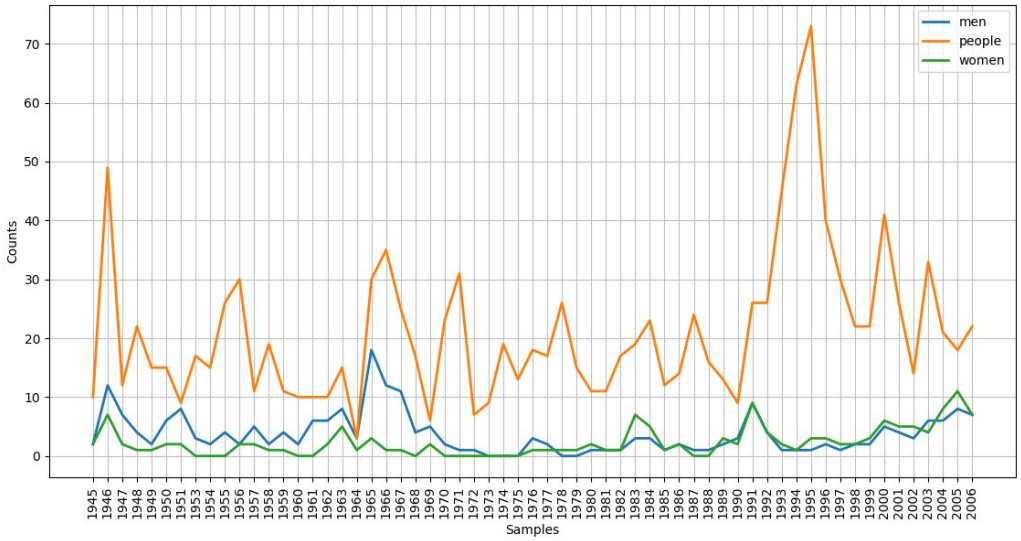
Name Jing Low
Section 27F3
Assignment 03
Due Date January 31, 2018

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?**
   See function problem_1() in hw3.py for the code to generate the plot below.



   Throughout the years included in this corpus, usage of the word “people” constantly remained higher than “men” and “women” and reached its peak in 1995, which could be marking the government’s emphasis on the citizens around that time period. On the other hand, the word “men” was used more frequently than the word “women” until the year 1973, when these two words reached an equilibrium. This shows that the country before 1973 was most probably more of a patriarchal society, where males held dominant power and women were not a part of politics. Afterwards, “men” and “women” were used nearly as frequently in the texts, with “women” sometimes occurring more frequently than “men” in the texts. This clearly shows a shift towards gender equality and the rise of feminism.

2. **What percentage of noun synsets have no hyponyms? You can get all noun synsets using wn.all_synsets('n').**
   See problem_2() in hw3.py for the code to get the result below.

   About 79.67% of noun synsets have no hyponyms.

3. **Write a program to create a table of word frequencies by genre, like the one given in 1 for modals. Choose your own words and try to find words whose presence (or absence) is typical of a genre. Discuss your findings.**

   See function problem_3() in hw3.py for the code to generate the tables below.

	angry	happy	sad	scared
adventure	8	2	3	2
belles_lettres	10	9	6	0
editorial	1	8	3	1
fiction	4	9	3	3
hobbies	0	5	0	0
humor	1	4	0	0
learned	0	4	1	1
lore	3	2	1	0
mystery	5	11	4	3
news	3	12	1	0
religion	0	4	2	0
reviews	0	5	1	0
romance	5	8	6	9
science_fiction	2	8	3	0

	anger	fear	joy	sadness
adventure	8	7	2	1
belles_lettres	12	37	10	1
editorial	1	9	3	1
fiction	3	12	5	1
government	0	1	0	0
hobbies	0	2	1	0
humor	1	5	1	0
learned	4	6	1	0
lore	4	5	5	0
mystery	3	5	2	0
news	0	4	0	0
religion	0	24	3	1
reviews	1	1	3	0
romance	9	4	4	1
science_fiction	0	1	0	0

Two ConditionalFreqDists, one for adjectives and one for nouns, were constructed to investigate the occurrences of the four primary emotions in each genre covered in the Brown corpus. Not surprisingly, all four of the primary emotion words occur in categories such as “adventure”, ”belles_lettres”, ”fiction”, and ”mystery”. On the other hand, more formal categories such as “government” contains almost none of these words. Additionally, the term “fear” appeared significantly frequently in genres such as belles_lettres and religion, which hints to the style of fine writing (what belles_lettres is) and religious texts. One possible explanation for this frequency could be that both of

these genres discuss human nature with respect to some higher, greater identity or force that humans would be afraid of. Also, interestingly, the term “sadness” is used relatively infrequently throughout all genres even though it is quite a common, recurring theme in many writings. This might be explained by the fact that writers often avoid stating the emotion directly; they would usually express the feeling through imagery or other literary devices. This explanation can also be applied to the absence of some other emotions in certain texts.

However, to fully investigate the presence of such emotions in the text semantically, we must consider all terms that constitute to these four emotions and build the frequency distribution chart based on that data.