# exp1

## January 23, 2024

Name: Hatim Sawai UID: 2021300108 Batch: A

##

Experiment 1

#### Aim:

- 1. Install NLTK and perform basic Corpus analysis using NLTK such as frequency distribution
- 2. Learn about morphological features of a word by analysing it.

# 0.0.1 1. Installig NLTK and downloading the required corpus

```
[]: import nltk
     from nltk import FreqDist
     from nltk.tokenize import word_tokenize, sent_tokenize
     from nltk.tag import pos_tag
     from nltk.stem import WordNetLemmatizer
     from nltk.corpus import wordnet
     from prettytable import PrettyTable
[]: nltk.download("punkt")
     nltk.download("averaged_perceptron_tagger")
    nltk.download("wordnet")
    [nltk_data] Downloading package punkt to
    [nltk_data]
                    C:\Users\hatim\AppData\Roaming\nltk_data...
                  Package punkt is already up-to-date!
    [nltk_data]
    [nltk_data] Downloading package averaged_perceptron_tagger to
                    C:\Users\hatim\AppData\Roaming\nltk_data...
    [nltk_data]
    [nltk_data]
                  Package averaged_perceptron_tagger is already up-to-
    [nltk_data]
                      date!
    [nltk_data] Downloading package wordnet to
    [nltk_data]
                    C:\Users\hatim\AppData\Roaming\nltk_data...
    [nltk_data]
                  Package wordnet is already up-to-date!
[]: True
```

## 0.0.2 2. Morphological analysis of a words

```
[]: pos_mapping = {
         "CC": "Coordinating conjunction",
         "CD": "Cardinal number",
         "DT": "Determiner",
         "EX": "Existential there",
         "FW": "Foreign word",
         "IN": "Preposition or subordinating conjunction",
         "JJ": "Adjective",
         "JJR": "Adjective, comparative",
         "JJS": "Adjective, superlative",
         "LS": "List item marker",
         "MD": "Modal",
         "NN": "noun, singular or mass",
         "NNS": "noun, plural",
         "NNP": "Proper noun, singular",
         "NNPS": "Proper noun, plural",
         "PDT": "Predeterminer",
         "POS": "Possessive ending",
         "PRP": "Personal pronoun",
         "PRP$": "Possessive pronoun",
         "RB": "Adverb",
         "RBR": "Adverb, comparative",
         "RBS": "Adverb, superlative",
         "RP": "Particle",
         "SYM": "Symbol",
         "TO": "to",
         "UH": "Interjection",
         "VB": "Verb, base form",
         "VBD": "Verb, past tense",
         "VBG": "Verb, gerund or present participle",
         "VBN": "Verb, past participle",
         "VBP": "Verb, non3rd person singular present",
         "VBZ": "Verb, 3rd person singular present",
         "WDT": "Whdeterminer",
         "WP": "Whpronoun",
         "WP$": "Possessive whpronoun",
         "WRB": "Whadverb",
     }
     def get_wordnet_pos(tag):
         if tag.startswith("N"):
             return wordnet.NOUN
         elif tag.startswith("V"):
             return wordnet. VERB
         elif tag.startswith("R"):
```

```
return wordnet.ADV
    elif tag.startswith("J"):
        return wordnet.ADJ
    else:
        return wordnet.NOUN
def get_category(tag):
    if tag.startswith("N"):
        return "Noun"
    elif tag.startswith("V"):
        return "Verb"
    elif tag.startswith("R"):
        return "Adverb"
    elif tag.startswith("J"):
        return "Adjective"
    else:
        return "Noun"
```

```
[]: def analyze_sentence(sentence):
         words = word_tokenize(sentence)
         tags = pos_tag(words)
         maleWords = [
             "he",
             "him",
             "his",
             "himself",
             "boy",
             "sir",
             "man",
             "gentleman",
             "father",
             "son",
             "brother",
             "uncle",
             "nephew",
             "grandfather",
             "grandson",
             "king",
             "prince",
             "husband",
             "groom",
         ]
         femaleWords = [
             "she",
             "her",
             "hers",
             "hersef",
```

```
"girl",
       "madam",
       "lady",
       "woman",
      "mother".
      "daughter",
      "sister",
      "aunt",
      "niece",
      "grandmother",
      "granddaughter",
      "queen",
      "princess",
      "wife",
      "bride".
      "widow",
  ]
  lemmatizer = WordNetLemmatizer()
  morphological_table = PrettyTable()
  morphological_table.field_names = ["Root", "Category", "Gender", "Number", u
⇔"Tense"]
  print(f"\nSentence: {sentence}")
  for i in range(len(words)):
      root = lemmatizer.lemmatize(words[i], get_wordnet_pos(tags[i][1]))
      category = pos_mapping[tags[i][1]] if tags[i][1] in pos_mapping else_u
→tags[i][1]
      if words[i].lower() in maleWords:
          gender = "male"
      elif words[i].lower() in femaleWords:
          gender = "female"
       # check if word is a pronoun
      elif "Proper noun" in category and (
          words[i].endswith("i")
          or words[i].endswith("a")
          or words[i].endswith("e")
          or words[i].endswith("y")
      ):
          gender = "female"
      elif "Proper noun" in category:
          gender = "male"
      elif "pronoun" in category:
          gender = "Can't Assume"
      else:
          gender = "neutral"
      # find frequency of word in corpus
      number = sentence.count(words[i])
```

```
# check if word is a verb
      if tags[i][1].startswith("V"):
           # determine tense of the verb
           if (
               words[i].endswith("ed")
               or words[i - 1].lower() == "had"
               or words[i - 1].lower() == "was"
               or words[i - 1].lower() == "were"
           ):
               tense = "past"
           elif (
               words[i].endswith("ing")
               or words[i].endswith("s")
               or words[i - 1].lower() == "is"
               or words[i - 1].lower() == "are"
           ):
               tense = "present"
           elif words[i - 1].lower() == "will" or words[i - 1].lower() ==__

¬"shall":

               tense = "future"
           else:
              tense = "present"
       else:
           tense = "NA"
       # print(f"Root: \{root\}, Category: \{category\}, Gender: \{gender\}, Number: \square
→{number}, Tense: {tense}")
      morphological_table add_row([root, category, gender, number, tense])
  print(morphological_table)
```

