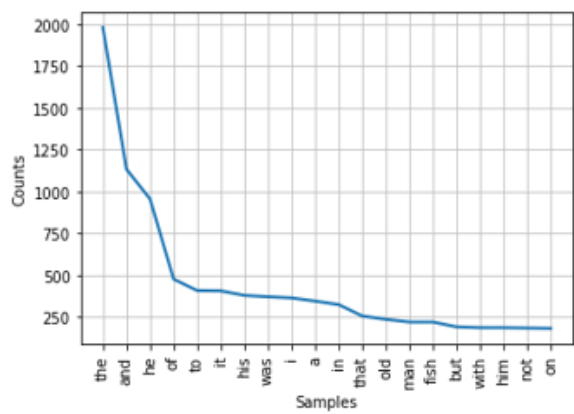
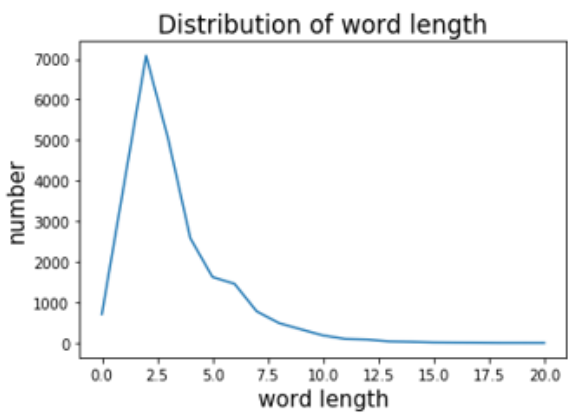
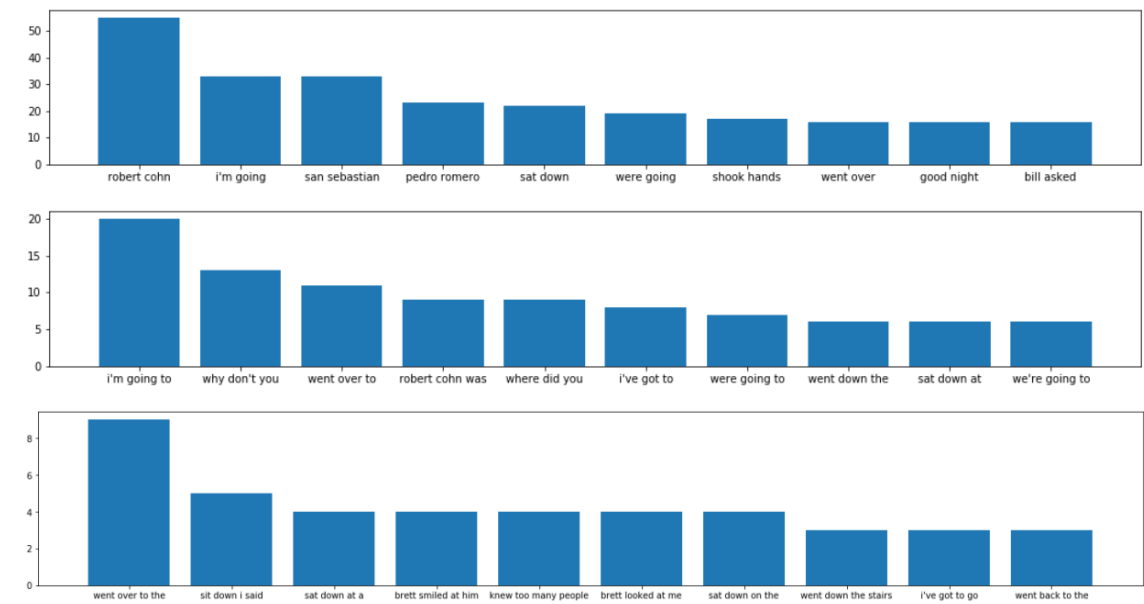
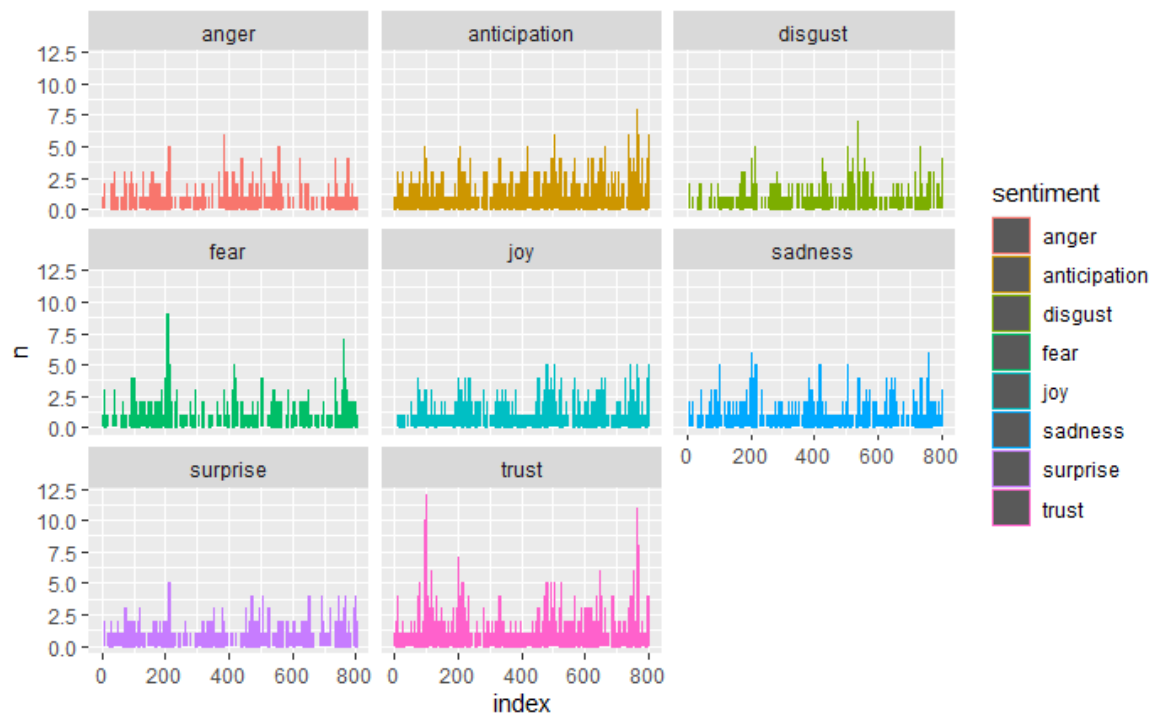
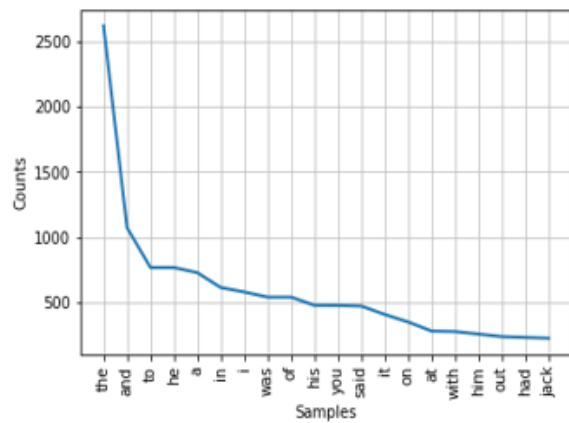
