

# LAB 1 : Understanding Large Text Files

suriya

225229140

```
In [1]: import nltk
```

```
In [2]: nltk.download("wordnet")
```

```
[nltk_data] Downloading package wordnet to  
[nltk_data]     C:\Users\1mscdsa01\AppData\Roaming\nltk_data...  
[nltk_data]   Package wordnet is already up-to-date!
```

```
Out[2]: True
```

```
In [3]: text="This is Andrew's text,isn't it?"
```

## Excercise-1

```
In [4]: tokenizer= nltk.tokenize.WhitespaceTokenizer()  
tokens=tokenizer.tokenize(text)  
print(len(tokens))  
print(tokens)
```

```
5  
['This', 'is', "Andrew's", "text,isn't", 'it?']
```

```
In [5]: tokenizer= nltk.tokenize.TreebankWordTokenizer()  
tokens=tokenizer.tokenize(text)  
print(len(tokens))  
print(tokens)
```

```
10  
['This', 'is', 'Andrew', "'s", 'text', ',', 'is', "n't", 'it', '?']
```

```
In [6]: tokenizer= nltk.tokenize.WordPunctTokenizer()  
tokens=tokenizer.tokenize(text)  
print(len(tokens))  
print(tokens)
```

```
12  
['This', 'is', 'Andrew', "'", 's', 'text', ',', 'isn', "'", 't', 'it', '?']
```

## Excercise-2

In [7]:

```
#1
filename = ("gift-of-magi.txt")
f=open (filename,'r')
text=f.read()
f.close()
```

In [8]: text

Out[8]: 'The Gift of the Magi  
O. Henry  
One dollar and eighty-seven cents. That was all. And sixty cents of it was in pennies. Pennies saved one and two at a time by bulldozing the grocer and the vegetable man and the butcher until one's cheeks burned with the silent imputation of parsimony that such close dealing implied. Three times Della counted it. One dollar and eighty-seven cents. And the next day would be Christmas.  
There was clearly nothing left to do but flop down on the shabby little couch and howl. So Della did it. Which instigates the moral reflection that life is made up of sobs, sniffles, and smiles, with sniffles predominating.  
While the mistress of the home is gradually subsiding from the first stage to the second, take a look at the home. A furnished flat at \$8 per week. It did not exactly beggar description, but it certainly had that word on the look-out for the mendicancy squad.  
In the vestibule below was a letter-box into which no letter would go, and an electric button from which no mortal finger could coax a ring. Also appertaining thereunto was a card bearing the name "Mr. James Dillingham Young."  
The "Dillingham" had been flung to the breeze during a former period of prosperity when its possessor was being paid \$30 per week. Now, when the income was shrunk to \$20, the letters of "Dillingham" looked blurred, as though they were thinking seriously of contracting to a modest and unassuming D. But whenever Mr. James Dillingham Young came home and reached his flat above he was called "Jim" and greatly hugged by Mrs. James Dillingham Young, already introduced to you as Della. Which is all very good.  
Della finished her cry and attended to her cheeks with the powder rag. She stood by the window and looked out dully at a grey cat walking a grey fence in a grey backyard. To-morrow would be Christmas Day, and she had only \$1.87 with which to buy Jim a present. She had been saving every penny she could for months, with this result. Twenty dollars a week doesn't go far. Expenses had been greater than she had calculated. They always are. Only \$1.87 to buy a present for Jim. Her Jim. Many a happy hour she had spent planning for something nice for him. Something fine and rare and sterling--something just a little bit near to being worthy of the honour of being owned by Jim.  
There was a pier-glass between the windows of the room. Perhaps you have seen a pier-glass in an \$8 flat. A very thin and very agile person may, by observing his reflection in a rapid sequence of longitudinal strips, obtain a fairly accurate conception of his looks. Della, being slender, had mastered the art.  
Suddenly she whirled from the window and stood before the glass. Her eyes were shining brilliantly, but her face had lost its colour within twenty seconds. Rapidly she pulled down her hair and let it fall to its full length.  
Now, there were two possessions of the James Dillingham Youngs in which they both took a mighty pride. One was Jim's gold watch that had been his father's and his grandfather's. The other was Della's hair. Had the Queen of Sheba lived in the flat across the airshaft, Della would have let her hair hang out of the window some day to dry just to depreciate Her Majesty's jewels and gifts. Had King Solomon been the janitor, with all his treasures piled up in the basement, Jim would have pulled out his watch every time he passed, just to see him pluck at his beard from envy.  
So now Della's beautiful hair fell about her, rippling and shining like a cascade of brown water. It reached below her knee and made itself almost a garment for her. And then she did it up again nervously and quickly. Once she faltered for a minute and stood still while a tear or two splashed on the worn red carpet.  
On went her old brown jacket; on went her old brown hat. With a whirl of skirts and with the brilliant sparkle still in her eyes, she clattered out of the door and down the stairs to the street.  
Where she stopped the sign read: "Mme Sofronie. Hair Goods of All Kinds." One Eight up Della ran, and collected herself, panting. Madame, large, too white, chilly, hardly looked the "Sofronie."  
"Will you buy my hair?" asked Della.  
"I buy hair," said Madame. "Take yer hat off and let's have a sight at the looks of it."  
Down rippled the brown cascade.  
"T

