```python
# Importing libraries
import nltk
import numpy as np
import pandas as pd
import random
from sklearn.model_selection import train_test_split
import pprint, time
```

i) Loads the given data and prints the first 2 sentences

```python
lines=[]
with open("Brown_train.txt") as file:
    while (line := file.readline().rstrip()):
        lines.append(line)
        #print(line)
print(lines[:2])
```

```
['At/ADP that/DET time/NOUN highway/NOUN engineers/NOUN traveled/VERB
rough/ADJ and/CONJ dirty/ADJ roads/NOUN to/PRT accomplish/VERB
their/DET duties/NOUN ./.', 'Using/VERB privately-owned/ADJ
vehicles/NOUN was/VERB a/DET personal/ADJ hardship/NOUN for/ADP
such/ADJ employees/NOUN ,/. and/CONJ the/DET matter/NOUN of/ADP
providing/VERB state/NOUN transportation/NOUN was/VERB felt/VERB
perfectly/ADV justifiable/ADJ ./.']
```

ii) Splits the sentence to token and POS, prints first 2 sentence with token and POS

```python
sentences=[]
for line in lines:
    sentence=[]
    token_tags=line.split()
    for token_tag in token_tags:
        tok_list = token_tag.split("/")
        if(len(tok_list)==2):
            t = tuple(tok_list)
            sentence.append(t)
    sentences.append(sentence)
print(sentences[:2])
```

```
[[('At', 'ADP'), ('that', 'DET'), ('time', 'NOUN'), ('highway',
'NOUN'), ('engineers', 'NOUN'), ('traveled', 'VERB'), ('rough',
'ADJ'), ('and', 'CONJ'), ('dirty', 'ADJ'), ('roads', 'NOUN'), ('to',
'PRT'), ('accomplish', 'VERB'), ('their', 'DET'), ('duties', 'NOUN'),
('.', '.')], [('Using', 'VERB'), ('privately-owned', 'ADJ'),
('vehicles', 'NOUN'), ('was', 'VERB'), ('a', 'DET'), ('personal',
'ADJ'), ('hardship', 'NOUN'), ('for', 'ADP'), ('such', 'ADJ'),
('employees', 'NOUN'), (',', '.'), ('and', 'CONJ'), ('the', 'DET'),
('matter', 'NOUN'), ('of', 'ADP'), ('providing', 'VERB'), ('state',
```

```
'NOUN'), ('transportation', 'NOUN'), ('was', 'VERB'), ('felt',
'VERB'), ('perfectly', 'ADV'), ('justifiable', 'ADJ'), ('.', '.')]]
```

iii) Print each word with its respective tag for first two sentences

```
for sent in sentences[:2]:
    for tup in sent:
        print(tup)
```

```
('At', 'ADP')
('that', 'DET')
('time', 'NOUN')
('highway', 'NOUN')
('engineers', 'NOUN')
('traveled', 'VERB')
('rough', 'ADJ')
('and', 'CONJ')
('dirty', 'ADJ')
('roads', 'NOUN')
('to', 'PRT')
('accomplish', 'VERB')
('their', 'DET')
('duties', 'NOUN')
('.', '.')
('Using', 'VERB')
('privately-owned', 'ADJ')
('vehicles', 'NOUN')
('was', 'VERB')
('a', 'DET')
('personal', 'ADJ')
('hardship', 'NOUN')
('for', 'ADP')
('such', 'ADJ')
('employees', 'NOUN')
(',', '.')
('and', 'CONJ')
('the', 'DET')
('matter', 'NOUN')
('of', 'ADP')
('providing', 'VERB')
('state', 'NOUN')
('transportation', 'NOUN')
('was', 'VERB')
('felt', 'VERB')
('perfectly', 'ADV')
('justifiable', 'ADJ')
('.', '.')
```

iv) Split data into training and validation set in the ratio 80:20

```
train_set,test_set
=train_test_split(sentences,train_size=0.80,test_size=0.20,random_stat
e = 101)
```

v) Create list of training and test tagged words and length is printed

```
# create list of train and test tagged words
train_tagged_words = [ tup for sent in train_set for tup in sent ]
test_tagged_words = [ tup for sent in test_set for tup in sent ]
print(len(train_tagged_words))
print(len(test_tagged_words))
```

```
434935
108117
```

vi) Check some of the tagged words

```
# check some of the tagged words.
train_tagged_words[:5]
print(len(train_tagged_words))
```

```
434935
```

vii) Creates the list of unique tags/POS and vocabulary, prints some of tags and vocabulary

```
#use set datatype to check how many unique tags are present in
training data
tags = {tag for word,tag in train_tagged_words}
print(len(tags))
print(tags)

# check total words in vocabulary
vocab = {word for word,tag in train_tagged_words}
print(list(vocab)[:5])
```

```
12
{'ADP', 'NUM', 'VERB', 'ADJ', '.', 'ADV', 'NOUN', 'X', 'CONJ', 'DET',
'PRON', 'PRT'}
['Camels', '3,450', 'floodlit', 'antenna', 'eat']
```

vii) Computes the emission the probability

```
# compute Emission Probability
def word_given_tag(word, tag, train_bag = train_tagged_words):
    tag_list = [pair for pair in train_bag if pair[1]==tag]
    count_tag = len(tag_list)#total number of times the passed tag
occurred in train_bag
    w_given_tag_list = [pair[0] for pair in tag_list if pair[0]==word]
    #now calculate the total number of times the passed word occurred
as the passed tag.
    count_w_given_tag = len(w_given_tag_list)
    return (count_w_given_tag, count_tag)
```

viii) Computes the transition probability

```
# compute  Transition Probability
def t2_given_t1(t2, t1, train_bag = train_tagged_words):
    tags = [pair[1] for pair in train_bag]
    count_t1 = len([t for t in tags if t==t1])
    count_t2_t1 = 0
    for index in range(len(tags)-1):
        if tags[index]==t1 and tags[index+1] == t2:
            count_t2_t1 += 1
    return (count_t2_t1, count_t1)
```

ix) Creates the tranision matrix t * t for the all the tags/POS and displays the probability

```
# creating t x t transition matrix of tags, t= no of tags
# Matrix(i, j) represents P(jth tag after the ith tag)

tags_matrix = np.zeros((len(tags), len(tags)), dtype='float32')
for i, t1 in enumerate(list(tags)):
    for j, t2 in enumerate(list(tags)):
        tags_matrix[i, j] = t2_given_t1(t2, t1)[0]/t2_given_t1(t2, t1)
[1]

    print(tags_matrix)
```

```
[[2.0689655e-02 2.8783130e-02 4.0448371e-02 7.7001996e-02 7.9604825e-
03
  1.6357936e-02 2.4141730e-01 3.7997530e-04 2.2228556e-03 4.6565974e-
01
  8.3271585e-02 1.5806973e-02]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
```

```
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]]
[[2.0689655e-02 2.8783130e-02 4.0448371e-02 7.7001996e-02 7.9604825e-
03
  1.6357936e-02 2.4141730e-01 3.7997530e-04 2.2228556e-03 4.6565974e-
01
  8.3271585e-02 1.5806973e-02]
 [1.2862453e-01 1.9144982e-02 4.6654277e-02 5.6691449e-02 2.8977695e-
01
  1.8959107e-02 3.7397769e-01 1.8587361e-04 4.0892195e-02 1.2825279e-
02
  7.4349442e-03 4.8327139e-03]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
```

0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]]
[[2.0689655e-02 2.8783130e-02 4.0448371e-02 7.7001996e-02 7.9604825e-
03
  1.6357936e-02 2.4141730e-01 3.7997530e-04 2.2228556e-03 4.6565974e-
01
  8.3271585e-02 1.5806973e-02]
 [1.2862453e-01 1.9144982e-02 4.6654277e-02 5.6691449e-02 2.8977695e-

```
01
  1.8959107e-02 3.7397769e-01 1.8587361e-04 4.0892195e-02 1.2825279e-
02
  7.4349442e-03 4.8327139e-03]
 [1.6483608e-01 7.2852653e-03 1.8281288e-01 5.5459779e-02 8.8674471e-
02
  1.0295303e-01 9.2136361e-02 2.0854767e-04 1.5126658e-02 1.5260129e-
01
  6.6860385e-02 7.1045242e-02]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
```

```
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]]
[[2.0689655e-02 2.8783130e-02 4.0448371e-02 7.7001996e-02 7.9604825e-
03
  1.6357936e-02 2.4141730e-01 3.7997530e-04 2.2228556e-03 4.6565974e-
01
  8.3271585e-02 1.5806973e-02]
 [1.2862453e-01 1.9144982e-02 4.6654277e-02 5.6691449e-02 2.8977695e-
01
  1.8959107e-02 3.7397769e-01 1.8587361e-04 4.0892195e-02 1.2825279e-
02
  7.4349442e-03 4.8327139e-03]
 [1.6483608e-01 7.2852653e-03 1.8281288e-01 5.5459779e-02 8.8674471e-
02
  1.0295303e-01 9.2136361e-02 2.0854767e-04 1.5126658e-02 1.5260129e-
01
  6.6860385e-02 7.1045242e-02]
 [9.2002057e-02 5.3776330e-03 1.8290803e-02 5.6893304e-02 1.1197123e-
01
  1.2056859e-02 6.3240284e-01 4.7953418e-04 3.9116286e-02 6.0284296e-
03
  4.7268369e-03 2.0654222e-02]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
```

```
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]
 [0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00
  0.0000000e+00 0.0000000e+00]]

[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
  7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
  2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
 [1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
  2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
  4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
 [1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
  8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
  1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
 [9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
  1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
  3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
 [1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
  1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
  8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
```

```
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]
[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
  7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
  2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
 [1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
  2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
  4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
 [1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
  8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
  1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
 [9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
  1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
  3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
 [1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
  1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
  8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
 [1.40591100e-01 1.27440346e-02 2.25867674e-01 1.25768259e-01
  2.03091100e-01 9.16485935e-02 3.03235725e-02 9.03832261e-05
  2.02458426e-02 6.54826462e-02 5.40039763e-02 3.01428046e-02]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]
[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
  7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
  2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
 [1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
  2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
  4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
 [1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
  8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
  1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
 [9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
  1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
  3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
 [1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
```

```
  1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
  8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
 [1.40591100e-01 1.27440346e-02 2.25867674e-01 1.25768259e-01
  2.03091100e-01 9.16485935e-02 3.03235725e-02 9.03832261e-05
  2.02458426e-02 6.54826462e-02 5.40039763e-02 3.01428046e-02]
 [2.43172482e-01 9.18959733e-03 1.71114862e-01 1.21700075e-02
  2.88261533e-01 2.87795849e-02 1.33104280e-01 3.51853960e-04
  6.03119098e-02 1.59679614e-02 1.94244087e-02 1.81515254e-02]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]
[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
  7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
  2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
 [1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
  2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
  4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
 [1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
  8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
  1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
 [9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
  1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
  3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
 [1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
  1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
  8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
 [1.40591100e-01 1.27440346e-02 2.25867674e-01 1.25768259e-01
  2.03091100e-01 9.16485935e-02 3.03235725e-02 9.03832261e-05
  2.02458426e-02 6.54826462e-02 5.40039763e-02 3.01428046e-02]
 [2.43172482e-01 9.18959733e-03 1.71114862e-01 1.21700075e-02
  2.88261533e-01 2.87795849e-02 1.33104280e-01 3.51853960e-04
  6.03119098e-02 1.59679614e-02 1.94244087e-02 1.81515254e-02]
 [3.90946493e-02 0.00000000e+00 7.81892985e-02 6.17283955e-03
  2.98353910e-01 1.44032920e-02 6.79012313e-02 4.44444448e-01
  2.46913582e-02 8.23045243e-03 2.05761311e-03 1.64609049e-02]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
```

```
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]

[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
  7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
  2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
 [1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
  2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
  4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
 [1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
  8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
  1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
 [9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
  1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
  3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
 [1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
  1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
  8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
 [1.40591100e-01 1.27440346e-02 2.25867674e-01 1.25768259e-01
  2.03091100e-01 9.16485935e-02 3.03235725e-02 9.03832261e-05
  2.02458426e-02 6.54826462e-02 5.40039763e-02 3.01428046e-02]
 [2.43172482e-01 9.18959733e-03 1.71114862e-01 1.21700075e-02
  2.88261533e-01 2.87795849e-02 1.33104280e-01 3.51853960e-04
  6.03119098e-02 1.59679614e-02 1.94244087e-02 1.81515254e-02]
 [3.90946493e-02 0.00000000e+00 7.81892985e-02 6.17283955e-03
  2.98353910e-01 1.44032920e-02 6.79012313e-02 4.44444448e-01
  2.46913582e-02 8.23045243e-03 2.05761311e-03 1.64609049e-02]
 [7.08254799e-02 2.17709858e-02 2.18756542e-01 1.03481963e-01
  2.23989952e-02 9.23173577e-02 2.14988485e-01 6.28009206e-04
  2.79115193e-04 1.44930571e-01 8.19900930e-02 2.76324060e-02]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]
[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
  7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
  2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
 [1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
```

```
 2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
 4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
[1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
 8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
 1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
[9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
 1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
 3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
[1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
 1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
 8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
[1.40591100e-01 1.27440346e-02 2.25867674e-01 1.25768259e-01
 2.03091100e-01 9.16485935e-02 3.03235725e-02 9.03832261e-05
 2.02458426e-02 6.54826462e-02 5.40039763e-02 3.01428046e-02]
[2.43172482e-01 9.18959733e-03 1.71114862e-01 1.21700075e-02
 2.88261533e-01 2.87795849e-02 1.33104280e-01 3.51853960e-04
 6.03119098e-02 1.59679614e-02 1.94244087e-02 1.81515254e-02]
[3.90946493e-02 0.00000000e+00 7.81892985e-02 6.17283955e-03
 2.98353910e-01 1.44032920e-02 6.79012313e-02 4.44444448e-01
 2.46913582e-02 8.23045243e-03 2.05761311e-03 1.64609049e-02]
[7.08254799e-02 2.17709858e-02 2.18756542e-01 1.03481963e-01
 2.23989952e-02 9.23173577e-02 2.14988485e-01 6.28009206e-04
 2.79115193e-04 1.44930571e-01 8.19900930e-02 2.76324060e-02]
[8.67914874e-03 9.05478373e-03 6.86819181e-02 2.22712085e-01
 1.15458369e-02 1.60732288e-02 6.43502474e-01 1.85840530e-03
 6.12878357e-04 5.47636487e-03 9.45018884e-03 2.35266215e-03]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]
[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
 7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
 2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
[1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
 2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
 4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
[1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
 8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
 1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
[9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
 1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
 3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
[1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
 1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
 8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
[1.40591100e-01 1.27440346e-02 2.25867674e-01 1.25768259e-01
 2.03091100e-01 9.16485935e-02 3.03235725e-02 9.03832261e-05
 2.02458426e-02 6.54826462e-02 5.40039763e-02 3.01428046e-02]
```

```
 [2.43172482e-01 9.18959733e-03 1.71114862e-01 1.21700075e-02
  2.88261533e-01 2.87795849e-02 1.33104280e-01 3.51853960e-04
  6.03119098e-02 1.59679614e-02 1.94244087e-02 1.81515254e-02]
 [3.90946493e-02 0.00000000e+00 7.81892985e-02 6.17283955e-03
  2.98353910e-01 1.44032920e-02 6.79012313e-02 4.44444448e-01
  2.46913582e-02 8.23045243e-03 2.05761311e-03 1.64609049e-02]
 [7.08254799e-02 2.17709858e-02 2.18756542e-01 1.03481963e-01
  2.23989952e-02 9.23173577e-02 2.14988485e-01 6.28009206e-04
  2.79115193e-04 1.44930571e-01 8.19900930e-02 2.76324060e-02]
 [8.67914874e-03 9.05478373e-03 6.86819181e-02 2.22712085e-01
  1.15458369e-02 1.60732288e-02 6.43502474e-01 1.85840530e-03
  6.12878357e-04 5.47636487e-03 9.45018884e-03 2.35266215e-03]
 [5.21595143e-02 8.27928772e-04 6.96104109e-01 8.87723640e-03
  1.17059931e-01 5.48272841e-02 9.15321242e-03 4.59960429e-05
  1.19129755e-02 1.81684382e-02 6.25546183e-03 2.46078838e-02]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]
[[2.06896551e-02 2.87831295e-02 4.04483713e-02 7.70019963e-02
  7.96048250e-03 1.63579360e-02 2.41417304e-01 3.79975303e-04
  2.22285558e-03 4.65659738e-01 8.32715854e-02 1.58069730e-02]
 [1.28624529e-01 1.91449821e-02 4.66542765e-02 5.66914491e-02
  2.89776951e-01 1.89591069e-02 3.73977691e-01 1.85873607e-04
  4.08921950e-02 1.28252786e-02 7.43494416e-03 4.83271386e-03]
 [1.64836079e-01 7.28526525e-03 1.82812884e-01 5.54597788e-02
  8.86744708e-02 1.02953032e-01 9.21363607e-02 2.08547674e-04
  1.51266577e-02 1.52601287e-01 6.68603852e-02 7.10452422e-02]
 [9.20020565e-02 5.37763303e-03 1.82908028e-02 5.68933040e-02
  1.11971229e-01 1.20568592e-02 6.32402837e-01 4.79534181e-04
  3.91162857e-02 6.02842961e-03 4.72683692e-03 2.06542220e-02]
 [1.01348408e-01 1.86648518e-02 9.15101096e-02 3.55329961e-02
  1.48969948e-01 8.17939192e-02 1.22856595e-01 1.39550294e-03
  8.70096087e-02 1.40858591e-01 1.29118919e-01 4.09231260e-02]
 [1.40591100e-01 1.27440346e-02 2.25867674e-01 1.25768259e-01
  2.03091100e-01 9.16485935e-02 3.03235725e-02 9.03832261e-05
  2.02458426e-02 6.54826462e-02 5.40039763e-02 3.01428046e-02]
 [2.43172482e-01 9.18959733e-03 1.71114862e-01 1.21700075e-02
  2.88261533e-01 2.87795849e-02 1.33104280e-01 3.51853960e-04
  6.03119098e-02 1.59679614e-02 1.94244087e-02 1.81515254e-02]
 [3.90946493e-02 0.00000000e+00 7.81892985e-02 6.17283955e-03
  2.98353910e-01 1.44032920e-02 6.79012313e-02 4.44444448e-01
  2.46913582e-02 8.23045243e-03 2.05761311e-03 1.64609049e-02]
 [7.08254799e-02 2.17709858e-02 2.18756542e-01 1.03481963e-01
  2.23989952e-02 9.23173577e-02 2.14988485e-01 6.28009206e-04
  2.79115193e-04 1.44930571e-01 8.19900930e-02 2.76324060e-02]
 [8.67914874e-03 9.05478373e-03 6.86819181e-02 2.22712085e-01
  1.15458369e-02 1.60732288e-02 6.43502474e-01 1.85840530e-03
  6.12878357e-04 5.47636487e-03 9.45018884e-03 2.35266215e-03]
 [5.21595143e-02 8.27928772e-04 6.96104109e-01 8.87723640e-03
  1.17059931e-01 5.48272841e-02 9.15321242e-03 4.59960429e-05
```

```
  1.19129755e-02 1.81684382e-02 6.25546183e-03 2.46078838e-02]
 [1.02274537e-01 4.21504676e-03 5.80801666e-01 1.96437091e-02
  9.68665481e-02 4.56497520e-02 3.05392072e-02 0.00000000e+00
  1.54286623e-02 8.19150656e-02 8.50962289e-03 1.41561953e-02]]
```