## 0.0.3 3. Reading a text file and Tokenizination

```
[]: # read input file
with open("input.txt", "r") as f:
    text = f.read()

# tokenise sentences
sentences = sent_tokenize(text)
```

```
[]: # Analyze each sentence
for sentence in sentences:
    analyze_sentence(sentence)
```

Sentence: The quick brown fox jumps over The lazy dog.

+-----

+

Root   	Category	Gender	Number	Tense
+		-+	-+	-+
+   The	Determiner	neutral	2	l NA
quick	Adjective	neutral	1	l NA
brown	noun, singular or mass	neutral	1	l NA
fox	noun, singular or mass	neutral	1	l NA
   jump   V	erb, 3rd person singular present	neutral	1	present
over   Prep	osition or subordinating conjunction	neutral	l 1	l NA
The I	Determiner	neutral	1 2	l NA
lazy   	Adjective	neutral	l 1	l NA
dog   	noun, singular or mass	neutral	l 1	l NA
i . I	•	neutral	l 1	l NA
++		-+	-+	-+
Sentence: She	plays The piano beautifully.			
+	+	+   Gender	•	•
+   She	+   Personal pronoun	+   female		++   NA
•	Verb, 3rd person singular present			
The		neutral		
l piano	noun, singular or mass			
beautifully	_			I NA I
1 .		neutral	1	I NA I
Sentence: The	sun sets in The west every evening.			
+   Root   	Category	Gend	er   Numb	er   Tense
+   The   	Determiner		al   2	

I	sun	noun, singular or mass	neutra	al	1	1	NA		
	set	noun, plural	neutra	al	1	1	NA		
	in	Preposition or subordinating conjunction	on   neutra	al	2	1	NA		
	The	Determiner	neutra	al	2	I	NA		
	west	noun, singular or mass	neutra	al	1	ı	NA		
1	every	Determiner	neutra	al	1	ı	NA		
 	evening	noun, singular or mass	neutra	al	1	ı	NA		
 			neutra		1	1	NA		
 +-		· +		+		· +			
+		•	•	•		•			
Se	ntence:	John and Mary are going to The beach tomo	orrow.						
+-	Root	9 0	Gender		+- er	Ten	+ se		
+-	John	-+	   male		+- 	 N	+ A		
1	and	-	neutral	1	- 1	N.	A		
İ	Mary		female		i	N.	A İ		
i	be	Verb, non3rd person singular present			i	pres			
i		8 1		•	·	1			
i	go	Verb, gerund or present participle	neutral	1	- 1	pres	ent		
i	to		neutral		i	N.			
i	The		neutral		i	N.			
1	beach		neutral	•		N.			
					1				
	tomorrow		neutral				A I		
- 1	•		neutral	1	1		A I		
Se	++ Sentence: The delicious aroma of freshly baked bread fills The air. ++								
	+								
	Root	Category	Ger	nder	Numb	er			
	nse	+							
		+	+	+		+			
1		Determiner	1 2000	tral	n	ı			
l nta	The	l pererminer	i neu	tral	2	ı			
NA	•		1	h7 '	4				
	deliciou	s   Adjective	) neu	tral	1	ı			
NA 	aroma	noun, singular or mass	neu	tral	1	1			