wenty dollars," said Madame, lifting the mass with a practised hand.  
"Give it to me quick" said Della.  
Oh, and the next two hours tripped by on rosy wings. Forget the hashed metaphor. She was ransacking the stores for Jim's present.  
She found it at last. It surely had been made for Jim and no one else. There was no other like it in any of the stores, and she had turned all of them inside out. It was a platinum fob chain simple and chaste in design, properly proclaiming its value by substance alone and not by meretricious ornamentation--as all good things should do. It was even worthy of The Watch. As soon as she saw it she knew that it must be Jim's. It was like him. Quietness and value--the description applied to both. Twenty-one dollars they took from her for it, and she hurried home with the 78 cents. With that chain on his watch Jim might be properly anxious about the time in any company. Grand as the watch was, he sometimes looked at it on the sly on account of the old leather strap that he used in place of a chain.  
When Della reached home her intoxication gave way a little to prudence and reason. She got out her curling irons and lighted the gas and went to work repairing the ravages made by generosity added to love. Which is always a tremendous task dear friends--a mammoth task.  
Within forty minutes her head was covered with tiny, close-lying curls that made her look wonderfully like a truant schoolboy. She looked at her reflection in the mirror long, carefully, and critically.  
"If Jim doesn't kill me," she said to herself, "before he takes a second look at me, he'll say I look like a Coney Island chorus girl. But what could I do--oh! what could I do with a dollar and eighty-seven cents?"  
At 7 o'clock the coffee was made and the frying-pan was on the back of the stove hot and ready to cook the chops.  
Jim was never late. Della doubted the fob chain in her hand and sat on the corner of the table near the door that he always entered. Then she heard his step on the stair away down on the first flight, and she turned white for just a moment. She had a habit of saying little silent prayers about the simplest everyday things, and now she whispered: "Please, God, make him think I am still pretty."  
The door opened and Jim stepped in and closed it. He looked thin and very serious. Poor fellow, he was only twenty-two--and to be burdened with a family! He needed a new overcoat and he was without gloves.  
Jim stepped inside the door, as immovable as a setter at the scent of quail. His eyes were fixed upon Della, and there was an expression in them that she could not read, and it terrified her. It was not anger, nor surprise, nor disapproval, nor horror, nor any of the sentiments that she had been prepared for. He simply stared at her fixedly with that peculiar expression on his face.  
Della wriggled off the table and went for him.  
"Jim, darling," she cried, "don't look at me that way. I had my hair cut off and sold it because I couldn't have lived through Christmas without giving you a present. It'll grow out again--you won't mind, will you? I just had to do it. My hair grows awfully fast. Say 'Merry Christmas!' Jim, and let's be happy. You don't know what a nice--what a beautiful, nice gift I've got for you."  
"You've cut off your hair?" asked Jim, laboriously, as if he had not arrived at that patent fact yet, even after the hardest mental labour.  
"Cut it off and sold it," said Della. "Don't you like me just as well, anyhow? I'm me without my hair, ain't I?"  
Jim looked about the room curiously.  
"You say your hair is gone?" he said, with an air almost of idiocy.  
"You needn't look for it," said Della. "It's sold, I tell you--sold and gone, too. It's Christmas Eve, boy. Be good to me, for it went for you. Maybe the hairs of my head were numbered," she went on with a sudden serious sweetness, "but nobody could ever count my love for you. Shall I put the chops on, Jim?"  
Out of his trance Jim seemed quickly to wake. He enfolded his Della. For ten seconds let us regard with discreet scrutiny some inconsequential object in the other direction. Eight dollars a week or a million a year--what is the difference? A mathematician or a wit would give you the wrong answer. The magi brought valuable gifts, but that was not among them. This dark assertion will be illuminated later on.  
Jim drew a package from his overcoat pocket and threw it upon the table.  
"Don't make an

