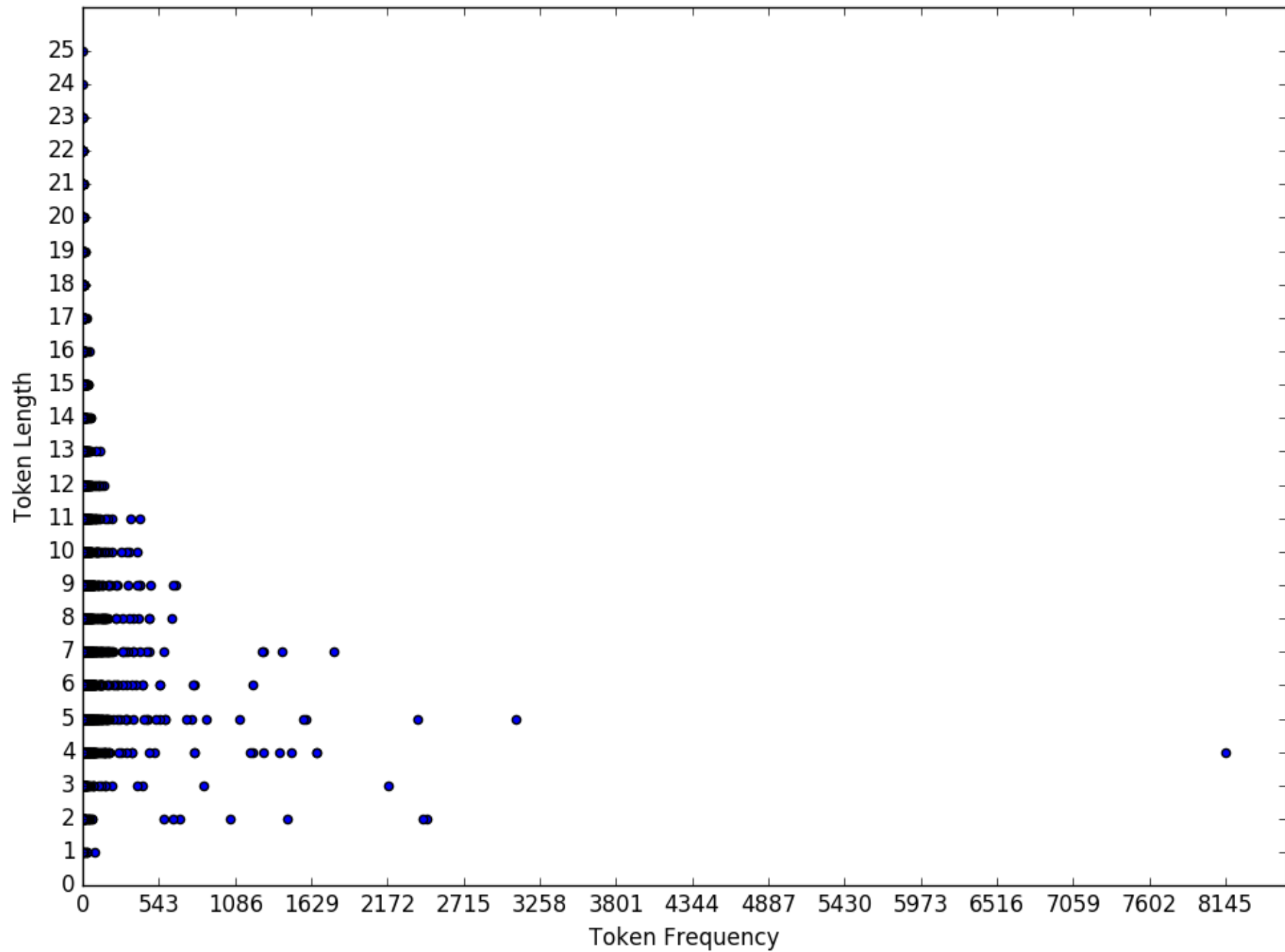
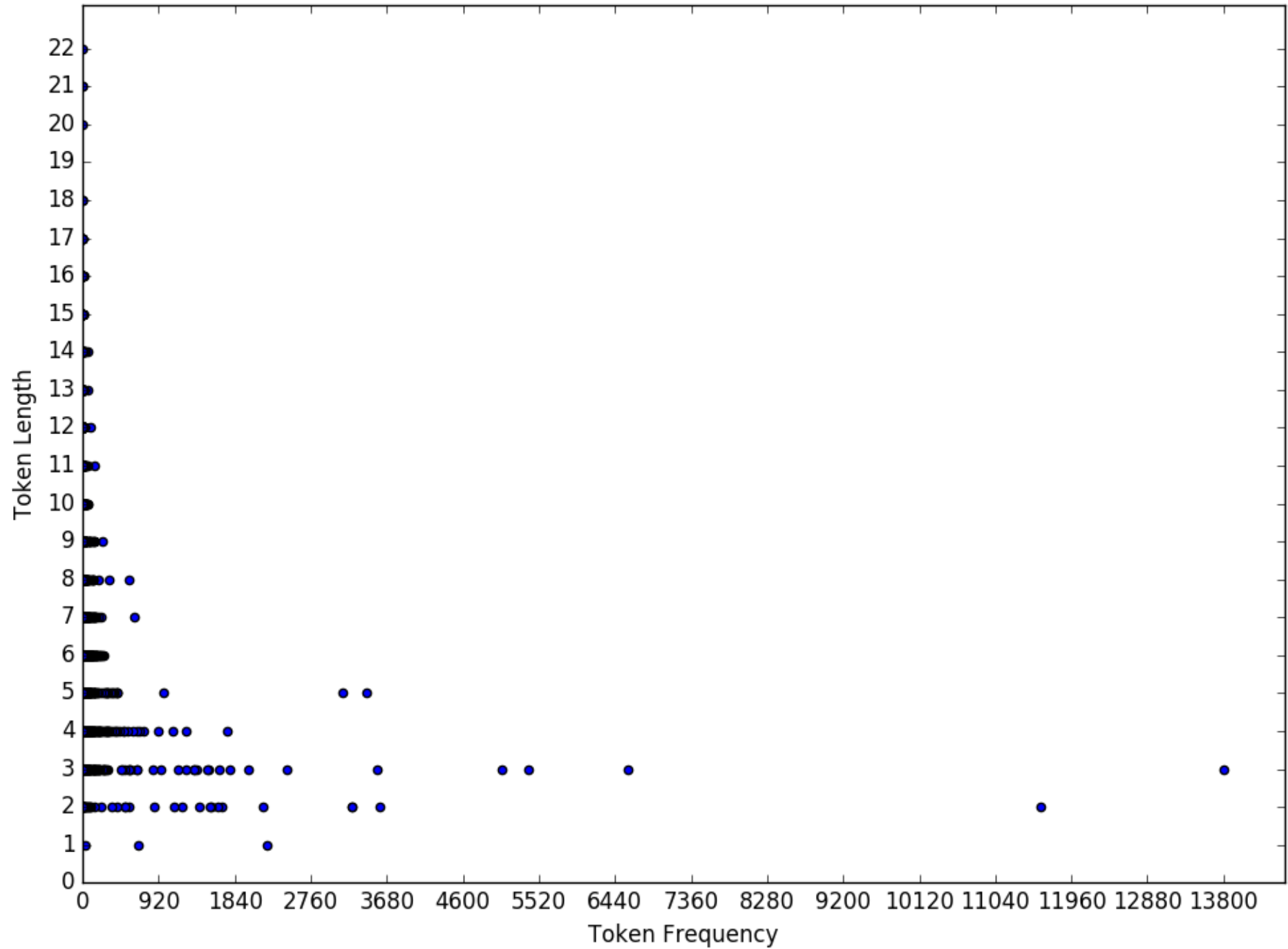


