

# CS 557 - Assignment 6 (Sequence Processing and Formal Grammars in English)

## Group 3 Members:

Parth Parab - CWID 10444835

Sejal Vyas - CWID 10450395

Shiwani Deo - CWID 10454959

1. Show your responses to the Your Turn practice of BKL Ch5 on pages 180. Try a variant of this exercise with the use of the two words “in” and “on”, as in “The book is not on the table, it is in my backpack”. Also tag the following sentence “Book the cooks who cook the books.” (a tagger based on the WSJ Corpus might be appropriate).

Q. Many words, like ski and race, can be used as nouns or verbs with no difference in pronunciation. Can you think of others? Hint: think of a commonplace object and try to put the word to before it to see if it can also be a verb, or think of an action and try to put the before it to see if it can also be a noun. Now make up a sentence with both uses of this word, and run the POS-tagger on this sentence.

```
In [1]: import nltk
nltk.download('punkt')
nltk.download('gutenberg')
nltk.download('brown')
nltk.download('averaged_perceptron_tagger')
from nltk.tokenize import word_tokenize
from nltk.corpus import gutenberg
from nltk.corpus import brown

[nltk_data] Downloading package punkt to
[nltk_data]   /Users/shivanideo/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package gutenberg to
[nltk_data]   /Users/shivanideo/nltk_data...
[nltk_data]   Unzipping corpora/gutenberg.zip.
[nltk_data] Downloading package brown to
[nltk_data]   /Users/shivanideo/nltk_data...
[nltk_data]   Unzipping corpora/brown.zip.
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /Users/shivanideo/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
```

```
[('It', 'PRP'), ('took', 'VBD'), ('him', 'PRP'), ('a', 'DT'), ('lot', 'NN'), ('of', 'IN'), ('time', 'NN'), ('to', 'TO'), ('combat', 'VB'), ('the', 'DT'), ('PTSD', 'NNP'), ('he', 'PRP'), ('suffered', 'VBD'), ('after', 'IN'), ('his', 'PRP$'), ('friends', 'NNS'), ('were', 'VBD'), ('killed', 'VBN'), ('in', 'IN'), ('combat', 'NN'), ('.', '.')]
[('He', 'PRP'), ('used', 'VBD'), ('his', 'PRP$'), ('brand', 'NN'), ('new', 'JJ'), ('hammer', 'NN'), ('to', 'TO'), ('hammer', 'VB'), ('the', 'DT'), ('nail', 'NN'), ('sticking', 'VBG'), ('out', 'RP'), ('on', 'IN'), ('his', 'PRP$'), ('staircase', 'NN'), ('.', '.')]
[('She', 'PRP'), ('told', 'VBD'), ('the', 'DT'), ('teacher', 'NN'), ('the', 'DT'), ('dog', 'NN'), ('ate', 'VBD'), ('her', 'PRP$'), ('homework', 'NN'), ('as', 'IN'), ('an', 'DT'), ('excuse', 'NN'), (',', ','), ('but', 'CC'), ('the', 'DT'), ('teacher', 'NN'), ('would', 'MD'), ('not', 'RB'), ('excuse', 'VB'), ('her', 'PRP'), ('.', '.')]
[('I', 'PRP'), ('had', 'VBD'), ('an', 'DT'), ('asthma', 'JJ'), ('attack', 'NN'), (',', ','), ('so', 'IN'), ('the', 'DT'), ('doctor', 'NN'), ('decided', 'VBD'), ('to', 'TO'), ('attack', 'VB'), ('the', 'DT'), ('problem', 'NN'), ('using', 'VBG'), ('medication', 'NN'), ('.', '.')]
[('When', 'WRB'), ('the', 'DT'), ('professor', 'NN'), ('asked', 'VBD'), ('her', 'PRP'), ('a', 'DT'), ('difficult', 'JJ'), ('question', 'NN'), (',', ','), ('it', 'PRP'), ('made', 'VBD'), ('her', 'PRP'), ('question', 'NN'), ('her', 'PRP$'), ('knowledge', 'NN'), ('.', '.')]
[('The', 'DT'), ('book', 'NN'), ('is', 'VBZ'), ('not', 'RB'), ('on', 'IN'), ('the', 'DT'), ('table', 'NN'), (',', ','), ('it', 'PRP'), ('is', 'VBZ'), ('in', 'IN'), ('my', 'PRP$'), ('backpack', 'NN')]
[('He', 'PRP'), ('likes', 'VBZ'), ('to', 'TO'), ('spend', 'VB'), ('his', 'PRP$'), ('evenings', 'NNS'), ('in', 'IN'), ('his', 'PRP$'), ('backyard', 'NN'), (',', ','), ('sitting', 'VBG'), ('on', 'IN'), ('the', 'DT'), ('swing', 'NN'), ('.', '.')]
[('We', 'PRP'), ('knocked', 'VBD'), ('on', 'IN'), ('his', 'PRP$'), ('door', 'NN'), ('but', 'CC'), ('no', 'DT'), ('one', 'NN'), ('was', 'VBD'), ('in', 'IN'), ('.', '.'), ('He', 'PRP'), ('was', 'VBD'), ('probably', 'RB'), ('vacationing', 'VBG'), ('on', 'IN'), ('his', 'PRP$'), ('yacht', 'NN'), ('.', '.')]
[('Book', 'IN'), ('the', 'DT'), ('cooks', 'NNS'), ('who', 'WP'), ('cook', 'VBP'), ('the', 'DT'), ('books', 'NNS'), ('.', '.')]

```

Q. Search the Web for “spoof newspaper headlines,” to find such gems as: British Left Waffles on Falkland Islands, and Juvenile Court to Try Shooting Defendant. Manually tag these headlines to see whether knowledge of the part-of-speech tags removes the ambiguity.