y mistake, Dell," he said, "about me. I don't think there's anything in the way of a haircut or a shave or a shampoo that could make me like my girl any less. But if you'll unwrap that package you may see why you had me going a while at first."  
White fingers and nimble tore at the string and paper. And then a n ecstatic scream of joy; and then, alas! a quick feminine change to hysterical tears and wails, necessitating the immediate employment of all the comforting powers of the lord of the flat.  
For there lay The Combs--the set of combs, side and back, that Della had worshipped for long in a Broadway window. Beautiful combs, pure tortoise-shell, with jewelled rims--just the shade to wear in the beautiful vanished hair. They were expensive combs, she knew, and her heart had simply craved and yearned over them without the least hope of possession. And now, they were hers, but the tresses that should have adorned the coveted adornments were gone.  
But she hugged them to her bosom, and at length she was able to look up with dim eyes and a smile and say: "My hair grows so fast, Jim!"  
And then Della leaped up like a little singed cat and cried, "Oh, oh!"  
Jim had not yet seen his beautiful present. She held it out to him eagerly upon her open palm. The dull precious metal seemed to flash with a reflection of her bright and ardent spirit.  
"Isn't it a dandy, Jim? I hunted all over town to find it. You'll have to look at the time a hundred times a day now. Give me your watch. I want to see how it looks on it."  
Instead of obeying, Jim tumbled down on the couch and put his hands under the back of his head and smiled.  
"Dell," said he, "let's put our Christmas presents away and keep 'em a while. They're too nice to use just at present. I sold the watch to get the money to buy Your combs. And now suppose you put the chops on."  
The magi, as you know, were wise men--wonderfully wise men--who brought gifts to the Babe in the manger. They invented the art of giving Christmas presents. Being wise, their gifts were no doubt wise ones, possibly bearing the privilege of exchange in case of duplication. And here I have lamely related to you the uneventful chronicle of two foolish children in a flat who most unwisely sacrificed for each other the greatest treasures of their house. But in a last word to the wise of these days let it be said that of all who give gifts these two were the wisest. Of all who give and receive gifts, such as they are wisest. Everywhere they are wisest. They are the magi.

In [10]: #2(i)

```
tokenizer= nltk.tokenize.WhitespaceTokenizer()
tokens=tokenizer.tokenize(text)
print(len(tokens))
```

2074

In [29]: #2(ii)

```
from nltk import *
data=FreqDist(tokens)
data

'''Jim,'': 1,
'''Mme': 1,
'''Mr.'': 1,
'''My': 1,
'''Oh,'': 1,
'''Please,'': 1,
'''Sofronie.''': 1,
'''Take': 1,
'''Twenty': 1,
'''Will': 1,
'''You': 2,
'''You\ve': 1,
'''about': 1,
'''before': 1,
'''but': 1,
'''don\t': 1,
'''let\s': 1,
'$1.87': 2,
'$20,'': 1,
'$30': 1,
```