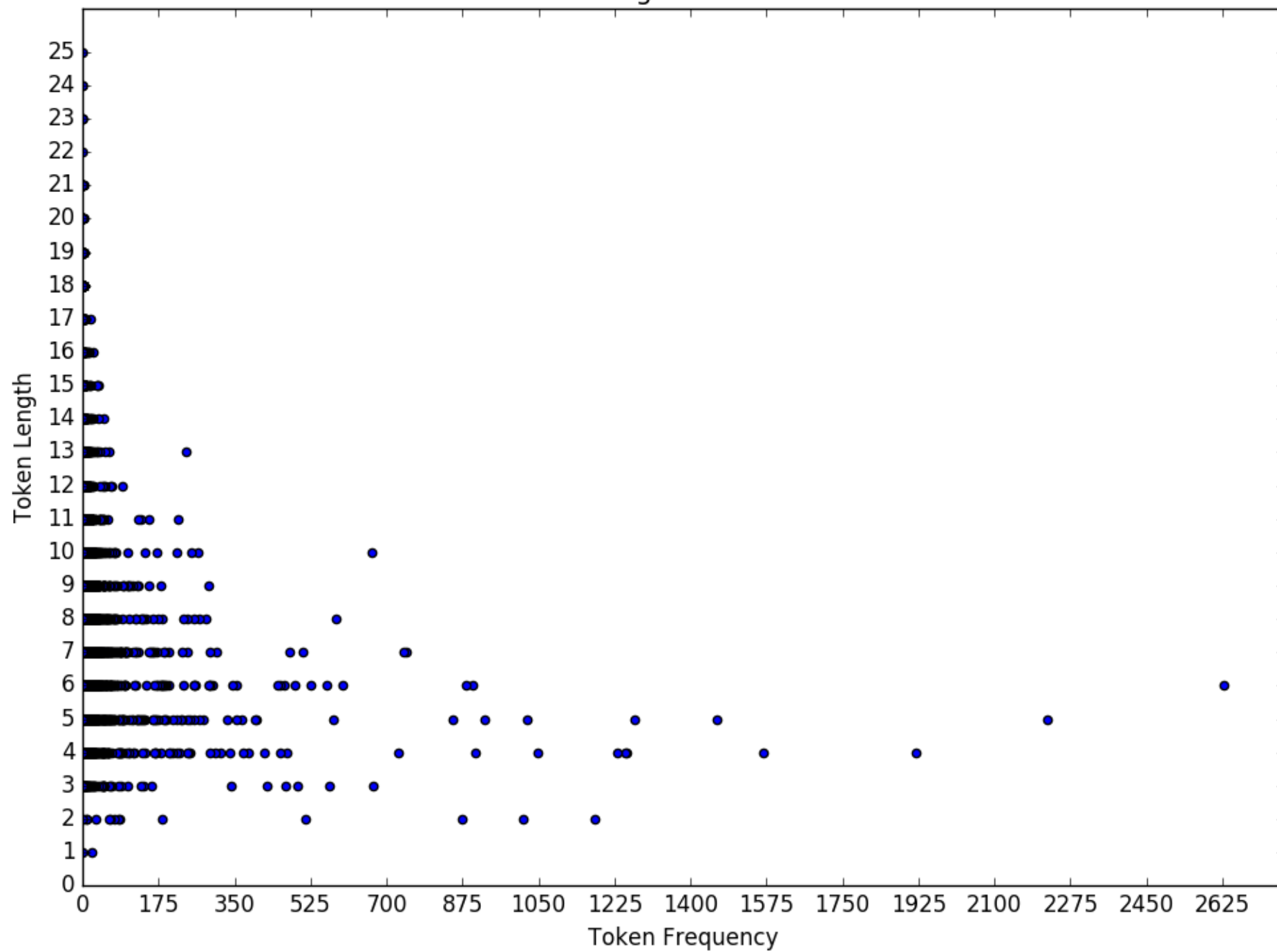
## Token Frequency



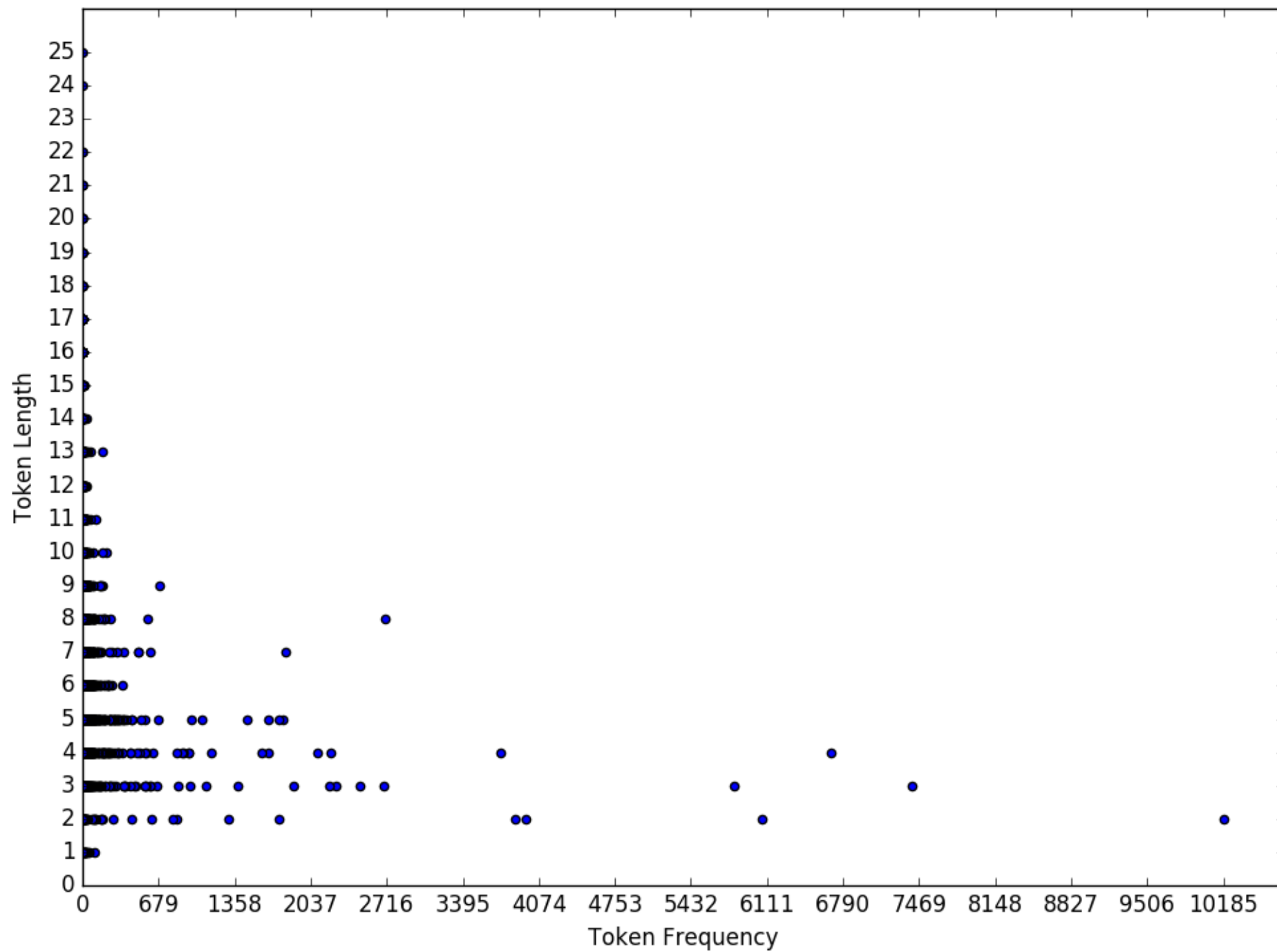
## Token Frequency



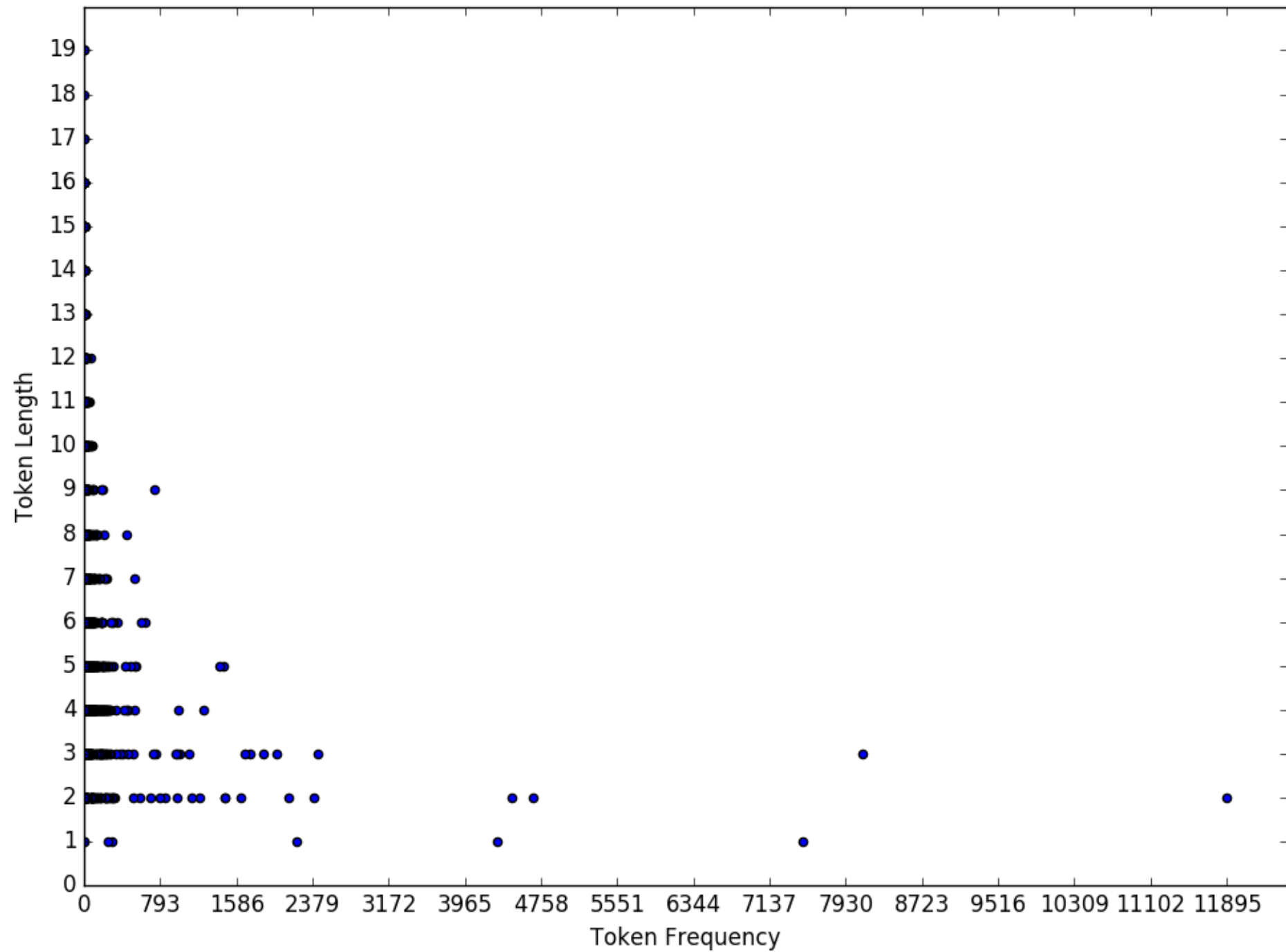
Aguaruna



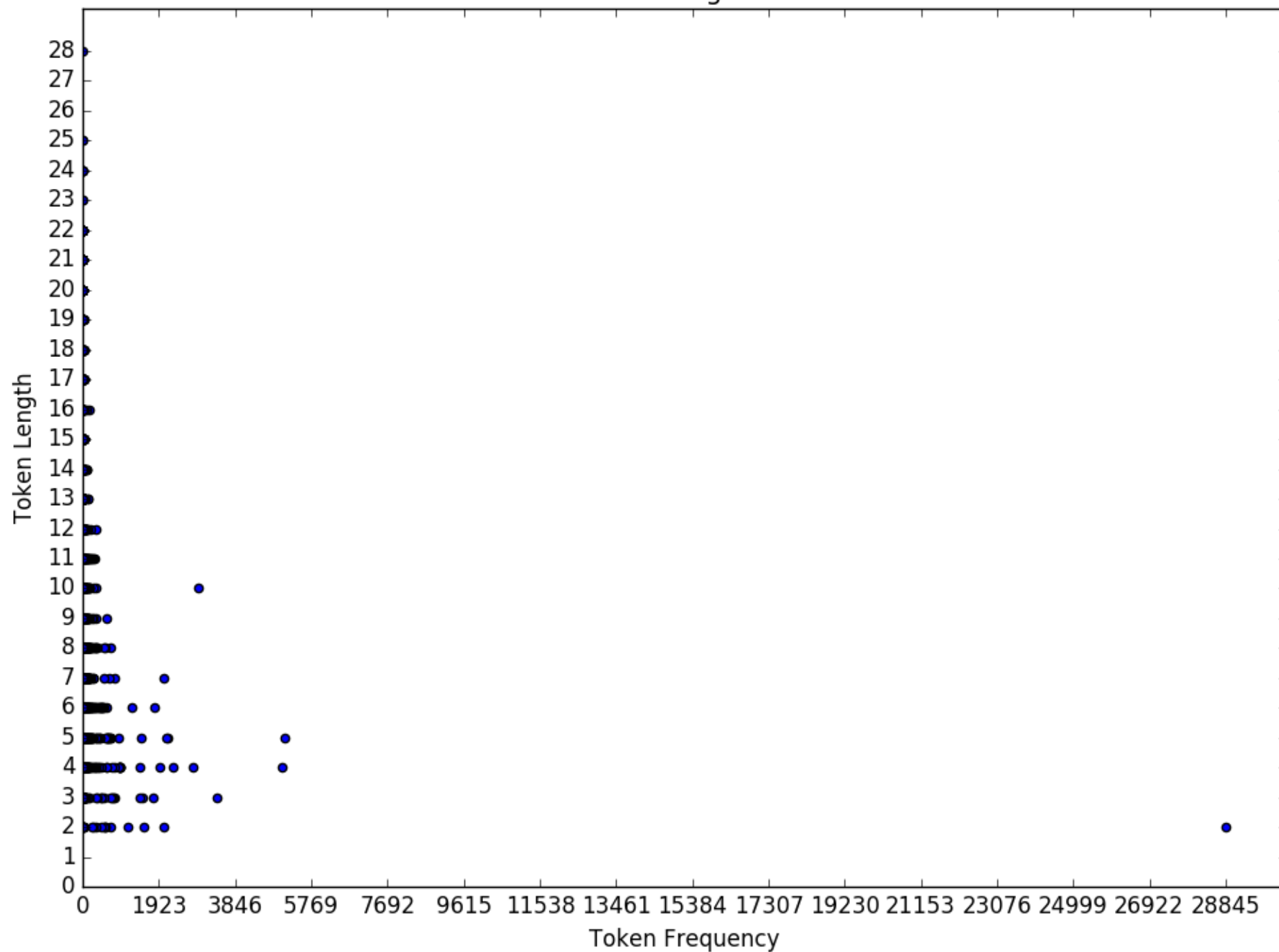
# Akawaio



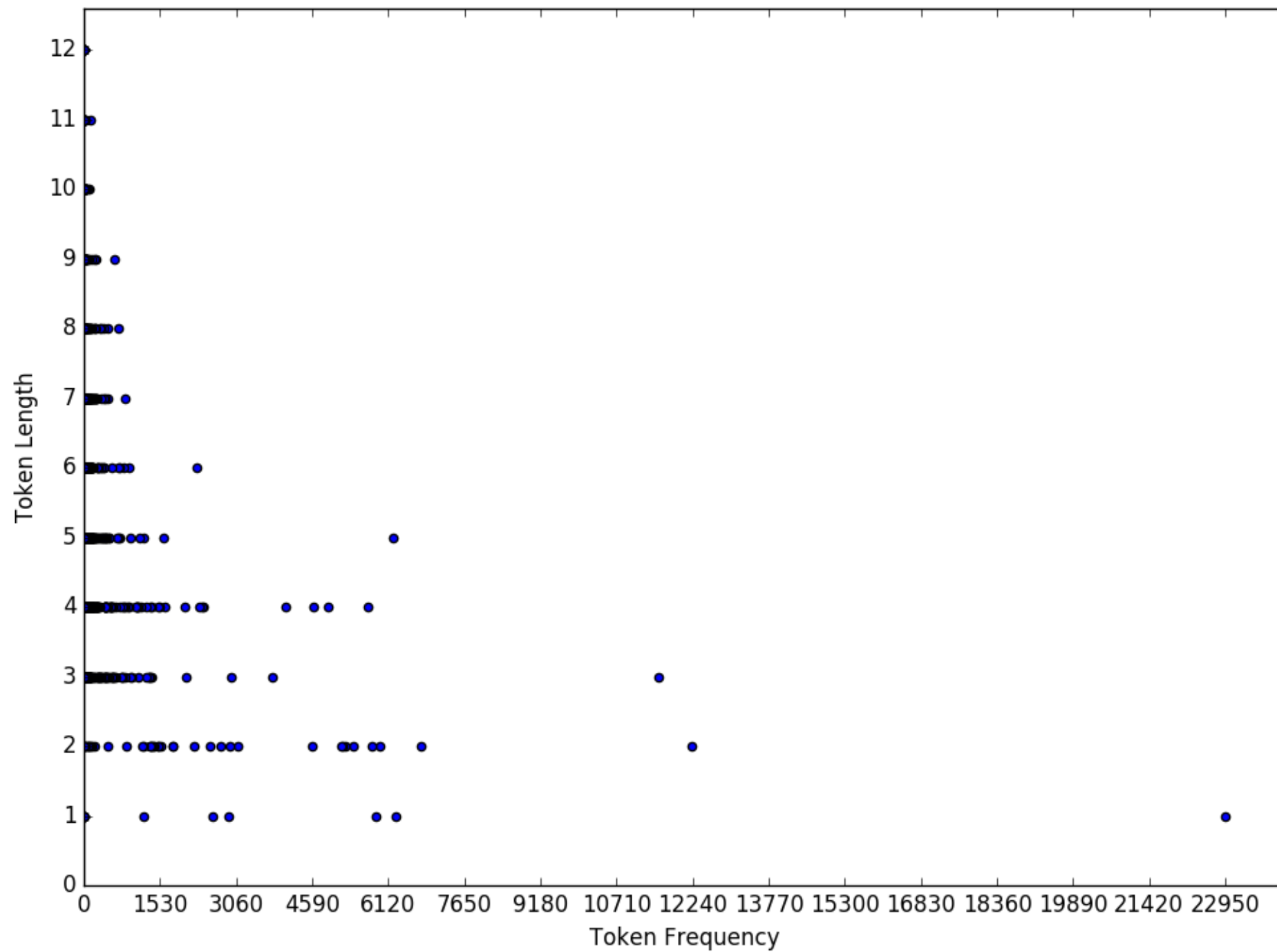
Albanian



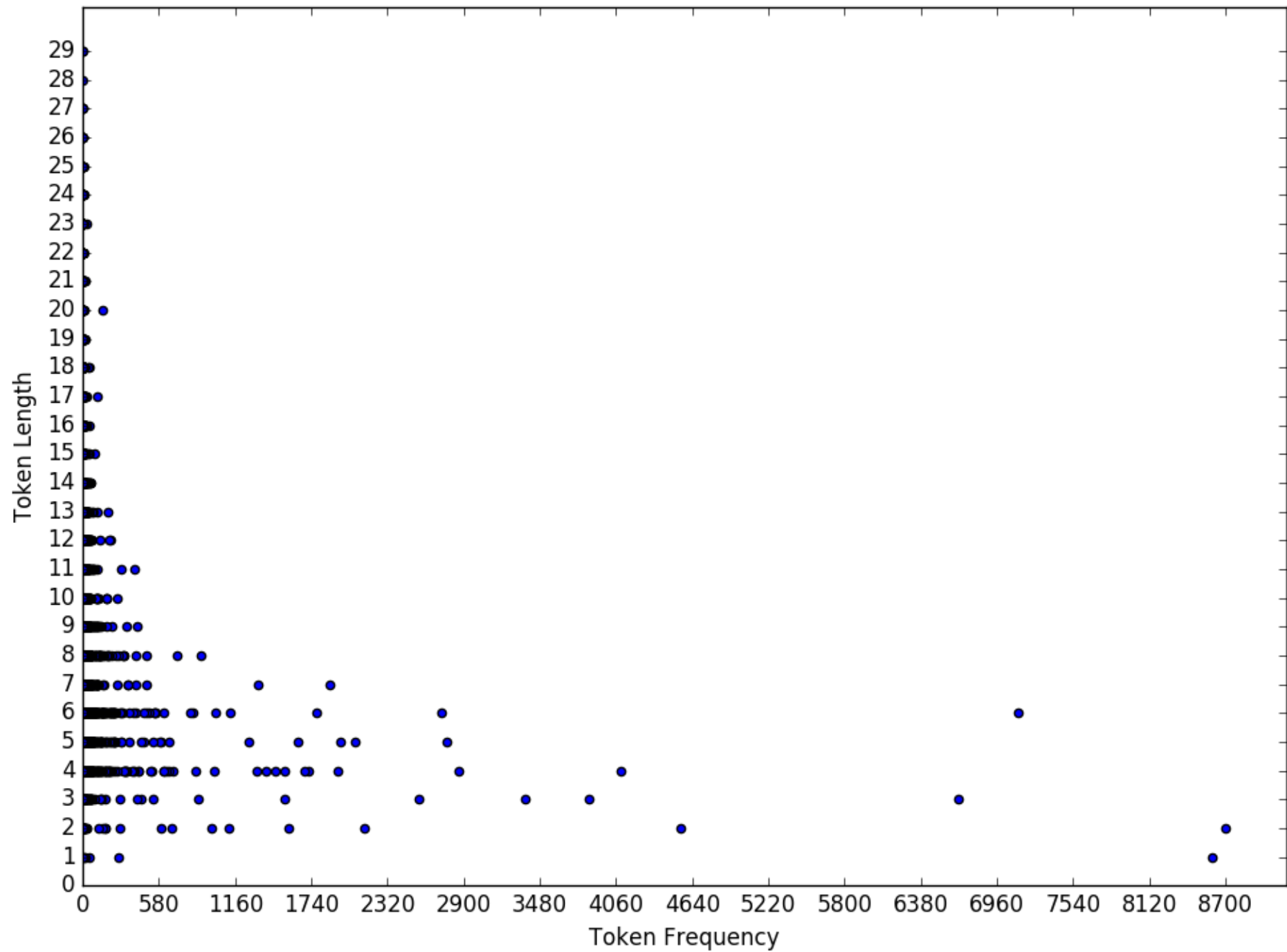
# Amuzgo



Aukan

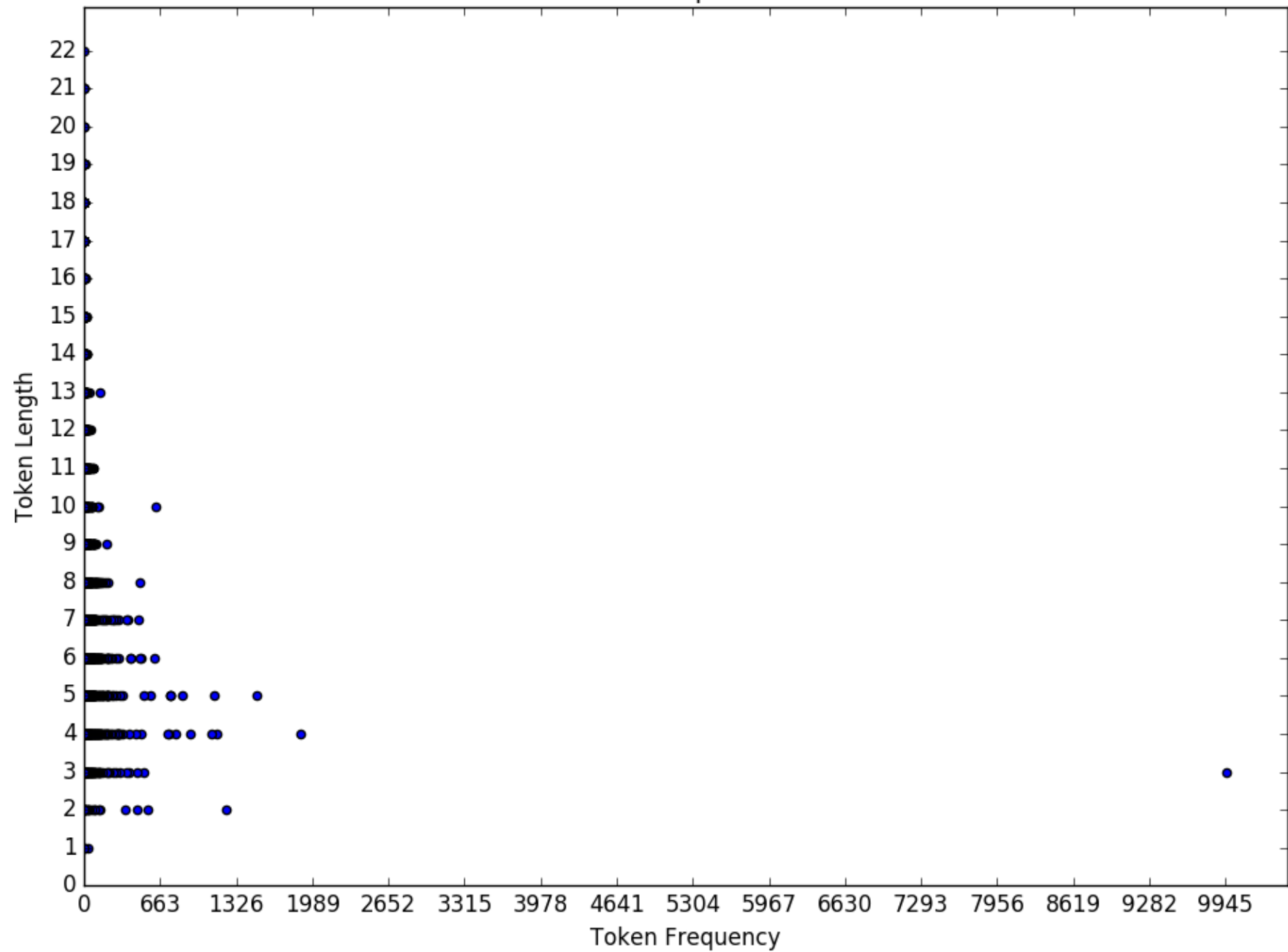


## Token Frequency





Basque

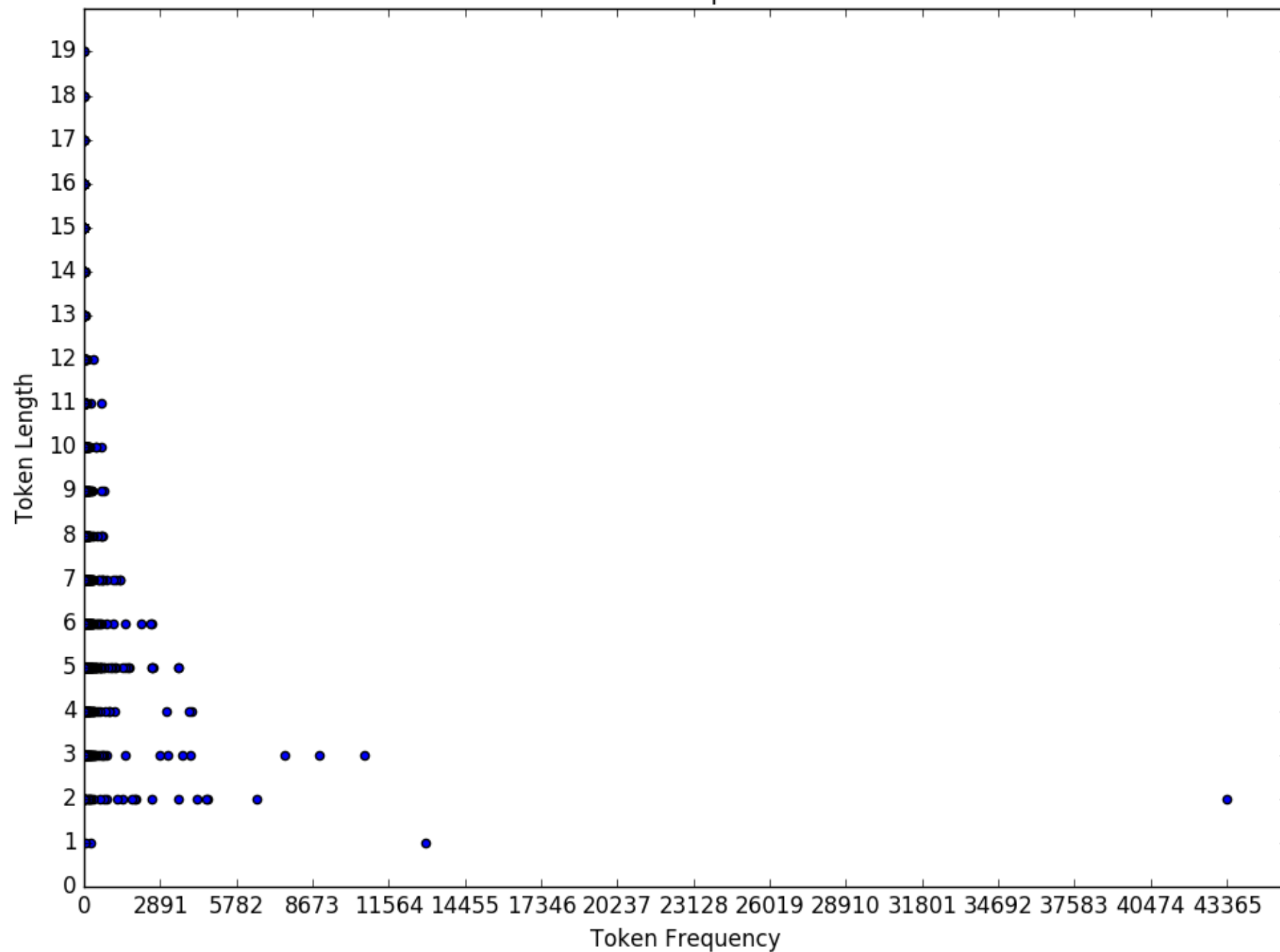


The scatter plot displays the relationship between the number of species ( $S$ ) and the number of samples ( $n$ ) for 1000 samples. The x-axis represents the number of samples ( $n$ ) and ranges from 0 to 1000. The y-axis represents the number of species ( $S$ ) and ranges from 0 to 1000. A horizontal line is drawn at  $S=1000$ . The data points show a decreasing trend in species richness as the number of samples increases, with a horizontal line at  $S=1000$ .

