Unit 2 Assignment

**ANLY:520-51 (Fall 2016)**

Dean D’souza

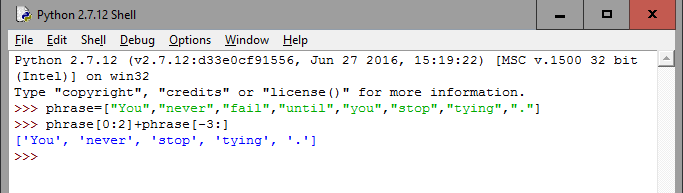
H.U. ID: 168424

# **Solutions:**

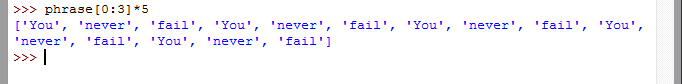
1. **Create a variable phrase containing a list of words. Experiment with the operations described in this chapter, including addition, multiplication, indexing, slicing, and sorting.**

We perform the commands to display the required techniques as follows:

1. Addition (also includes slicing):

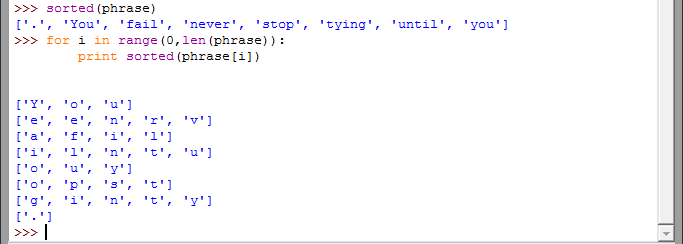


1. Multiplication (also includes slicing):



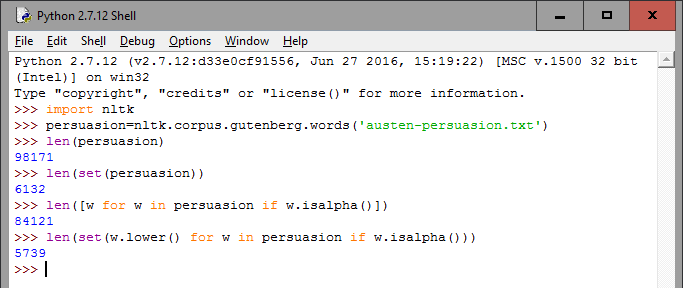
3. Indexing:

4. Slicing:

5. Sorting: 

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

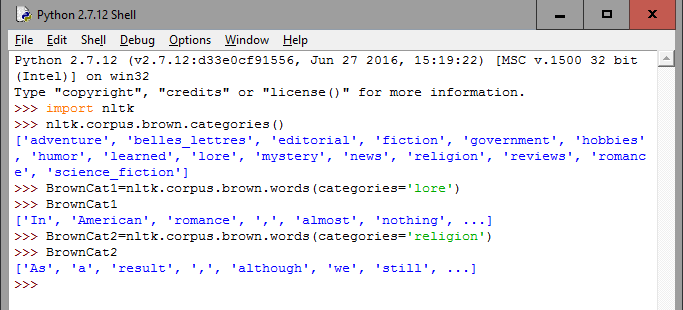
We perform the required commands as follows:

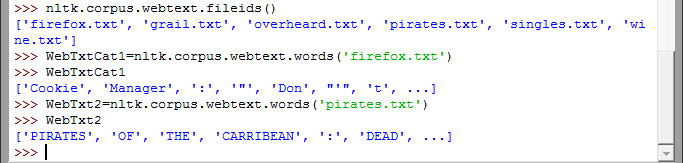
****

It should be noted that the first two counts represent total number of types and total number of unique types which includes numeric and punctuations. The next two counts represent the total number of types and the total number of unique types which are alphabetical.