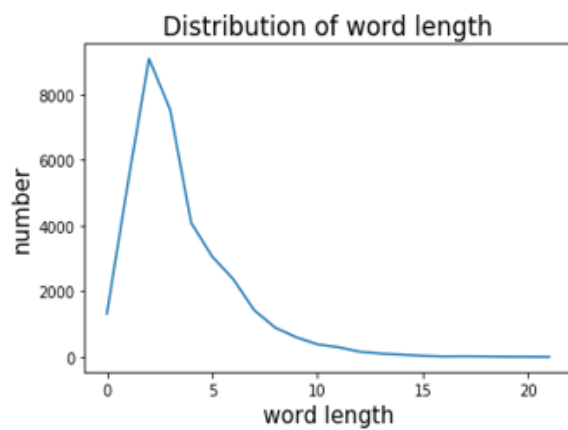
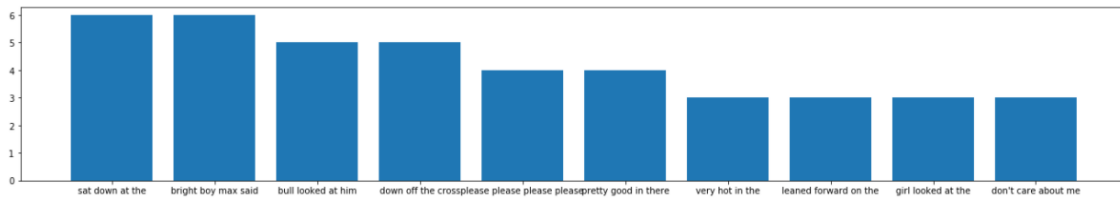
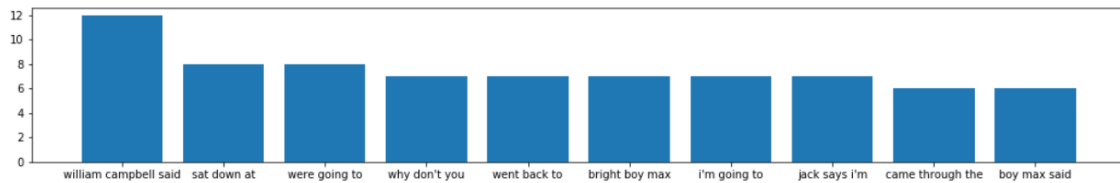
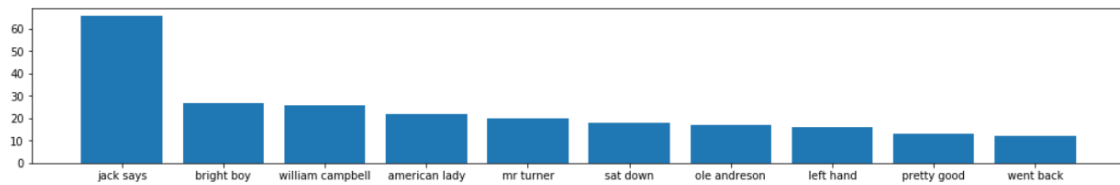


