CSCI-8450

2018 Spring

Yuchen Wang

1. Exercise 2

☼ Use the corpus module to explore austen-persuasion.txt. How many word tokens does this book have? How many word types?

Code:

print(**"word tokens: "**, len(nltk.corpus.gutenberg.words(**'austen-persuasion.txt'**)))  
print(**"word types: "** , len(set(nltk.corpus.gutenberg.words(**'austen-persuasion.txt'**))))

Result:

Exercise 2

word tokens: 98171

word types: 6132

1. exercise 5 In this question, consider at least 10 pairs of nouns that you believe lexically related via holonym-meronym relation. List these pairs. Then verify your assumption using Wordnet related methods of NLTK. Provide output by NLTK. Provide summary on how often NLTK supported your conjecture.

☼ Investigate the holonym-meronym relations for some nouns. Remember that there are three kinds of holonym-meronym relation, so you need to use: member\_meronyms(), part\_meronyms(), substance\_meronyms(), member\_holonyms(), part\_holonyms(), and substance\_holonyms().

Code:

list\_computer = [(**"computer"**,**"busbar"**),(**"computer"**,**"cathode-ray\_tube"**), (**"computer"**,**"central\_processing\_unit"**),  
 (**"computer"**, **"chip"**), (**"computer"**,**"computer\_accessory"**), (**"computer"**,**"computer\_circuit"**),  
 (**"computer"**, **"data\_converter"**), (**"computer"**, **"disk\_cache"**), (**"computer"**, **"diskette"**),  
 (**"computer"**, **"hardware"**)]  
**for** group **in** list\_computer:  
 print(group[0],group[1])  
 **for** sa **in** wn.synsets(group[0]):  
 **for** sb **in** wn.synsets(group[1]):  
 **if** sb **in** sa.member\_meronyms():  
 print(**"member meronyms"**)  
 **if** sb **in** sa.part\_meronyms():  
 print(**"part\_meronyms"**)  
 **if** sb **in** sa.substance\_meronyms():  
 print(**"substance\_meronyms"**)  
 **if** sb **in** sa.member\_holonyms():  
 print(**"member\_holonyms"**)  
 **if** sb **in** sa.part\_holonyms():  
 print(**"part\_holonyms"**)  
 **if** sb **in** sa.substance\_holonyms():  
 print(**"substance\_holonyms"**)

Result:

computer busbar

part\_meronyms

computer cathode-ray\_tube

part\_meronyms

computer central\_processing\_unit

part\_meronyms

computer chip

part\_meronyms

computer computer\_accessory

part\_meronyms

computer computer\_circuit

part\_meronyms

computer data\_converter

part\_meronyms

computer disk\_cache

part\_meronyms

computer diskette

part\_meronyms

computer hardware

part\_meronyms

From the results, we can see that the relationship of computer and other elements of list\_computer are part\_meronyms.

1. exercise 9 In this question, consider text1 and text7. Find at least three pairs of words which have quite different meanings across the two texts. Support your conjecture by your NLTK-based findings.

◑ Pick a pair of texts and study the differences between them, in terms of vocabulary, vocabulary richness, genre, etc. Can you find pairs of words which have quite different meanings across the two texts, such as *monstrous* in Moby Dick and in Sense and Sensibility?

Code:

print(**"text1 vocabulary: "** , len(text1))  
print(**"text1 vocabulary richness: "** , len(set(text1))/len(text1))  
print(**"text7 vocabulary: "** , len(text7))  
print(**"text7 vocabulary richness: "** , len(set(text7))/len(text7))  
print(**"light in text 1: "**)  
text1.concordance(**"light"**)  
print(**"light in text 7: "**)  
text7.concordance(**"light"**)  
print(**"right in text 1: "**)  
text1.concordance(**"right"**)  
print(**"right in text 7: "**)  
text7.concordance(**"right"**)  
print(**"hand in text 1: "**)  
text1.concordance(**"hand"**)  
print(**"hand in text 7: "**)  
text7.concordance(**"hand"**)

Result:

text1 vocabulary: 260819

text1 vocabulary richness: 0.07406285585022564

text7 vocabulary: 100676

text7 vocabulary richness: 0.12324685128531129

light in text 1:

Displaying 25 of 79 matches:

ed . But presently I came to a smoky light proceeding from a low , wide buildin

on , I at last came to a dim sort of light not far from the docks , and heard a

e is an emigrant from there . As the light looked so dim , and the place , for

d about a table , examining by a dim light divers specimens of SKRIMSHANDER . I

ook it up , and held it close to the light , and felt it , and smelt it , and t

and boots , and then blowing out the light tumbled into bed , and commended mys

ong time . At last I slid off into a light doze , and had pretty nearly made a

n the passage , and saw a glimmer of light come into the room from under the do

ay a word till spoken to . Holding a light in one hand , and that identical New

nced to turn his face so towards the light , that I plainly saw they could not

igh time , now or never , before the light was put out , to break the spell in

instant , and then holding it to the light , with his mouth at the handle , he

tobacco smoke . The next moment the light was extinguished , and this wild can

o keep quiet , and let me get up and light the lamp again . But his guttural re

ment the landlord came into the room light in hand , and leaping from the bed I

ached to think of it . And it was so light too ; the sun shining in at the wind

se marble tablets , and by the murky light of that darkened , doleful day read

from off his swarthy brow , and the light leaping from his eye , made all his

on me . I have read ye by what murky light may be mine the lesson that Jonah te

per element of our essences , though light be more congenial to our clayey part

hat perhaps it were best to strike a light , seeing that we were so wide awake

d against mine , and blowing out the light , we rolled over from each other , t

lf , I , Ishmael , should infallibly light upon , for all the world as though i

slide of the scuttle open . Seeing a light , we went down , and found only an o

t workman ' s arm with some ethereal light ; if I shall spread a rainbow over h

light in text 7:

Displaying 17 of 17 matches:

Mrs. Yeargin 's case also casts some light on the dark side of school reform ,

ectively gives the deal `` the green light . '' Mr. Jennison said 0 Northeast B

took further steps 0 \* to impose on light trucks and vans the safety requireme

posed \* requiring stronger roofs for light trucks and minivans , \* beginning wi

rule requiring auto makers to equip light trucks and minivans with lap-shoulde

o promote vehicle occupant safety in light trucks and minivans through its exte

-1 to impose car safety standards on light trucks and vans . Safety advocates ,

to extend car-safety requirements to light trucks and vans , which now \*T\*-157

any vehicles classed \* as commercial light trucks actually carry more people th

ditional safety equipment \*ICH\*-1 in light trucks and minivans , including air

r-seat lap and shoulder belts in its light trucks , but plans \*-1 to begin \*-2

es seem rather modest , however , in light of other French wines from current v

o send continuously changing colored light to dozens of fiber-end bunches . MED

12-year average life . Activity was light in derivative markets , with no new

mostly unchanged to up 1\/8 point in light , cautious trading prior to tomorrow

s closed mostly higher in relatively light trading as farmers continued \*-1 to

and chief designer , Seymour Cray . Light trucks and vans will face the same s

right in text 1:

Displaying 25 of 151 matches:

as from pirates and robbers , is the right to royal fish , which are whale and

go ." -- IBID . " No , Sir , ' tis a Right Whale ," answered Tom ; " I saw his

than the True Whale " ( Greenland or Right Whale ) " in possessing a formidable

be A giant in might , where might is right , And King of the boundless sea ." -

ommerce surrounds it with her surf . Right and left , the streets take you wate

o to sea , I go as a simple sailor , right before the mast , plumb down into th

tisfaction of knowing that it is all right ; that everybody else is one way or

en -- the bar -- a rude attempt at a right whale ' s head . Be that how it may

n ' s blanket . " I thought so . All right ; take a seat . Supper ?-- you want

ning like an ape . The shavings flew right and left ; till at last the plane -

y plank thunders with trampling feet right over Jonah ' s head ; in all this ra

y it , seemed madness . It flew from right to left , and back again , almost in

empires ; other seamen having but a right of way through it . Merchant ships a

Come on , Queequeg ," said I , " all right . There ' s Mrs . Hussey ." And so i

and highest part of the jaws of the right - whale . Planted with their broad e

thou mayest as well sign the papers right off ," he added --" come along with

ou want of Captain Ahab ? It ' s all right enough ; thou art shipped ." " Yes ,

together cool and self - collected ; right in the middle of the room ; squattin

eg , " and he hasn ' t been baptized right either , or it would have washed som

stuff that ; and he handles it about right . I say , Quohog , or whatever your

sharp aim at it , he darted the iron right over old Bildad ' s broad brim , cle

is getting better , and will be all right again before long ." " All right aga

all right again before long ." " All right again before long !" laughed the str

" Look ye ; when Captain Ahab is all right , then this left arm of mine will be

en this left arm of mine will be all right ; not before ." " What do you know a

right in text 7:

Displaying 25 of 36 matches:

stock holdings , gives Faulding the right \* to increase its interest to 70 % o

y products company has the exclusive right \* to file a reorganization plan \*T\*-

's so wild about the funds ' frenzy right now is that many are trading at hist

e . The other has opposed a woman 's right \* to choose . '' `` Funny thing , ''

Coleman wants \*-1 to take away your right \* to choose and \*-1 give it to the p

