

1. In "austen-persuasion.txt"  
word tokens : 98171  
word types : 6132
2. The 10 noun pairs which are in holonym-meronym relation are [('house', 'library'), ('university', 'college'), ('school', 'staff'), ('tree', 'wood'), ('tree', 'forest'), ('foot', 'leg'), ('feather', 'bird'), ('ice', 'glacier'), ('melanin', 'skin'), ('river', 'water')]

I selected most pairs of nouns after analysis and understanding of the topic, except ('melanin', 'skin'), ('river', 'water'), but wordnet was able to find all of my noun pairs.

The noun pairs selected are :

(['house', 'library'], ('university', 'college'), ('school', 'staff'), ('tree', 'wood'), ('tree', 'forest'), ('foot', 'leg'), ('feather', 'bird'), ('ice', 'glacier'), ('melanin', 'skin'), ('river', 'water'))

Finding relations for given noun pairs...

```
part_meronyms : [('house', 'library'), ('river', 'water')]
member_meronyms : [('university', 'college'), ('school', 'staff')]
substance_meronyms : [('tree', 'wood')]
member_holonyms : [('tree', 'forest')]
part_holonyms : [('foot', 'leg'), ('feather', 'bird')]
substance_holonyms : [('ice', 'glacier'), ('melanin', 'skin')]
total relations identified : 10
```

3. Vocabulary size of text1 = 19317  
Vocabulary size of text7 = 12408  
Text1 has more vocabulary than text7.

Text richness of text1 : 0.07406285585022564

Text richness of text7 : 0.12324685128531129

Text7 has used more distinct words than text1

Content fraction of text 1 : 0.07275926983847035

Content fraction of text 7 : 0.12114108625690333

Almost both text1 and text7 has used equal amount of stop words.

Collocations used in text1 :

Sperm Whale; Moby Dick; White Whale; old man; Captain Ahab; sperm whale; Right Whale; Captain Peleg; New Bedford; Cape Horn; cried Ahab; years ago; lower jaw; never mind; Father Mapple; cried Stubb; chief mate; white whale; ivory leg; one hand

Collocations used in text7 :

## CSCI 8450 Chapter 2 : Assignment

Mohan Sai Ambati  
Collaborated with Sai Tarun Battula

million \*U\*; New York; billion \*U\*; Wall Street; program trading; Mrs.  
Yeargin; vice president; Stock Exchange; Big Board; Georgia Gulf;  
chief executive; Dow Jones; S&P 500; says \*T\*-1; York Stock; last  
year; Sea Containers; South Korea; American Express; San Francisco

Top 5 common words used in text1 : [('the', 13721), ('of', 6536), ('and', 6024), ('a', 4569), ('to', 4542)]

Top 5 common words used in text7 : [('the', 4045), ('of', 2319), ('to', 2164), ('a', 1878), ('in', 1572)]

To find the pairs of words which have different meaning in two texts, this is my approach

*Step 1:* I found common words in text1 and text7.

*Step2:* I found context of each word in two texts. To find the context of the word I found bigrams of the word in two texts.

*Step3:* check if the word in text1 and text7 are in same context or not.

To do that I have compared bigrams obtained from text1 with bigrams obtained from text7, I removed all the filter only the words that have different bigrams in two texts.

*Step 4:* Further filtered words by checking if these bigrams are normally used or not.

To do that I have created my own context dictionary using 'english-kjv.txt' and 'english-web.txt' in genesis corpus. I have filtered all the bigrams that are commonly used in my dictionary.

After this processing I found 395 words that are used with different bigrams. In that I analyzed manually and selected first four : ['condemned', 'Intermediate', 'sketching', 'promotion']

Condemned:

Text1: inding one of the bulwarks of some condemned old craft . On one side hung a ver

Meaning: disapproval

Text7 : e rebels . A White House spokesman condemned the truce suspension as `` deplora

Meaning : punish

'Intermediate':

Text1: e halving an apple ; there is no intermediate remainder . Now as this law , un

Meaning: in between thing

Text 7: 1 push up crude oil . West Texas Intermediate crude for December delivery rose

Meaning: "West Texas Intermediate" name of a company

'sketching':

Text1: concerned , is about tantamount to sketching the profile of a pyramid . Even Sc

Meaning: general outline

Text7: s routine . He spends his days \*-1 sketching passers-by , or \*-1 trying \*-2 to

Meaning: drawing

'promotion' :

Text 1: ements to embark , fine chance for promotion , it seems -- aye , a stove boat w

