CSCE 5290 - NLP - ICE1

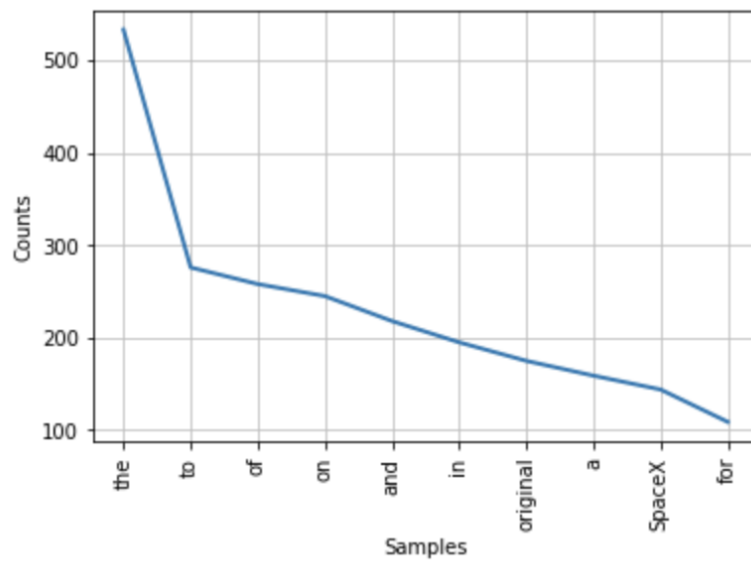**Name:** Rafael Alves Moreira

Group Members:
Kevin Raj
Robert Linhart
Alejandro Munoz
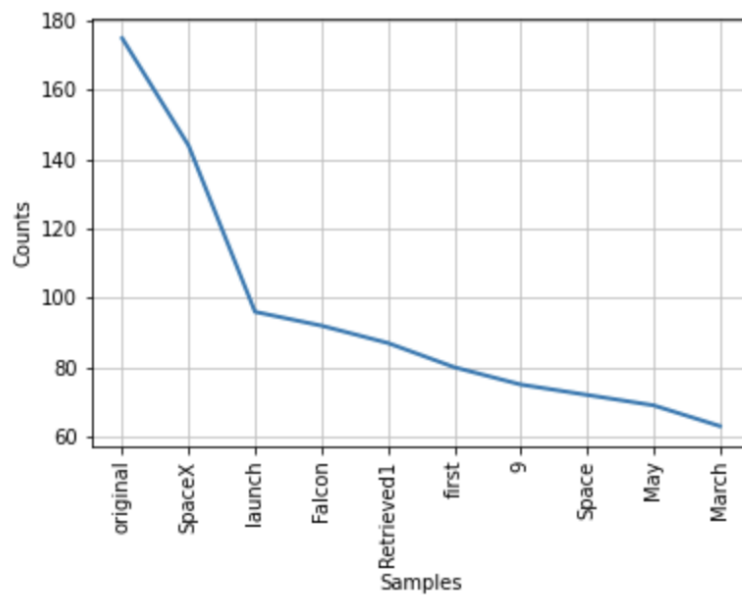
1. Why we use stopwords? Why stopwords are not necessary for NLP frequency distribution.
    a. We use stopwords to give proper meaning and context to a sentence.
    b. Stopwords are not necessary for NLP frequency distribution because it likely won't capture the main subject of the text, since stopwords appear more frequently.

2. Based on high frequency words what information you can extract from the graph?
    a. From the graph with stopwords, very little was able to indicate what SpaceX is involved with.
    b. From the graph without stopwords, I was able to gather that SpaceX was going to launch, where it was going (Space), and possibly the months (May or March)

All Tokens

Clean Tokens

3. Can you provide different visualization for frequency distribution? If yes, please perform. If no, why?
   a. Yes, we can tabulate the sample distribution.

```
# With Stopwords
freq1.tabulate(10, cumulative=False)
```

```
     the      to      of      on     and      in original       a  SpaceX     for
     533     276     258     245     218     195      175     159     144     109
```

```
# Without Stopwords
freq2.tabulate(10, cumulative=False)
```

```
  original    SpaceX   launch   Falcon Retrieved1    first        9   Space     May   March
       175       144       96       92        87       80       75      72      69      63
```

Link to github: https://github.com/rmoreira/NLP
Link to Wiki: https://github.com/rmoreira/NLP/wiki
Link to assignment: https://github.com/rmoreira/NLP/tree/main/ICE-1