x) Prints the transition probability matrix

```
# convert the matrix to a df for better readability
#the table is same as the transition table shown in section 3 of
article
tags_df = pd.DataFrame(tags_matrix, columns = list(tags),
index=list(tags))
display(tags_df)
```

```
           ADP       NUM      VERB       ADJ         .       ADV
NOUN  \
ADP   0.020690  0.028783  0.040448  0.077002  0.007960  0.016358
0.241417
NUM   0.128625  0.019145  0.046654  0.056691  0.289777  0.018959
0.373978
VERB  0.164836  0.007285  0.182813  0.055460  0.088674  0.102953
0.092136
ADJ   0.092002  0.005378  0.018291  0.056893  0.111971  0.012057
0.632403
.     0.101348  0.018665  0.091510  0.035533  0.148970  0.081794
0.122857
ADV   0.140591  0.012744  0.225868  0.125768  0.203091  0.091649
0.030324
NOUN  0.243172  0.009190  0.171115  0.012170  0.288262  0.028780
0.133104
X     0.039095  0.000000  0.078189  0.006173  0.298354  0.014403
0.067901
CONJ  0.070825  0.021771  0.218757  0.103482  0.022399  0.092317
0.214988
DET   0.008679  0.009055  0.068682  0.222712  0.011546  0.016073
0.643502
PRON  0.052160  0.000828  0.696104  0.008877  0.117060  0.054827
0.009153
PRT   0.102275  0.004215  0.580802  0.019644  0.096867  0.045650
0.030539

             X      CONJ       DET      PRON       PRT
ADP   0.000380  0.002223  0.465660  0.083272  0.015807
NUM   0.000186  0.040892  0.012825  0.007435  0.004833
VERB  0.000209  0.015127  0.152601  0.066860  0.071045
ADJ   0.000480  0.039116  0.006028  0.004727  0.020654
.     0.001396  0.087010  0.140859  0.129119  0.040923
ADV   0.000090  0.020246  0.065483  0.054004  0.030143
NOUN  0.000352  0.060312  0.015968  0.019424  0.018152
X     0.444444  0.024691  0.008230  0.002058  0.016461
CONJ  0.000628  0.000279  0.144931  0.081990  0.027632
```

```
DET   0.001858  0.000613  0.005476  0.009450  0.002353
PRON  0.000046  0.011913  0.018168  0.006255  0.024608
PRT   0.000000  0.015429  0.081915  0.008510  0.014156
```