## Manual Tagging of the sentences:

1) British Left Waffles on Falkland Islands.

#1: NNP VB NNS IN NNP NNS

#2: NNP NN VB IN NNP NNS

'Left': Political Party (NN), Leave (VB)

'Waffles': Breakfast food (NNS), Unable to decide (VB)

2) Squad Helps Dog Bite Victim.

#1: NN VB NN VB NN

#2: NN VB NN NN NN

'Bite': Action of biting (VB), wound/bruise (NN)

3) Dealers Will Hear Car Talk at Noon.

#1: NNS MD VB NN VB IN NN

#2: NNS MD VB NN NN IN NN

'Talk': Conversing (VB), lecture on something / topic (NN)

4) Stolen painting found by tree.

#1: NNP VBG VBN IN NN

#2: NNP NN VBN IN NN

'Painting': To paint (VB), Art (NN)

5) Milk drinkers are turning to powder.

#1: NN NNS VBP VBP TO NN

'Turning': Semantic ambiguity

6) Book the cooks who cook the books.

#1: NN DT NNS WP VB DT NNS

#2: VB DT NNS WP VB DT NNS

'Book': Book (NN), Enter charges against (VB)

```
In [14]: #Tagging using the POS tagger

sentence1 = word_tokenize("British Left waffles on Falkland islands.")
print(nltk.pos_tag(sentence1)) #The POS tagger is trained on WSJ

sentence2 = word_tokenize("Squad helps dog bite victim.")
print(nltk.pos_tag(sentence2))

sentence3 = word_tokenize("Dealers will hear car talk at noon.")
print(nltk.pos_tag(sentence3))

sentence4 = word_tokenize("Stolen painting found by tree.")
print(nltk.pos_tag(sentence4))

sentence5 = word_tokenize("Milk drinkers are turning to powder.")
print(nltk.pos_tag(sentence5))

sentence6 = word_tokenize("Book the cooks who cook the books.")
print(nltk.pos_tag(sentence6))

[('British', 'JJ'), ('Left', 'NNP'), ('waffles', 'VBZ'), ('on', 'IN'), ('Falkland', 'NNP'), ('islands', 'NNS'), ('.', '.')]
[('Squad', 'NNP'), ('helps', 'VBZ'), ('dog', 'VB'), ('bite', 'JJ'), ('victim', 'NN'), ('.', '.')]
[('Dealers', 'NNS'), ('will', 'MD'), ('hear', 'VB'), ('car', 'NN'), ('talk', 'NN'), ('at', 'IN'), ('noon', 'NN'), ('.', '.')]
[('Stolen', 'NNP'), ('painting', 'VBG'), ('found', 'VBN'), ('by', 'IN'), ('tree', 'NN'), ('.', '.')]
[('Milk', 'NN'), ('drinkers', 'NNS'), ('are', 'VBP'), ('turning', 'VBG'), ('to', 'TO'), ('powder', 'VB'), ('.', '.')]
[('Book', 'IN'), ('the', 'DT'), ('cooks', 'NNS'), ('who', 'WP'), ('cook', 'VBP'), ('the', 'DT'), ('books', 'NNS'), ('.', '.')]
```

As we can see above, knowledge of part-of-speech taggers can help remove ambiguity when tagged manually. Although, semantic ambiguity cannot be avoided. Running the pos tagger will usually give either one of the results mentioned above, and it's not always the correct one.

**3. Respond to Exercise 12.10 from J&M 2rd Ch12 for at least one Wh- terminal. The lexicon and grammar in Figures 10.2 and 10.3 should be enough. If not indicate any other terminals and rules necessary. Obviously the Wh- you chose needs to be a terminal. You are welcome to search the WWW for help here; be sure to show references to the sites you use.**

Q. Page 397 discussed the need for a Wh-NP constituent. The simplest Wh-NP is one of the Wh-pronouns (who, whom, whose, which). The Wh-words what and which can be determiners: which four will you have?, what credit do you have with the Duke? Write rules for the different types of Wh-NPs.

Rules for Wh- terminal with examples are given as follows,

**Rule : S -> Wh- NP VP**

Example : Which flights serve breakfast?, What airlines fly from New York to Halifax?

**Rule : S -> Wh-NP Aux NP VP**

Example : What flights do you have from Burbank to Tacoma Washington?

**Rule : S -> Wh- Aux NP VP**

Example : What do you think of this?

**4. Chapter 10 of J&M 3rd and Chapter 5 of BKL demonstrate several taggers. Make a list of them and any others you can find with a brief statement as to what they do, what is distinctive about them, and what is like the other taggers in the chapter. You can show them with a table of at least 4 columns: ID, Description, Distinctive Feature(s), and “Similar To”. The Description should include words used in both J&M 2nd and 3rd and BKL, such as “bidirectional MEMM-style”, “regular expression tagger”, “lookup tagger”, etc. to distinguish the taggers.**



ID	Description	Distinctive Feature(s)	Similar To
RegexTagger	Regular Expression Tagger	Assigns tags to tokens by matching based on user-provided pattern	Rule-based taggers - The regular expression is similar to matching a rule
UnigramTagger	Lookup Tagger	User chooses tagged text and create a lookup table, i.e., user determines how tagged text will tag other text. If no match in lookup table, use default tagger (backoff)	Stochastic Tagger - Since we use frequency to determine the tag of words, this can be compared to the word frequency approach in stochastic tagging
Brill Tagger	Transformation Tagger	This tagger guesses the tag of each word, once done, go back and fix mistakes. This successively transform bad text tagging into a better one.	Rule-based taggers - Brill tagging templates define tagging rules.
HMM Tagger	Probabilistic Sequence Model	Given a sequence of units (words, letters, morphemes, sentences, whatever), it computes a probability distribution over possible sequences of labels and chooses the best label sequence.	Stochastic Tagger (Due to tagging of sequence probabilities)
MEMM Tagger	Maximum Entropy Markov Model	Uses the class assigned to the prior word as a feature in the classification of the next word, i.e., applies logistic regression.	HMM Tagger, Stochastic Tagger