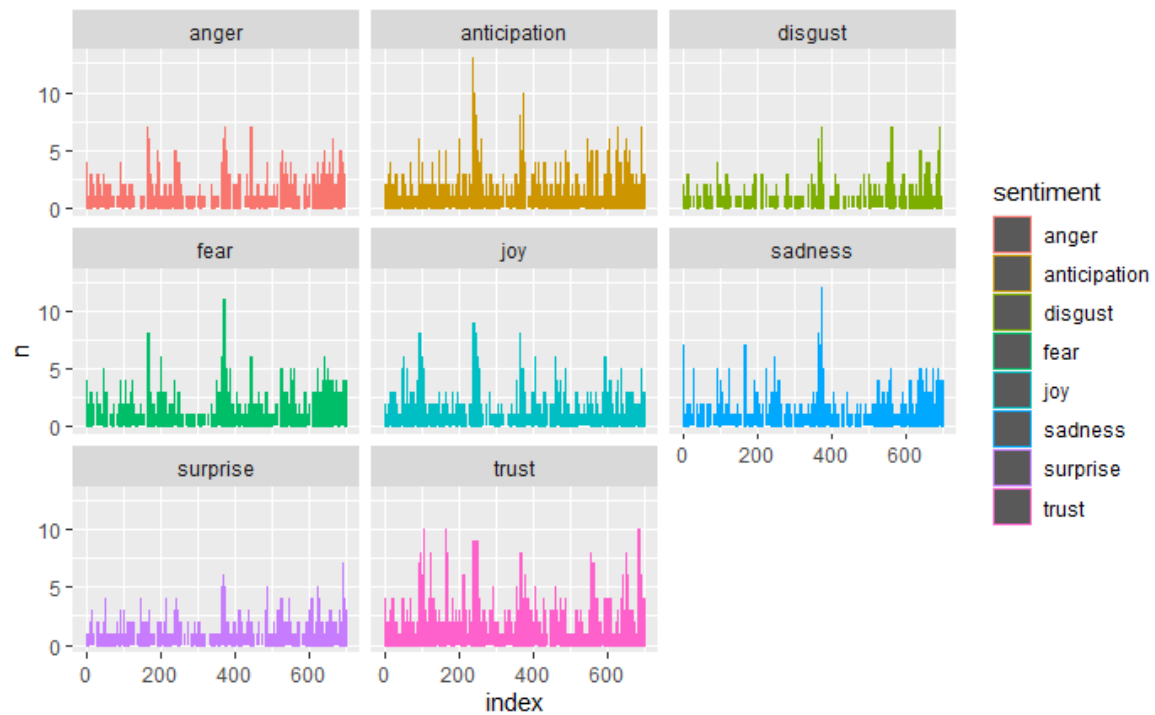
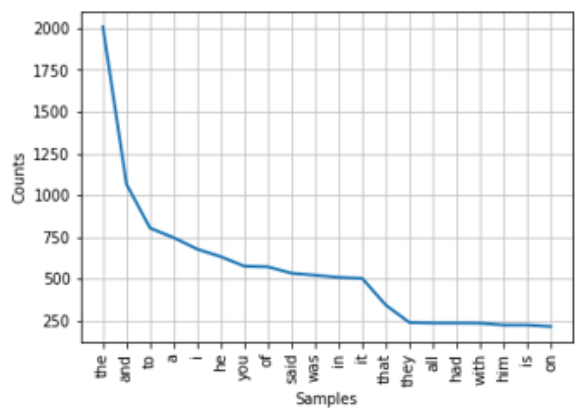
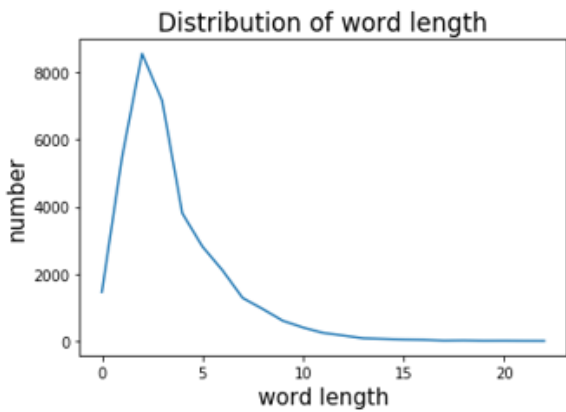
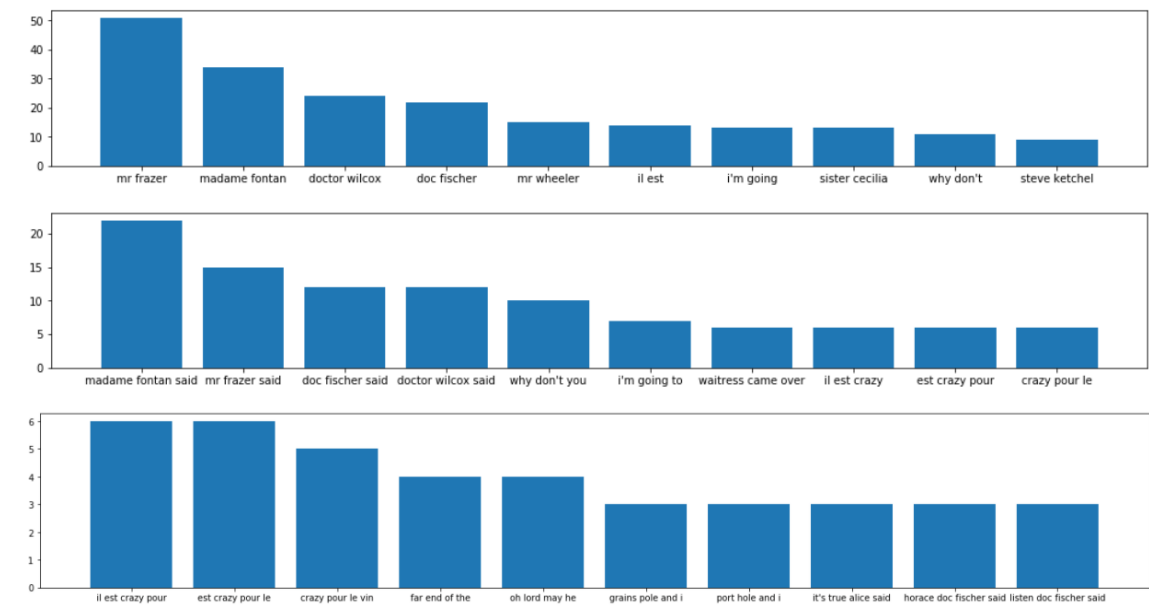
The Sun Also Rises (1926)