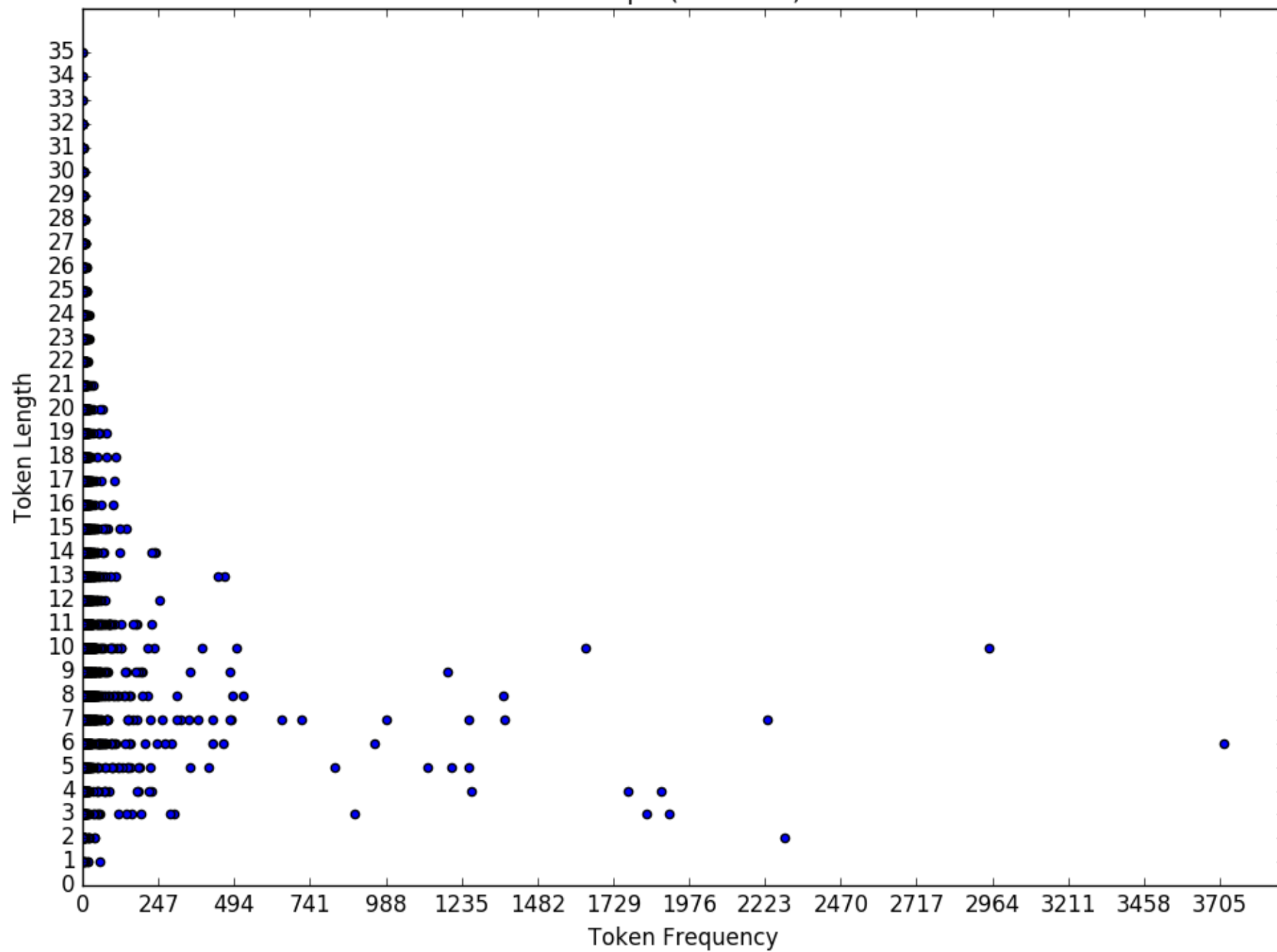
18  
17  
16  
15  
14  
13  
12  
11  
10  
9  
8  
7  
6  
5  
4  
3  
2  
1  
0

0	532	1064	1596	2128	2660	3192	3724	4256	4788	5320	5852	6384	6916	7448	7980
---	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------

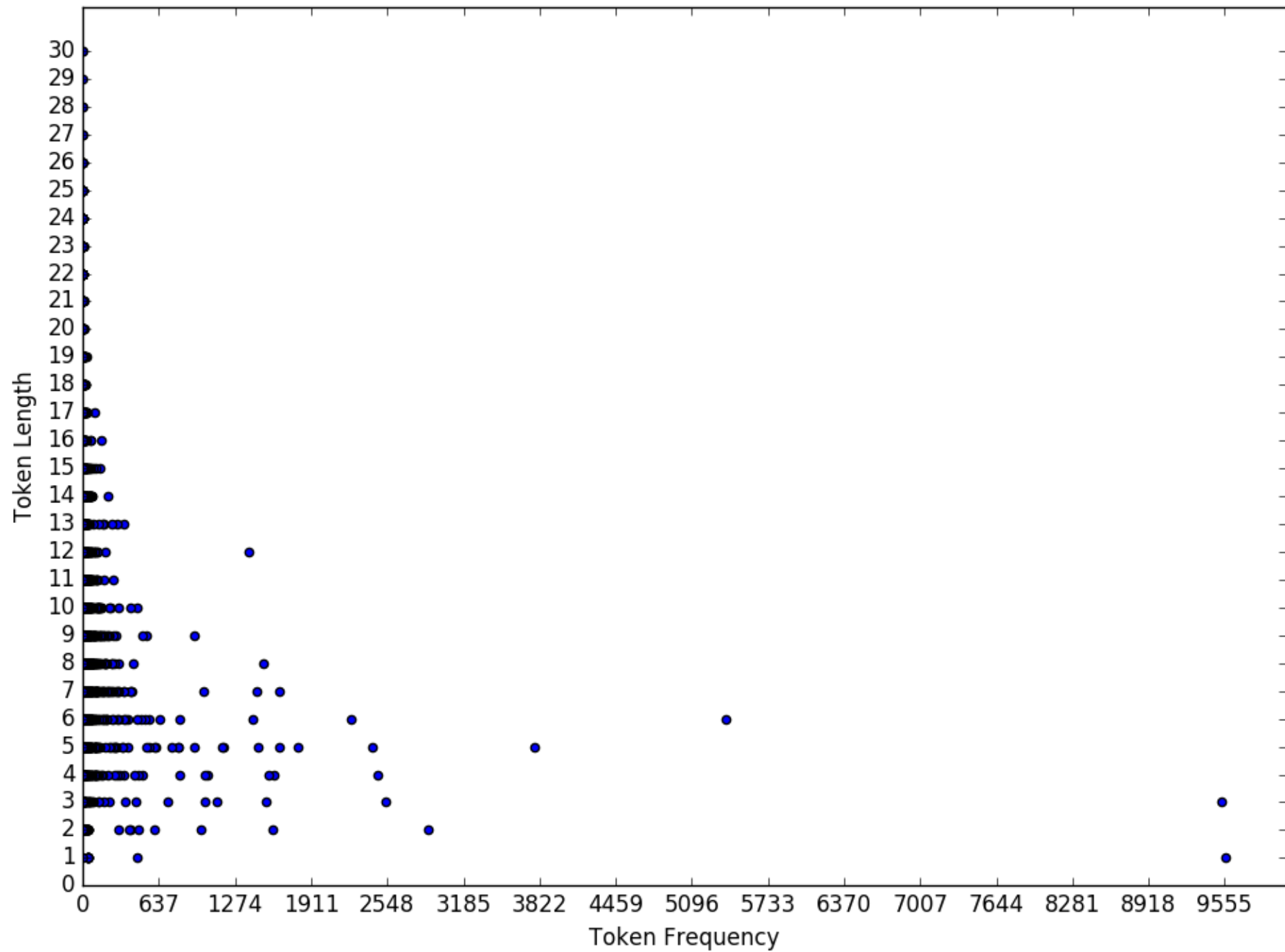
Cakchiquel



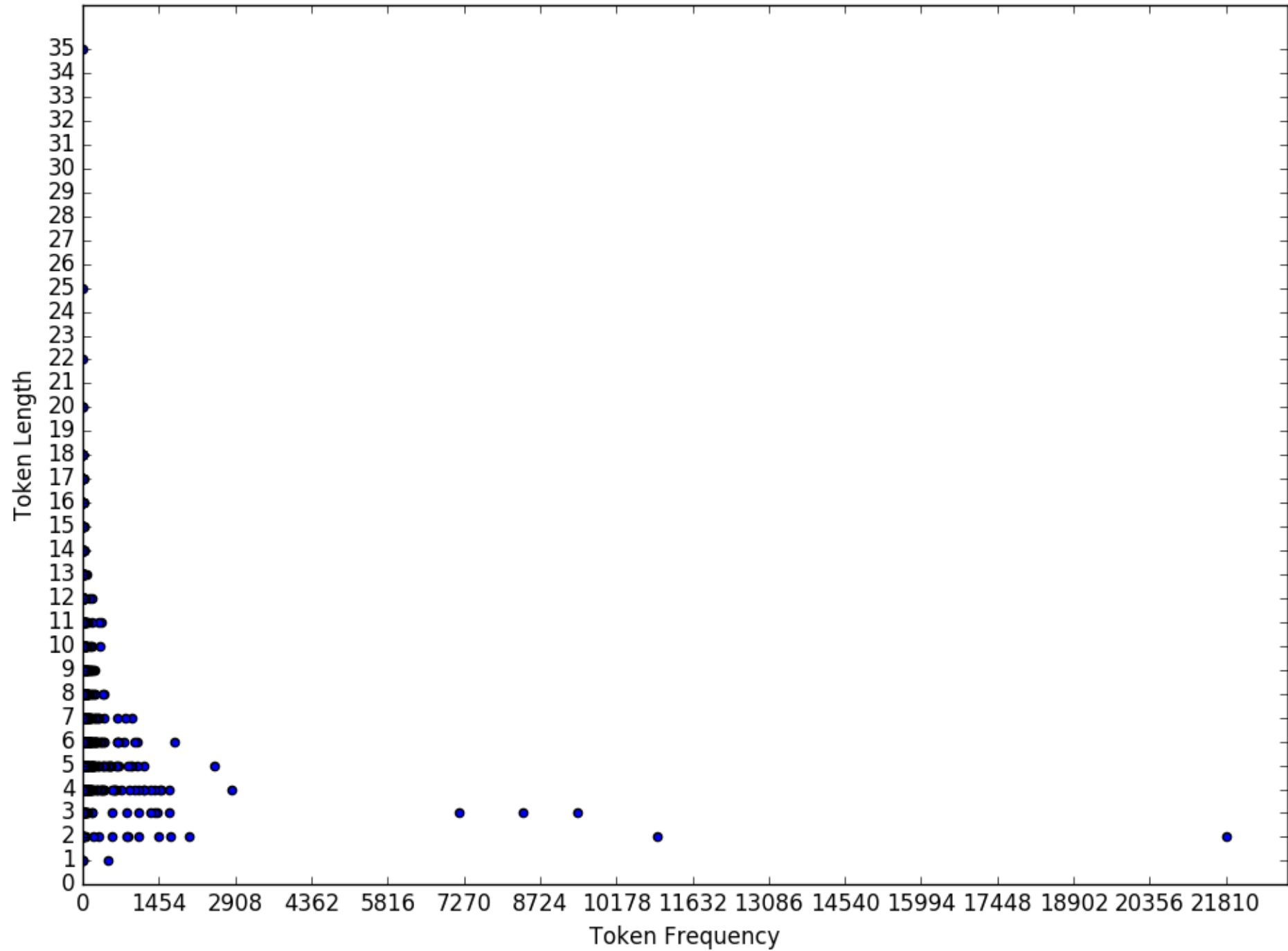
Campa (Axininca)



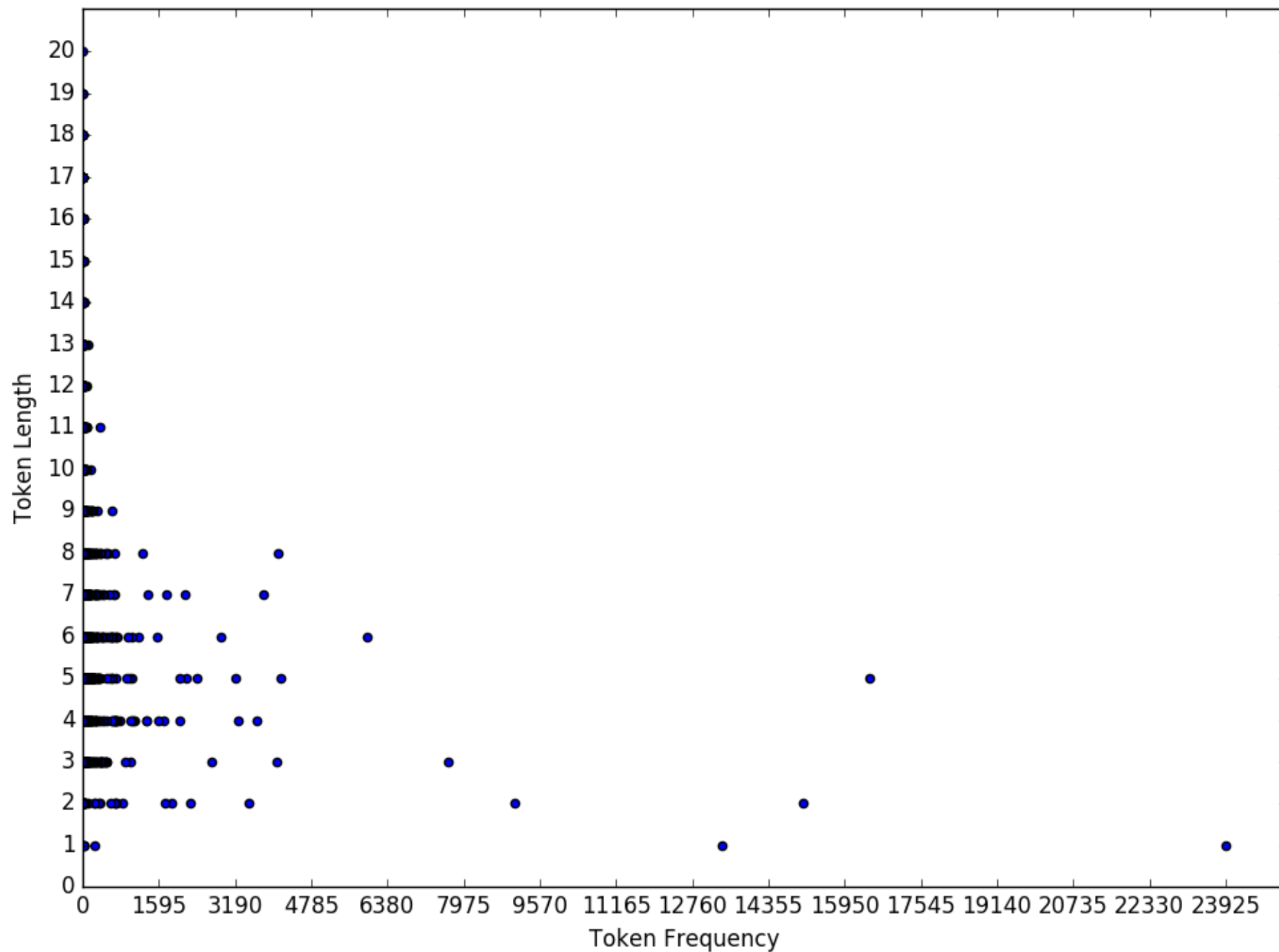
Camsa



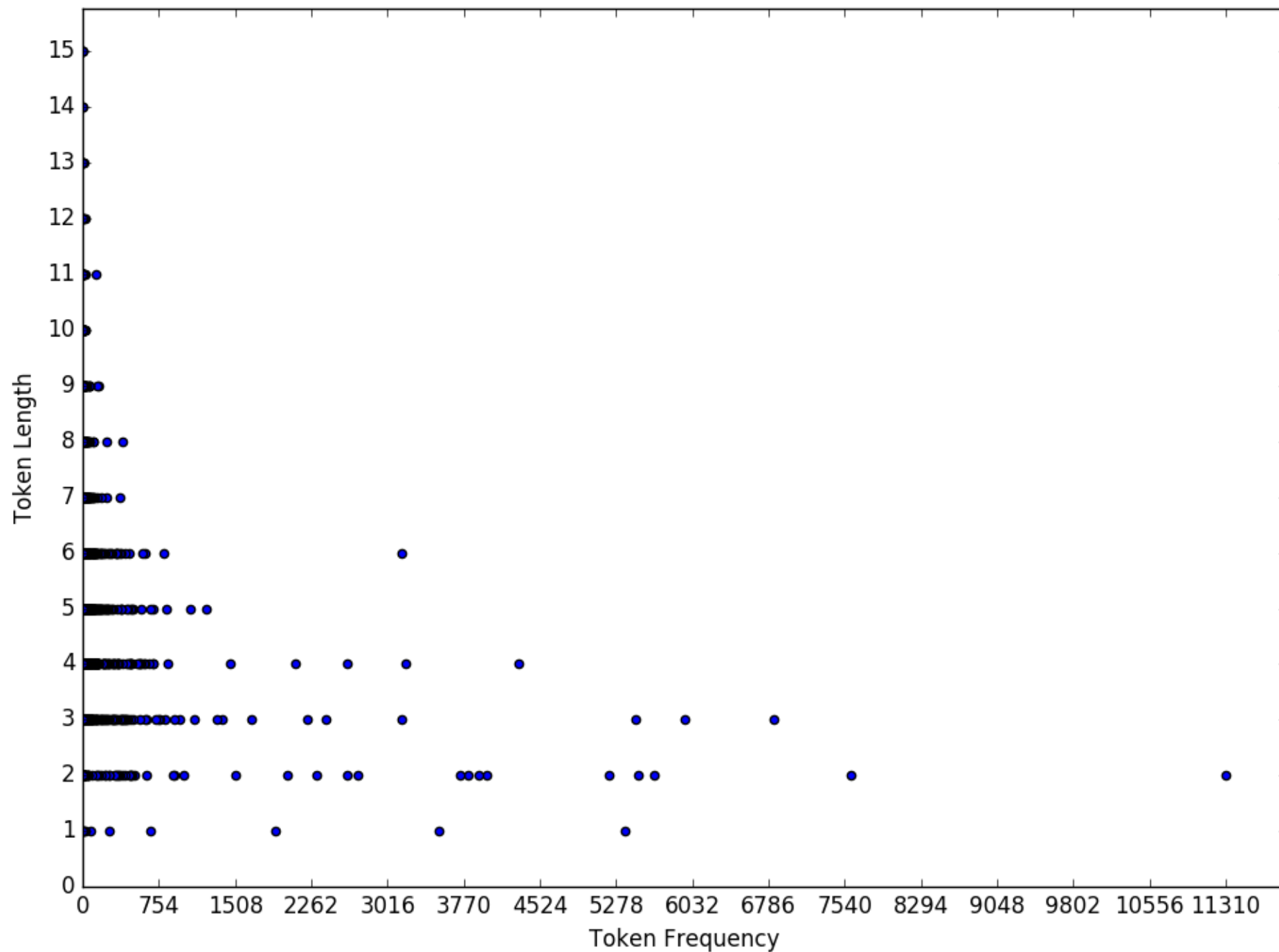
# Cebuano



# Chinantec

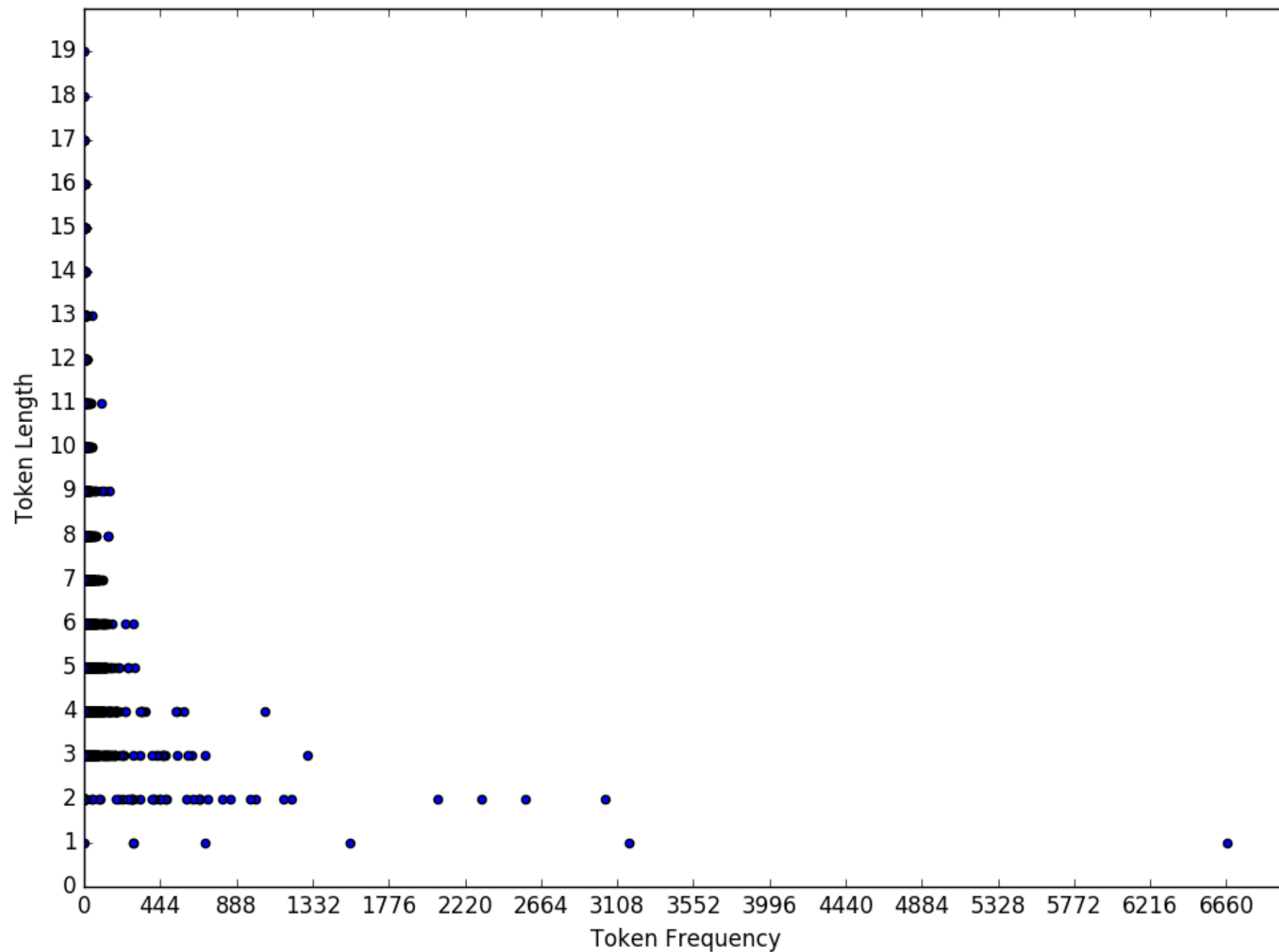