5. Tag the longest “sentence” you found from Assignments 2. Compare the results for at least two different taggers. Briefly discuss what applying embeddings to this might produce.

```
In [3]: #In Assignment 2, the corpus chosen was Moby Dick.
moby = gutenbergsents('melville-moby_dick.txt')
longest = max(len(s) for s in moby)
a = [s for s in moby if len(s) == longest]
longestSentence = ' '.join(str(x) for x in a[0]) #Find longest sentence
print("The longest sentence in the Moby Dick corpus is - ", longestSentence)
```

The longest sentence in the Moby Dick corpus is - Though in many natural objects , whiteness refiningly enhances beauty , as if imparting some special virtue of its own , as in marbles , japonicas , and pearls ; and though various nations have in some way recognised a certain royal preeminence in this hue ; even the barbaric , grand old kings of Pegu placing the title " Lord of the White Elephants " above all their other magniloquent ascriptions of dominion ; and the modern kings of Siam unfurling the same snow - white quadruped in the royal standard ; and the Hanoverian flag bearing the one figure of a snow - white charger ; and the great Austrian Empire , Caesarian , heir to overlording Rome , having for the imperial colour the same imperial hue ; and though this pre - eminence in it applies to the human race itself , giving the white man ideal mastership over every dusky tribe ; and though , besides , all this , whiteness has been even made significant of gladness , for among the Romans a white stone marked a joyful day ; and though in other mortal sympathies and symbolizings , this same hue is made the emblem of many touching , noble things -- the innocence of brides , the benignity of age ; though among the Red Men of America the giving of the white belt of wampum was the deepest pledge of honour ; though in many climes , whiteness typifies the majesty of Justice in the ermine of the Judge , and contributes to the daily state of kings and queens drawn by milk - white steeds ; though even in the higher mysteries of the most august religions it has been made the symbol of the divine spotlessness and power ; by the Persian fire worshippers , the white forked flame being held the holiest on the altar ; and in the Greek mythologies , Great Jove himself being made incarnate in a snow - white bull ; and though to the noble Iroquois , the midwinter sacrifice of the sacred White Dog was by far the holiest festival of their theology , that spotless , faithful creature being held the purest envoy they could send to the Great Spirit with the annual tidings of their own fidelity ; and though directly from the Latin word for white , all Christian priests derive the name of one part of their sacred vesture , the alb or tunic , worn beneath the cassock ; and though among the holy pomps of the Romish faith , white is specially employed in the celebration of the Passion of our Lord ; though in the Vision of St . John , white robes are given to the redeemed , and the four - and - twenty elders stand clothed in white before the great - white throne , and the Holy One that sitteth there white like wool ; yet for all these accumulated associations , with whatever is sweet , and honourable , and sublime , there yet lurks an elusive something in the innermost idea of this hue , which strikes more of panic to the soul than that redness which affrights in blood .

```
In [4]: #Lookup Tagger
brownTaggedSents = brown.tagged_sents()
unigramTagger = nltk.UnigramTagger(brownTaggedSents)
print(unigramTagger.tag(nltk.tokenize.word_tokenize(longestSentence))) #Print the tags
unigramTagger.evaluate(brownTaggedSents) #Print the accuracy of the lookup tagger
```

```
[('Though', 'CS'), ('in', 'IN'), ('many', 'AP'), ('natural', 'JJ'), ('objects', 'NNS'), (',', ','), ('whiteness', 'NN'), ('refiningly', None), ('enhances', 'VBZ'), ('beauty', 'NN'), (',', ','), ('as', 'CS'), ('if', 'CS'), ('imparting', None), ('some', 'DTI'), ('special', 'JJ'), ('virtue', 'NN'), ('of', 'IN'), ('its', 'PP$'), ('own', 'JJ'), (',', ','), ('as', 'CS'), ('in', 'IN'), ('marbles', 'NNS'), (',', ','), ('japonicas', None), (',', ','), ('and', 'CC'), ('pearls', 'NNS'), (';', ';'), ('and', 'CC'), ('though', 'CS'), ('various', 'JJ'), ('nations', 'NNS'), ('have', 'HV'), ('in', 'IN'), ('some', 'DTI'), ('way', 'NN'), ('recognised', 'VBD'), ('a', 'AT'), ('certain', 'JJ'), ('royal', 'JJ'), ('preeminence', None), ('in', 'IN'), ('this', 'DT'), ('hue', 'NN'), (';', ';'), ('even', 'RB'), ('the', 'AT'), ('barbaric', 'JJ'), (',', ','), ('grand', 'JJ'), ('old', 'JJ'), ('kings', 'NNS'), ('of', 'IN'), ('Pegu', None), ('placing', 'VBG'), ('the', 'AT'), ('title', 'NN'), (''', '''), ('Lord', 'NN-TL'), ('of', 'IN'), ('the', 'AT'), ('White', 'JJ-TL'), ('Elephants', None), (''', '''), ('above', 'IN'), ('all', 'ABN'), ('their', 'PP$'), ('other', 'AP'), ('magniloquent', None), ('ascriptions', None), ('of', 'IN'), ('dominion', 'NN')]
```