In [30]: #2(iii)

```
data.most_common(20)
```

Out[30]: [('the', 107),  
 ('and', 74),  
 ('a', 64),  
 ('of', 51),  
 ('to', 41),  
 ('was', 26),  
 ('she', 25),  
 ('in', 24),  
 ('had', 21),  
 ('her', 21),  
 ('that', 20),  
 ('it', 19),  
 ('at', 19),  
 ('with', 19),  
 ('for', 19),  
 ('his', 17),  
 ('on', 16),  
 ('I', 14),  
 ('Jim', 13),  
 ('were', 11)]

```
In [32]: #2(iv)
from nltk import *
test=[w for w in tokens if len(w) >10]
freq=FreqDist(test)
freq
```

```
Out[32]: FreqDist({'Dillingham': 2,
                    'Sofronie.': 1,
                    'Christmas!': 1,
                    'appertaining': 1,
                    'brilliantly,': 1,
                    'calculated.': 1,
                    'close-lying': 1,
                    'contracting': 1,
                    'critically.': 1,
                    'description': 1,
                    'description,' : 1,
                    'difference?': 1,
                    'disapproval,' : 1,
                    'duplication.': 1,
                    'eighty-seven': 3,
                    "grandfather's.": 1,
                    'illuminated': 1,
                    'inconsequential': 1,
                    'intoxication': 1,
                    'laboriously,' : 1,
                    'longitudinal': 1,
                    'mathematician': 1,
                    'men--wonderfully': 1,
                    'meretricious': 1,
                    'necessitating': 1,
                    'ornamentation--as': 1,
                    'possession.': 1,
                    'possessions': 1,
                    'predominating.': 1,
                    'proclaiming': 1,
                    'sterling--something': 1,
                    'tortoise-shell,' : 1,
                    'twenty-two--and': 1,
                    'wonderfully': 1})
```

```
In [34]: #2(v)
for i,j in freq.items():
    if len(i) > 10 and j>2:
        print(i,j)
```

```
eighty-seven 3
```

## Exercise 3

**step -1**

```
In [39]: ## download the document Austen's Emma("austen-emma.txt")
```

## step-2

### list-comprehend emma

```
In [38]: fname = "austen-emma.txt"
f = open(fname, 'r')
etxt= f.read()
print(etxt)
f.close()
```

[Emma by Jane Austen 1816]

VOLUME I

CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy disposition, seemed to unite some of the best blessings of existence; and had lived nearly twenty-one years in the world with very little to distress or vex her.

She was the youngest of the two daughters of a most affectionate, indulgent father; and had, in consequence of her sister's marriage, been mistress of his house from a very early period. Her mother had died too long ago for her to have more than an indistinct remembrance of her caresses; and her place had been supplied by an excellent woman as governess, who had fallen little short of a mother in affection.

```
In [40]: fname = "austen-emma.txt"
```

```
In [41]: f = open(fname, 'r')
etxt = f.read()
f.close()
etxt[-200:]
```

Out[41]: 'e deficiencies, the wishes,\nthe hopes, the confidence, the predictions of the small band\nof true friends who witnessed the ceremony, were fully answered\nin the perfect happiness of the union.\n\n\nFINIS\n'

```
In [45]: nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\1mscda01\AppData\Roaming\nltk_data...
[nltk_data]     Unzipping tokenizers\punkt.zip.
```

Out[45]: True

```
In [46]: etoks = nltk.word_tokenize(etxt.lower())
etoks [-20:]
```

```
Out[46]: ['of',
'true',
'friends',
'who',
'witnessed',
'the',
'ceremony',
',',
'were',
'fully',
'answered',
'in',
'the',
'perfect',
'happiness',
'of',
'the',
'union',
'.',
'finis']
```

```
In [47]: etypes =sorted(set(etoks))
etypes[-10:]
```

```
Out[47]: ['younger',
'youngest',
'your',
'yours',
'yourself',
'yourself.',
'youth',
'youthful',
'zeal',
'zigzags']
```

```
In [48]: len(etypes)
```

```
Out[48]: 8000
```

```
In [49]: efreq = nltk.FreqDist(etoks)
```

```
In [50]: efreq['the']
```

```
Out[50]: 5198
```

