

Name: Jimmy Tan

Matrikelnummer: 00819296

## NLTK-Corpora

Exercises:

- 1) Install NLTK and test installation (done)
- 2) Install the german packages of NLTK and corpora (done)
- 3) Choose a large german (or other non-english) corpus (> 1 or 10 MB).  
Load it in NLTK and use it for the following exercises (done)
- 4) Calculate the lexical richness of the selected corpus.

Output:

```
lexical_richness()
[8] ✓ 0.1s
... exercise 4 (lexical richness)
0.12779596721385508
```

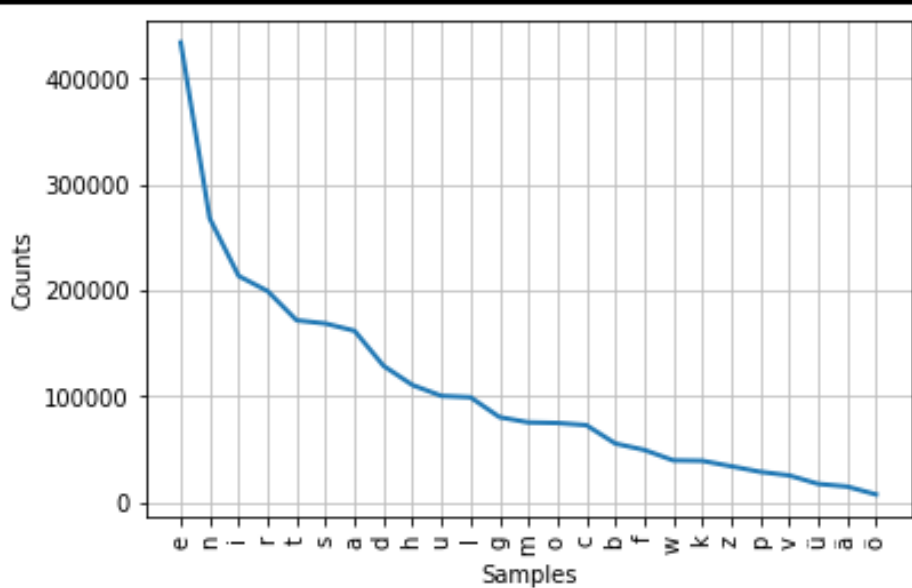
5) Plot the character distribution of your selected corpus.

Output:

```
plot_character_distribution()
```

✓ 1.4s

exercise 5 (character distribution)



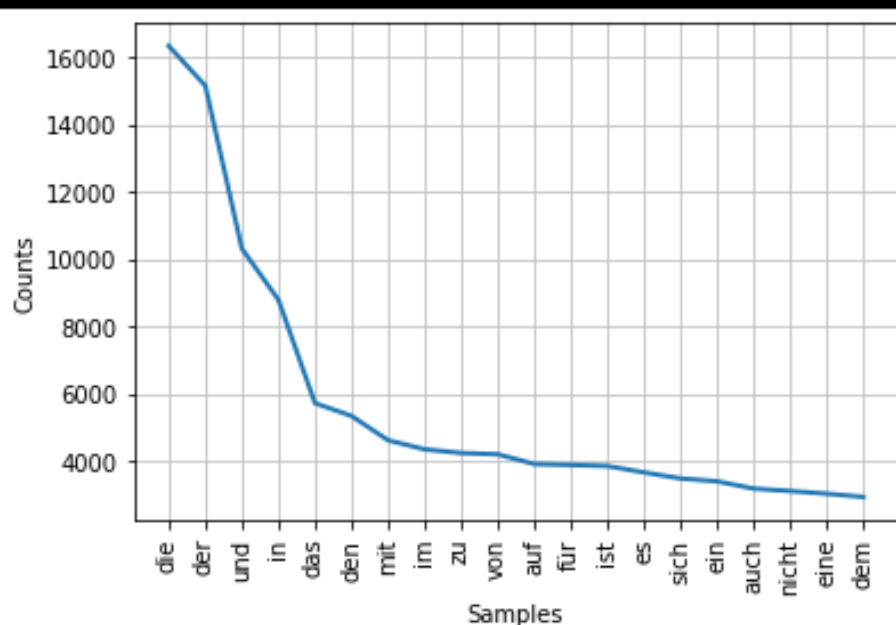
6) Plot the word distribution of the 20 most often used words of your selected corpus.

Output:

```
plot_word_distribution()
```

✓ 0.4s

exercise 6 (word distribution)



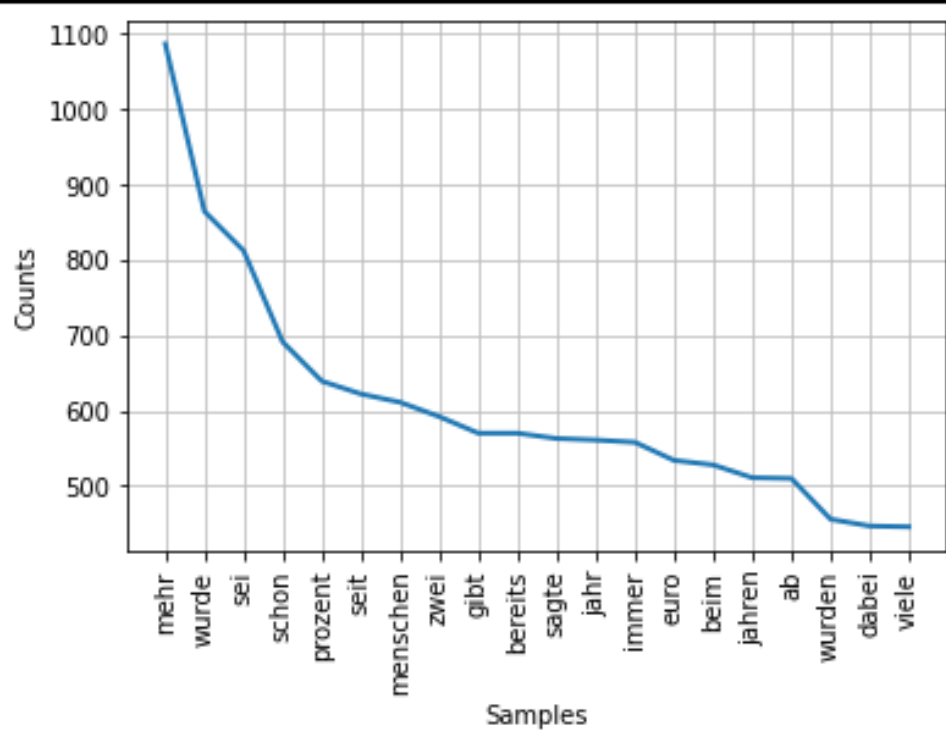
7) Remove stopwords and repeat exercise 6 (word distribution).

Output:

```
plot_word_distribution_wo_stopwords()
```

✓ 0.4s

exercise 7 (word distribution without stopwords)



8) Print a wordcloud without your stopwords.

Output:

