans → Emission probability

Word	noun	verb	Model
Martin	4/9	0	0
Justin	2/9	0	0
Can	0	0	1/4
Watch	0	2/4	0
Will	1/9	0	3/4
Spot	2/9	1/4	0
pat	0	1/4	0

Transition probability

	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



1) $S \to N$

$V_{11} = P(S \to N) \times P(\text{Can} \mid \text{Noun})$

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

$= 0$

$V_{12} = P(S \to M) \times P(\text{Can} | \text{Modal})$

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

$= 0.062$

$V_{13} = P(S \to V) \times P(\text{Can} | \text{Verb})$

$= 0$

2) $V_{21} = V_{12} \times P(M \to N) \times (\text{Tuition} | \text{noun})$

$= \frac{1}{16} \times \frac{1}{4} \times \frac{2}{9}$

$= 0.0034$

$V_{22} = V_{12} \times P(M \to M) \times (\text{Tuition} | \text{Modal})$

$= \frac{1}{16} \times \frac{3}{9} \times 0$

$= 0.0015 \times 0$

$= 0$

$V_{23} = V_{12} \times P(N \to V) \times (\text{Tuition} | \text{Verb})$

$= 0$

$V_{31} = V_{22} \times P(N \to N) \times P(\text{Watch} | \text{noun})$

$= 0.0035 \times \frac{1}{9} \times 0$

$= 0$

$V_{32} = V_{22} \times P(N \to M) \times P(\text{Watch} \mid \text{model})$
$= 0$

$V_{33} = V_{22} \times P(N \to V) \times P(\text{Watch} \mid \text{Verb})$
$= 0.0035 \times \frac{1}{9} \times \frac{2}{4}$

$= 0.000086$

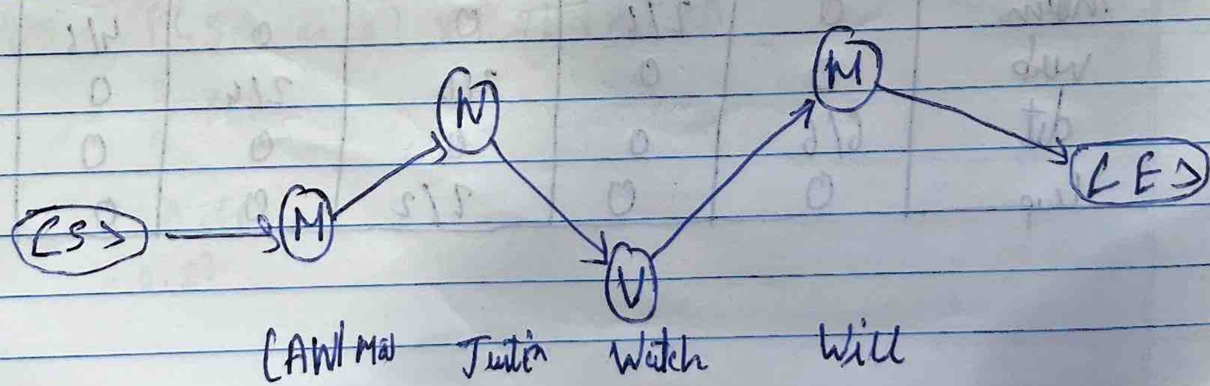
$V_{41} = V_{33} \times P(V \to N) \times (\text{Will} \mid \text{noun})$
$= 0.000086 \times 1 \times \frac{1}{9}$

$= 9.55 \times 10^{-6}$

$V_{42} = V_{33} \times P(V \to M) \times (\text{Will} \mid \text{model})$
$= 0.00086 \times 0$
$= 0$

$V_{43} = V_{33} \times P(V \to V)$
$= 0$

final probability is $9.55 \times 10^{-6} \times \frac{4}{2}$
$= 4.24 \times 10^{-6}$



