Simon Fraser University

SDA 250 – Assignment 1 - NLTK and corpus functions

Instructor - Dr. Maite Taboada

Name (Last, First)	Student ID	Section contributed	Section edited
Jokonya, Zvikomborero Kennedy	301449517	300.txt collection and organisation	300.txt - code, write up and edited
Park, Hyungjoo	301566942	Hyungjoo_Park_Assignment.txt collection	Hyungjoo_Park_Assignment.txt - code, write up and edited
Zhou, William (FeiLian)	301406145	Odyssey.txt collection	Odyssey.txt - code, write up and edited

In "hyungjoo-park-assignment-fin-txt", which consists of historical works that can be complex and deep in language and often indicated by long words with prefixes, suffixes, and complex morphological structures. Using long sentences to weave complex plots, convey elaborate ideas, or describe vivid scenes, while having a lexical diversity of '0.07'. A work such as "Odyssey.txt" tells a story from ancient times whose punctuation is different from today's language which can increase the longest sentence to 168 words as it has been translated. Whose longest words contain many words not used commonly in today's speech, while having a lexical diversity of '0.05'. On a less educational side, "300.txt" is a movie script whose most frequent words are the lead character name, 'Leonidas', and conjunction words and punctuation marks. The longest words have a relatively low frequency of repetition, and most are unique words, and the longest sentence has a higher word count than a standard sentence but is lower compared to the Odyssey. 300's lexical diversity is '0.15'.

These texts are of a similar nature tied to history but mainly differ in audience. The historical work will use descriptions used in the past just like Homer's poem and 300 will use today's lingo to describe the past. One is for a scholar, the second is for a reader and the third is purely for spectacle. Their lexical diversity can vary accordingly in their respective genres; it is difficult to interpret between different genres.

Link to GitHub repository:

https://github.com/butterbumbs/Group7-Assignment1.git

hyungjoo-park-assignment-fin-txt

October 15, 2023

[186]:
```python
import nltk
```

[187]:
```python
Text = "Hyungjoo_Park_Assignment.txt"
```

[188]:
```python
with open(Text,'r', encoding = "utf8" ) as f:
    text = f.read()
```

[189]:
```python
len(text)
```

[189]: 42955

[190]:
```python
def lexical_diversity(text):
    return len(set(text)) / len(text)
```

[191]:
```python
lexical_diversity(text1)
```

[191]: 0.07406285585022564

[192]:
```python
textTokens = nltk.word_tokenize(text)
```

[194]:
```python
textFreq = nltk.FreqDist(textTokens)
```

[195]:
```python
textFreq
```

[195]: FreqDist({'the': 617, ',': 385, 'of': 272, '.': 248, 'to': 173, 'and': 148, 'a': 145, 'in': 143, ''': 115, '(': 108, …})

[197]:
```python
for w in textTokens:
    if (len(w) >= 10):
        print(w)
```

```
Collections
particularly
institutions
institutions
democratic
assemblies
assemblies
```

assemblies
assemblies
overridden
democratic
exceptional
assemblies
traditional
conventions
structured
assemblies
assemblies
developing
democratic
especially
conscripti
encouraged
accordance
BretDevereaux
occasional
assemblies
assemblies
magistrates
circumstances
assemblies
magistrates
magistrates
assemblies
Magistrates
centuriata
magistrate
up-or-down
candidates
magistrate
Candidates
dispensation
requirement
surrounded
magistrates
effectively
gatherings
face-to-face
influential
politicians
supporters
auctoritas
influential
politicians
assemblies

```
recommends
traditional
magistrates
senatorial
legislation
Legislation
recommended
collective
auctoritas
traditional
addressing
democratic
centerpiece
somewhat-democratic
Morstein-Marx
democratic
expression
opportunity
Morstein-Marx
discussion
undervalues
situations
candidates
trinundinum
candidates
assemblies
effecively
supporters
magistrates
magistrate
accompliushed
interrupted
assemblies
dismissing
assemblies
ritualized
Technically
magistrates
rubber-stamp
formalizing
Centuriata
functioning
assemblies
centuriata
centuriate
assemblies
centuriata
cumbersome
```

centuriata
army-as-voting-body
competence
occasionally
centuriata
particular
centuriata
magistrates
centuriata
assemblies
cumbersome
centuriata
attendance
assemblies
centuriata
centuriata
presidential
ballot-measures
centuriata
cumbersome
effectively
Hallicarnasus
schematics
understood
structured
reconstructions
considerable
anachronistic
reconstructions
description
understand
distribution
wealthiest
distinction
distribution
80/20/20/20/30
non-combat
establishing
designated
praerogativa
considered
implications
reconstructions
conservative
individuals
presumably
outnumbered
infantrymen

distributed
reconstruction
correlates
centuriata
unfortunate
fifty-thousand
expectancy
discussion
principally
fundamentally
computations
proportion
proportion
data-point
infantrymen
respectively
represented
respectively
graduating
connection
body-armor
requirement
assessment
anachronism
projecting
constitution
immediately
practically
constructs
oligarchies
concentrates
preferring
significantly
obligations
proportional
non-crisis
absolutely
understand
concentrate
decision-making
particular
competences
explaining
aggressively
disenfranchises
centuriata
worthwhile
democratic

well-to-do
super-rich
infantrymen
bewildering
complexity
centuriata
refreshingly
substantially
democratically
assemblies
legislation
cumbersome
centuriata
substantial
implications
competence
assemblies
legislation
technically
plebiscita
remarkable
tribunates
legislation
particular
legislation
registration
tribe-of-residence
Originally
apportioned
disconnected
Consequently
countryside
settlements
countryside
presumably
intentional
consisting
landholders
prestigious
downmarket
substantially
democratic
arrangement
centuriata
mega-wealthy
freeholders
structured
freeholding