## Question 1: Words with prefix and suffix

```
In [51]: [word for word in etoks if word.startswith("un") & word.endswith("able")]
```

```
Out[51]: ['unexceptionable',
          'unsuitable',
          'unreasonable',
          'unreasonable',
          'uncomfortable',
          'unfavourable',
          'unexceptionable',
          'unexceptionable',
          'uncomfortable',
          'unpersuadable',
          'unavoidable',
          'unreasonable',
          'uncomfortable',
          'unsuitable',
          'unmanageable',
          'unexceptionable',
          'unreasonable',
          'unobjectionable',
          'unpersuadable',
          'unsuitable',
          'unreasonable',
          'uncomfortable',
          'unexceptionable',
          'unpardonable',
          'unmanageable',
          'unanswerable',
          'unfavourable',
          'unpersuadable',
          'unaccountable',
          'undesirable',
          'unable',
          'unable',
          'unpardonable',
          'unexceptionable',
          'unreasonable',
          'unreasonable',
          'uncomfortable',
          'unreasonable',
          'unpardonable',
          'unaccountable',
          'unexceptionable',
          'unreasonable',
          'unaccountable']
```

## Question 2: Length

```
In [52]: tokenizer = nltk.tokenize.WordPunctTokenizer()
toke = tokenizer.tokenize (etxt)
```

```
In [53]: [word for word in toke if len(word)>15]
```

```
Out[53]: ['companionableness',
'misunderstanding',
'incomprehensible',
'undistinguishing',
'unceremoniousness',
'Disingenuousness',
'disagreeableness',
'misunderstandings',
'misunderstandings',
'misunderstandings',
'misunderstandings',
'disinterestedness',
'unseasonableness']
```

### Question 3: Average word length

```
In [54]: average=sum(len (word) for word in toke)/len (toke)
average
```

```
Out[54]: 3.755268231589122
```

### Question 4: Word frequency

```
In [56]: from nltk import*
fdiemm = FreqDist (toke)
```

```
In [59]: for i,j in fdiemm.items():
    if j > 200:
        print(i,j)
```

```
; 2199
had 1606
- 574
one 413
in 2118
very 1151
little 354
or 490
her 2381
. 6928
She 562
was 2385
most 243
father 207
' 1007
s 933
been 759
his 1088
from 535
too 253
```

## STEP 3: bigrams in Emma

```
In [60]: e2grams = list(nltk.bigrams (toke))
e2gramfd = nltk.FreqDist(e2grams)
```

```
In [61]: e2gramfd
```

```
, , clever, . ,
('clever', ','): 7,
(,',', 'and'): 1879,
('and', 'rich'): 1,
('rich', ','): 3,

(,',', 'with'): 177,
('with', 'a'): 119,
('a', 'comfortable'): 3,
('comfortable', 'home'): 2,
('home', 'and'): 4,
('and', 'happy'): 8,
('happy', 'disposition'): 1,
('disposition', ','): 5,
(,',', 'seemed'): 13,
('seemed', 'to'): 44,
('to', 'unite'): 1,
('unite', 'some'): 1,
('some', 'of'): 19,
('of', 'the'): 556,
```

## Question 6: Bigrams

```
In [63]: last_ten = FreqDist(dict(e2gramfd.most_common()[-10:]))
last_ten
```

```
Out[63]: FreqDist({('.', 'FINIS'): 1,
                   ('answered', 'in'): 1,
                   ('fully', 'answered'): 1,
                   ('the', 'ceremony'): 1,
                   ('the', 'perfect'): 1,
                   ('the', 'union'): 1,
                   ('union', '.'): 1,
                   ('were', 'fully'): 1,
                   ('who', 'witnessed'): 1,
                   ('witnessed', 'the'): 1})
```