Q. 10) Using the training corpus, Identify the tag sentence "The park is a book".

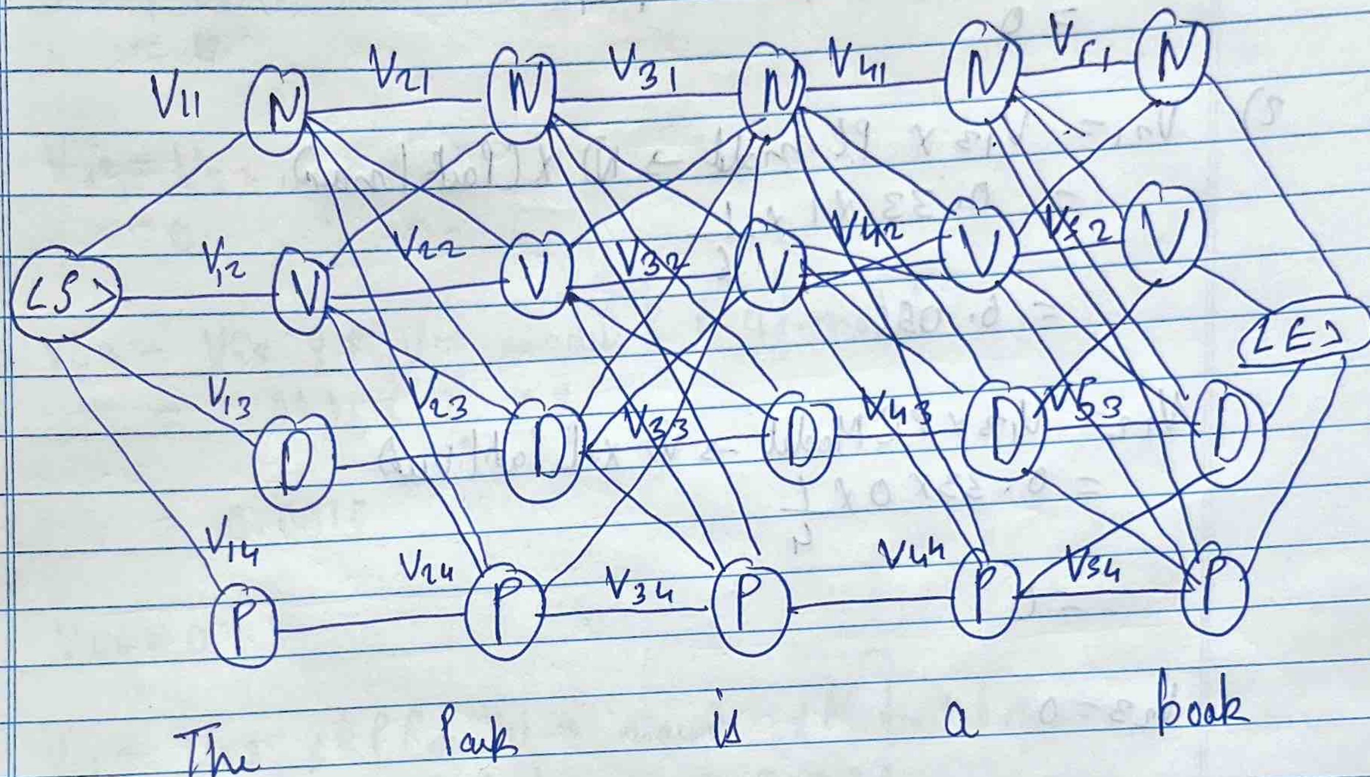
<S> Book a car </S>
<S> Park the car </S>
<S> The book is in the car </S>
<S> The car is in a park </S>

→ Emission probability matrix

	Noun	Verb	Model	Prep
Book	1/6	1/4	0	0
a	0	0	2/6	0
Car	4/6	0	0	0
the	0	0	4/6	0
Park	1/6	1/4	0	0
is	0	2/4	0	0
in	0	0	0	2/2

Transition

	noun	verb	model	prep	</S>
<S>	0	2/4	2/4	0	0
noun	0	2/6	0	0	4/6
verb	0	0	2/4	2/4	0
det	6/6	0	0	0	0
prep	0	0	2/2	0	0