man-one-vote
particular
represented
determined
exceptional
praerogativa
centuriata
assemblies
centuriata
superiority
democratic
deliberate
constraining
opportunities
discontent
critically
freeholding
intentionally
physicality
accommodate
undetermined
procedures
consistently
time-saving
delivering
comisioned
interested
time-saving
tremendous
unmanageable
assemblies
Consequently
single-digits
percentages
infrastructure
democracies
accommodate
assemblies
situations
popularity
perception
electorate
substantially
electorate
simplicity
discussing
afterwards
prominence

particularly
influential
compromise
exclusively
magistrates
counter-balance
magistrates
eventually
assemblies
plebiscita
magistrates
magistrates
membership
non-patricians
plebiscita
plebiscita
significance
plebeian-patrician
distinction
meaninglessness
traditional
institutions
staggering
patricians
plebiscita
meaningful
distinction
understood
patricians
magistrate
tremendously
legislation
patricians
influential
reputation
Conclusions
assemblies
contradiction
assemblies
magistrates
conclusion
assemblies
magistrates
Magistrates
themselves
democratic
institutions
centuriata

explicitly
comfortable
contention
assemblies
traditional
privileges
prerogatives
magistrates
traditional
long-standing
assemblies
rubber-stamps
assemblies
unreasonably
Mediterranean
effectively
threatening
centuriata
declaration
compromise
declaration
controversy
restricted
successful
assemblies
Aemilianus
consulship
Aemilianus
consulship
centuriata
Aemilianus
Aemilianus
Aemilianus
ineligable
Numantia.7
assemblies
occasionally
displeasure
particulars
importance
frequently
Consequently
understand
assemblies
census-ratification
institutions
determining
functioned

```
importantly
magistrates
assemblies
competitive
elimination
unimportance
figurative
installment
magistrates
```

```python
# 4. Words that are at least 10 characters long and their counts.
word_counts = {}
for word in textTokens:
    if (len(word) >= 10):
        word_counts[word] = word_counts.get(word, 0) + 1

for word, count in word_counts.items():
    print(f'{word}: {count}')
```

```
Collections: 1
particularly: 2
institutions: 5
democratic: 9
assemblies: 35
overridden: 1
exceptional: 2
traditional: 6
conventions: 1
structured: 3
developing: 1
especially: 1
conscripti: 1
encouraged: 1
accordance: 1
BretDevereaux: 1
occasional: 1
magistrates: 17
circumstances: 1
Magistrates: 2
centuriata: 21
magistrate: 4
up-or-down: 1
candidates: 3
Candidates: 1
dispensation: 1
requirement: 2
surrounded: 1
effectively: 3
```

```
gatherings: 1
face-to-face: 1
influential: 4
politicians: 2
supporters: 2
auctoritas: 2
recommends: 1
senatorial: 1
legislation: 6
Legislation: 1
recommended: 1
collective: 1
addressing: 1
centerpiece: 1
somewhat-democratic: 1
Morstein-Marx: 2
expression: 1
opportunity: 1
discussion: 2
undervalues: 1
situations: 2
trinundinum: 1
effecively: 1
accompliushed: 1
interrupted: 1
dismissing: 1
ritualized: 1
Technically: 1
rubber-stamp: 1
formalizing: 1
Centuriata: 1
functioning: 1
centuriate: 1
cumbersome: 4
army-as-voting-body: 1
competence: 2
occasionally: 2
particular: 4
attendance: 1
presidential: 1
ballot-measures: 1
Hallicarnasus: 1
schematics: 1
understood: 2
reconstructions: 3
considerable: 1
anachronistic: 1
description: 1
```

```
understand: 3
distribution: 2
wealthiest: 1
distinction: 3
80/20/20/20/30: 1
non-combat: 1
establishing: 1
designated: 1
praerogativa: 2
considered: 1
implications: 2
conservative: 1
individuals: 1
presumably: 2
outnumbered: 1
infantrymen: 3
distributed: 1
reconstruction: 1
correlates: 1
unfortunate: 1
fifty-thousand: 1
expectancy: 1
principally: 1
fundamentally: 1
computations: 1
proportion: 2
data-point: 1
respectively: 2
represented: 2
graduating: 1
connection: 1
body-armor: 1
assessment: 1
anachronism: 1
projecting: 1
constitution: 1
immediately: 1
practically: 1
constructs: 1
oligarchies: 1
concentrates: 1
preferring: 1
significantly: 1
obligations: 1
proportional: 1
non-crisis: 1
absolutely: 1
concentrate: 1
```

```
decision-making: 1
competences: 1
explaining: 1
aggressively: 1
disenfranchises: 1
worthwhile: 1
well-to-do: 1
super-rich: 1
bewildering: 1
complexity: 1
refreshingly: 1
substantially: 3
democratically: 1
substantial: 1
technically: 1
plebiscita: 5
remarkable: 1
tribunates: 1
registration: 1
tribe-of-residence: 1
Originally: 1
apportioned: 1
disconnected: 1
Consequently: 3
countryside: 2
settlements: 1
intentional: 1
consisting: 1
landholders: 1
prestigious: 1
downmarket: 1
arrangement: 1
mega-wealthy: 1
freeholders: 1
freeholding: 2
man-one-vote: 1
determined: 1
superiority: 1
deliberate: 1
constraining: 1
opportunities: 1
discontent: 1
critically: 1
intentionally: 1
physicality: 1
accommodate: 2
undetermined: 1
procedures: 1
```

```
consistently: 1
time-saving: 2
delivering: 1
comisioned: 1
interested: 1
tremendous: 1
unmanageable: 1
single-digits: 1
percentages: 1
infrastructure: 1
democracies: 1
popularity: 1
perception: 1
electorate: 2
simplicity: 1
discussing: 1
afterwards: 1
prominence: 1
compromise: 2
exclusively: 1
counter-balance: 1
eventually: 1
membership: 1
non-patricians: 1
significance: 1
plebeian-patrician: 1
meaninglessness: 1
staggering: 1
patricians: 3
meaningful: 1
tremendously: 1
reputation: 1
Conclusions: 1
contradiction: 1
conclusion: 1
themselves: 1
explicitly: 1
comfortable: 1
contention: 1
privileges: 1
prerogatives: 1
long-standing: 1
rubber-stamps: 1
unreasonably: 1
Mediterranean: 1
threatening: 1
declaration: 2
controversy: 1
```

```
restricted: 1
successful: 1
Aemilianus: 5
consulship: 2
ineligable: 1
Numantia.7: 1
displeasure: 1
particulars: 1
importance: 1
frequently: 1
census-ratification: 1
determining: 1
functioned: 1
importantly: 1
competitive: 1
elimination: 1
unimportance: 1
figurative: 1
installment: 1
```