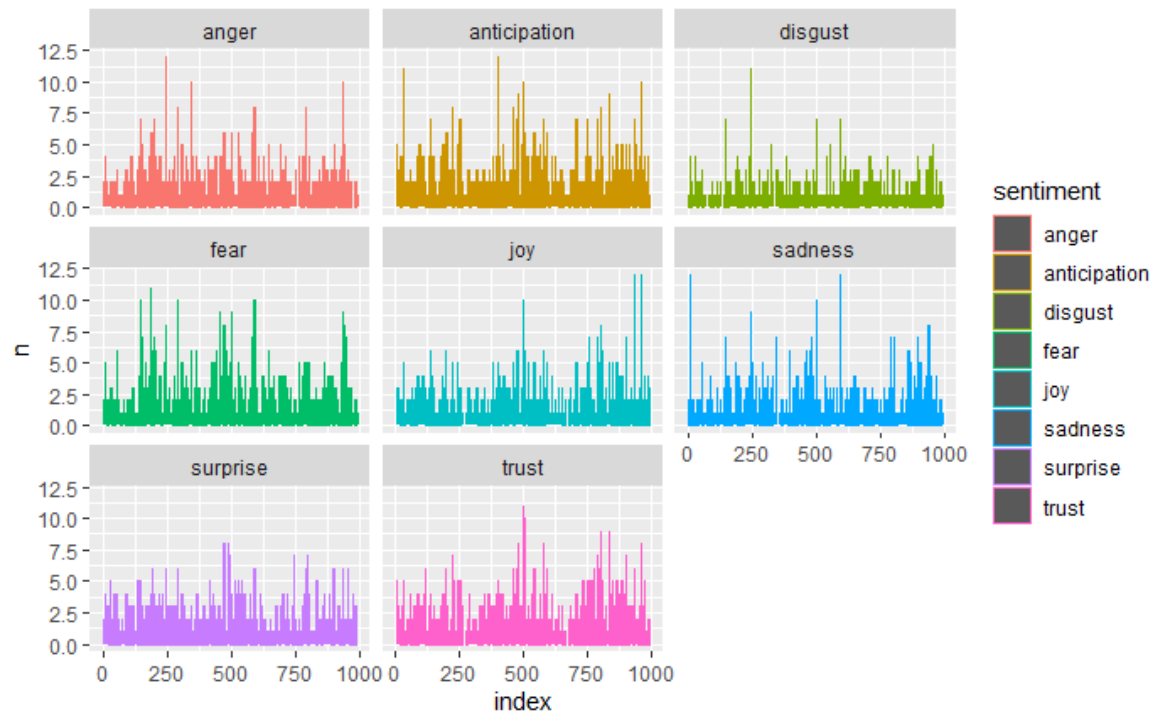
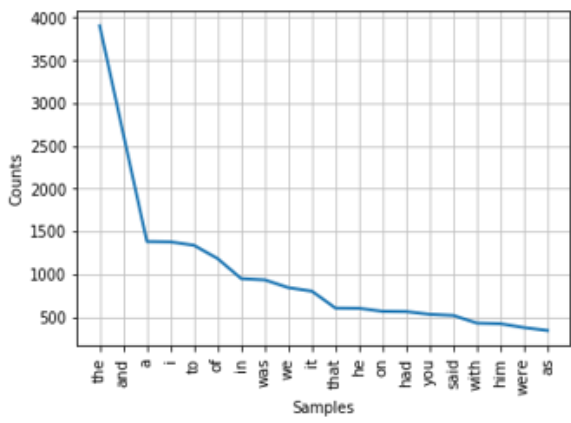
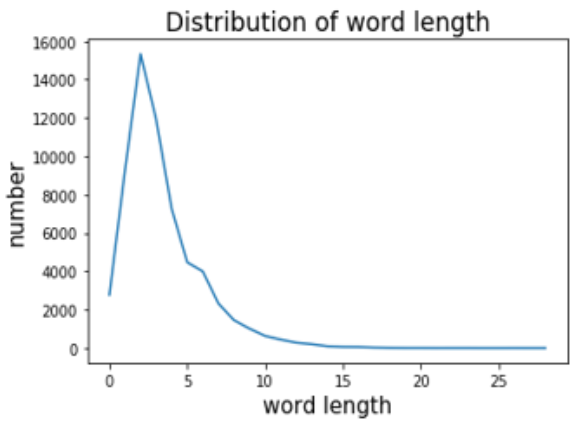
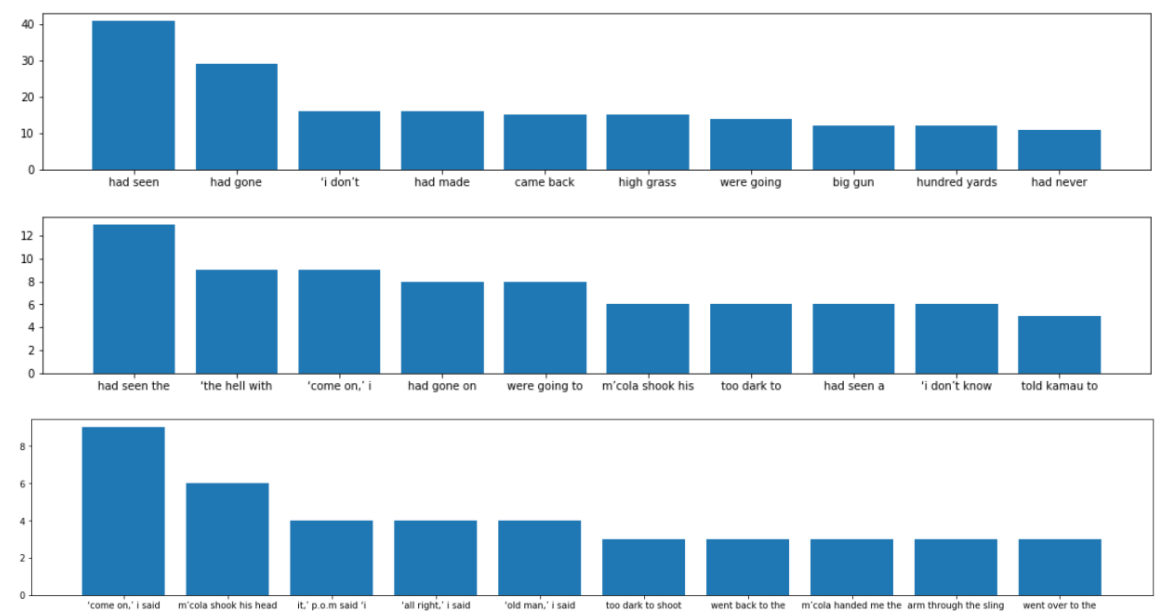
Men Without Women (1927)