1) $V_{11} = S \to N$

$(S \to N) \times P(\text{The} \mid \text{noun})$

$= 0$

$V_{12} = P(S \to V) \times P(\text{The} \mid \text{Verb})$

$= 0$

$V_{13} = P(S \to \text{model}) \times P(\text{The} \mid \text{model})$

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

$= 0.66$

$= 0.33$

$V_{14} = P(S \to pbup) \times P(The \mid pbup)$
$= 0$

2) $V_{21} = V_{13} \times P(Model \to N) \times (Park \mid noun)$
$= 0.33 \times 1 \times \frac{1}{6}$
$= 0.055$

$V_{22} = V_{13} \times P(Model \to V) \times (Park \mid Verb)$
$= 0.33 \times 0 \times \frac{1}{4}$
$= 0$

$V_{23} = 0$

$V_{24} = 0$

3) $V_{31} = V_{22} \times P(N \to N) \times P(is \mid noun)$
$= 0.055 \times 0$
$= 0$

$V_{32} = V_{21} \times P(N \to V) \times P(is \mid verb)$
$= 0.055 \times \frac{2}{7} \times \frac{2}{4}$
$= 0.0091$

$V_{33} = 0$

$V_{34} = 0$

4) $V_{41} = V_{32} \times P(V \to N)$
$= 0$

$V_{42} = V_{32} \times P(V \to V)$
$= 0$

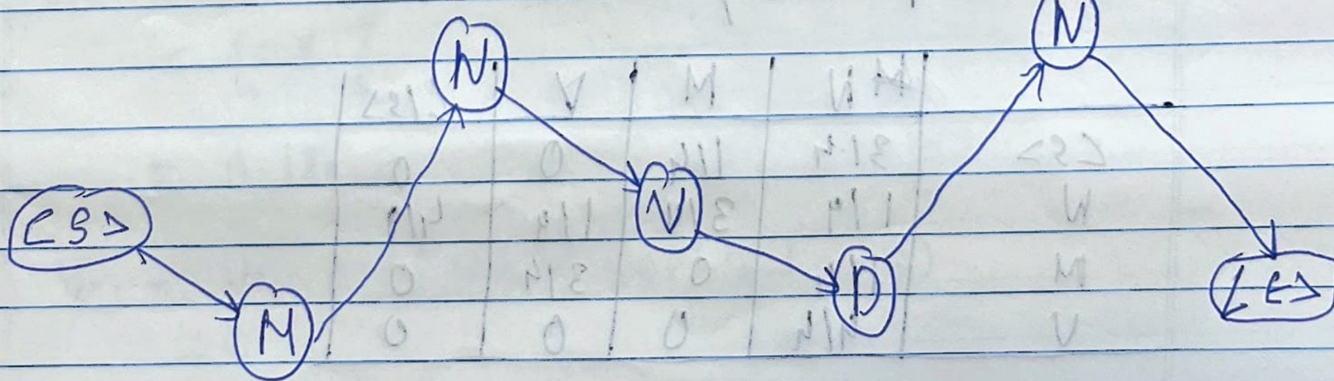
$V_{43} = V_{32} \times P(V \to model) \times P(a \mid model)$
$= 0.0091 \times \frac{2}{4} \times \frac{2}{6}$

$= 0.0015$

$V_{44} = 0$

5) $V_{51} = V_{43} \times P(model \to noun) \times P(book \mid noun)$
$= 0.0015 \times 1 \times \frac{1}{6} = 0.00025$

$V_{52} = 0$
$V_{53} = 0$
$V_{54} = 0$



The | model park | noun is | verb a | Model book | noun

Q.11 Calculate the transition and emission probability for set of sentences below

- Mary Jane will pat spot
- Mary Jane can see will
- Spot will see mary
- Mary will pat spot