[231]:
```python
nltk.download('gutenberg')
```

```
[nltk_data] Downloading package gutenberg to
[nltk_data]     C:\Users\LG\AppData\Roaming\nltk_data…
[nltk_data]   Package gutenberg is already up-to-date!
```

[231]: True

[179]:
```python
gutenberg.fileids()
```

[179]:
```
['austen-emma.txt',
 'austen-persuasion.txt',
 'austen-sense.txt',
 'bible-kjv.txt',
 'blake-poems.txt',
 'bryant-stories.txt',
 'burgess-busterbrown.txt',
 'carroll-alice.txt',
 'chesterton-ball.txt',
 'chesterton-brown.txt',
 'chesterton-thursday.txt',
 'edgeworth-parents.txt',
 'melville-moby_dick.txt',
 'milton-paradise.txt',
 'shakespeare-caesar.txt',
 'shakespeare-hamlet.txt',
 'shakespeare-macbeth.txt',
 'whitman-leaves.txt']
```

```
[232]:  # 5. The longest sentence
        from nltk.corpus import gutenberg
        from nltk.tokenize import sent_tokenize
        from nltk.tokenize import word_tokenize

        sentences = sent_tokenize(text)
        longest_sentence = ""
        max_sentence_length = 0

        for sentence in sentences:
            sentence_length = len(sentence.split())

            if sentence_length > max_sentence_length:
                max_sentence_length = sentence_length
                longest_sentence = sentence

        print("Longest Sentence:")
        print(longest_sentence)
        print("Length:", max_sentence_length, "words")
```

Longest Sentence:
Still this alternate venue with its alternate convening magistrate becomes
tremendously important in the late second century as a favorable venue for
popularis ('popular' in the sense of 'of the people' or 'against the elite')
legislation, perhaps in part because of the exclusion of the patricians (who
still made up many influential senators and so you might not want to have them
there working the crowd) and in part because of the reputation and cultural
cachet behind the notion of the concilium plebis as a venue 'for the people,' as
it were.
Length: 92 words

```
[233]:  # 6. A stemmed version of the longest sentence
        from nltk.stem import PorterStemmer
        stemmer = PorterStemmer()

        stemmed_words = [stemmer.stem(word) for word in word_tokenize(longest_sentence)]

        stemmed_longest_sentence = ' '.join(stemmed_words)
        print("\nStemmed Longest Sentence:")
        print(stemmed_longest_sentence)
```

Stemmed Longest Sentence:
still thi altern venu with it altern conven magistr becom tremend import in the
late second centuri as a favor venu for populari (' popular ' in the sens of '
of the peopl ' or ' against the elit ') legisl , perhap in part becaus of the
exclus of the patrician (who still made up mani influenti senat and so you
might not want to have them there work the crowd) and in part becaus of the
```

reput and cultur cachet behind the notion of the concilium plebi as a venu ' for the peopl , ' as it were .

# odyssey-txt-assignment-1

October 15, 2023

```python
[323]: import nltk
       import numpy
       import matplotlib
       import os
       from nltk.corpus import PlaintextCorpusReader
       from nltk.tokenize import word_tokenize
       from nltk.tokenize import sent_tokenize
       from nltk.stem import PorterStemmer
```

```python
[324]: corpus_root = "C:/Users/william/SDA250 D100 Assignment"
```

```python
[325]: corpus_root
```

```python
[325]: 'C:/Users/william/SDA250 D100 Assignment'
```

```python
[326]: pwd
```

```python
[326]: 'C:\\Users\\william\\Desktop\\sda-binder-main'
```

```python
[327]: texts = PlaintextCorpusReader(corpus_root, '.*', encoding = "utf8")
```

```python
[328]: Odyssey = "Odyssey.txt"
```

```python
[329]: with open(Odyssey,'r', encoding = "utf8" ) as f:
           Odyssey1 = f.read()
```

```python
[330]: odyssey_words = texts.words('Odyssey.txt')
```

```python
[331]: odyssey_words
```

```python
[331]: ['HOMER', 'THE', 'ODYSSEY', 'TRANSLATED', 'BY', …]
```

```python
[332]: #1 The length (in words).
       len(odyssey_words)
```

```python
[332]: 149901
```

```
[333]:  #2 The lexical diversity
        def lexical_diversity(text):
            return len(set(text)) / len(odyssey_words)
```

```
[334]:  lexical_diversity(odyssey_words)
```

[334]: 0.056730775645259204

```
[335]:  #3 Top 10 most frequent words and their counts
        fdlist1 = FreqDist(odyssey_words)
        fdlist1.most_common(10)
```

```
[335]: [(',', 11357),
        ('the', 6704),
        ('.', 3676),
        ('and', 3537),
        ('to', 2884),
        (''', 2846),
        ('of', 2275),
        ('a', 2134),
        ('-', 1992),
        ('in', 1971)]
```

```
[337]:  fdlist1.pprint(maxlen = 10, stream = None)
```

```
FreqDist({',': 11357, 'the': 6704, '.': 3676, 'and': 3537, 'to': 2884, ''':
2846, 'of': 2275, 'a': 2134, '-': 1992, 'in': 1971, …})
```