NA I							
l of	Preposition or subordinating con	iunction	neutral	1			
NA I	Trapobleton of buseralmating conjunction   neutral   T						
freshly	Adjective	1					
NA		·					
l bake	Verb, past participle	1 l					
past		·					
bread	noun, singular or mass	noun, singular or mass   neutral					
NA	, , , , , , , , , , , , , , , , , , , ,	•	,	·			
fill	Verb, 3rd person singular pre	sent	neutral	1			
present							
The	Determiner		neutral	2			
NA							
air	noun, singular or mass		neutral	1			
NA	, ,						
1			neutral	1			
NA							
+	+	+-	+	+			
+							
Sentence: 5	The old library is a quiet and peace	ful place.					
+	-+	-+	+	+	+		
Root	Category	Gender	Number	Tense	l		
+	-+	-+	+	+	F		
•		•					
The	Determiner	neutral	1	l NA			
The old	Determiner Adjective	neutral neutral	_	NA NA	   		
•			1		   		
old	Adjective	neutral   neutral	1   1	l NA	     		
old   library	Adjective noun, singular or mass	neutral   neutral	1   1   1	I NA I	       		
old   library   be	Adjective noun, singular or mass Verb, 3rd person singular present	neutral   neutral   neutral	1   1   1   5	NA NA NA NA NA NA NA			
old library be a	Adjective noun, singular or mass Verb, 3rd person singular present Determiner	neutral   neutral   neutral   neutral	1   1   1   5   1	NA NA NA NA NA NA NA NA NA	         		
old library be a quiet	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction	neutral   neutral   neutral   neutral	1   1   1   5   1	NA NA present NA NA			
old library be a quiet and	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction Adjective	neutral   neutral   neutral   neutral   neutral	1   1   1   5   1   1	NA NA Present NA NA NA NA			
old library be a quiet and peaceful	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction	neutral   neutral   neutral   neutral   neutral   neutral   neutral	1   1   1   5   1   1   1	NA NA Present NA NA NA NA NA			
old library be a quiet and peaceful place	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction Adjective	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral	1   1   5   1   1   1   1	NA NA Present NA NA NA NA NA NA NA NA NA NA NA	             		
old library be a quiet and peaceful place	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction Adjective noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral	1   1   5   1   1   1   1	NA NA Present NA NA NA NA NA NA NA NA NA NA NA			
old library be a quiet and peaceful place	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction Adjective noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral	1   1   5   1   1   1   1	NA NA Present NA NA NA NA NA NA NA NA NA NA NA			
old   library   be   a   quiet   and   peaceful   place   .	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction Adjective noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral	1	NA NA Present NA NA NA NA NA NA NA NA NA			
old   library   be   a   quiet   and   peaceful   place   .	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral	1	NA NA Present NA NA NA NA NA NA NA NA NA			
old   library   be   a   quiet   and   peaceful   place   . +	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction Adjective noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   ts of their	1	NA NA Present NA NA NA NA NA NA NA NA NA NA	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral	1	NA NA Present NA NA NA NA NA NA NA NA NA NA	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective noun, singular or mass Verb, 3rd person singular present Determiner Adjective Coordinating conjunction Adjective noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   ts of their	1	NA NA Present NA NA NA NA NA NA NA NA NA NA NA NA NA	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   ts of their	1	NA NA Present NA NA NA NA NA NA NA NA NA NA NA NA NA	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   tral   neutral	1	NA   NA   NA   Present   NA   NA   NA   NA   NA   NA   NA   N	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   tral   neutral	1	NA   NA   NA   Present   NA   NA   NA   NA   NA   NA   NA   N	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   to of their	1	NA	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   to of their	1	NA	-+		
old   library   be   a   quiet   and   peaceful   place   . +	Adjective   noun, singular or mass   Verb, 3rd person singular present   Determiner   Adjective   Coordinating conjunction   Adjective   noun, singular or mass	neutral   neutral   neutral   neutral   neutral   neutral   neutral   neutral   to of their	1	NA   NA   NA   Present   NA   NA   NA   NA   NA   NA   NA   N	-+		