## Question 7: Bigram top frequency

```
In [66]: tokenizer = nltk.tokenize.WhitespaceTokenizer()
tokes = tokenizer.tokenize(etxt)
```

```
In [67]: e2grams = list(nltk.bigrams(tokes))
e2gramfd = nltk.FreqDist(e2grams)
```

```
In [68]: e2gramfd.most_common(20)
```

```
Out[68]: [((('to', 'be'), 562),
            (('of', 'the'), 556),
            (('in', 'the'), 431),
            (('I', 'am'), 302),
            (('had', 'been'), 299),
            (('could', 'not'), 270),
            (('it', 'was'), 253),
            (('she', 'had'), 242),
            (('to', 'the'), 236),
            (('have', 'been'), 233),
            (('of', 'her'), 230),
            (('I', 'have'), 214),
            (('and', 'the'), 208),
            (('would', 'be'), 208),
            (('she', 'was'), 206),
            (('do', 'not'), 196),
            (('of', 'his'), 182),
            (('that', 'she'), 178),
            (('to', 'have'), 176),
            (('such', 'a'), 176)]
```

## Question 8: Bigram frequency count

```
In [69]: for i, j in e2gramfd.items():
    if i == ('so', 'happy'):
        print(i,j)
('so', 'happy') 3
```

## Question 9: Word following 'so'

```
In [70]: import re
from collections import Counter
```

```
In [74]: words = re.findall(r'so+ \w+', open('austen-emma.txt').read())
ab = Counter(zip(words))
print(ab)
```