1. **Use the Brown corpus reader nltk.corpus.brown.words() or the Web text corpus reader nltk.corpus.webtext.words() to access some sample text in two different genres.**

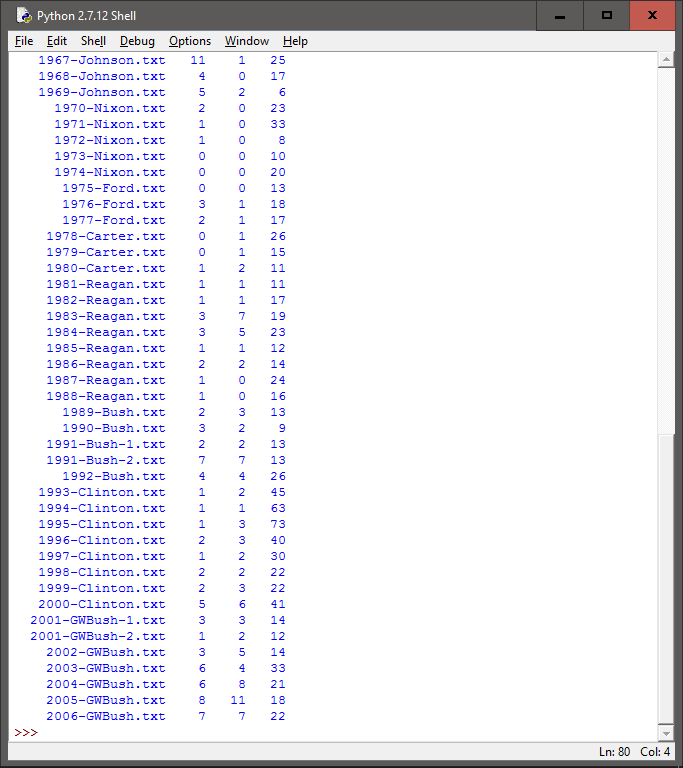
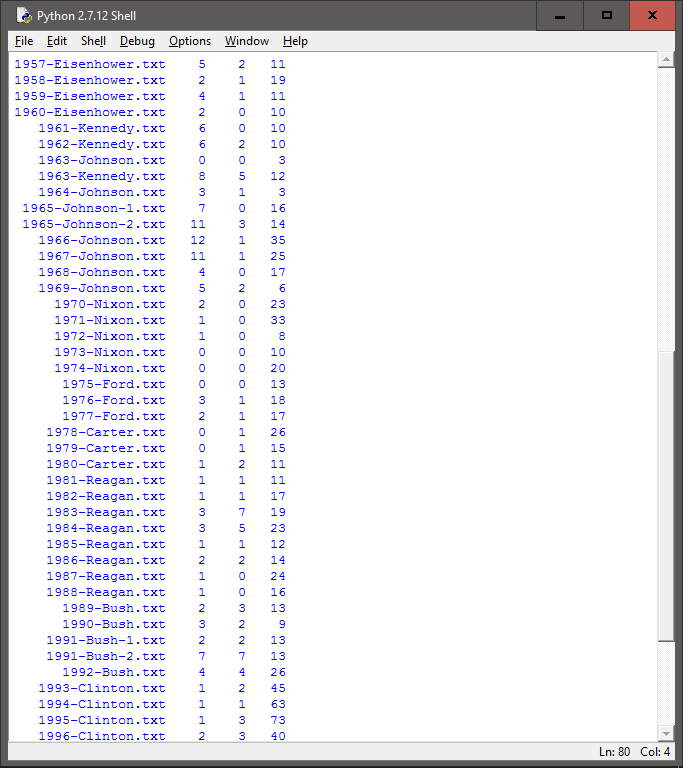
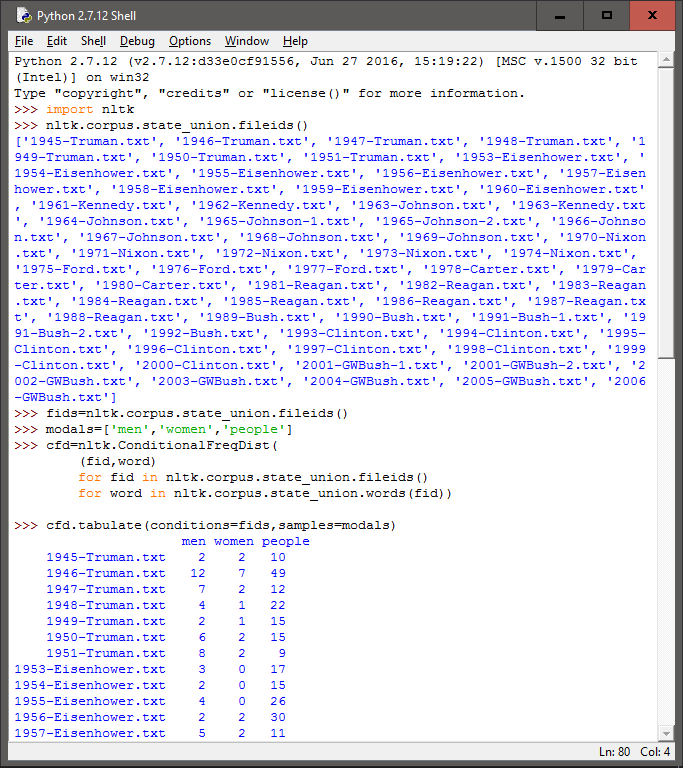
We perform the necessary commands as follows for the brown corpus:

****

We can also do the same for the webtext corpus, however, it should be noted that there are no clearly defined categories and categories() function. Hence we choose to compare over fileids, which are representative of the source of the texts (and may be considered as categories?).****

1. **Read in the texts of the State of the Union addresses, using the state\_union corpus reader. Count occurrences of men, women, and people in each document. What has happened to the usage of these words over time?**

We perform the necessary commands as follows:

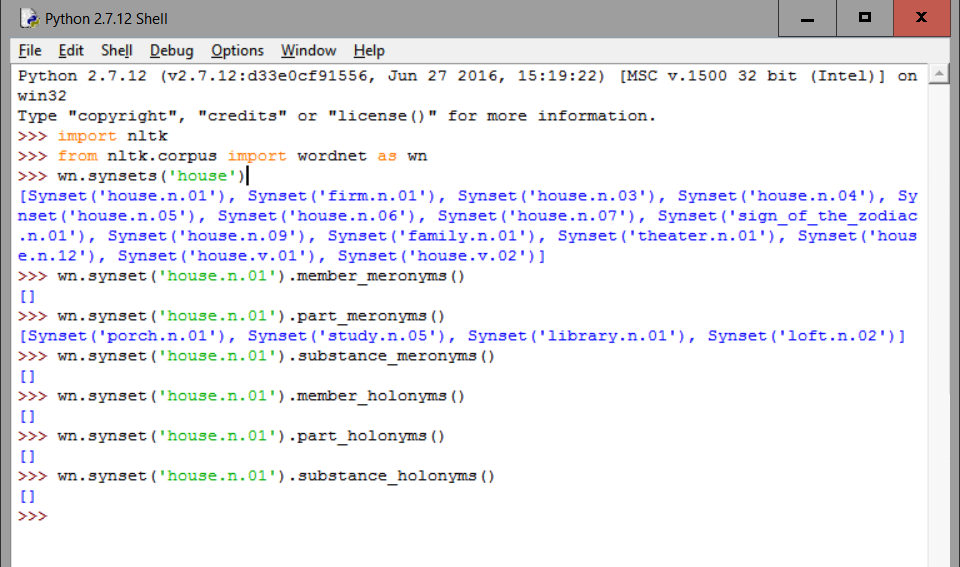
****

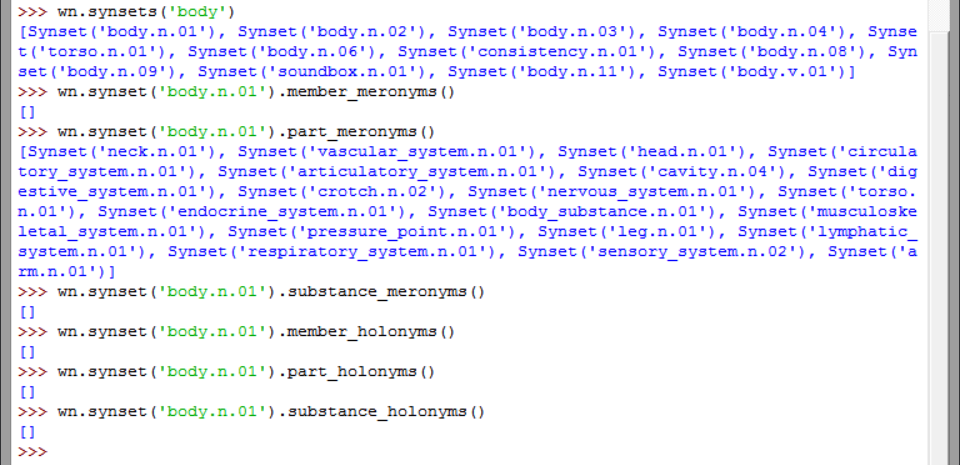
One thing to take into consideration while examining the output is that it does not take into consideration those forms of the word which have punctuations attached to it, so our output might not be completely accurate. However, based on the output obtained we can conclude that the words ‘men’ has been used to maximum times by Truman in 1946 and Johnson