, (';', '.'), ('and', 'CC'), ('the', 'AT'), ('modern', 'JJ'), ('kings', 'NNS'), ('of', 'IN'), ('Siam', 'None'), ('unfurling', 'None'), ('the', 'AT'), ('same', 'AP'), ('snow', 'NN'), ('-', 'IN'), ('white', 'JJ'), ('quadruped', 'None'), ('in', 'IN'), ('the', 'AT'), ('royal', 'JJ'), ('standard', 'JJ'), (';', '.'), ('and', 'CC'), ('the', 'AT'), ('Hanoverian', 'JJ'), ('flag', 'NN'), ('bearing', 'VBG'), ('the', 'AT'), ('one', 'CD'), ('figure', 'NN'), ('of', 'IN'), ('a', 'AT'), ('snow', 'NN'), ('-', 'IN'), ('white', 'JJ'), ('charger', 'None'), (';', '.'), ('and', 'CC'), ('the', 'AT'), ('great', 'JJ'), ('Austrian', 'JJ'), ('Empire', 'NN-TL'), ('', ' '), ('Caesarian', 'None'), ('', ' '), ('heir', 'NN'), ('to', 'TO'), ('overlording', 'None'), ('Rome', 'NP'), ('', ' '), ('having', 'HVG'), ('for', 'IN'), ('the', 'AT'), ('imperial', 'JJ'), ('colour', 'None'), ('the', 'AT'), ('same', 'AP'), ('imperial', 'JJ'), ('hue', 'NN'), (';', '.'), ('and', 'CC'), ('though', 'CS'), ('this', 'DT'), ('pre', 'None'), ('-', 'IN'), ('eminence', 'NN'), ('in', 'IN'), ('it', 'PPS'), ('applies', 'VBZ'), ('to', 'TO'), ('the', 'AT'), ('human', 'JJ'), ('race', 'NN'), ('itself', 'PPL'), ('', ' '), ('giving', 'VBG'), ('the', 'AT'), ('white', 'JJ'), ('man', 'NN'), ('ideal', 'JJ'), ('mastership', 'None'), ('over', 'IN'), ('every', 'AT'), ('dusky', 'JJ'), ('tribe', 'NN'), (';', '.'), ('and', 'CC'), ('though', 'CS'), ('', ' '), ('besides', 'IN'), ('', ' '), ('all', 'ABN'), ('this', 'DT'), ('', ' '), ('whiteness', 'NN'), ('has', 'HVZ'), ('been', 'BEN'), ('even', 'RB'), ('made', 'VBN'), ('significant', 'JJ'), ('of', 'IN'), ('gladness', 'NN'), ('', ' '), ('for', 'IN'), ('among', 'IN'), ('the', 'AT'), ('Romans', 'NPS'), ('a', 'AT'), ('white', 'JJ'), ('stone', 'NN'), ('marked', 'VBN'), ('a', 'AT'), ('joyful', 'JJ'), ('day', 'NN'), (';', '.'), ('and', 'CC'), ('though', 'CS'), ('in', 'IN'), ('other', 'AP'), ('mortal', 'JJ'), ('sympathies', 'NNS'), ('and', 'CC'), ('symbolizings', 'None'), ('', ' '), ('this', 'DT'), ('same', 'AP'), ('hue', 'NN'), ('is', 'BEZ'), ('made', 'VBN'), ('the', 'AT'), ('emblem', 'None'), ('of', 'IN'), ('many', 'AP'), ('touching', 'VBG'), ('', ' '), ('noble', 'JJ'), ('things', 'NNS'), ('--', '--'), ('the', 'AT'), ('innocence', 'NN'), ('of', 'IN'), ('brides', 'NNS'), ('', ' '), ('the', 'AT'), ('benignity', 'None'), ('of', 'IN'), ('age', 'NN'), (';', '.'), ('though', 'CS'), ('among', 'IN'), ('the', 'AT'), ('Red', 'JJ-TL'), ('Men', 'NNS'), ('of', 'IN'), ('America', 'NP'), ('the', 'AT'), ('giving', 'VBG'), ('of', 'IN'), ('the', 'AT'), ('white', 'JJ'), ('belt', 'NN'), ('of', 'IN'), ('wampum', 'None'), ('was', 'BEDZ'), ('the', 'AT'), ('deepest', 'JJT'), ('pledge', 'NN'), ('of', 'IN'), ('honour', 'NN'), (';', '.'), ('though', 'CS'), ('in', 'IN'), ('many', 'AP'), ('climes', 'NNS'), ('', ' '), ('whiteness', 'NN'), ('typifies', 'None'), ('the', 'AT'), ('majesty', 'NN'), ('of', 'IN'), ('Justice', 'NN-TL'), ('in', 'IN'), ('the', 'AT'), ('ermine', 'None'), ('of', 'IN'), ('the', 'AT'), ('Judge', 'NN-TL'), ('', ' '), ('and', 'CC'), ('contributes', 'VBZ'), ('to', 'TO'), ('the', 'AT'), ('daily', 'RB'), ('state', 'NN'), ('of', 'IN'), ('kings', 'NNS'), ('and', 'CC'), ('queens', 'NNS'), ('drawn', 'VBN'), ('by', 'IN'), ('milk', 'NN'), ('-', 'IN'), ('white', 'JJ'), ('steeds', 'None'), (';', '.'), ('though', 'CS'), ('even', 'RB'), ('in', 'IN'), ('the', 'AT'), ('higher', 'JJR'), ('mysteries', 'NNS'), ('of', 'IN'), ('the', 'AT'), ('most', 'QL'), ('august', 'JJ'), ('religions', 'NNS'), ('it', 'PPS'), ('has', 'HVZ'), ('been', 'BEN'), ('made', 'VBN'), ('the', 'AT'), ('symbol', 'NN'), ('of', 'IN'), ('the', 'AT'), ('divine', 'JJ'), ('spotlessness', 'None'), ('and', 'CC'), ('power', 'NN'), (';', '.'), ('by', 'IN'), ('the', 'AT'), ('Persian', 'JJ'), ('fire', 'NN'), ('worshippers', 'NNS'), ('', ' '), ('the', 'AT'), ('white', 'JJ'), ('forked', 