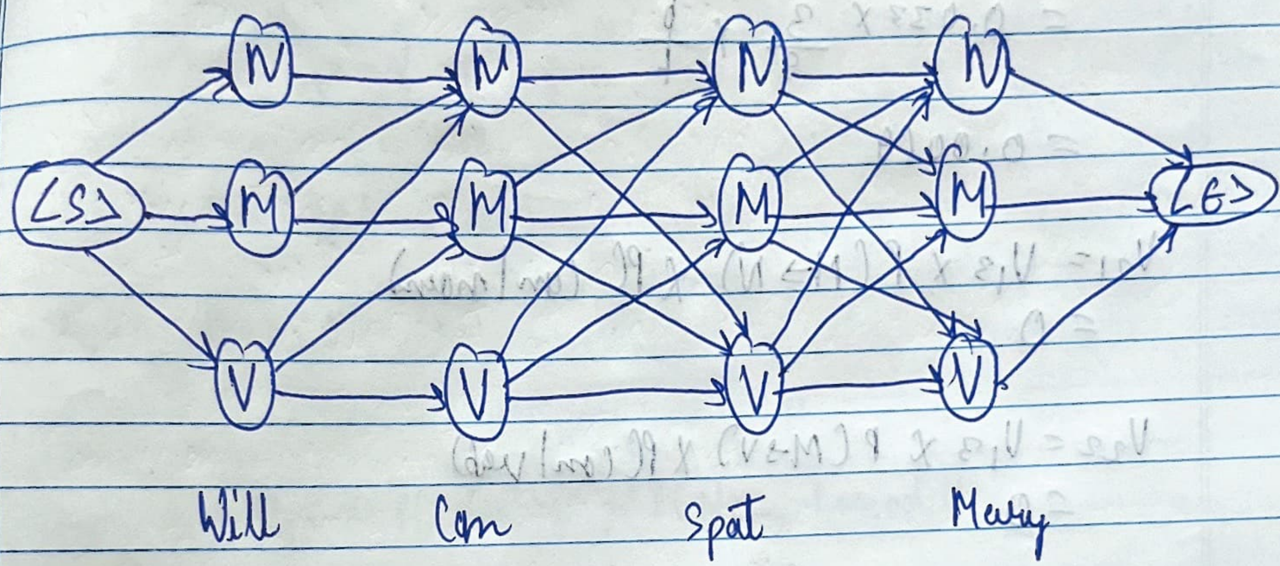
Ans Emission probability algorithm

Words	noun	model	Verb
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

Transition probability matrix

	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

Tag the sentence "Will can spot Mary" using ~~...~~ Viterbi



1) $V_{11} = P(S \to N) \times P(\text{Will} \mid \text{noun})$

$= \frac{3}{4} \times \frac{1}{9}$

$= 0.083$

$V_{12} = P(S \to V) \times P(\text{Will} \mid \text{verb})$

$= 0$

$V_{13} = P(S \to \text{Model}) \times P(\text{Will} \mid \text{model})$

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

$= 0.18$

2) $V_{21} = V_{11} \times P(N \to N) \times P(\text{Can} \mid \text{noun})$

$= 0$

$V_{22} = V_{11} \times P(N \to V) \times P(\text{Can} \mid \text{verb})$

$= 0$

$V_{23} = V_{11} \times P(N \to M) \times P(\text{spot} \mid \text{model})$

$= 0.033 \times \frac{3}{9} \times \frac{1}{4}$

$= 0.0069$

$V_{21} = V_{13} \times P(M \to N) \times P(\text{can} \mid \text{noun})$

$= 0$

$V_{22} = V_{13} \times P(M \to V) \times P(\text{can} \mid \text{verb})$