```
[338]:  #4 Words at least 10 and their counts
        words_at_least_10 = sorted(w for w in set(odyssey_words) if len (w) >= 10)
        print(words_at_least_10)
        len(words_at_least_10)
        counts = {}
        for word in words_at_least_10:
            if (len(word) >= 10):
                counts[word] = counts.get(word, 0) + 1

        for word, count in counts.items():
            print(f'{word}: {count}')
```

```
['Amphiaraus', 'Amphilochus', 'Amphimedon', 'Amphinomus', 'Amphitrite',
'Amphitryon', 'Antilochus', 'Antiphates', 'Approaching', 'Astonished',
'Bashfulness', 'Bewitching', 'Cauconians', 'Cephallenian', 'Cephallenians',
'Clambering', 'Clytemnestra', 'Compassion', 'Daydreaming', 'Demoptolemus',
'Destruction', 'Embarrassed', 'Encouraging', 'Erechtheus', 'Erymanthus',
'Ethiopians', 'Eurylochus', 'Eurymachus', 'Eurymedusa', 'Falsehoods',
'Favoritism', 'Greatfleet', 'Halitherses', 'Hellespont', 'Hephaestus',
'Hippodamia', 'Impossible', 'Incredible', 'Iphimedeia', 'Lacedaemon',
```

'Laestrygonian', 'Laestrygonians', 'Megapenthes', 'Melanthius', 'Nausithous',
'Neoptolemus', 'Nevertheless', 'Orchomenos', 'Orsilochus', 'Ortilochus',
'Outrageous', 'Pelasgians', 'Periclymenus', 'Persephone', 'Phaeacians',
'Phaethousa', 'Philoctetes', 'Philoetius', 'Philomelides', 'Phoenician',
'Phoenicians', 'Pisistratus', 'Polydeuces', 'Polyphemus', 'Polyphides',
'Polytherses', 'Quicksilver', 'Reconsider', 'Rhadamanthys', 'Scrounging',
'Shipwrightson', 'Straightaway', 'Struggling', 'Suppliants', 'Surrounding',
'Suspicious', 'TRANSLATED', 'Telemachus', 'Theoclymenus', 'Thesprotia',
'Thesprotian', 'Thesprotians', 'Thrasymedes', 'Unforgettable', 'Unthinkable',
'Wonderstruck', 'absolutely', 'accomplished', 'accusations', 'addressing',
'adventures', 'affliction', 'affronting', 'appearance', 'appetizers',
'approached', 'approaching', 'assemblies', 'astonished', 'astounding',
'attendants', 'backsprung', 'banqueters', 'banqueting', 'bargainers',
'befriended', 'bellwether', 'bewitching', 'birthright', 'blacksmith',
'blithering', 'blustering', 'boisterous', 'borderland', 'bottomless',
'branchtops', 'brandished', 'brandishing', 'breathless', 'breathtaking',
'bridegroom', 'brightening', 'campaigner', 'captivating', 'catastrophe',
'cautiously', 'celebrating', 'celebration', 'challenged', 'challenging',
'chambermaid', 'charioteer', 'charioteers', 'chattering', 'childbirth',
'circumspect', 'clambering', 'clattering', 'clustering', 'collecting',
'colonnades', 'commandeer', 'commandeered', 'commanding', 'companions',
'compassion', 'competition', 'concentrate', 'concentrating', 'concluding',
'conditioning', 'confronting', 'connoisseur', 'conquering', 'conscience',
'conscientious', 'considerate', 'consorting', 'constructed', 'construction',
'consummated', 'contenders', 'contingent', 'contribute', 'conversation',
'convulsively', 'countryman', 'countrymen', 'countryside', 'countrywomen',
'crossbeams', 'cutthroats', 'declaration', 'defenseless', 'deliberate',
'deliberately', 'delicacies', 'delivering', 'demolished', 'destroying',
'destruction', 'determined', 'devastates', 'devastating', 'devastation',
'diplomatic', 'disastrous', 'discreetly', 'discretion', 'disembarked',
'disgusting', 'dispatched', 'dispatching', 'displacing', 'disrespect',
'dissolving', 'distinction', 'distraught', 'downstream', 'dumbfounded',
'dumbfounds', 'dumbstruck', 'earthquake', 'earthquakes', 'eighteenth',
'embarrassed', 'embarrassment', 'emblazoned', 'embroidery', 'enchanting',
'entertained', 'entertainment', 'enthralled', 'enthralling', 'entrancing',
'especially', 'evanescent', 'everlasting', 'everything', 'exclaiming',
'exclamations', 'exhaustion', 'expectation', 'expectations', 'expedition',
'exploiting', 'faithfully', 'falsehoods', 'fatherland', 'fieldhands',
'fleldhands', 'flickering', 'flittering', 'fluttering', 'footstools',
'foreboding', 'forebodings', 'forgetting', 'forthright', 'foundering',
'friendship', 'furthering', 'generation', 'generations', 'glistening',
'glittering', 'goatflocks', 'groundswell', 'handclasps', 'handsomest',
'handsprings', 'handstraps', 'headstrong', 'heartbreak', 'heartbreaking',
'horrendous', 'hospitality', 'households', 'housekeeper', 'humiliating',
'husbandAnd', 'hysterical', 'ignominious', 'illustrious', 'imagination',
'improviser', 'incessantly', 'inconceivable', 'indestructible', 'indignation',
'infiltrated', 'information', 'insatiable', 'inspiration', 'instructed',
'instructions', 'insufferable', 'intentions', 'interlopers', 'interrogating',

'interrupting', 'intolerable', 'introductions', 'invincible', 'invitation',
'irreproachable', 'irresistible', 'landlocked', 'lighthearted', 'lionhearted',
'livelihood', 'longhorned', 'magnanimous', 'magnificent', 'maneuvering',
'manhandled', 'mastermind', 'masterpiece', 'masterwork', 'merchantman',
'mercilessly', 'misfortune', 'misgivings', 'mortifying', 'mountaintops',
'multitudes', 'mysterious', 'nightingale', 'nonchalantly', 'outperformed',
'outrageous', 'outstretched', 'outstripped', 'overbearing', 'overflowing',
'overpowered', 'overpowering', 'overshadowed', 'overweening', 'overwhelmed',
'overwhelming', 'overwhelms', 'pandemonium', 'patrolling', 'perfection',
'pernicious', 'persuading', 'pomegranates', 'possessions', 'pretending',
'profiteers', 'prophecies', 'prophesied', 'protesting', 'provisions',
'punishment', 'punishments', 'questioned', 'questioning', 'reassurance',
'recklessness', 'recognized', 'recognizing', 'reembarked', 'refreshing',
'refreshment', 'reinforced', 'reinforcements', 'relentless', 'remembered',
'remembering', 'reparations', 'reproached', 'resemblance', 'resounding',
'resourceful', 'resplendent', 'returnEven', 'revelations', 'ridiculing',
'ringleaders', 'roundhouse', 'sacrificed', 'sacrifices', 'sacrificing',
'safekeeping', 'salutations', 'scampering', 'scandalized', 'scattering',
'scavenging', 'scoundrels', 'scrabbling', 'scrambling', 'scrounging',
'shamelessly', 'shattering', 'shearwater', 'sheepflocks', 'shimmering',
'shipwrecked', 'shipwright', 'shipwrights', 'shouldered', 'shouldering',
'shriveling', 'shuddering', 'skillfully', 'slaughtered', 'slaughtering',
'sonAgelaus', 'spearpoint', 'spearshaft', 'speechless', 'spellbinding',
'spellbound', 'splattered', 'sprinkling', 'squabbling', 'squandered',
'stampeding', 'standstill', 'stormclouds', 'stormwinds', 'straightaway',
'straightway', 'strangling', 'streambeds', 'stronghold', 'struggling',
'submission', 'sufferedBut', 'superhuman', 'suppliants', 'supplication',
'surrendered', 'suspecting', 'suspicious', 'swaggering', 'swineherds',
'sympathies', 'techniques', 'terrifying', 'themselves', 'thighbones',
'thirteenth', 'thoughtful', 'thoughtfulness', 'threatened', 'throttling',
'throughout', 'thundercloud', 'thunderhead', 'thunderheads', 'thundering',
'torchlight', 'tormenting', 'transforming', 'transgressors', 'treacherous',
'tremendous', 'triumphant', 'unappeased', 'unbearable', 'uncallused',
'uncontrollable', 'uncontrollably', 'underneath', 'understanding', 'underworld',
'undisguised', 'undisturbed', 'unendurable', 'unfamiliar', 'unforgettable',
'unfortunate', 'ungracious', 'unluckiest', 'unmistakable', 'unprotected',
'unrecognized', 'unstaggered', 'unstinting', 'unswerving', 'unwavering',
'upbraiding', 'upstanding', 'volunteered', 'volunteers', 'wanderings',
'weatherbeaten', 'wellspring', 'wheatlands', 'whirlwinds', 'widespread',
'withdrawing', 'wonderstruck', 'youngsters', 'yourselves']
Amphiaraus: 1
Amphilochus: 1
Amphimedon: 1
Amphinomus: 1
Amphitrite: 1
Amphitryon: 1
Antilochus: 1
Antiphates: 1