# Creole

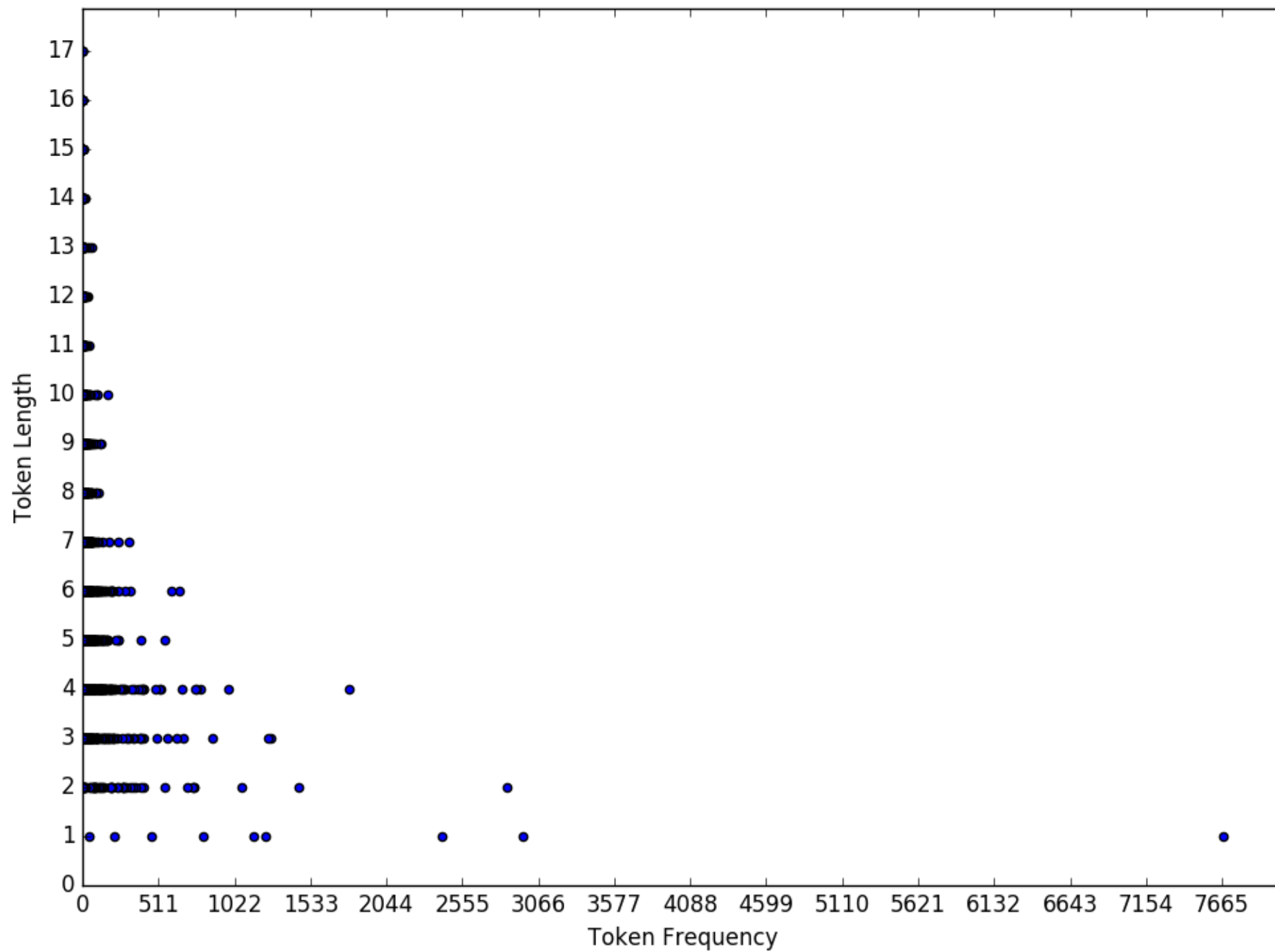




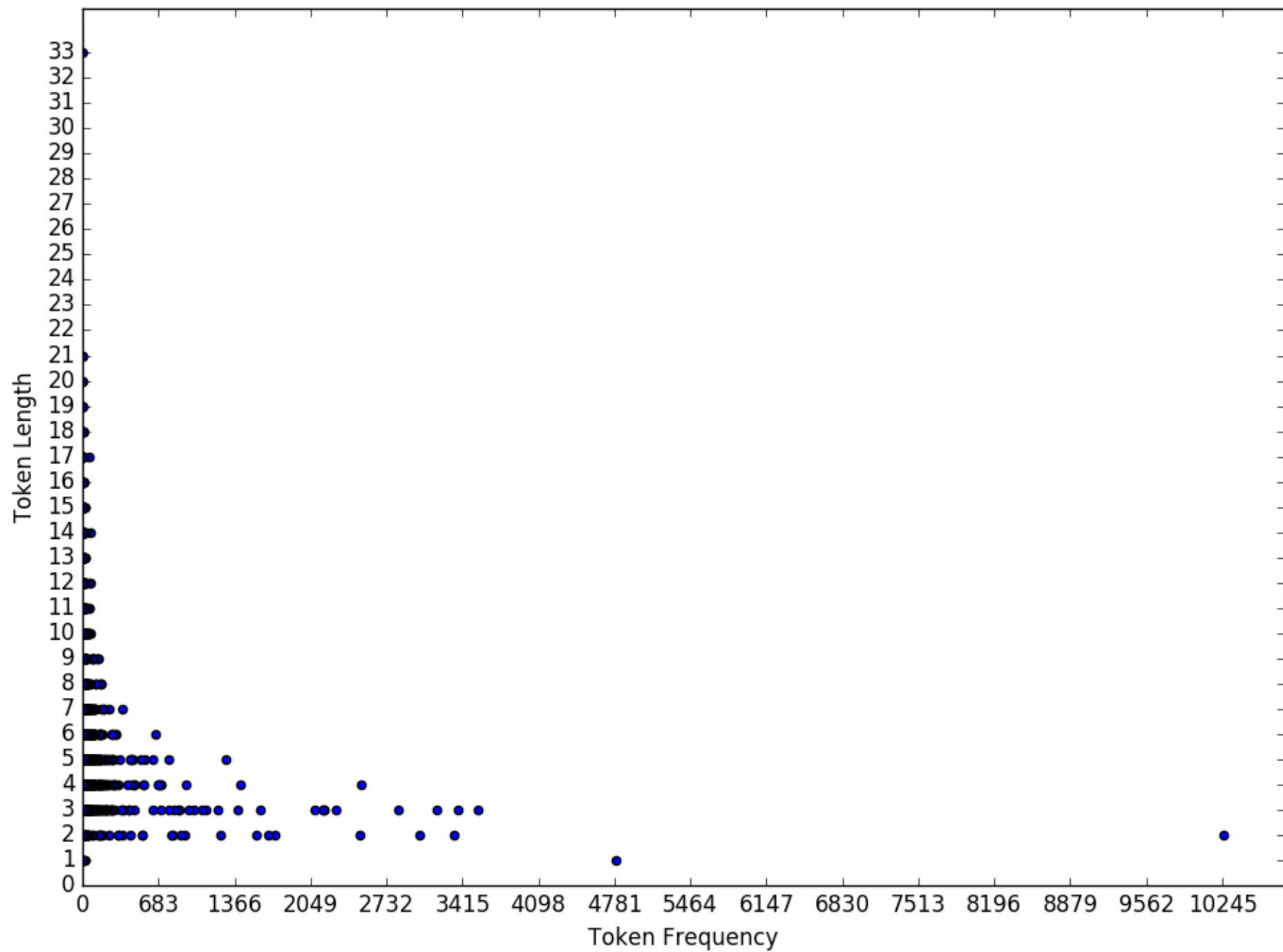
# Croatian



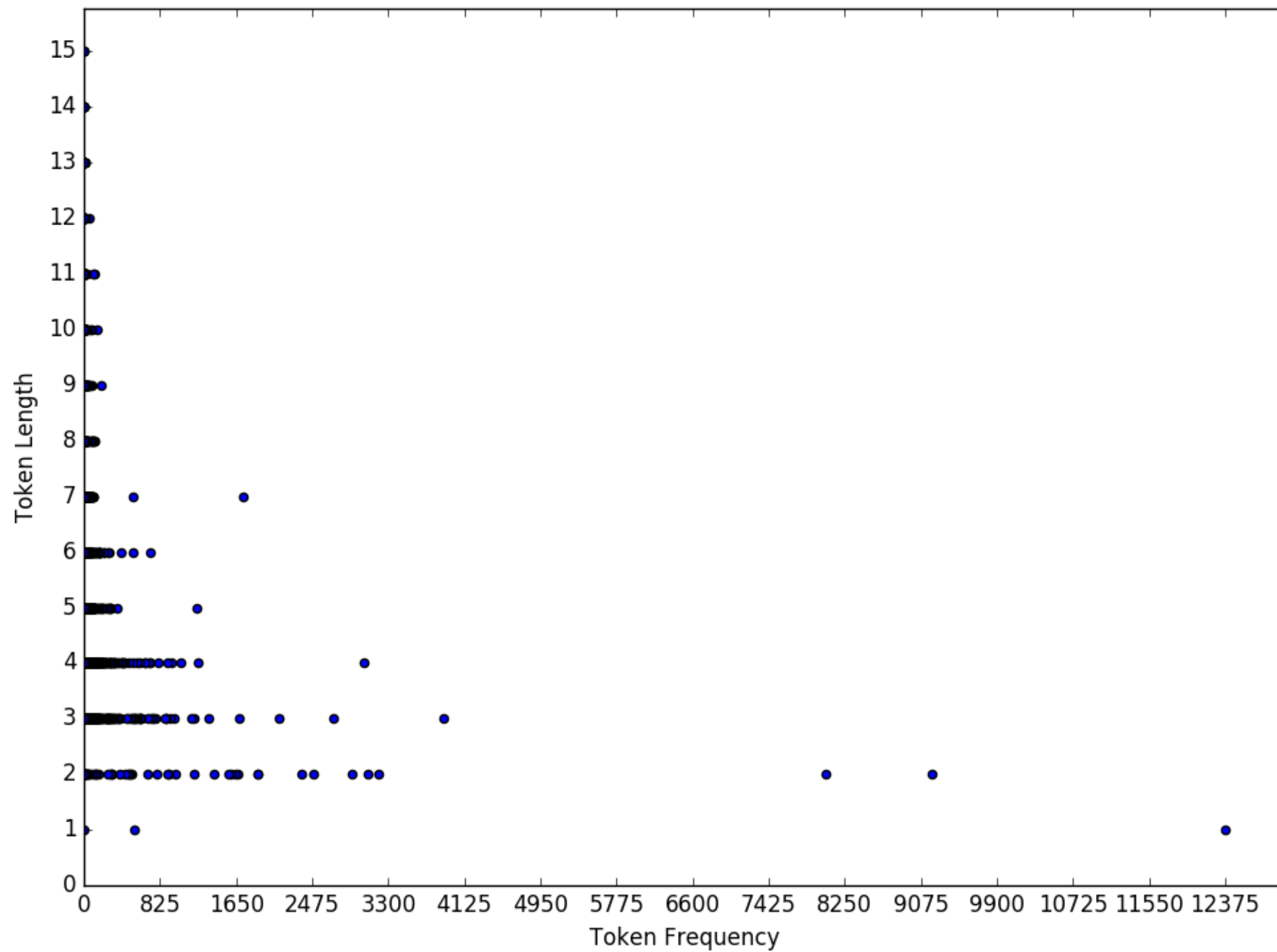
Czech



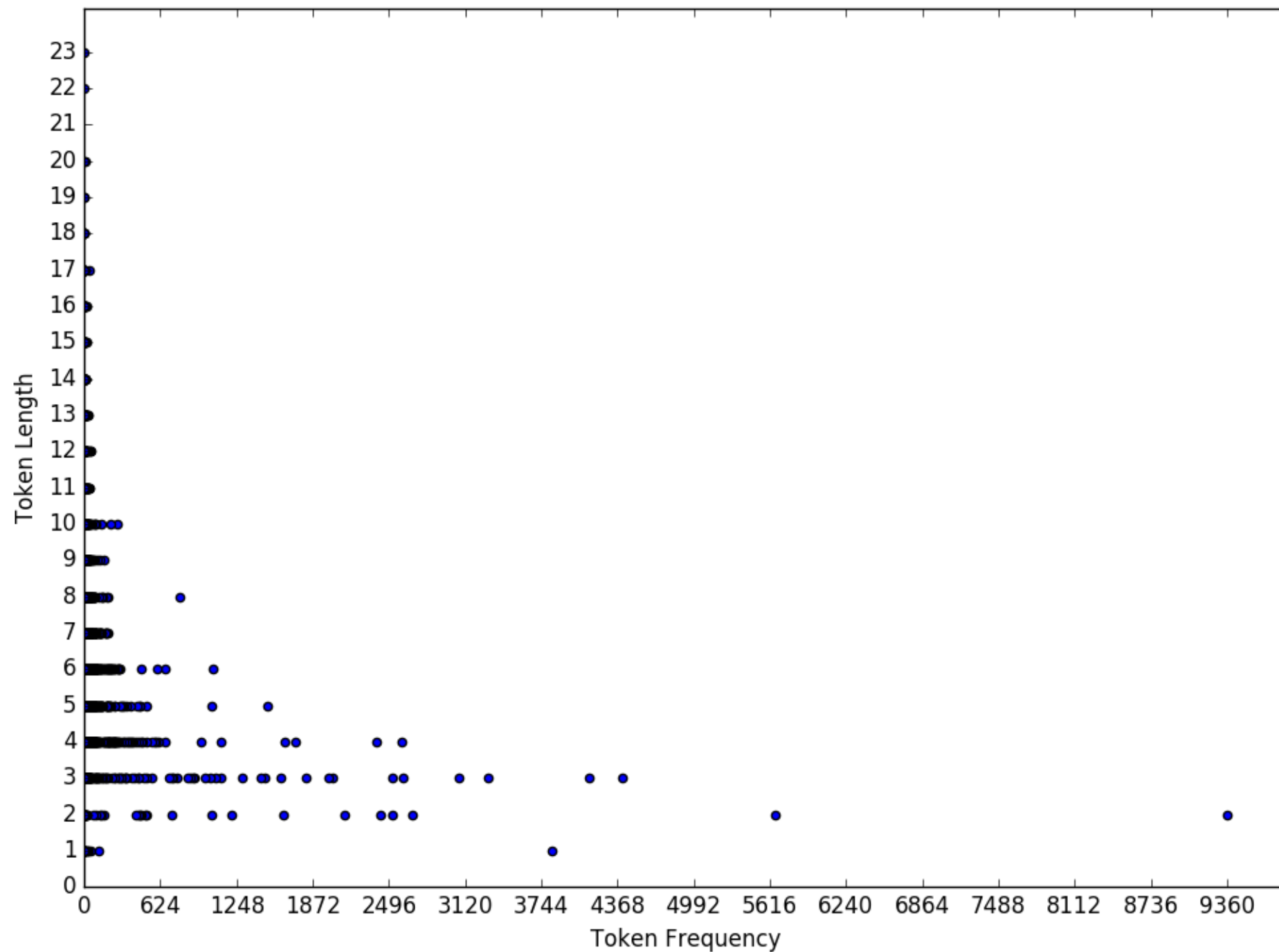