$= 0.14$

$V_{23} = V_{13} \times P(M \to M) \times P(\text{can} \mid \text{model})$

$= 0$

3) $V_{31} = V_{23} \times P(M \to N) \times P(\text{spot} \mid \text{noun})$

$= 0.006 \times \frac{1}{4} \times \frac{2}{9}$

$= 0.00033$

$V_{32} = V_{23} \times P(M \to V) \times P(\text{spot} \mid \text{verb})$

$= 0.006 \times \frac{3}{4} \times \frac{1}{4}$

$= 0.0011$

$V_{33} = 0$

$V_{41} = V_{31} \times P(N \to N) \times P(\text{mary} \mid \text{noun})$

$= 0.0003 \times \frac{1}{9} \times \frac{4}{9}$

$= 0.00001$

$V_{42} = V_{31} \times 0$

$= 0$

$V_{42} = V_{32} \times P(N \to M) \times P(\text{Mary} \mid \text{model})$

$= 0$

With V_{32}

$V_{41} = V_{32} \times P(V \to N) \times P(\text{Mary} \mid \text{noun})$

$= 0.0011 \times 1 \times \frac{4}{9} = 0.00048$

$V_{42} = V_{32} \times P(V \to V)$

$= 0$

$V_{43} = V_{32} \times P(V \to M)$

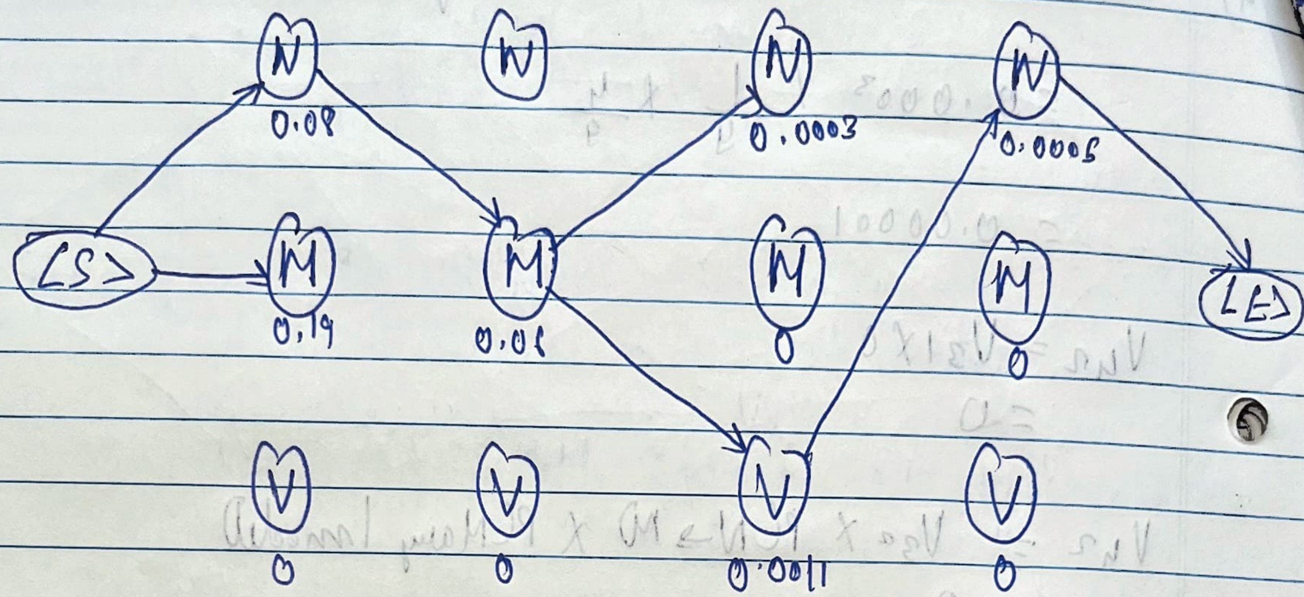
$= 0$

final probability

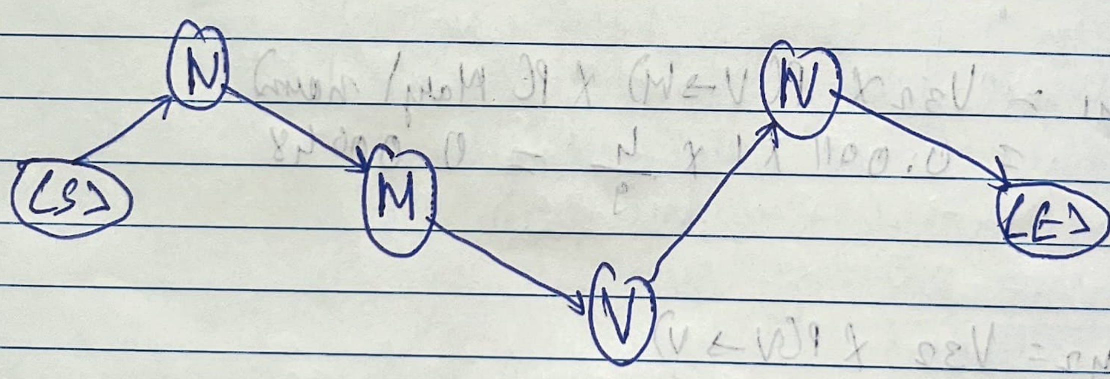
$= 0.00048 \times P(N \to \langle E \rangle)$