NTA I						
	Verb, non3rd person singular present	I	neutral	1	1	I
present     The	Determiner	I	neutral	1	2	1
NA     result	noun, plural	I	neutral	1	1	I
NA     of	Preposition or subordinating conjunction	I	neutral	1	1	1
NA     their	Possessive pronoun	I	Can't Assu	ıme	1	1
NA     exam	noun, plural	1	neutral	1	1	I
NA     .	1 .	1	neutral	1	1	I
NA I	-+	-+		+		+
+						
Sentence:	The dog chased The playful cat around The	co.	om.			
+		-+	·		·	
Root 	<b>5</b> •	I	,			
++	-+	-+	+-		+	
The	Determiner	I	neutral	3		NA
l dog	noun, singular or mass	I	neutral	1	1	NA
chase	Verb, past tense	I	neutral	1	1	past
The	Determiner	I	neutral	3	1	NA
playful	Adjective	I	neutral	1	I	NA
   cat	noun, singular or mass	I	neutral	1	I	NA
   around	Preposition or subordinating conjunction	I	neutral	1	1	NA
   The	Determiner	ı	neutral	3	I	NA
   room	noun, singular or mass	ı	neutral	1	1	NA
1 1 .	1 .	1	neutral	1	I	NA
 +	-+		+-			
			•		•	

9

+

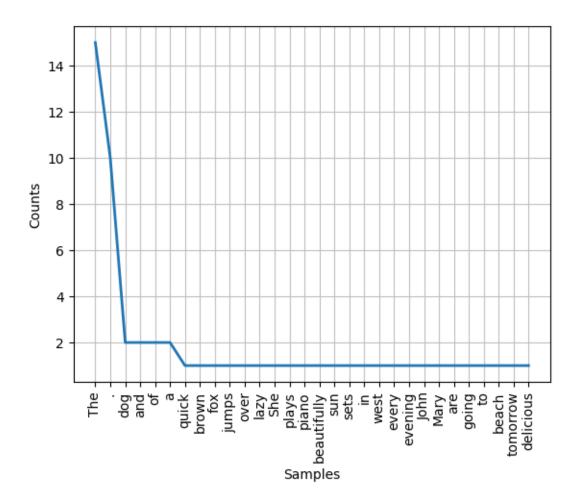
+					
Root Tense		Gender	Numbe		
·+	+	+	+		
We	Personal pronoun	Can't Assume	e   1		
A					
enjoy	Verb, past tense	neutral	1		
ast	Dotorminor	noutrol	l 6		
a A l	Decerminer	Determiner   neutral			
relaxing	Adjective	1			
A	· •	neutral			
vacation	noun, singular or mass	neutral	1		
A					
on	Preposition or subordinating conjunction	neutral	2		
A	Data	l manutur-3	4		
The A	Determiner	neutral	1		
tropical	Adjective	neutral	1		
A	najosorvo	, moderal	, -		
island	noun, singular or mass	neutral	1		
I A					
•		neutral	1		
ſΛ I					
•		1			
·	+	+	+		
·	+	+	+		
+	am Hatim, taking this NLP course taught by	Reeta mam.			
	am Hatim, taking this NLP course taught by	Reeta mam.			
+ entence: ] +	am Hatim, taking this NLP course taught by	Reeta mam. +		+-	
+ entence: I+ Root	am Hatim, taking this NLP course taught by	Reeta mam. +		+-	
Gentence: I+ Root   Cense	am Hatim, taking this NLP course taught by	Reeta mam. + Gender	Number	+-	
Gentence: I+ Root   Cense	am Hatim, taking this NLP course taught by	Reeta mam. + Gender	Number	+-	
entence: I+ Root	am Hatim, taking this NLP course taught by  Category	Reeta mam. + Gender	Number	+-	
entence: I+ Root   ense  + I	Category   Personal pronoun	Reeta mam+ Gender  + Can't Assume	Number	+-	
entence: I+ Root   ense  + I   A   be	am Hatim, taking this NLP course taught by  Category	Reeta mam. 	Number 1	+-	
entence: I+ Root   ense  + I   A   be   resent	Category   Personal pronoun   Verb, non3rd person singular present	Reeta mam+ Gender  + Can't Assume   neutral	Number  1 2	+-	
Gentence: I+ Root   Gense  + I   A   be   Dresent   Hatim	Category   Personal pronoun	Reeta mam+ Gender  + Can't Assume	Number	+-	
Sentence: I	Category  Personal pronoun  Verb, non3rd person singular present  Proper noun, singular	Reeta mam	Number  1 2 1	+-	
Gentence: I+ Root   Cense  + I   IA   be   Dresent   Hatim	Category   Personal pronoun   Verb, non3rd person singular present	Reeta mam+ Gender  + Can't Assume   neutral	Number  1 2	+-	
Sentence: I	Category  Personal pronoun  Verb, non3rd person singular present  Proper noun, singular  ,	Reeta mam.  Gender    Can't Assume    neutral    male    neutral	Number  1 2 1	+-	
Sentence: I+ Root   Tense  + I     JA   Hatim   JA   JA	Category  Personal pronoun  Verb, non3rd person singular present  Proper noun, singular	Reeta mam	Number  1 2 1	+-	