# Danish



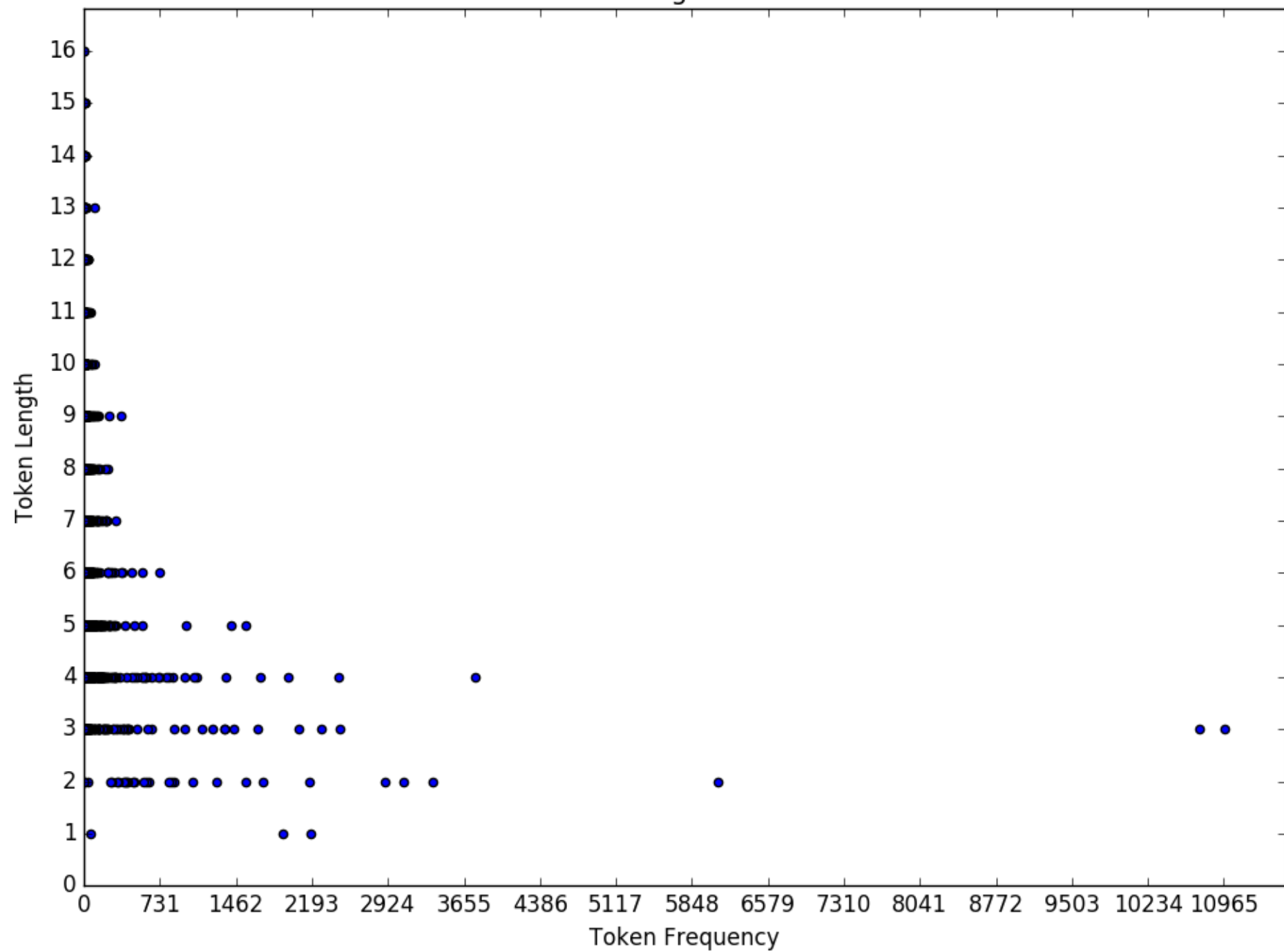
Dinka



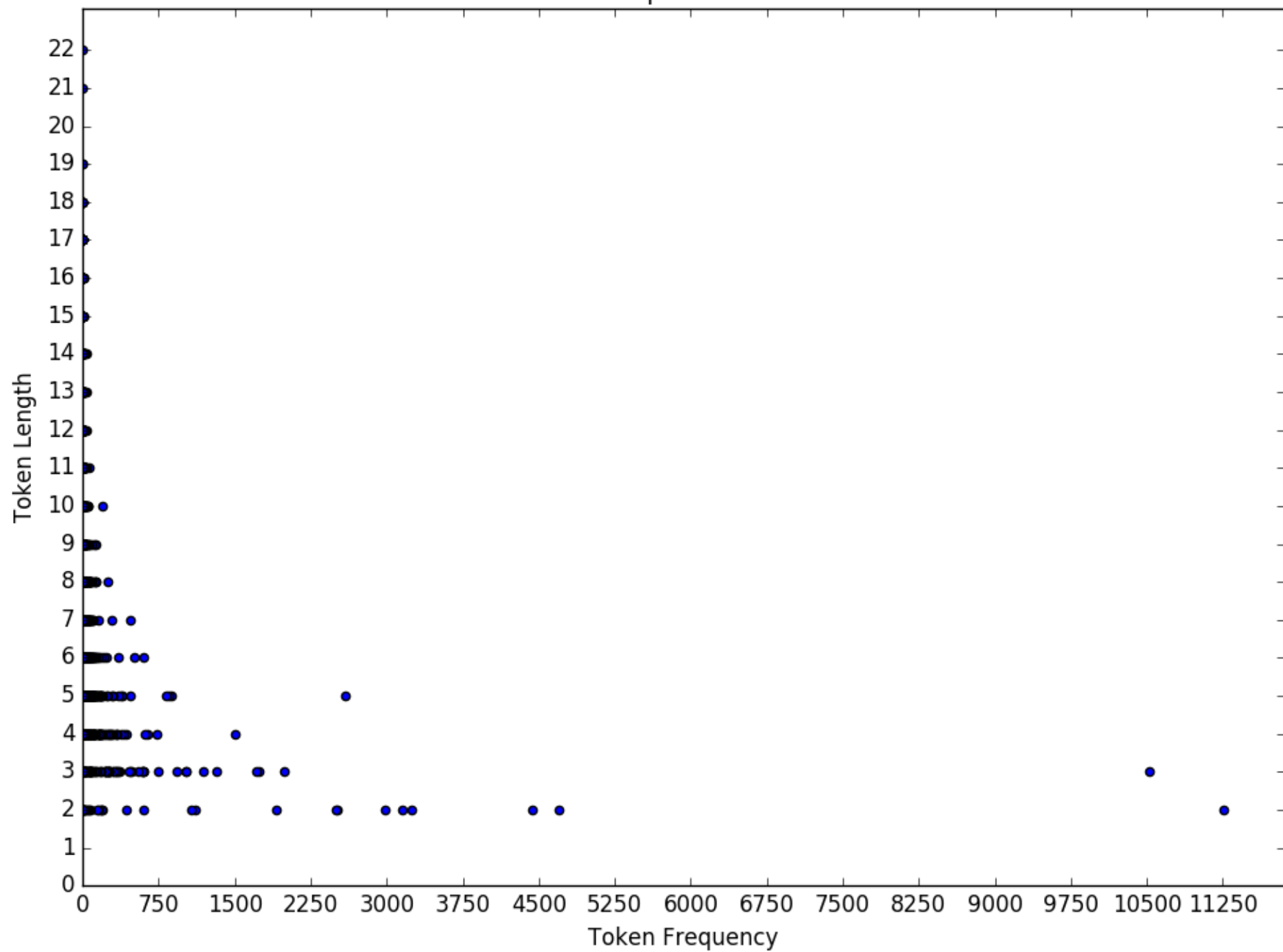
Dutch



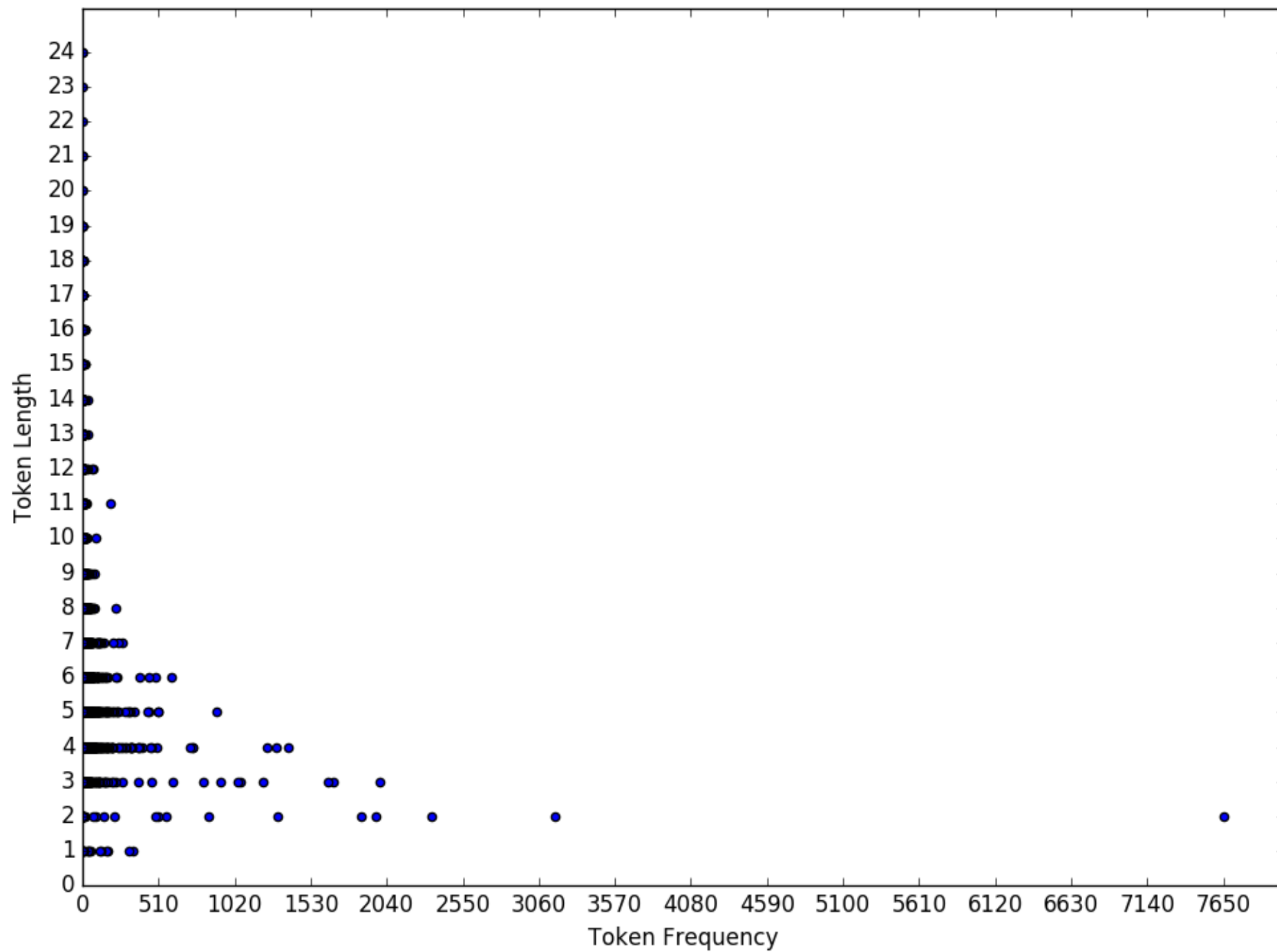
# English



## Token Frequency