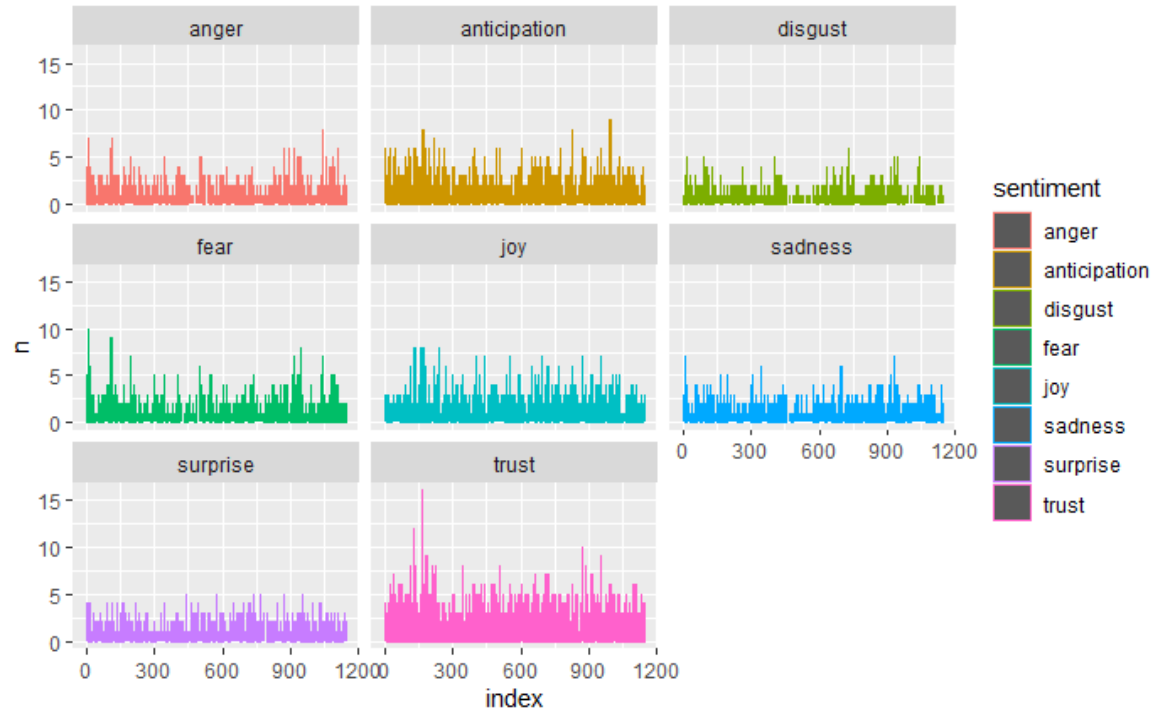
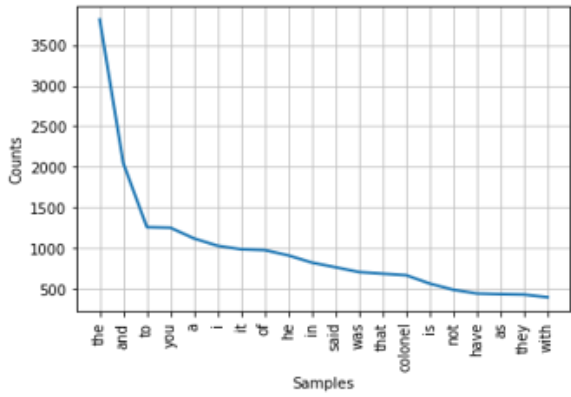
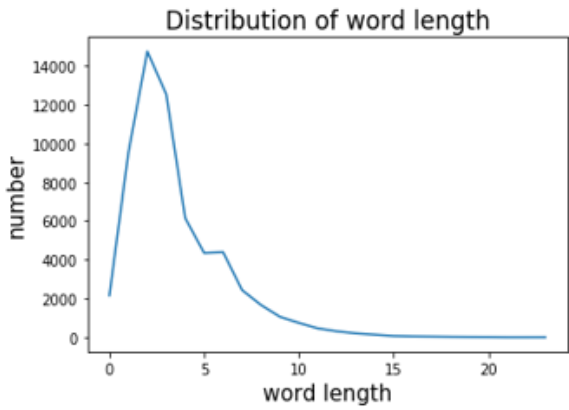
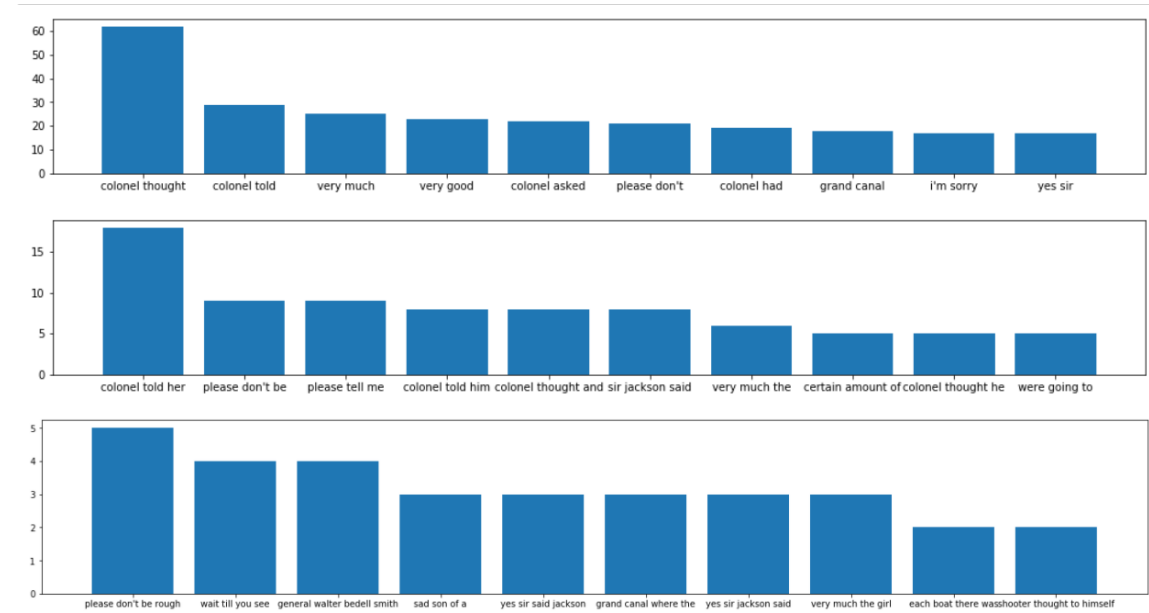
Winner Take Nothing (1933)