NA I						
NLP	Proper noun, singular	1	male	1	1	1
NA I						
course	noun, singular or mass	ı	neutral	ı	1	l
NA	. <u>.</u>					
taught	noun, singular or mass	ı	neutral	ı	1	I
NA I			_			
•	Preposition or subordinating conjunction	ı	neutral	ı	1	I
NA						
Reeta	Proper noun, singular	ı	female	ı	1	ı
NA I						
mam	noun, singular or mass	ı	neutral	l	1	ı
NA I						
1 . 1	•	ı	neutral	l	1	ı
NA						
++-		-+		+		+
+						

# 0.0.4 4. Performing frequency distribution on the tokens

```
[]: tokens = word_tokenize(text)
  fdist = FreqDist(tokens)
  freq_table = fdist.tabulate()
  # Display frequency distribution as a bar graph
  fdist.plot(30, cumulative=False)
```

	The	•	dog	and	of	a
quick	brown	fox	jumps	over	lazy	She
plays	piano	beautifully	sun	sets	in	west
every	evening	John	Mary	are	going	to
beach	tomorrow	delicious	aroma	freshly	baked	bread
fills	air	old	library	is	quiet	peaceful
place	students	eagerly	await	results	their	exams
chased	playfu	l cat	around	room	We	enjoyed
relaxing	g vacat	ion	on tropical	l island		I am
Hatim	,	taking	this	NLP	course	taught
by	Reeta	$\mathtt{mam}$				
	15	10	2	2	2	2
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1				



[]: <Axes: xlabel='Samples', ylabel='Counts'>

## 0.0.5 5. Curiosity Questions

Q1. What is Natural language processing? Discuss various levels of analysis under it?

Ans: NLP bridges the gap between human language and the digital world. It's like giving computers the ears and brain to understand and respond to spoken and written language, mimicking how humans communicate.

Levels of NLP analysis progressively dig deeper into the meaning:

- Morphological analysis: Breaks down words into their smallest meaningful units (morphemes), like "un-happy" into "un-" (negation) and "happy" (root word).
- Syntactic analysis: Examines the structure of sentences, identifying parts of speech and their relationships. For example, recognizing "The dog chased the cat" as Subject-Verb-Object structure.

- Semantic analysis: Deals with the meaning of words and sentences, including context and relationships between words. Understanding that "bark" can refer to sound, tree part, or a team's performance depending on context.
- Pragmatic analysis: Considers the speaker's intent and the context in which language is used. Knowing that "It's cold" might be a statement, a request for warmth, or sarcasm depending on the situation.
- Q2. What do you mean by ambiguity in NLP? Discuss various ways to resolve ambiguity in NLP.

Ans: Ambiguity in NLP refers to the presence of multiple possible interpretations of a word, phrase, or sentence. This can hinder the ability of machines to understand the true meaning.

### Ways to resolve ambiguity:

- Context analysis: Using surrounding information to determine the intended meaning.
- Part-of-speech tagging: Identifying the grammatical roles of words (e.g., noun, verb) to understand their function.
- Word sense disambiguation: Determining the specific meaning of a word based on context and usage.
- ML algorithms: Trained on large datasets of text, these algorithms can statistically predict the most likely meaning in ambiguous situations.
- Q3. What is morphology with examples?

Ans: Morphology is the branch of linguistics that studies the internal structure of words, focusing on how morphemes combine to form meaningful units.

### Examples:

- "Unhappy" is composed of "un-" (prefix meaning negation) and "happy" (root).
- "Books" is the plural form of "book" by adding the suffix "-s".
- "Teacher" combines the root "teach" with the suffix "-er" to indicate someone who does the action of teaching.

#### Uses of Morphology in NLP:

- Stemming and lemmatization: reducing words to their base form for better indexing and comparison. - Part-of-speech tagging: identifying the grammatical function of words based on their morphological features. - Spelling correction: suggesting correct spellings by analyzing possible morpheme combinations.

#### 0.0.6 6. Conclusion

In this experiment we learnt about the basic corpus analysis using NLTK and also learnt about the morphological features of a word. We also learnt about the tokenizination of a text file and performing frequency distribution on the tokens.