Coleman wanted \*-1 to take away the right of abortion `` even in cases of rape

that \*T\*-85 changed wrong answers to right ones on a statewide test . After num

plan now because we feel 0 it 's the right time . '' But , says 0 \*T\*-1 the gen

ffer , shareholders will receive one right for each 105 common shares owned \* .

ach 105 common shares owned \* . Each right entitles the shareholder to buy $ 10

r amount of punitive damages and the right \* to license Retin-A elsewhere . In

sful , says 0 \*T\*-1 Mr. Baldwin . `` Right now , we 're lucky if after five yea

et . Panama was stripped \*-8 of this right because of U.S. differences with the

ary citizen of this democracy had no right \* to see what your government was te

as the law . And I apparently had no right \* to print hither what the Voice was

y and publicly declared the absolute right \*ICH\*-2 of everyone except the USIA

-1 . I had sought , in my suit , the right \* to print Voice material , which \*T

denied me \*-36 , and I had sought a right \* to receive the information , \*-3 a

ation , \*-3 arguing in effect that a right \* to print government information is

ion is n't very helpful if I have no right \* to get the information . But the c

from \*-2 passing laws abridging the right to free speech , '' Judge Donald O'B

w the situation is this : You have a right \*-2 to read Voice of America scripts

ice during business hours . I have a right \*-1 to print those scripts if I go t

`` You 've got two champions sitting right before you , '' said \*T\*-1 Mr. Baum

n \*-1 to say , `` It just seemed the right thing 0 \* to do \*T\*-2 at this minute

hand in text 1:

Displaying 25 of 214 matches:

his mortality . " While you take in hand to school others , and to teach them

whenever my hypos get such an upper hand of me , that it requires a strong mo

l , if just previous to putting your hand into the tar - pot , you have been l

- deck gets his atmosphere at second hand from the sailors on the forecastle .

f blackness , not houses , on either hand , and here and there a candle , like

between his legs . He was trying his hand at a ship under full sail , but he d

l spoken to . Holding a light in one hand , and that identical New Zealand hea

landlord came into the room light in hand , and leaping from the bed I ran up

was to be heard ; but a supernatural hand seemed placed in mine . My arm hung

ilent form or phantom , to which the hand belonged , seemed closely seated by

l fears , not daring to drag away my hand ; yet ever thinking that if I could

nsations at feeling the supernatural hand in mine were very similar , in their

ent was to crush himself -- boots in hand , and hat on -- under the bed ; when

ke but still reverential dexterity , hand over hand , mounted the steps as if

ll reverential dexterity , hand over hand , mounted the steps as if ascending

tween the marble cenotaphs on either hand of the pulpit , the wall which forme

rolling off -- serenest azure is at hand ." Nor was the pulpit itself without

he Bible , and at last , folding his hand down upon the proper page , said : "

wer is forced from Jonah by the hard hand of God that is upon him . "' I am a

gale howls louder ; then , with one hand raised invokingly to God , with the

: " Shipmates , God has laid but one hand upon you ; both his hands press upon

ut oh ! shipmates ! on the starboard hand of every woe , there is a sure delig

eet on the stove hearth , and in one hand was holding close up to his face tha

ore my splintered heart and maddened hand were turned against the wolfish worl

at down in the stern , paddle low in hand ; and when the ship was gliding by ,

hand in text 7:

Displaying 8 of 8 matches:

\*U\* 0 it posted \*T\*-1 . On the other hand , had it existed then , Cray Compute

gs . High test scores , on the other hand , bring recognition and extra money

eptitiously -- copy them out in long hand . But neither of us can copy the mat

. The program traders , on the other hand , portray old-fashioned stock picker

s 0 I 'm losing \*T\*-1 . On the other hand , if it goes way sky high , I always

0 \*T\*-2 . Speculation , on the other hand , sparked buying in certain incentiv

s also very complicated . On the one hand , Brazil started an ethanol program

` It has \*-1 to weigh , on the other hand , the relatively high price of sugar

We can see that for “light”, there are meanings of brightness, gently in two texts. “right” has meanings of “the right side”, “right now”. “hand” : apparatus of hand, “on the other hand” in two texts.

1. exercise 11

◑ Investigate the table of modal distributions and look for other patterns. Try to explain them in terms of your own impressionistic understanding of the different genres. Can you find other closed classes of words that exhibit significant differences across different genres?