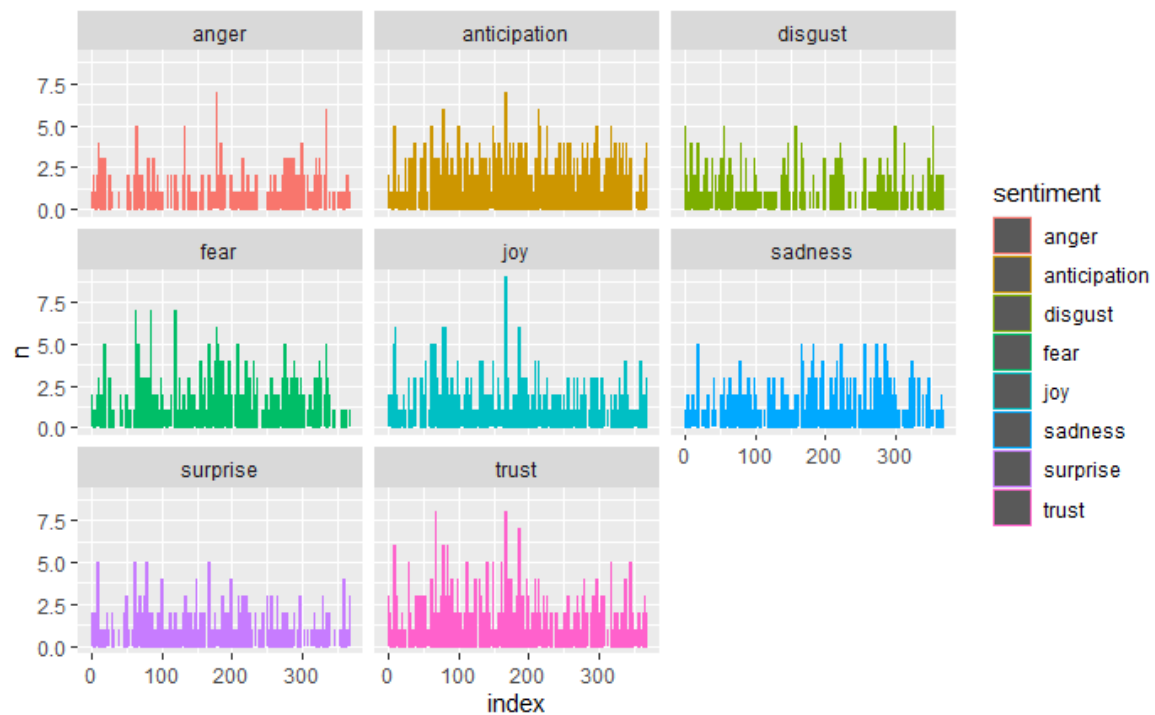
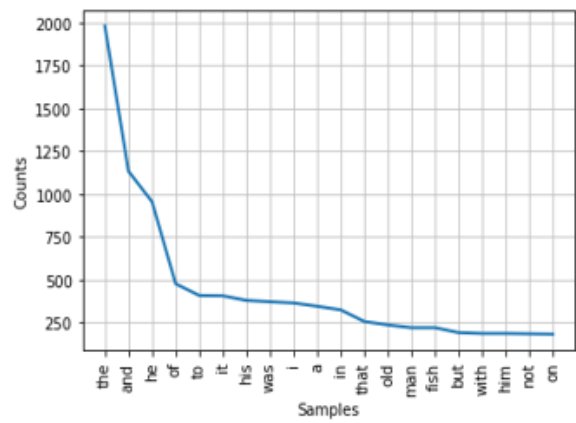