Counter({('so much',): 95, ('so very',): 76, ('so well',): 30, ('so many',): 27, ('so long',): 27, ('so little',): 20, ('so far',): 17, ('so I',): 14, ('so kind',): 13, ('so good',): 12, ('so often',): 10, ('so soon',): 9, ('so great',): 8, ('so to',): 7, ('so fond',): 7, ('so she',): 7, ('so it',): 6, ('so anxious',): 6, ('so as',): 6, ('so you',): 6, ('so truly',): 6, ('so completely',): 5, ('so obliging',): 5, ('so extremely',): 5, ('so entirely',): 4, ('so happy',): 4, ('so interesting',): 4, ('so fast',): 4, ('so near',): 4, ('so pleased',): 4, ('so few',): 4, ('so that',): 4, ('so strong',): 4, ('so liberal',): 4, ('so miserable',): 4, ('so happily',): 3, ('so proper',): 3, ('so pleasantly',): 3, ('so superior',): 3, ('so warmly',): 3, ('so bad',): 3, ('so odd',): 3, ('so ill',): 3, ('so delighted',): 3, ('so particularly',): 3, ('so easily',): 3, ('so on',): 3, ('so attentive',): 3, ('so fortunate',): 3, ('so glad',): 3, ('so shocked',): 3, ('so at',): 3, ('so obliged',): 2, ('so perfectly',): 2, ('so dear',): 2, ('so busy',): 2, ('so did',): 2, ('so forth',): 2, ('so totally',): 2, ('so remarkably',): 2, ('so plainly',): 2, ('so charming',): 2, ('so surprised',): 2, ('so early',): 2, ('so too',): 2, ('so easy',): 2, ('so decidedly',): 2, ('so absolutely',): 2, ('so particular',): 2, ('so deceived',): 2, ('so palpably',): 2, ('so clever',): 2, ('so short',): 2, ('so cold',): 2, ('so high',): 2, ('so happened',): 2, ('so full',): 2, ('so thoroughly',): 2, ('so equal',): 2, ('so off',): 2, ('so naturally',): 2, ('so afraid',): 2, ('so deep',): 2, ('so kindly',): 2, ('so pale',): 2, ('so noble',): 2, ('so lovely',): 2, ('so mad',): 2, ('so nearly',): 2, ('so sorry',): 2, ('so cheerful',): 2, ('so unfeeling',): 2, ('so ready',): 2, ('so unperceived',): 1, ('so mild',): 1, ('so constantly',): 1, ('so comfortably',): 1, ('so avowed',): 1, ('so deservedly',): 1, ('so convenient',): 1, ('so just',): 1, ('so apparent',): 1, ('so sorrowful',): 1, ('so spent',): 1, ('so artlessly',): 1, ('so plain',): 1, ('so firmly',): 1, ('so genteel',): 1, ('so\_then',): 1, ('so brilliant',): 1, ('so seldom',): 1, ('so nervous',): 1, ('so indeed',): 1, ('so pack',): 1, ('so doubtful',): 1, ('so with',): 1, ('so contemptible',): 1, ('so slightly',): 1, ('so by',): 1, ('so loudly',): 1, ('so materially',): 1, ('so hard',): 1, ('so delightful',): 1, ('so pointed',): 1, ('so equalled',): 1, ('so evidently',): 1, ('so immediately',): 1, ('so sought',): 1, ('so excellent',): 1, ('so prettily',): 1, ('so extreme',): 1, ('so wonder',): 1, ('so always',): 1, ('so silly',): 1, ('so satisfied',): 1, ('so smiling',): 1, ('so prosing',): 1, ('so undistinguishing',): 1, ('so apt',): 1, ('so dreadful',): 1, ('so respected',): 1, ('so tenderly',): 1, ('so grieved',): 1, ('so shocking',): 1, ('so conceited',): 1, ('so before',): 1, ('so prevalent',): 1, ('so heavy',): 1, ('so swiftly',): 1, ('so spoken',): 1, ('so or',): 1, ('so overcharge',): 1, ('so pleasant',): 1, ('so fenced',): 1, ('so hospitable',): 1, ('so interested',): 1, ('so sanguine',): 1, ('so sure',): 1, ('so careless',): 1, ('so rapidly',): 1, ('so frequent',): 1, ('so sensible',): 1, ('so misled',): 1, ('so blind',): 1, ('so complaisant',): 1, ('so misinterpreted',): 1, ('so active',): 1, ('so pointedly',): 1, ('so striking',): 1, ('so sudden',): 1, ('so indistinctly',): 1, ('so partial',): 1, ('so natural',): 1, ('so inevitable',): 1, ('so lately',): 1, ('so beautifully',): 1, ('so distinct',): 1, ('so considerate',): 1, ('so light',): 1, ('so intimate',): 1, ('so magnified',): 1, ('so cautious',): 1, ('so confined',): 1, ('so wish',): 1, ('so he',): 1, ('so glorious',): 1, ('so quick',): 1, ('so sweetly',): 1, ('so inseparably',): 1, ('so deserving',): 1, ('so disappointed',): 1, ('so ended',): 1, ('so sluggish',): 1, ('so amiable',): 1, ('so quiet',): 1, ('so idolized',): 1, ('so cried',): 1, ('so acceptable',): 1, ('so properly',): 1, ('so reasonable',): 1, ('so delightfully',): 1, ('so rich',): 1, ('so warm',): 1, ('so large',): 1, ('so handsome',): 1, ('so abundant',): 1, ('so outree',): 1, ('so thoughtful',): 1, ('so mu

```
st'): 1, ('so effectually'): 1, ('so beautiful'): 1, ('so Patty'): 1, ('so honoured'): 1, ('so close'): 1, ('so imprudent'): 1, ('so limited'): 1, ('so from'): 1, ('so amusing'): 1, ('so indifferent'): 1, ('so indignant'): 1, ('so said'): 1, ('so right'): 1, ('so wretched'): 1, ('so now'): 1, ('so occupied'): 1, ('so unhappy'): 1, ('so highly'): 1, ('so generally'): 1, ('so exactly'): 1, ('so double'): 1, ('so secluded'): 1, ('so regular'): 1, ('so determined'): 1, ('so motherly'): 1, ('so the'): 1, ('so glibly'): 1, ('so calculated'): 1, ('so thrown'): 1, ('so exclusively'): 1, ('so disgusting'): 1, ('so needlessly'): 1, ('so does'): 1, ('so resolutely'): 1, ('so would'): 1, ('so infinitely'): 1, ('so fluently'): 1, ('so they'): 1, ('so impatient'): 1, ('so briskly'): 1, ('so vigorously'): 1, ('so young'): 1, ('so hardened'): 1, ('so gratified'): 1, ('so received'): 1, ('so then'): 1, ('so and'): 1, ('so gratefully'): 1, ('so found'): 1, ('so placed'): 1, ('so lain'): 1, ('so his'): 1, ('so arranged'): 1, ('so moving'): 1, ('so walking'): 1, ('so when'): 1, ('so favourable'): 1, ('so late'): 1, ('so silent'): 1, ('so dull'): 1, ('so irksome'): 1, ('so agitated'): 1, ('so brutal'): 1, ('so cruel'): 1, ('so depressed'): 1, ('so no'): 1, ('so justly'): 1, ('so astonished'): 1, ('so will'): 1, ('so simple'): 1, ('so dignified'): 1, ('so suddenly'): 1, ('so a'): 1, ('so herself'): 1, ('so peremptorily'): 1, ('so uneasy'): 1, ('so wonderful'): 1, ('so _very_'): 1, ('so expressly'): 1, ('so angry'): 1, ('so anxiously'): 1, ('so strange'): 1, ('so stoutly'): 1, ('so mistake'): 1, ('so mistaken'): 1, ('so dreadfully'): 1, ('so voluntarily'): 1, ('so satisfactory'): 1, ('so disinterested'): 1, ('so foolishly'): 1, ('so ingeniously'): 1, ('so entreated'): 1, ('so like'): 1, ('so cordially'): 1, ('so essential'): 1, ('so designedly'): 1, ('so hasty'): 1, ('so richly'): 1, ('so grateful'): 1, ('so tenaciously'): 1, ('so feeling'): 1, ('so engaging'): 1, ('so engaged'): 1, ('so hot'): 1, ('so useful'): 1, ('so attached'): 1, ('so peculiarly'): 1, ('so singularly'): 1, ('so taken'): 1, ('so recently'): 1, ('so fresh'): 1, ('so hateful'): 1, ('so heartily'): 1, ('so steady'): 1, ('so complete'): 1, ('so in'): 1, ('so suffered'): 1})
```