Code:

cfd = nltk.ConditionalFreqDist(  
 (genre, word)  
 **for** genre **in** brown.categories()  
 **for** word **in** brown.words(categories=genre))  
genres = [**'adventure'**, **'editorial'**, **'fiction'**, **'government'**, **'mystery'**]  
modals = [**'can'**, **'could'**, **'may'**, **'might'**, **'must'**, **'will'**]  
cfd.tabulate(conditions=genres, samples=modals)

Result:

can could may might must will

adventure 46 151 5 58 27 50

editorial 121 56 74 39 53 233

fiction 37 166 8 44 55 52

government 117 38 153 13 102 244

mystery 42 141 13 57 30 20

1. Exercise 13

◑ What percentage of noun synsets have no hyponyms? You can get all noun synsets using wn.all\_synsets('n').

Code:

allword = len([synset **for** synset **in** list(wn.all\_synsets(**'n'**))])  
n = len([synset **for** synset **in** list(wn.all\_synsets(**'n'**)) **if** len(list(synset.hyponyms()))==0])  
percentage = n / allword  
print(**"the percentage is "**, percentage)

Result:

the percentage is 0.7967119283931072

1. exercise 18 In addition to omitting stopwords also omit punctuation. Run your function on Brown corpus. What are the first 5 bigrams your function outputs.

◑ Write a program to print the 50 most frequent bigrams (pairs of adjacent words) of a text, omitting bigrams that contain stopwords.

Code:

emma = nltk.corpus.gutenberg.words(**'austen-emma.txt'**)  
stopen = stopwords.words(**'english'**)  
print(nltk.FreqDist(nltk.bigrams(w **for** w **in** emma  
 **if** w **not in** stopen  
 **and** w.isalpha())).most\_common(50))

Result:

[(('Mr', 'Knightley'), 299), (('Mrs', 'Weston'), 256), (('Mr', 'Elton'), 229), (('Miss', 'Woodhouse'), 173), (('Mr', 'Weston'), 167), (('Frank', 'Churchill'), 151), (('Mrs', 'Elton'), 150), (('Mr', 'Woodhouse'), 135), (('I', 'think'), 132), (('Miss', 'Fairfax'), 125), (('I', 'shall'), 122), (('I', 'know'), 119), (('every', 'thing'), 118), (('I', 'sure'), 116), (('Miss', 'Bates'), 113), (('Jane', 'Fairfax'), 111), (('every', 'body'), 92), (('I', 'hope'), 88), (('I', 'could'), 85), (('I', 'must'), 85), (('young', 'man'), 84), (('I', 'never'), 80), (('I', 'cannot'), 67), (('said', 'Emma'), 65), (('great', 'deal'), 64), (('I', 'would'), 64), (('Emma', 'could'), 61), (('Mrs', 'Goddard'), 59), (('said', 'I'), 59), (('Miss', 'Smith'), 58), (('John', 'Knightley'), 58), (('I', 'believe'), 56), (('But', 'I'), 54), (('I', 'dare'), 54), (('dare', 'say'), 51), (('I', 'assure'), 51), (('Mr', 'Frank'), 50), (('Miss', 'Taylor'), 48), (('I', 'thought'), 48), (('My', 'dear'), 43), (('said', 'Mr'), 42), (('Mrs', 'Churchill'), 42), (('I', 'suppose'), 41), (('I', 'I'), 40), (('I', 'wish'), 39), (('Mr', 'Mrs'), 39), (('Mr', 'Perry'), 39), (('Mr', 'Martin'), 38), (('Emma', 'I'), 38), (('know', 'I'), 37)]

In the codes, stopen omits the stopwords, w.isalpha() omit punctuation. Therefore, in the code, I omit stopwords and punctuation in my code.

1. exercise 27 What are the values you have computed. Note that the same lemma may occur in different synsets. Your code should account for this fact

★ The polysemy of a word is the number of senses it has. Using WordNet, we can determine that the noun dog has 7 senses with: len(wn.synsets('dog', 'n')). Compute the average polysemy of nouns, verbs, adjectives and adverbs according to WordNet.

Code:

polysemy = 0.00  
num = 0.00  
partsofspeech = **''**english = [**'n'**, **'v'**, **'a'**, **'r'**]  
print(**"n = noun; v = verb, a = adjective, r = adverb"**)  
  
**for** e **in** english:  
 **for** synset **in** wn.all\_synsets(e):  
 lename = []  
  
 **for** lemma **in** synset.lemma\_names():  
 lename.append(lemma)  
 name = set(lename)  
  
 **for** n **in** lename:  
 polysemy += len(wn.synsets(n, e))  
 num += 1  
  
 average = 0.000  
 average = polysemy / num  
 print(**"The average polysemy of {} is {:.3f}"**.format(e, average))  
 polysemy = 0.00  
 num = 0.00

Result:

n = noun; v = verb, a = adjective, r = adverb

The average polysemy of n is 1.853

The average polysemy of v is 5.098

The average polysemy of a is 2.234

The average polysemy of r is 1.699