Short answers

Question 1

State of the Union Speech

What has happened to the usage of these words over time?

When we look at the usage of the word ‘men’, it remained relatively random, spiking some years and dipping other years. The word ‘women’ remained relatively low throughout the years until around President Ford’s speech in 1976 because after that year the word ‘women’ was used quite consistently. This makes sense because the feminist movement occurred during this period and as a result women became more prominent in the United States and therefore were addressed more consistently following this period. The usage of the word ‘people’ remain relatively constant throughout, it is the most heavily used word out of the three probably because it addresses everyone.

Question 3

Genre

For this exercise, I used news, religion, hobbies, romance and humor as my genres and president, God, sports, love and funny as my words. I expected ‘president’ to show up many times in the news obviously because presidents are world leaders and therefore their smallest actions are scrutinized in the news. The word ‘God’ was obviously going to pop up lots of times in religion, I chose this word just to see how many times it would appear in the religion genre and it appears the most out of any of the other words in any other category. I was quite surprised to find 23 occurrences of God in romance genres, but it may just be part of a dialogue if someone finds out about something dramatic happened and then says “Oh God!” It also made sense that God didn’t appear too many times in humor because people often don’t poke at religion since it is a sensitive topic. When I looked at the genre ‘hobbies’ I had to test several words before getting one which came up with a reasonable number of occurrences. I did this by thinking of all the possible activities which may possibly be hobbies and used them. I figured ‘sports’ would be one general enough to get several occurrences. It also made sense that the word ‘sports’ popped up in the news given that there is a comprehensive coverage of sports in the United States. The word ‘love’ was expected to be in the romance genre. Also there were a few hits in hobbies because the book may have contained phrases like ‘I love to play tennis’. Lastly, when we look at the word ‘funny’, we see that there are many hits in the humor genre which makes sense given that humor is supposed to be funny. I noticed a few occurrences of ‘funny’ in romance, this may be in those romantic-comedy books or maybe someone in the book tells someone else they’re funny.

Question 4

Upon observing the list of words that accounted for a third of all word tokens, it was obvious that articles such as ‘the’ and ‘an’ would appear a lot. Words like ‘of’, ‘that’, ‘in’ and ‘had’ were also displayed because these words often are used as ‘helper’ words in sentences. I noticed that a lot of these words were related to government or the judicial system such as ‘election’ or ‘executive’ or ‘jury’. This type of program which indicates a third of all word tokens can provide us with a good idea of what the overall genre of the corpus is going to be about. Although there were some words that could be categorized under topics related to the government or judicial system, there were some words which seemed somewhat out of place such as ‘Friday’ or ‘produced’. I can’t seem to comprehend where or why the word ‘Friday’ appears so much in the brown corpus.