'VBN'), ('flame', 'NN'), ('being', 'BEG'), ('held', 'VBN'), ('the', 'AT'), ('holiest', 'None'), ('on', 'IN'), ('the', 'AT'), ('altar', 'NN'), (';', '.'), ('and', 'CC'), ('in', 'IN'), ('the', 'AT'), ('Greek', 'JJ'), ('mythologies', 'NNS'), ('', ' '), ('Great', 'JJ-TL'), ('Jove', 'None'), ('himself', 'PPL'), ('being', 'BEG'), ('made', 'VBN'), ('incarnate', 'VB'), ('in', 'IN'), ('a', 'AT'), ('snow', 'NN'), ('-', 'IN'), ('white', 'JJ'), ('bull', 'NN'), (';', '.'), ('and', 'CC'), ('though', 'CS'), ('to', 'TO'), ('the', 'AT'), ('noble', 'JJ'), ('Iroquois', 'NP-TL'), ('', ' '), ('the', 'AT'), ('midwinter', 'None'), ('sacrifice', 'NN'), ('of', 'IN'), ('the', 'AT'), ('sacred', 'JJ'), ('White', 'JJ-TL'), ('Dog', 'NN-TL'), ('was', 'BEDZ'), ('by', 'IN'), ('far', 'RB'), ('the', 'AT'), ('holiest', 'None'), ('festival', 'NN'), ('of', 'IN'), ('their', 'PP\$'), ('theology', 'NN'), ('', ' '), ('that', 'CS'), ('spotless', 'JJ'), ('', ' '), ('faithful', 'JJ'), ('creature', 'NN'), ('being', 'BEG'), ('held', 'VBN'), ('the', 'AT'), ('purest', 'JJT'), ('envoy', 'None'), ('they', 'PPSS'), ('could', 'MD'), ('send', 'VB'), ('to', 'TO'), ('the', 'AT'), ('Great', 'JJ-TL'), ('Spirit', 'NN-TL'), ('with', 'IN'), ('the', 'AT'), ('annual', 'JJ'), ('tidings', 'NNS'), ('of', 'IN'), ('their', 'PP\$'), ('own', 'JJ'), ('fidelity', 'NN'), (';', '.'), ('and', 'CC'), ('though', 'CS'), ('directly', 'RB'), ('from', 'IN'), ('the', 'AT'), ('Latin', 'JJ'), ('word', 'NN'), ('for', 'IN'), ('white', 'JJ'), ('', ' '), ('all', 'ABN'), ('Christian', 'JJ'), ('priests', 'NNS'), ('derive', 'VB'), ('the', 'AT'), ('name', 'NN'), ('of', 'IN'), ('one', 'CD'), ('part', 'NN'), ('of', 'IN'), ('their', 'PP\$'), ('sacred', 'JJ'), ('vesture', 'None'), ('', ' '), ('the', 'AT'), ('alb', 'None'), ('or', 'CC'), ('tunic', 'NN'), ('', ' '), ('worn', 'VBN'), ('beneath', 'IN'), ('the', 'AT'), ('cassock', 'None'), (';', '.'), ('and', 'CC'), ('though', 'CS'), ('among', 'IN'), ('the', 'AT'), ('holy', 'JJ'), ('pomps', 'None'), ('of', 'IN'), ('the', 'AT'), ('Romish', 'None'), ('faith', 'NN'), ('', ' '), ('white', 'JJ'), ('is', 'BEZ'), ('specially', 'RB'), ('employed', 'VBN'), ('in', 'IN'), ('the', 'AT'), ('celebration', 'NN'), ('of', 'IN'), ('the', 'AT'), ('Passion', 'NN-TL'), ('of', 'IN'), ('our', 'PP\$'), ('Lord', 'NN-TL'), (';', '.'), ('though', 'CS'), ('in', 'IN'), ('the', 'AT'), ('Vision', 'NN-TL'), ('of', 'IN'), ('St', 'None'), ('.', '.'), ('John', 'NP'), ('', ' '), ('white', 'JJ'), ('robes', 'NNS'), ('are', 'BER'), ('given', 'VBN'), ('to', 'TO'), ('the', 'AT'), ('redeemed', 'VBN'), ('', ' '), ('and', 'CC'), ('the', 'AT'), ('four', 'CD'), ('-', 'IN'), ('and', 'CC'), ('-', 'IN'), ('twenty', 'CD'), ('elders', 'NNS'), ('stand', 'VB'), ('clothed', 'VBN'), ('in', 'IN'), ('white', 'JJ'), ('before', 'CS'), ('the', 'AT'), ('great', 'JJ'), ('-', 'IN'), ('white', 'JJ'), ('throne', 'NN'), ('', ' '), ('and', 'CC'), ('the', 'AT'), ('Holy', 'JJ-TL'), ('One', 'CD'), ('that', 'CS'), ('sitteth', 'None'), ('there', 'EX'), ('white', 'JJ'), ('like', 'CS'), ('wool', 'NN'), (';', '.'), ('yet', 'RB'), ('for', 'IN'), ('all', 'ABN'), ('these', 'DTS'), ('accumulated', 'VBN'), ('associations', 'NNS'), ('', ' '), ('with', 'IN'), ('whatever', 'WDT'), ('is', 'BEZ'), ('sweet', 'JJ'), ('', ' '), ('and', 'CC'), ('honourable', 'None'), ('', ' '), ('and', 'CC'), ('sublime', 'NN'), ('', ' '), ('there', 'EX'), ('yet', 'RB'), ('lurks', 'VBZ'), ('an', 'AT'), ('elusive', 'JJ'), ('something', 'PN'), ('in', 'IN'), ('the', 'AT'), ('innermost', 'JJS'), ('idea', 'NN'), ('of', 'IN'), ('this', 'DT'), ('hue', 'NN'), ('', ' '), ('which', 'WDT'), ('strikes', 'NNS'), ('more', 'QL'), ('of', 'IN'), ('panic', 'NN'), ('to', 'TO'), ('the', 'AT'), ('soul', 'NN'), ('than', 'CS'), ('that', 'CS'), ('redness', 'None'), ('which', 'WDT'), ('affrights', 'None'), ('in', 'IN'), ('blood', 'NN'), ('.', '.))]