```
Approaching: 1
Astonished: 1
Bashfulness: 1
Bewitching: 1
Cauconians: 1
Cephallenian: 1
Cephallenians: 1
Clambering: 1
Clytemnestra: 1
Compassion: 1
Daydreaming: 1
Demoptolemus: 1
Destruction: 1
Embarrassed: 1
Encouraging: 1
Erechtheus: 1
Erymanthus: 1
Ethiopians: 1
Eurylochus: 1
Eurymachus: 1
Eurymedusa: 1
Falsehoods: 1
Favoritism: 1
Greatfleet: 1
Halitherses: 1
Hellespont: 1
Hephaestus: 1
Hippodamia: 1
Impossible: 1
Incredible: 1
Iphimedeia: 1
Lacedaemon: 1
Laestrygonian: 1
Laestrygonians: 1
Megapenthes: 1
Melanthius: 1
Nausithous: 1
Neoptolemus: 1
Nevertheless: 1
Orchomenos: 1
Orsilochus: 1
Ortilochus: 1
Outrageous: 1
Pelasgians: 1
Periclymenus: 1
Persephone: 1
Phaeacians: 1
Phaethousa: 1
```

```
Philoctetes: 1
Philoetius: 1
Philomelides: 1
Phoenician: 1
Phoenicians: 1
Pisistratus: 1
Polydeuces: 1
Polyphemus: 1
Polyphides: 1
Polytherses: 1
Quicksilver: 1
Reconsider: 1
Rhadamanthys: 1
Scrounging: 1
Shipwrightson: 1
Straightaway: 1
Struggling: 1
Suppliants: 1
Surrounding: 1
Suspicious: 1
TRANSLATED: 1
Telemachus: 1
Theoclymenus: 1
Thesprotia: 1
Thesprotian: 1
Thesprotians: 1
Thrasymedes: 1
Unforgettable: 1
Unthinkable: 1
Wonderstruck: 1
absolutely: 1
accomplished: 1
accusations: 1
addressing: 1
adventures: 1
affliction: 1
affronting: 1
appearance: 1
appetizers: 1
approached: 1
approaching: 1
assemblies: 1
astonished: 1
astounding: 1
attendants: 1
backsprung: 1
banqueters: 1
banqueting: 1
```

```
bargainers: 1
befriended: 1
bellwether: 1
bewitching: 1
birthright: 1
blacksmith: 1
blithering: 1
blustering: 1
boisterous: 1
borderland: 1
bottomless: 1
branchtops: 1
brandished: 1
brandishing: 1
breathless: 1
breathtaking: 1
bridegroom: 1
brightening: 1
campaigner: 1
captivating: 1
catastrophe: 1
cautiously: 1
celebrating: 1
celebration: 1
challenged: 1
challenging: 1
chambermaid: 1
charioteer: 1
charioteers: 1
chattering: 1
childbirth: 1
circumspect: 1
clambering: 1
clattering: 1
clustering: 1
collecting: 1
colonnades: 1
commandeer: 1
commandeered: 1
commanding: 1
companions: 1
compassion: 1
competition: 1
concentrate: 1
concentrating: 1
concluding: 1
conditioning: 1
confronting: 1
```