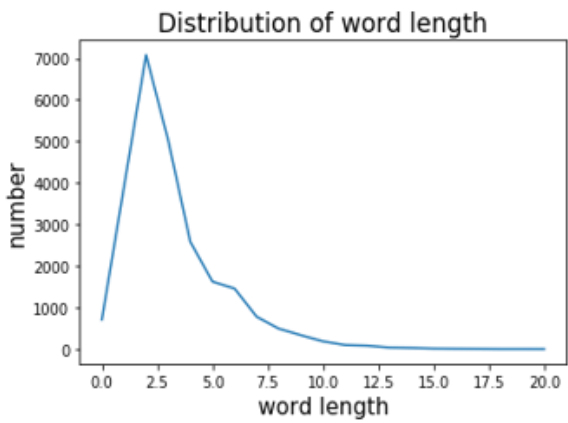
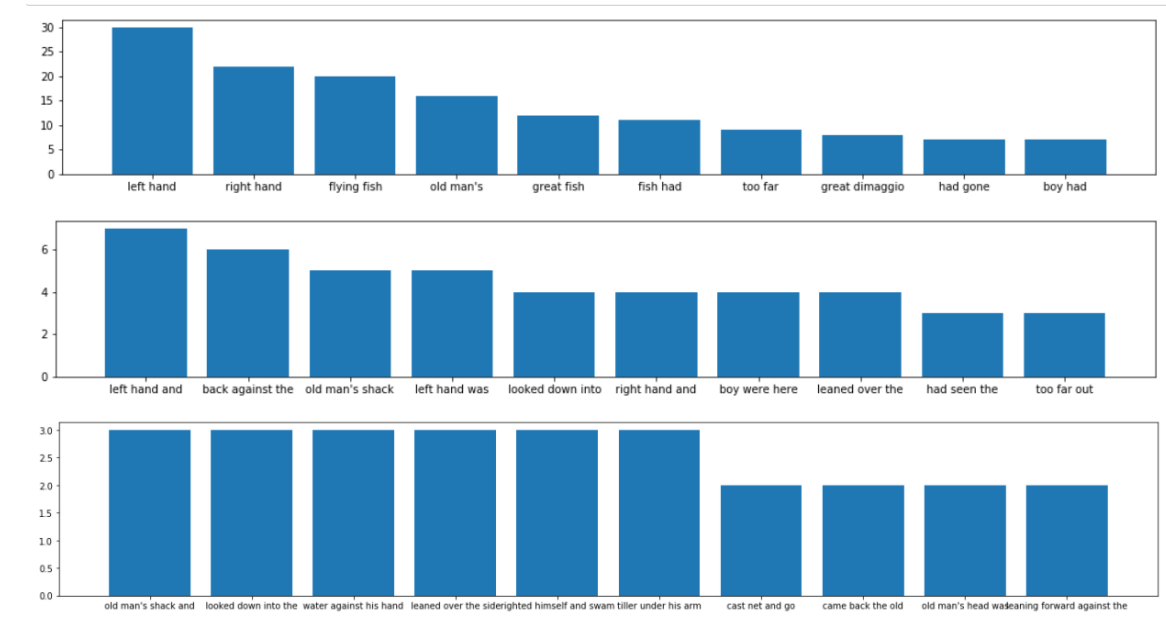
Green Hills of Africa (1935)