## Question 10: Trigrams

In [75]:

```
e3grams = list(nltk.trigrams(tokes))
e3gramfd = nltk.FreqDist(e3grams)
```

In [76]:

```
last_ten = FreqDist(dict(e3gramfd.most_common()[-10:]))
last_ten
```

Out[76]:

```
FreqDist({('answered', 'in', 'the'): 1,
          ('ceremony', 'were', 'fully'): 1,
          ('fully', 'answered', 'in'): 1,
          ('in', 'the', 'perfect'): 1,
          ('of', 'the', 'union.'): 1,
          ('perfect', 'happiness', 'of'): 1,
          ('the', 'ceremony', 'were'): 1,
          ('the', 'perfect', 'happiness'): 1,
          ('the', 'union.', 'FINIS'): 1,
          ('were', 'fully', 'answered'): 1})
```

## Question 11: Trigram top frequency

```
In [77]: e3gramfd.most_common(10)
```

```
Out[77]: [((‘I’, ‘do’, ‘not’), 94),
((‘I’, ‘am’, ‘sure’), 75),
((‘would’, ‘have’, ‘been’), 55),
((‘a’, ‘great’, ‘deal’), 55),
((‘she’, ‘could’, ‘not’), 49),
((‘could’, ‘not’, ‘be’), 45),
((‘she’, ‘had’, ‘been’), 44),
((‘it’, ‘would’, ‘be’), 43),
((‘do’, ‘not’, ‘know’), 43),
((‘Mr.’, ‘and’, ‘Mrs.’), 37)]
```

## Question 12: trigram frequency count

```
In [89]: words1 = re.findall(r’so happy to \w+’, open(‘austen-emma.txt’).read())
print(words1)
```

```
[]
```

```
In [ ]:
```

```
In [ ]:
```