```
connoisseur: 1
conquering: 1
conscience: 1
conscientious: 1
considerate: 1
consorting: 1
constructed: 1
construction: 1
consummated: 1
contenders: 1
contingent: 1
contribute: 1
conversation: 1
convulsively: 1
countryman: 1
countrymen: 1
countryside: 1
countrywomen: 1
crossbeams: 1
cutthroats: 1
declaration: 1
defenseless: 1
deliberate: 1
deliberately: 1
delicacies: 1
delivering: 1
demolished: 1
destroying: 1
destruction: 1
determined: 1
devastates: 1
devastating: 1
devastation: 1
diplomatic: 1
disastrous: 1
discreetly: 1
discretion: 1
disembarked: 1
disgusting: 1
dispatched: 1
dispatching: 1
displacing: 1
disrespect: 1
dissolving: 1
distinction: 1
distraught: 1
downstream: 1
dumbfounded: 1
```

```
dumbfounds: 1
dumbstruck: 1
earthquake: 1
earthquakes: 1
eighteenth: 1
embarrassed: 1
embarrassment: 1
emblazoned: 1
embroidery: 1
enchanting: 1
entertained: 1
entertainment: 1
enthralled: 1
enthralling: 1
entrancing: 1
especially: 1
evanescent: 1
everlasting: 1
everything: 1
exclaiming: 1
exclamations: 1
exhaustion: 1
expectation: 1
expectations: 1
expedition: 1
exploiting: 1
faithfully: 1
falsehoods: 1
fatherland: 1
fieldhands: 1
fleldhands: 1
flickering: 1
flittering: 1
fluttering: 1
footstools: 1
foreboding: 1
forebodings: 1
forgetting: 1
forthright: 1
foundering: 1
friendship: 1
furthering: 1
generation: 1
generations: 1
glistening: 1
glittering: 1
goatflocks: 1
groundswell: 1
```

```
handclasps: 1
handsomest: 1
handsprings: 1
handstraps: 1
headstrong: 1
heartbreak: 1
heartbreaking: 1
horrendous: 1
hospitality: 1
households: 1
housekeeper: 1
humiliating: 1
husbandAnd: 1
hysterical: 1
ignominious: 1
illustrious: 1
imagination: 1
improviser: 1
incessantly: 1
inconceivable: 1
indestructible: 1
indignation: 1
infiltrated: 1
information: 1
insatiable: 1
inspiration: 1
instructed: 1
instructions: 1
insufferable: 1
intentions: 1
interlopers: 1
interrogating: 1
interrupting: 1
intolerable: 1
introductions: 1
invincible: 1
invitation: 1
irreproachable: 1
irresistible: 1
landlocked: 1
lighthearted: 1
lionhearted: 1
livelihood: 1
longhorned: 1
magnanimous: 1
magnificent: 1
maneuvering: 1
manhandled: 1
```

```
mastermind: 1
masterpiece: 1
masterwork: 1
merchantman: 1
mercilessly: 1
misfortune: 1
misgivings: 1
mortifying: 1
mountaintops: 1
multitudes: 1
mysterious: 1
nightingale: 1
nonchalantly: 1
outperformed: 1
outrageous: 1
outstretched: 1
outstripped: 1
overbearing: 1
overflowing: 1
overpowered: 1
overpowering: 1
overshadowed: 1
overweening: 1
overwhelmed: 1
overwhelming: 1
overwhelms: 1
pandemonium: 1
patrolling: 1
perfection: 1
pernicious: 1
persuading: 1
pomegranates: 1
possessions: 1
pretending: 1
profiteers: 1
prophecies: 1
prophesied: 1
protesting: 1
provisions: 1
punishment: 1
punishments: 1
questioned: 1
questioning: 1
reassurance: 1
recklessness: 1
recognized: 1
recognizing: 1
reembarked: 1
```

refreshing: 1
refreshment: 1
reinforced: 1
reinforcements: 1
relentless: 1
remembered: 1
remembering: 1
reparations: 1
reproached: 1
resemblance: 1
resounding: 1
resourceful: 1
resplendent: 1
returnEven: 1
revelations: 1
ridiculing: 1
ringleaders: 1
roundhouse: 1
sacrificed: 1
sacrifices: 1
sacrificing: 1
safekeeping: 1
salutations: 1
scampering: 1
scandalized: 1
scattering: 1
scavenging: 1
scoundrels: 1
scrabbling: 1
scrambling: 1
scrounging: 1
shamelessly: 1
shattering: 1
shearwater: 1
sheepflocks: 1
shimmering: 1
shipwrecked: 1
shipwright: 1
shipwrights: 1
shouldered: 1
shouldering: 1
shriveling: 1
shuddering: 1
skillfully: 1
slaughtered: 1
slaughtering: 1
sonAgelaus: 1
spearpoint: 1

```
spearshaft: 1
speechless: 1
spellbinding: 1
spellbound: 1
splattered: 1
sprinkling: 1
squabbling: 1
squandered: 1
stampeding: 1
standstill: 1
stormclouds: 1
stormwinds: 1
straightaway: 1
straightway: 1
strangling: 1
streambeds: 1
stronghold: 1
struggling: 1
submission: 1
sufferedBut: 1
superhuman: 1
suppliants: 1
supplication: 1
surrendered: 1
suspecting: 1
suspicious: 1
swaggering: 1
swineherds: 1
sympathies: 1
techniques: 1
terrifying: 1
themselves: 1
thighbones: 1
thirteenth: 1
thoughtful: 1
thoughtfulness: 1
threatened: 1
throttling: 1
throughout: 1
thundercloud: 1
thunderhead: 1
thunderheads: 1
thundering: 1
torchlight: 1
tormenting: 1
transforming: 1
transgressors: 1
treacherous: 1
```