Out[4]: 0.9245482228606466

In [17]: *#Regex tagger , Patterns referred from BKL Chapter 5*

```
patterns = [
    (r'.*ing$', 'VBG'),          # Gerunds
    (r'.*ed$', 'VBD'),          # Simple Past Tense
    (r'.*es$', 'VBZ'),          # 3rd Singular Present Tense
    (r'.*ould$', 'MD'),         # Modals
    (r'.*\'s$', 'NN$'),          # Possessive Nouns
    (r'.*s$', 'NNS'),           # Plural Nouns
    (r'^-?[0-9]+(\.[0-9]+)?$', 'CD'), # Cardinal Numbers
    (r'.*', 'NN')               # Nouns (Default)
]
regexTagger = nltk.RegexpTagger(patterns)
print(regexTagger.tag(nltk.tokenize.word_tokenize(longestSentence)))
regexTagger.evaluate(brownTaggedSents) #Print the accuracy of the regex tagger
```

```
[('Though', 'NN'), ('in', 'NN'), ('many', 'NN'), ('natural', 'NN'), ('objects', 'NNS'), (',', 'NN'), ('
whiteness', 'NNS'), ('refiningly', 'NN'), ('enhances', 'VBZ'), ('beauty', 'NN'), (',', 'NN'), ('as', 'N
NS'), ('if', 'NN'), ('imparting', 'VBG'), ('some', 'NN'), ('special', 'NN'), ('virtue', 'NN'), ('of', '
NN'), ('its', 'NNS'), ('own', 'NN'), (',', 'NN'), ('as', 'NNS'), ('in', 'NN'), ('marbles', 'VBZ'), (',',
', 'NN'), ('japonicas', 'NNS'), (',', 'NN'), ('and', 'NN'), ('pearls', 'NNS'), (';', 'NN'), ('and', 'NN'
), ('though', 'NN'), ('various', 'NNS'), ('nations', 'NNS'), ('have', 'NN'), ('in', 'NN'), ('some', 'NN
'), ('way', 'NN'), ('recognised', 'VBD'), ('a', 'NN'), ('certain', 'NN'), ('royal', 'NN'), ('preeminenc
e', 'NN'), ('in', 'NN'), ('this', 'NNS'), ('hue', 'NN'), (';', 'NN'), ('even', 'NN'), ('the', 'NN'), ('
barbaric', 'NN'), (',', 'NN'), ('grand', 'NN'), ('old', 'NN'), ('kings', 'NNS'), ('of', 'NN'), ('Pegu',
'NN'), ('placing', 'VBG'), ('the', 'NN'), ('title', 'NN'), ('`', 'NN'), ('Lord', 'NN'), ('of', 'NN'),
('the', 'NN'), ('White', 'NN'), ('Elephants', 'NNS'), ('`', 'NN'), ('above', 'NN'), ('all', 'NN'), ('t
heir', 'NN'), ('other', 'NN'), ('magniloquent', 'NN'), ('ascriptions', 'NNS'), ('of', 'NN'), ('dominion
', 'NN'), (';', 'NN'), ('and', 'NN'), ('the', 'NN'), ('modern', 'NN'), ('kings', 'NNS'), ('of', 'NN'),
('Siam', 'NN'), ('unfurling', 'VBG'), ('the', 'NN'), ('same', 'NN'), ('snow', 'NN'), ('-', 'NN'), ('whi
te', 'NN'), ('quadruped', 'VBD'), ('in', 'NN'), ('the', 'NN'), ('royal', 'NN'), ('standard', 'NN'), (';
', 'NN'), ('and', 'NN'), ('the', 'NN'), ('Hanoverian', 'NN'), ('flag', 'NN'), ('bearing', 'VBG'), ('the
', 'NN'), ('one', 'NN'), ('figure', 'NN'), ('of', 'NN'), ('a', 'NN'), ('snow', 'NN'), ('-', 'NN'), ('wh
ite', 'NN'), ('charger', 'NN'), (';', 'NN'), ('and', 'NN'), ('the', 'NN'), ('great', 'NN'), ('Austrian'
, 'NN'), ('Empire', 'NN'), (',', 'NN'), ('Caesarian', 'NN'), (',', 'NN'), ('heir', 'NN'), ('to', 'NN'),
('overlording', 'VBG'), ('Rome', 'NN'), (',', 'NN'), ('having', 'VBG'), ('for', 'NN'), ('the', 'NN'), ('
imperial', 'NN'), ('colour', 'NN'), ('the', 'NN'), ('same', 'NN'), ('imperial', 'NN'), ('hue', 'NN'),
(';', 'NN'), ('and', 'NN'), ('though', 'NN'), ('this', 'NNS'), ('pre', 'NN'), ('-', 'NN'), ('eminence',
'NN'), ('in', 'NN'), ('it', 'NN'), ('applies', 'VBZ'), ('to', 'NN'), ('the', 'NN'), ('human', 'NN'), ('
race', 'NN'), ('itself', 'NN'), (',', 'NN'), ('giving', 'VBG'), ('the', 'NN'), ('white', 'NN'), ('man',
'NN'), ('ideal', 'NN'), ('mastership', 'NN'), ('over', 'NN'), ('every', 'NN'), ('dusky', 'NN'), ('tribe
', 'NN'), (';', 'NN'), ('and', 'NN'), ('though', 'NN'), (',', 'NN'), ('besides', 'VBZ'), (',', 'NN'), ('