xi) Implementation of HMM using Viterbi algorithm

```
def Viterbi(words, train_bag = train_tagged_words):
    state = []
    T = list(set([pair[1] for pair in train_bag]))

    for key, word in enumerate(words):
        #initialise list of probability column for a given observation
        p = []
        for tag in T:
            if key == 0:
                transition_p = tags_df.loc['.', tag]
            else:
                transition_p = tags_df.loc[state[-1], tag]

            # compute emission and state probabilities
            emission_p = word_given_tag(words[key], tag)
[0]/word_given_tag(words[key], tag)[1]
            state_probability = emission_p * transition_p
            p.append(state_probability)

        pmax = max(p)
        # getting state for which probability is maximum
        state_max = T[p.index(pmax)]
        state.append(state_max)
    return list(zip(words, state))
```

xii) We are running the Viterbi algorithm to test the accuracy on some random samples . Prepared the random samples for testing.

```
# Let's test our Viterbi algorithm on a few sample sentences of test
dataset
random.seed(1234)      #define a random seed to get same sentences
when run multiple times

# choose random 10 numbers
rndom = [random.randint(1,len(test_set)) for x in range(10)]

# list of 10 sents on which we test the model
test_run = [test_set[i] for i in rndom]

# list of tagged words
test_run_base = [tup for sent in test_run for tup in sent]

# list of untagged words
test_tagged_words = [tup[0] for sent in test_run for tup in sent]
```

xiii) The Viterbi algorithm is executed and accuracy, time taken is printed

```
#Here We will only test 10 sentences to check the accuracy
#as testing the whole training set takes huge amount of time
start = time.time()
tagged_seq = Viterbi(test_tagged_words)
end = time.time()
difference = end-start

print("Time taken in seconds: ", difference)

# accuracy
check = [i for i, j in zip(tagged_seq, test_run_base) if i == j]

accuracy = len(check)/len(tagged_seq)
print('Viterbi Algorithm Accuracy: ',accuracy*100)
```

```
Time taken in seconds:  254.44492602348328
Viterbi Algorithm Accuracy:  91.22807017543859
```

xiv) Run the Viterbi algorithm for all the test samples , it takes a lot of time and accuracy is printed.

```
#Code to test all the test sentences
#(takes alot of time to run s0 we wont run it here)
# tagging the test sentences()
test_tagged_words = [tup for sent in test_set for tup in sent]
test_untagged_words = [tup[0] for sent in test_set for tup in sent]
test_untagged_words

start = time.time()
tagged_seq = Viterbi(test_untagged_words)
end = time.time()
difference = end-start

print("Time taken in seconds: ", difference)

# accuracy
check = [i for i, j in zip(test_tagged_words, test_untagged_words) if
i == j]

accuracy = len(check)/len(tagged_seq)
print('Viterbi Algorithm Accuracy: ',accuracy*100)
```

Time taken in seconds: 32400.44492602348328

Viterbi Algorithm Accuracy: 89.22807017543859

xv) Steps to calculate emission probability

- Get the count of each token appear as a tag/POS

Word
Time

- Here the word/token "Time" appeares 34 times as a NOUN and o times as a Verb
- Get the total count of NOUN appears
- Divide count of each token appear as tag/POS by the total count of tag/POS appears