in 1966, meanwhile women has been used the maximum times in recent years (specifically in 2005 by George W. Bush), also ‘people’ was used the maximum number of times by Clinton in 1995. Overall we can see that ‘men’ has been used relatively consistently, while ‘women’ has been used more in recent times, also while ‘people’ was used rather frequently in the 90’s, it has seen a decline since then.

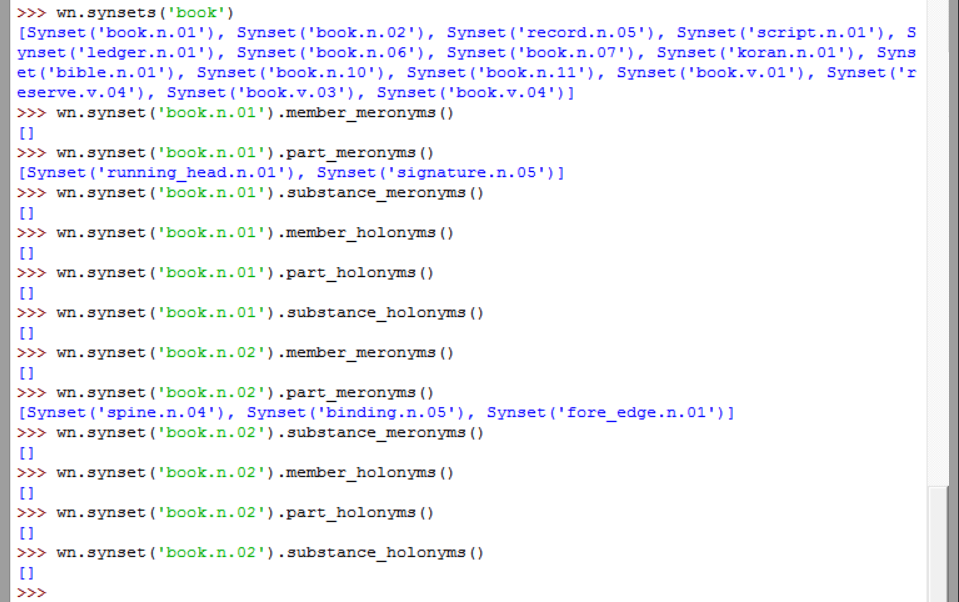
1. **Investigate the holonym-meronym relations for some nouns. Remember that there are three kinds of holonym-meronym relation, so you need to use: member\_meronyms(), part\_meronyms(), substance\_meronyms(), member\_holonyms(), part\_holonyms(), and substance\_holonyms().**

We perform the necessary commands as follows:

1. House:



1. Body:
2. Book:



In the above investigation of holonyms and meronyms of the word book we try to take a look at the two different listings of the word book in word net to try to get a better idea of how such listings work.