all', 'NN'), ('this', 'NNS'), (',', 'NN'), ('whiteness', 'NNS'), ('has', 'NNS'), ('been', 'NN'), ('eve
n', 'NN'), ('made', 'NN'), ('significant', 'NN'), ('of', 'NN'), ('gladness', 'NNS'), (',', 'NN'), ('for
', 'NN'), ('among', 'NN'), ('the', 'NN'), ('Romans', 'NNS'), ('a', 'NN'), ('white', 'NN'), ('stone', 'N
N'), ('marked', 'VBD'), ('a', 'NN'), ('joyful', 'NN'), ('day', 'NN'), (';', 'NN'), ('and', 'NN'), ('tho
ugh', 'NN'), ('in', 'NN'), ('other', 'NN'), ('mortal', 'NN'), ('sympathies', 'VBZ'), ('and', 'NN'), ('s
ymbolizings', 'NNS'), (',', 'NN'), ('this', 'NNS'), ('same', 'NN'), ('hue', 'NN'), ('is', 'NNS'), ('mad
e', 'NN'), ('the', 'NN'), ('emblem', 'NN'), ('of', 'NN'), ('many', 'NN'), ('touching', 'VBG'), (',', 'N
N'), ('noble', 'NN'), ('things', 'NNS'), ('--', 'NN'), ('the', 'NN'), ('innocence', 'NN'), ('of', 'NN')
, ('brides', 'VBZ'), (',', 'NN'), ('the', 'NN'), ('benignity', 'NN'), ('of', 'NN'), ('age', 'NN'), (';
', 'NN'), ('though', 'NN'), ('among', 'NN'), ('the', 'NN'), ('Red', 'VBD'), ('Men', 'NN'), ('of', 'NN'),
('America', 'NN'), ('the', 'NN'), ('giving', 'VBG'), ('of', 'NN'), ('the', 'NN'), ('white', 'NN'), ('be
lt', 'NN'), ('of', 'NN'), ('wampum', 'NN'), ('was', 'NNS'), ('the', 'NN'), ('deepest', 'NN'), ('pledge'
, 'NN'), ('of', 'NN'), ('honour', 'NN'), (';', 'NN'), ('though', 'NN'), ('in', 'NN'), ('many', 'NN'), ('
climes', 'VBZ'), (',', 'NN'), ('whiteness', 'NNS'), ('typifies', 'VBZ'), ('the', 'NN'), ('majesty', 'N
N'), ('of', 'NN'), ('Justice', 'NN'), ('in', 'NN'), ('the', 'NN'), ('ermine', 'NN'), ('of', 'NN'), ('th
e', 'NN'), ('Judge', 'NN'), (',', 'NN'), ('and', 'NN'), ('contributes', 'VBZ'), ('to', 'NN'), ('the', '
NN'), ('daily', 'NN'), ('state', 'NN'), ('of', 'NN'), ('kings', 'NNS'), ('and', 'NN'), ('queens', 'NNS'
), ('drawn', 'NN'), ('by', 'NN'), ('milk', 'NN'), ('-', 'NN'), ('white', 'NN'), ('steeds', 'NNS'), (';
', 'NN'), ('though', 'NN'), ('even', 'NN'), ('in', 'NN'), ('the', 'NN'), ('higher', 'NN'), ('mysteries'
, 'VBZ'), ('of', 'NN'), ('the', 'NN'), ('most', 'NN'), ('august', 'NN'), ('religions', 'NNS'), ('it', 'N
N'), ('has', 'NNS'), ('been', 'NN'), ('made', 'NN'), ('the', 'NN'), ('symbol', 'NN'), ('of', 'NN'), ('th
e', 'NN'), ('divine', 'NN'), ('spotlessness', 'NNS'), ('and', 'NN'), ('power', 'NN'), (';', 'NN'), ('by
', 'NN'), ('the', 'NN'), ('Persian', 'NN'), ('fire', 'NN'), ('worshippers', 'NNS'), (',', 'NN'), ('the
', 'NN'), ('white', 'NN'), ('forked', 'VBD'), ('flame', 'NN'), ('being', 'VBG'), ('held', 'NN'), ('the',
'NN'), ('holiest', 'NN'), ('on', 'NN'), ('the', 'NN'), ('altar', 'NN'), (';', 'NN'), ('and', 'NN'), ('i
n', 'NN'), ('the', 'NN'), ('Greek', 'NN'), ('mythologies', 'VBZ'), (',', 'NN'), ('Great', 'NN'), ('Jove
', 'NN'), ('himself', 'NN'), ('being', 'VBG'), ('made', 'NN'), ('incarnate', 'NN'), ('in', 'NN'), ('a',
'NN'), ('snow', 'NN'), ('-', 'NN'), ('white', 'NN'), ('bull', 'NN'), (';', 'NN'), ('and', 'NN'), ('thou
gh', 'NN'), ('to', 'NN'), ('the', 'NN'), ('noble', 'NN'), ('Iroquois', 'NNS'), (',', 'NN'), ('the', 'NN
'), ('midwinter', 'NN'), ('sacrifice', 'NN'), ('of', 'NN'), ('the', 'NN'), ('sacred', 'VBD'), ('White',
'NN'), ('Dog', 'NN'), ('was', 'NNS'), ('by', 'NN'), ('far', 'NN'), ('the', 'NN'), ('holiest', 'NN'), ('
festival', 'NN'), ('of', 'NN'), ('their', 'NN'), ('theology', 'NN'), (',', 'NN'), ('that', 'NN'), ('spo