$= 0.00048 \times \frac{4}{9}$

$= 0.00021$



Applying Viterbi and choosing most probable



Will/Noun Can/model Spot/Verb Mary/noun

13) For given grammar using CYK or CKY algorithm parse the statement.

"The man read this book"

Rules

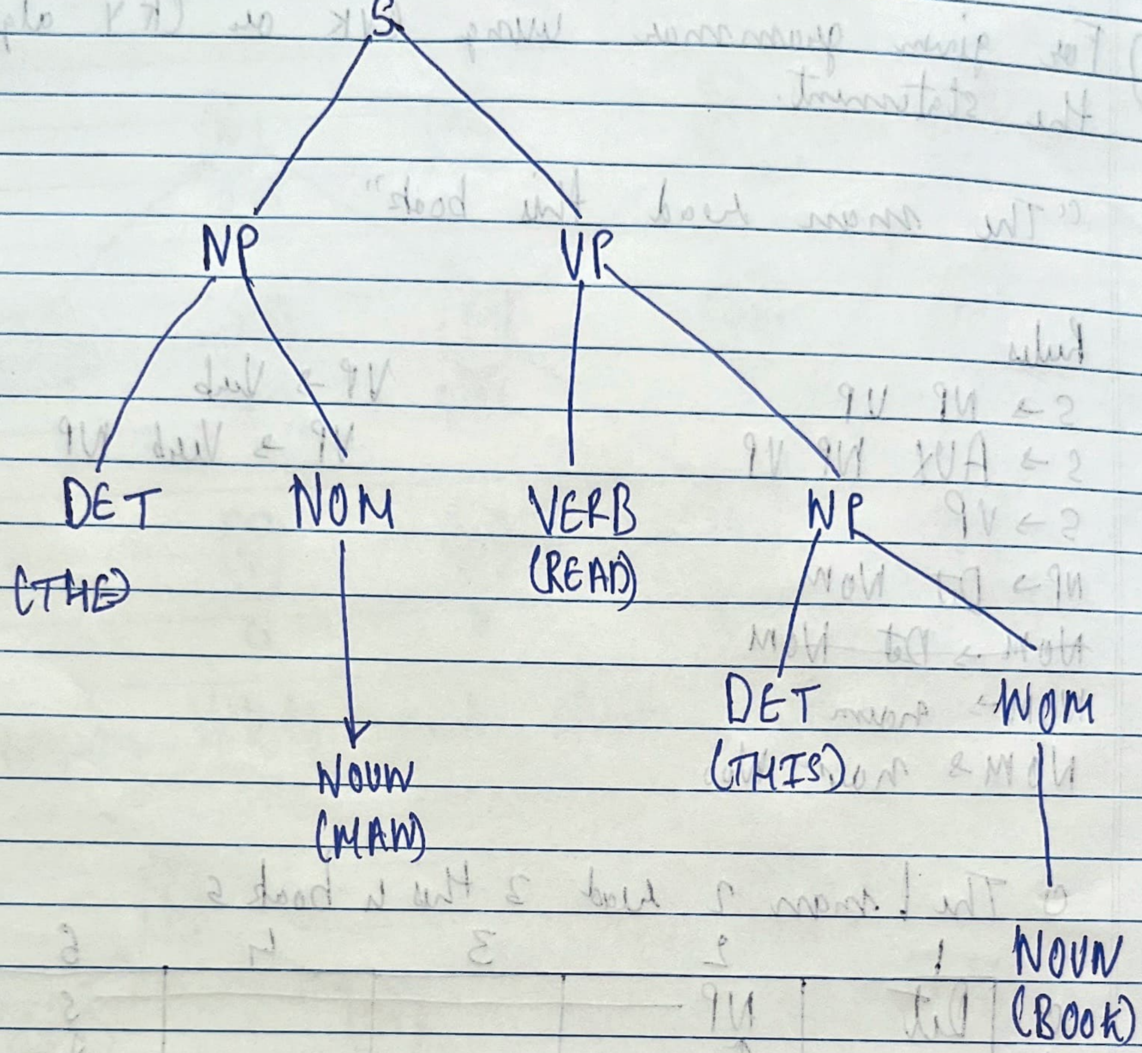
S → NP VP
S → AUX NP VP
S → VP
NP → Det NOM
~~NOM → Det NOM~~
NOM → noun
NOM → noun NOM