```
tremendous: 1
triumphant: 1
unappeased: 1
unbearable: 1
uncallused: 1
uncontrollable: 1
uncontrollably: 1
underneath: 1
understanding: 1
underworld: 1
undisguised: 1
undisturbed: 1
unendurable: 1
unfamiliar: 1
unforgettable: 1
unfortunate: 1
ungracious: 1
unluckiest: 1
unmistakable: 1
unprotected: 1
unrecognized: 1
unstaggered: 1
unstinting: 1
unswerving: 1
unwavering: 1
upbraiding: 1
upstanding: 1
volunteered: 1
volunteers: 1
wanderings: 1
weatherbeaten: 1
wellspring: 1
wheatlands: 1
whirlwinds: 1
widespread: 1
withdrawing: 1
wonderstruck: 1
youngsters: 1
yourselves: 1
```

[339]: `len(words_at_least_10)`

[339]: 479

[349]:
```python
#5 The longest sentence
sentences = sent_tokenize(Odyssey1)
longest_sent = ""
```

```
max_sent_len = 0
for sentence in sentences:
    sentence_len = len(sentence.split())
    if sentence_len > max_sent_len:
        max_sent_len = sentence_len
        longest_sent = sentence

print("The Longest Sentence is :", longest_sent)
print("There are", max_sent_len, "words in the longest sentence")
```

The Longest Sentence is : They slipped the cable free of the drilled stone post
and soon as they swung back and the blades tossed up the spray
an irresistible sleep fell deeply on his eyes, the sweetest,
soundest oblivion, still as the sleep of death itself …
And the ship like a four-horse team careering down the plain,
all breaking as one with the whiplash cracking smartly,
leaping with hoofs high to run the course in no time-
so the stern hove high and plunged with the seething rollers
crashing dark in her wake as on she surged unwavering,
never flagging, no, not even a darting hawk,
the quickest thing on wings, could keep her pace
as on she ran, cutting the swells at top speed,
bearing a man equipped with the gods' own wisdom,
one who had suffered twenty years of torment, sick at heart,
cleaving his way through wars of men and pounding waves at sea
but now he slept in peace, the memory of his struggles
laid to rest.
There are 168 words in the longest sentence

[348]:
```
#6 A stemmed version of the longest sentence
stem = PorterStemmer()
stem_words = [stem.stem(word) for word in word_tokenize(longest_sentence)]

stemmed_longest_sentence = ' '.join(stem_words)
print("The Stemmed longest sentence is:", stemmed_longest_sentence)
```

The Stemmed longest sentence is: they slip the cabl free of the drill stone post
and soon as they swung back and the blade toss up the spray an irresist sleep
fell deepli on hi eye , the sweetest , soundest oblivion , still as the sleep of
death itself … and the ship like a four-hors team career down the plain , all
break as one with the whiplash crack smartli , leap with hoof high to run the
cours in no time- so the stern hove high and plung with the seeth roller crash
dark in her wake as on she surg unwav , never flag , no , not even a dart hawk ,
the quickest thing on wing , could keep her pace as on she ran , cut the swell
at top speed , bear a man equip with the god ' own wisdom , one who had suffer
twenti year of torment , sick at heart , cleav hi way through war of men and
pound wave at sea but now he slept in peac , the memori of hi struggl laid to
rest .

```
[ ]:
```

300-txt-assignment-1-1

October 15, 2023

```python
[1]: import nltk
     import numpy
     import matplotlib
     import os
```

```python
[2]: FilmCorpus = "C:\\Users\\Kenn\\GitHub\\300.txt"
```

```python
[3]: Film = open(FilmCorpus).read()
```

```python
[4]: len(Film)
```

```
[4]: 176534
```

```python
[5]: FilmTokens = nltk.word_tokenize(Film)
```

```python
[6]: len(FilmTokens)
```

```
[6]: 27957
```

```python
[7]: def lexical_diversity(FilmTokens):
         return len(set(FilmTokens)) / len(FilmTokens)
```

```python
[8]: lexical_diversity(FilmTokens)
```

```
[8]: 0.1526987874235433
```

```python
[9]: FilmFreq = nltk.FreqDist(FilmTokens)

     FilmFreq.most_common(10)
```

```
[9]: [('.', 2210),
      ('the', 1388),
      (',', 1191),
      ('of', 640),
      ('and', 618),
      ('his', 518),
      ('to', 486),
```

```
  ('a', 404),
  ('The', 342),
  ('Leonidas', 242)]
```

[10]:
```python
LongWords = []

for word in FilmTokens:
    if len(word) >= 10:
        LongWords.append(word)

LongWord_dist = nltk.FreqDist(LongWords)
for word, frequency in LongWord_dist.items():
    print(word, frequency)
```

```
extinguishes 1
battlefield 9
three-foot 1
uncontrollably 1
SPARTANCOURTYARD 3
Manufactured 1
grim-faced 1
emotionless 1
Constantly 1
initiation 1
surrounded 4
defenseless 1
unforgiving 1
heightened 2
snow-covered 1
thick-muscled 1
SPARTANGENERAL 1
SPARTANHOPLITES 1
transfixed 1
punctuation 1
storyteller 2
approaches 7
COUNTRYSIDE 1
BANNERSSNAP 1
overturned 1
embroidered 2
long-shouldered 1
men-at-arms 1
bone-white 1
half-dozen 1
headdresses 1
HUMMINGBIRDS 1
six-year-old 1
```

```
PLEISTARCHOS 2
hesitation 4
Hesitation 1
continuing 1
QUEENGORGO 1
broad-shouldered 1
Pleistarchos 9
MARKETPLACE 2
terra-cotta 1
Councilman 2
entertaining 1
accountable 1
annihilation 2
everything 3
bodyguards 6
submission 1
Submission 1
philosophers 1
diplomatic 2
reputation 1
themselves 9
ROCKYCLIFF 1
inaccessible 1
sixty-pound 1
torchlight 1
impossible 1
ungracious 1
exaggerate 1
interrupts 1
perpendicular 1
representing 1
demoralized 1
blasphemies 2
hypnotically 1
extinguishing 1
translates 1
crisscrossed 1
superstitious 1
manifested 1
BEDCHAMBER 1
second-story 1
overlooking 2
adolescent 1
collarbone 2
expression 5
personified 1
hard-bodied 1
baby-faced 1
```

```
COUNCILMEN 1
COUNCILMAN 3
disciplined 1
councilmen 1
Councilmen 2
city-states 1
complaints 1
commitment 1
moss-covered 1
profession 3
small-framed 1
blacksmith 2
ARCADIAN114 1
collective 4
responsibility 4
recklessness 1
understood 1
silhouetted 4
slaughtered 1
destruction 1
Footprints 1
footprints 2
blood-soaked 3
dark-skinned 2
nightmares 1
wind-blown 1
preparation 1
concubines 1
friendship 1
motherless 1
thunderbolts 1
SLOWMOTION 1
illuminates 1
splintered 1
jubilation 2
motionless 1
impossibly 1
encampment 3
half-man/half-unknown 1
determination 2
questioning 1
BODYGUARDS 1
calisthenics 1
ambassador 1
escarpments 1
half-stands 1
approaching 1
frustrated 1
```

```
sharpening 2
his-horrified 1
protection 1
relentlessly 1
Ambassador 1
channeling 1
impenetrable 2
demonstrates 1
disappearing 2
outcropping 1
shade-dappled 1
formations 1
ENCAMPMENT 2
destroying 1
GROWINGCLOSER 1
Blue-turbaned 1
juggernaut 1
ivory-handled 1
snow-white 1
Apocalypse 1
razor-tipped 1
Cappadocians 1
Babylonians 1
puncturing 1
incalculable 1
Electrified 1
relentless 3
High-pitched 1
caterwauls 1
continuous 1
momentarily 2
recognition 1
sand-colored 1
quickening 1
tremendous 1
half-naked 1
hand-tooled 1
marketplace 1
BLACKSMITHS 1
disappeared 1
HOUSEMAIDS 1
comfortable 1
Loyalist's 1
remembering 1
passageway 1
unfortunate 2
contingent 1
assassinate 1
```

```
impressive 2
androgynous 1
adornments 1
regrettable 1
misunderstanding 1
fascinating 1
fundamental 1
difference 1
vigorously 1
slaughtering 2
reasonable 1
punishable 1
toenail-like 1
shark-like 1
DRUMSFLOURISH 1
MOVEACROSS 1
unblinking 1
IMMORTALCOMMANDER 3
retreating 2
recovering 1
hand-to-hand 1
collapsing 1
decapitating 1
dispatches 2
recognizes 1
thundering 2
bronze-covered 1
desperation 1
distracted 1
descending 1
Commander's 1
frightened 2
heavily-weighted 1
invincible 1
nigrescent 1
Haaaawoooo 1
smoldering 1
cauterizing 1
half-filled 1
inspecting 1
Silhouetted 1
projectiles 1
capillaries 1
unstoppable 1
infantrymen 1
ENCAMPMENTS 2
displeased 1
executioner 2
```

grotesquely 1
disciplines 1
disappointing 1
TRUMPETING 1
THUNDERING 1
constricted 1
surrendering 1
effortless 1
promontory 1
intimidation 1
blood-drunk 1
frightening 1
illuminating 1
conversation 2
disappoint 1
politician 1
willingness 1
opportunist 1
reinforces 1
Regardless 1
reinforcements 1
goat-headed 1
instruments 1
honeycombed 1
procession 1
Everything 1
Concubines 1
temptation 1
celebration 1
blacksmiths 1
contention 1
straighten 1
disappears 2
star-filled 1
impassioned 1
flickering 1
Transported 1
circumstance 1
SPARTANWARRIORS 1
transition 1
penetralia 1
appraising 1
principles 1
challenged 1
preservation 2
conviction 1
importantly 1
confidences 1

```
passionate 1
depression 1
antechamber 1
Queen-like 1
struggling 2
inglorious 1
midsection 1
compliments 1
congratulations 1
insufferable 1
generosity 1
completely 1
proclaimed 1
Answerable 1
stretching 2
restlessness 1
complaining 1
bloodthirsty 1
waist-high 1
unquenchable 1
surrounding 1
blood-running 1
splattered 1
waist-deep 1
SLOWLYACROSS 1
outstretched 1
Christ-like 1
confidence 1
obliteration 1
barbarians 1
commanding 1
thunderclap 1
GROUNDSHAKING 1
razor-sharp 1
```

[11]:
```python
FilmSentences = nltk.sent_tokenize(Film)

LongestSentence = (max(FilmSentences, key=len))

print(LongestSentence)
print("Count:", len(LongestSentence.split()))
```

WE SEE: the greatest gathering of men and animals the world has ever seen, for the valley below him, which stretches five miles across, bordered on one side by the sea and on the other by the mountains, contains from edge to edge a city of tents complete with roads that team with people and carts and horse and creatures no Greek eye has seen before.

```
Count: 66
```

```python
[14]: from nltk.stem import PorterStemmer

      Longest_Sentence_Token = nltk.word_tokenize(LongestSentence)

      Filmstemmer = PorterStemmer()

      Film_Sentences_Stem = []

      for word in Longest_Sentence_Token:
          Film_Sentences_Stem.append(Filmstemmer.stem(word))
      Film_Sentences_Stem = " ". join(Film_Sentences_Stem)
      print(Film_Sentences_Stem)
```

we see : the greatest gather of men and anim the world ha ever seen , for the
valley below him , which stretch five mile across , border on one side by the
sea and on the other by the mountain , contain from edg to edg a citi of tent
complet with road that team with peopl and cart and hors and creatur no greek
eye ha seen befor .