tless', 'NNS'), (',', 'NN'), ('faithful', 'NN'), ('creature', 'NN'), ('being', 'VBG'), ('held', 'NN'),
('the', 'NN'), ('purest', 'NN'), ('envoy', 'NN'), ('they', 'NN'), ('could', 'MD'), ('send', 'NN'), ('to
', 'NN'), ('the', 'NN'), ('Great', 'NN'), ('Spirit', 'NN'), ('with', 'NN'), ('the', 'NN'), ('annual', '
NN'), ('tidings', 'NNS'), ('of', 'NN'), ('their', 'NN'), ('own', 'NN'), ('fidelity', 'NN'), (';', 'NN')
, ('and', 'NN'), ('though', 'NN'), ('directly', 'NN'), ('from', 'NN'), ('the', 'NN'), ('Latin', 'NN'),
('word', 'NN'), ('for', 'NN'), ('white', 'NN'), (',', 'NN'), ('all', 'NN'), ('Christian', 'NN'), ('prie
```

```

sts', 'NNS'), ('derive', 'NN'), ('the', 'NN'), ('name', 'NN'), ('of', 'NN'), ('one', 'NN'), ('part', 'N
N'), ('of', 'NN'), ('their', 'NN'), ('sacred', 'VBD'), ('vesture', 'NN'), ('', 'NN'), ('the', 'NN'), ('
alb', 'NN'), ('or', 'NN'), ('tunic', 'NN'), ('', 'NN'), ('worn', 'NN'), ('beneath', 'NN'), ('the', 'N
N'), ('cassock', 'NN'), (';', 'NN'), ('and', 'NN'), ('though', 'NN'), ('among', 'NN'), ('the', 'NN'), ('
holy', 'NN'), ('pomps', 'NNS'), ('of', 'NN'), ('the', 'NN'), ('Romish', 'NN'), ('faith', 'NN'), ('', '
NN'), ('white', 'NN'), ('is', 'NNS'), ('specially', 'NN'), ('employed', 'VBD'), ('in', 'NN'), ('the',
'NN'), ('celebration', 'NN'), ('of', 'NN'), ('the', 'NN'), ('Passion', 'NN'), ('of', 'NN'), ('our', 'NN
'), ('Lord', 'NN'), (';', 'NN'), ('though', 'NN'), ('in', 'NN'), ('the', 'NN'), ('Vision', 'NN'), ('of'
, 'NN'), ('St', 'NN'), ('.', 'NN'), ('John', 'NN'), ('', 'NN'), ('white', 'NN'), ('robes', 'VBZ'), ('a
re', 'NN'), ('given', 'NN'), ('to', 'NN'), ('the', 'NN'), ('redeemed', 'VBD'), ('', 'NN'), ('and', 'NN
'), ('the', 'NN'), ('four', 'NN'), ('-', 'NN'), ('and', 'NN'), ('-', 'NN'), ('twenty', 'NN'), ('elders'
, 'NNS'), ('stand', 'NN'), ('clothed', 'VBD'), ('in', 'NN'), ('white', 'NN'), ('before', 'NN'), ('the',
'NN'), ('great', 'NN'), ('-', 'NN'), ('white', 'NN'), ('throne', 'NN'), ('', 'NN'), ('and', 'NN'), ('t
he', 'NN'), ('Holy', 'NN'), ('One', 'NN'), ('that', 'NN'), ('sitteth', 'NN'), ('there', 'NN'), ('white'
, 'NN'), ('like', 'NN'), ('wool', 'NN'), (';', 'NN'), ('yet', 'NN'), ('for', 'NN'), ('all', 'NN'), ('th
ese', 'NN'), ('accumulated', 'VBD'), ('associations', 'NNS'), ('', 'NN'), ('with', 'NN'), ('whatever',
'NN'), ('is', 'NNS'), ('sweet', 'NN'), ('', 'NN'), ('and', 'NN'), ('honourable', 'NN'), ('', 'NN'), ('
and', 'NN'), ('sublime', 'NN'), ('', 'NN'), ('there', 'NN'), ('yet', 'NN'), ('lurks', 'NNS'), ('an',
'NN'), ('elusive', 'NN'), ('something', 'VBG'), ('in', 'NN'), ('the', 'NN'), ('innermost', 'NN'), ('ide
a', 'NN'), ('of', 'NN'), ('this', 'NNS'), ('hue', 'NN'), ('', 'NN'), ('which', 'NN'), ('strikes', 'VBZ
'), ('more', 'NN'), ('of', 'NN'), ('panic', 'NN'), ('to', 'NN'), ('the', 'NN'), ('soul', 'NN'), ('than'
, 'NN'), ('that', 'NN'), ('redness', 'NNS'), ('which', 'NN'), ('affrights', 'NNS'), ('in', 'NN'), ('blo
od', 'NN'), ('.', 'NN')]

```

**Out[17]:** 0.19485235861080683

**Accuracy of the Lookup Tagger = 92%**

**Accuracy of the Regex Tagger = 19%**

The lookup tagger is definitely more accurate than the regex tagger. The regex tagger accuracy can be improved by better rules for assigning the tags. It can also be improved by adding more rules.

Applying embeddings would improve tagging accuracy. Since word embeddings help map semantically similar words to each other, it would increase the tagging accuracy in cases of ambiguity.