VP → Verb
VP → Verb NP

Ans 0 The 1 man 2 read 3 this 4 book 5

	1	2	3	4	5
0	Det	NP			S
1		NOM → noun			
2			Verb		VP
3				Det	NP
4					NOM

This can be seen tree as follow



S → NP VP
S → AUX NP VP
S → VP
NP → Det NOM
NOM → Noun
NOM → Noun NOM
VP → Verb
VP → Verb NP

Det → that | this | a | the
Noun → book | flight | meal | man
Verb → book | include | read
Aux → does

Calculating the probability of non-terminal, using

$$P(\alpha \to \beta \mid \alpha) = \frac{\text{Count}(\alpha \to \beta)}{\sum_{\gamma} \text{Count}(\alpha \to \gamma)} = \frac{\text{Count}(\alpha \to \beta)}{\text{Count}(\alpha)}$$

S → NPVP 0.33
S → AUX NPVP 0.3
S → VP 0.32
NP → Det Nom 1
NOM → noun 0.5
NOM → noun non 0.5
VP → Verb 0.5
VP → Verb NP 0.5

Det → 0.25
Det → 0.25
Det → 0.25
Det → 0.25
noun → 0.25
noun → 0.25
noun → 0.25
noun → 0.26
verb → 0.30
verb → 0.35
verb → 0.30
aux → 1

Det .25

NP → 1 × 0.25 × 0.12 = ~~0.03~~

Noun 0.25 × Nom 0.5

Verb 0.30

S (S) 0.0001

Det 0.25

VP → 0.30 × 0.5 × 0.5 × 0.2 × 0.25

N: 0.2 × 0.5 × 0.25

Nom = 0.5 × 0.25

Noun 0.25

The Man reads this book

Q.14] Identify the correct semantic relation between following

1) Write, Right
2) Big, Large
3) Dark, light
4) Car, vehicle

Ans 1. Write, Right: They are homophone, which means they sound same but have different meaning and spelling. The relation is phonological

2. Big, large: These words have synonymic relationship. They both mean "having a greater size".

3. Dark, light: These ~~wor~~ words have an antonymic relationship. Dark is opposite of dark / light

4. Car, Vehicle: These words have hypernymic relationship. A "car" is a specific type of vehicle. The relationship is one of generality and specifity.

Q.15] Which of following regular expression can be used to identify date(s) present in text? explain details.

Ans Ans is (c)

(19|20) \d{2} - (0[1-9]|1[0-2]) - ([02][1-9]|3[0-1])

This regular expression is similiar to option (B), but further specific that the days can range from 01-31, which covers all possible day value. It can correctly match the date "2017-09-21" in given text

Q. 1c) Consider the fallowing given sentences - Match lexical relation between first word

(W1) to second word (W2) i.e W1 is a lexical relation of (W2)

Ans Holonym → Hyponym

"Invention" (W1) is a holonym of "the wheel" (W2) in sentence 1

Hyponym → Hypernym

In second sentence "driverless car" (W1) is a hyponym of "companies" (W2)

Meronym → Holonym

In third sentence "fluttering" and "dancing" (W_1) are meronyms of "golden daffodils" (W_2)

Hypernym → Holonym

In fourth sentence "flower park" (W_1) is a hypernym of "Mumbai" (W_2)

So, to ~~sta~~ summarize

1) Holonym → hyponym

2) Hyponym → Hypernym

3) Meronym → Holonym

4) Hypernym → Holonym