Meaning: raising to higher position

Text7: hed American Express about a joint promotion because its card holders generally

Meaning: publicity

4. In the table “will” is used in news, “can” is used in religion genre , “can” and “will” are used most in hobbies, “could” is used in science\_fiction and “could” is mostly used in romance and humor. Similar to this I have created modals for pronouns, adverbs and most common words. The results are as follows:

Modal used : pronouns

	i	me	mine	you	yours	his	her	hers	we	they	them
news	179	31	4	66	0	428	121	0	107	267	96
religion	155	33	1	123	3	177	8	0	218	133	74
hobbies	154	16	2	434	1	258	17	0	159	217	127
science_fiction	98	21	1	96	3	102	82	0	45	67	47
romance	951	194	8	558	6	603	680	3	109	237	142
humor	239	56	3	148	2	145	68	0	50	98	49

Modal used : adverbs

	up	so	out	just	now	how	then	more	also	here
news	168	81	162	49	87	42	73	183	129	72
religion	37	99	54	24	45	28	59	65	57	17
hobbies	103	121	87	74	75	72	88	194	110	41
science_fiction	33	26	36	13	27	16	18	21	2	11
romance	212	192	223	128	112	77	143	89	16	73
humor	51	56	44	24	28	25	37	37	13	7

Modal used : most\_common\_words

	be	to	of	and	a	in	that	have	i	it	for	not	on
news	526	2144	2861	2186	2130	2020	829	265	179	478	969	257	691
religion	243	904	1505	962	694	774	492	194	155	337	315	236	170
hobbies	524	1836	2412	2178	1858	1543	526	376	154	605	857	223	537
science_fiction	80	306	329	294	236	164	131	61	98	158	100	95	90
romance	292	1517	1202	1905	1383	930	612	264	951	717	441	273	375
humor	79	465	518	529	523	360	252	56	239	210	158	84	130

5. Total noun synsets : 82115  
 noun synsets with no hyponyms : 65422  
 Percentage of noun synsets with no hyponyms : 79.67119283931072
6. Most common 50 bigrams : [ (('United', 'States'), 392), (('New', 'York'), 296), (('per', 'cent'), 146), (('years', 'ago'), 136), (('I', 'could'), 130), (('I', 'think'), 123), (('I', 'said'), 122), (('I', 'would'), 109), (('I', 'know'), 105), (('The', 'first'), 96), (('He', 'said'), 92), (('Rhode', 'Island'), 90), (('Of', 'course'), 87), (('could', 'see'), 87), (('In', 'addition'), 85), (('I', 'thought'), 70), (('last', 'year'), 68), (('first',

'time'), 67), (('It', 'would'), 65), (('White', 'House'), 65), (('For', 'example'), 64), (('even', 'though'), 63), (('I', 'saw'), 62), (('I', 'knew'), 61), (('World', 'War'), 60), (('He', 'could'), 58), (('two', 'years'), 58), (('U.', 'S.'), 57), (('In', 'fact'), 57), (('fiscal', 'year'), 57), (('last', 'night'), 56), (('But', 'I'), 56), (('I', 'want'), 55), (('high', 'school'), 54), (('old', 'man'), 52), (('I', 'told'), 52), (('Peace', 'Corps'), 52), (('He', 'would'), 51), (('I', 'believe'), 50), (('No', 'one'), 49), (('New', 'England'), 49), (('United', 'Nations'), 49), (('The', 'man'), 48), (('I', 'felt'), 47), (('I', 'asked'), 47), (('one', 'thing'), 47), (('young', 'man'), 47), (('Los', 'Angeles'), 47), (('I', 'can't'), 46), (('I', 'went'), 45)]

Most common 5 bigrams : [ (('United', 'States'), 392), (('New', 'York'), 296), (('per', 'cent'), 146), (('years', 'ago'), 136), (('I', 'could'), 130)]

**7. There are two ways to compute that**

1. Using “wordnet.synsets(word\_class)” we will get all synsets of that particular word class, we need to find the lemma names for each lemma in synset and take care of duplicates and find the average.
2. Using “wordnet.all\_lemma\_names(word\_class)” we will get all lemma names directly and we can directly calculate the average.

Second method is taking less time when compared to first one. Both are written in python file.

The average polysemy of nouns : 1.2610825311125826

The average polysemy of verbs : 2.1865729898516784

The average polysemy of adjectives : 1.406536617160948

The average polysemy of adverbs : 1.2532916759651864