Across the River and Into the Trees (1950)



The Old Man and the Sea (1952)



Introduction to four graphs

I apply n-grams method, word length distribution, word frequency, and emotional analysis for each of the six books. Each page before represents the figures for certain book (for example, the first page is for the first book, *The Sun Also Rises* (1926)). The first three graphs are Python-generated, and the last colorful one is R-generated.

I build the n-grams model based on deleting common words for unigram, but for bigrams, trigrams and four-grams, common words are included. Below that, the distribution of word length illustrates an aspect of writing style of certain book, and the graph next to it is the frequency counts of certain book.

Emotional analysis is a more specific one than sentiment analysis since it tells the fluctuations of different emotions as the story of book develops, rather than simple positive or negative figures. I mainly use an emotional documentary called 'nrc' in 'tidytext', and delete many stop-words to precisely analyze.

Overview and details

In a whole view, we find the distribution of word length and counts among the six books above are quite similar, which indicates the fundamental use style of vocabulary of Hemingway is stable. All the peaks of word length is approximately 2.5, which means Hemingway does not prefer to use gorgeous words in his writing. As for the word counts, *The Sun Also Rises* (1926) and *The Old Man and the Sea* (1952) are different with other four books, since the 'elbows' of these two figures are at the fourth word rather than the third one.

The n-grams and emotions figures are more important, so I will discuss these two figures for six books as follows in detail.

The Sun Also Rises (1926):

The bigrams bar chart illustrates that 'robert cohn', the character, is most frequent. 'san sebastian' is the main place in this book, these two grams give us basic information of 'who and where' of this book. According to trigrams and four-grams, we find relationship between robert and brett maybe the main plot in *The Sun Also Rises*.

As for emotional analysis, we find the anticipation, joy, surprise and trust reach the peak at same time and then drop, which implies the dramatic gain and loss at that main plot. We also find negative emotions like disgust, fear, sadness reach the peak at the beginning of this book.

Men Without Women (1927):

The bigrams bar chart illustrates that 'jack', 'william campbell', 'ole anderson', the characters, are most frequent. I could guess this book tells more about males due to these names.

As for emotional analysis, we find most figures really fluctuate several times. The peaks of fear and trust are attractive, and characters may experience several difficulties and lose trust and finally overcome these and build great relationship, since anticipation, surprise and trust are relatively higher than other negative emotions.

Winner Take Nothing (1933):

From bigrams chart, we can find the story really mentions more names, and it may include many plots related to these characters. According to trigrams and four-grams, we find interesting gram 'crazy pour', which implies the vocabulary may contains more dramatic emotions.

As for emotional analysis, we find many extremely emotional ups and downs. Intricate emotions are intertwined in the book, and twisted plots may be captivating and thought-provoking.

Green Hills of Africa (1935):

Compared with previous three books, the bigrams figure of this one does not show any name, and sometimes it implies the story is well-plotted without many actions from various characters. As for trigrams and four-grams, it illustrates more specific actions, and I also find the frequent chat between 'I' and 'POM', which represent a story Hemingway experienced himself.

It is the most emotional books among the all. Hemingway tells a true story of him and this contains more twisted plots. We also find its sad beginning and relatively happy ending.

Across the River and Into the Trees (1950):

It describe the story between colonel and an ordinary girl due to 'sir', 'girl'.

This books is not well-plotted because of less emotional fluctuations.

The Old Man and the Sea (1952), the most famous and popular book of Hemingway:

The whole n-grams figures shows the main plot happened on an old man, paying great efforts to catch a flying fish, and I find most of the n-grams in the graph are more representative and useful compared with other five books.

We find some extreme emotional fluctuations, since the key special feature of these eight subplots is the frequent 'close to zero' parts of negative emotions, which means the efforts old man pays are corresponding to his inner thoughts. It is clear that trends of anticipation, trust, and joy are similar, and changes of them are contrary to that of anger, fear and disgust.