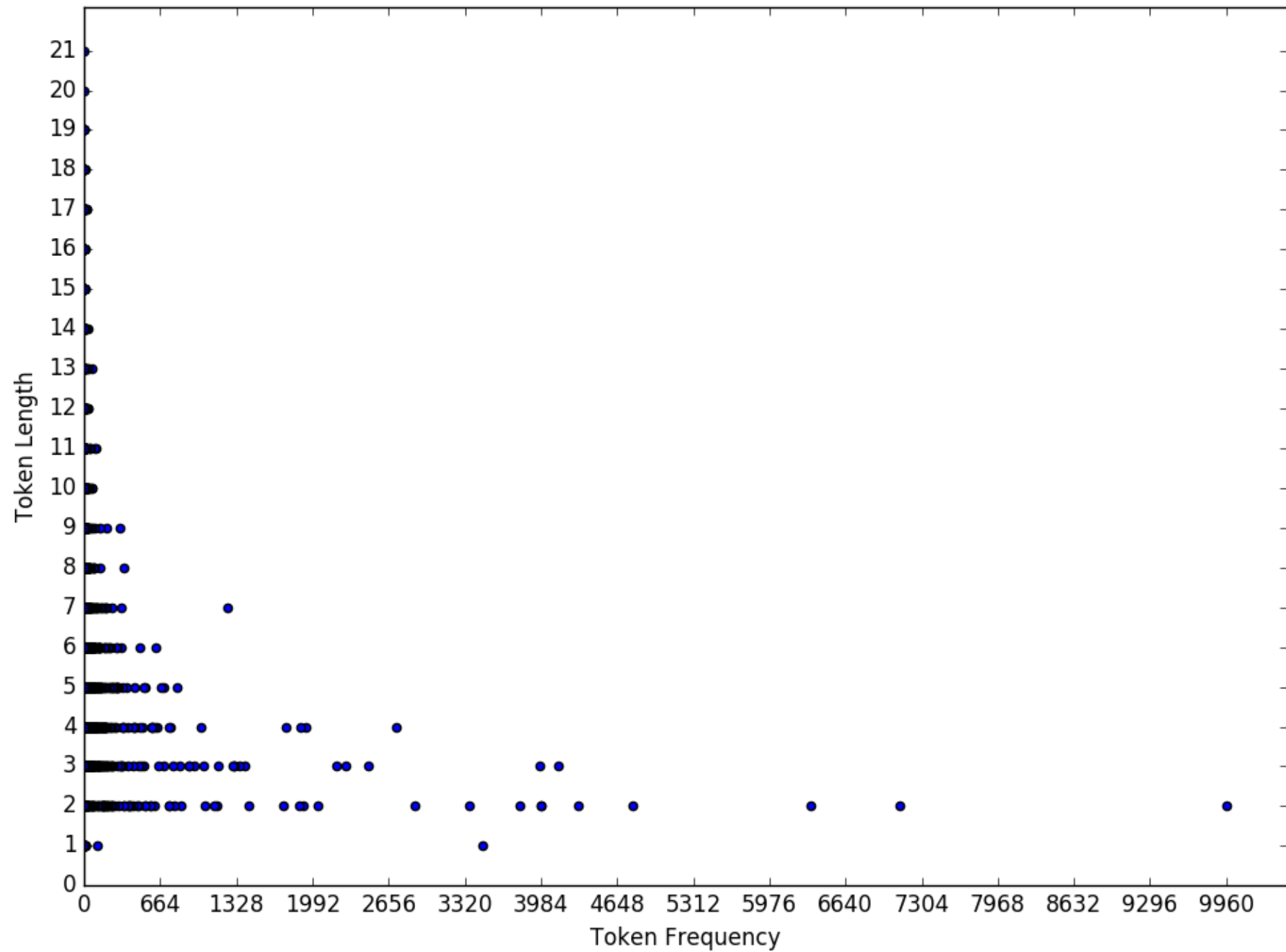


# Estonian

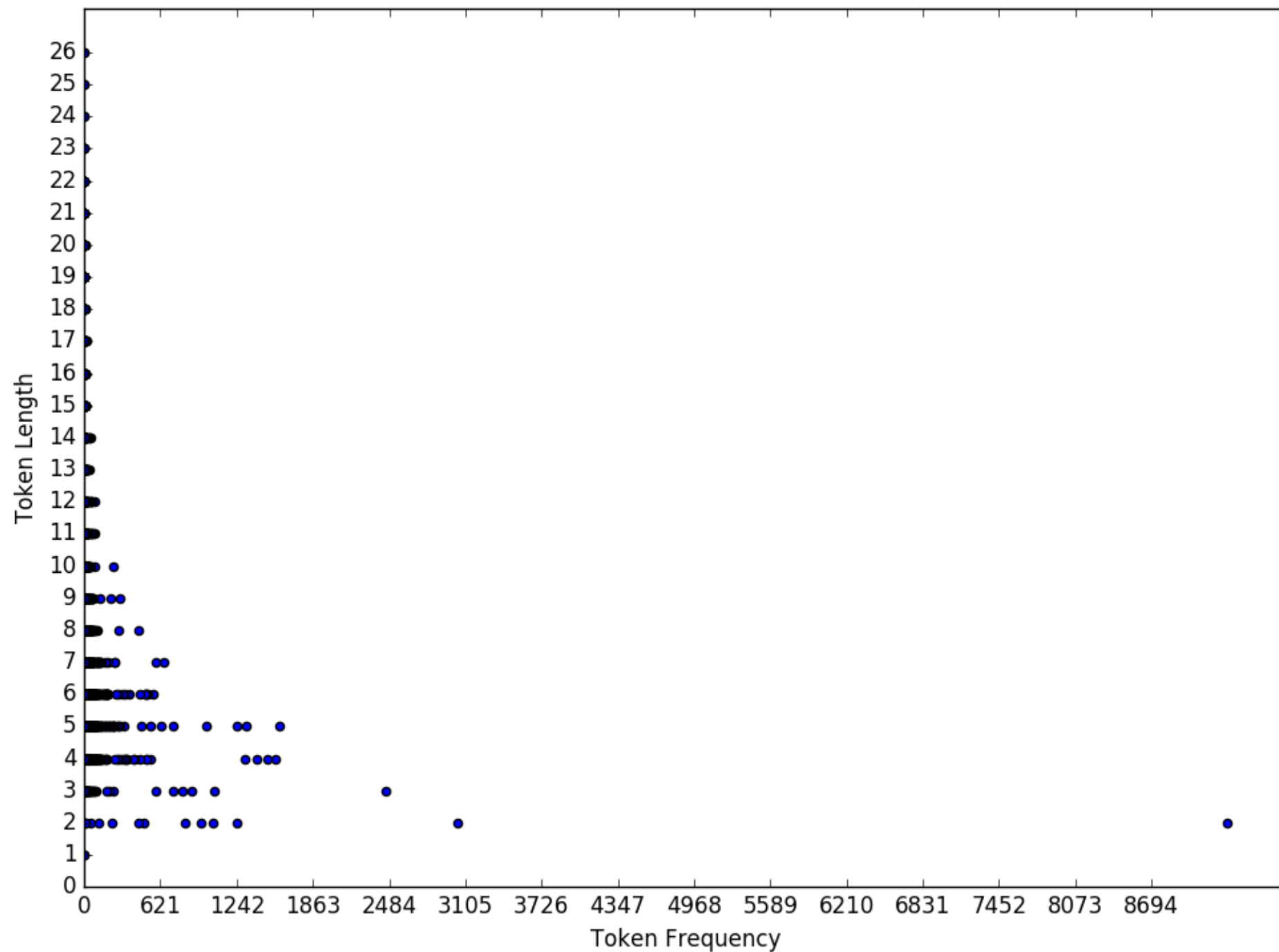




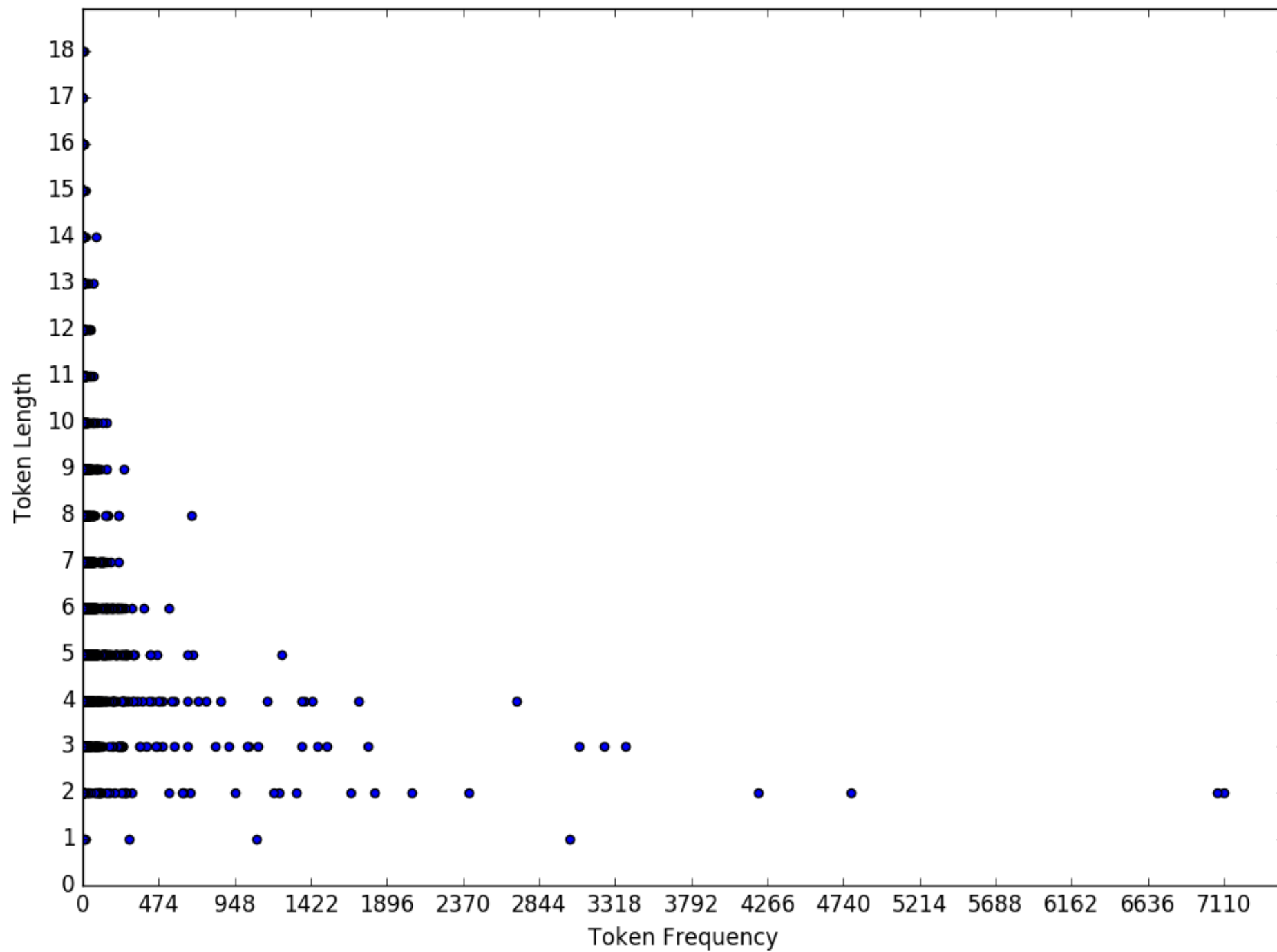
Ewe



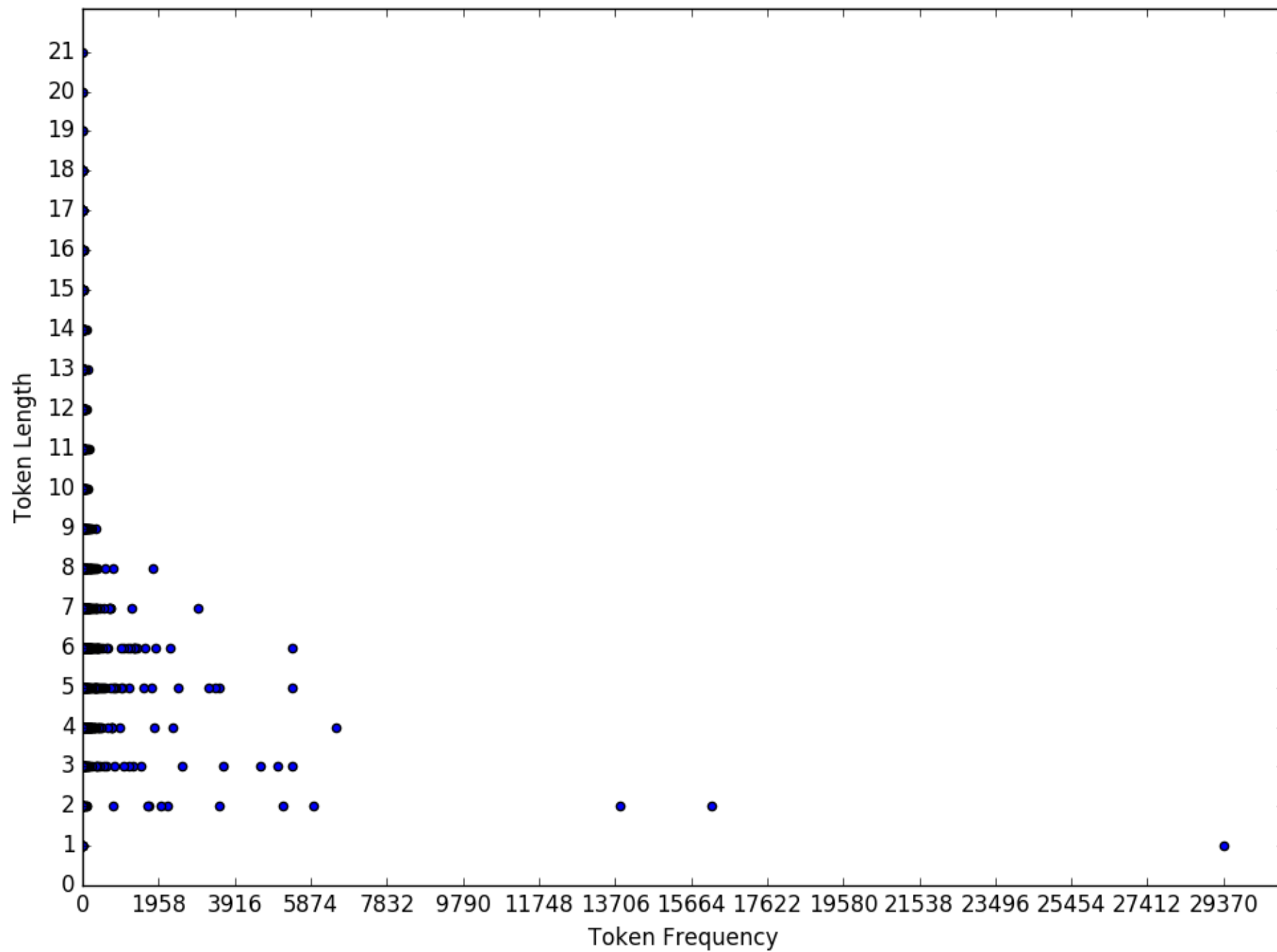
# Finnish



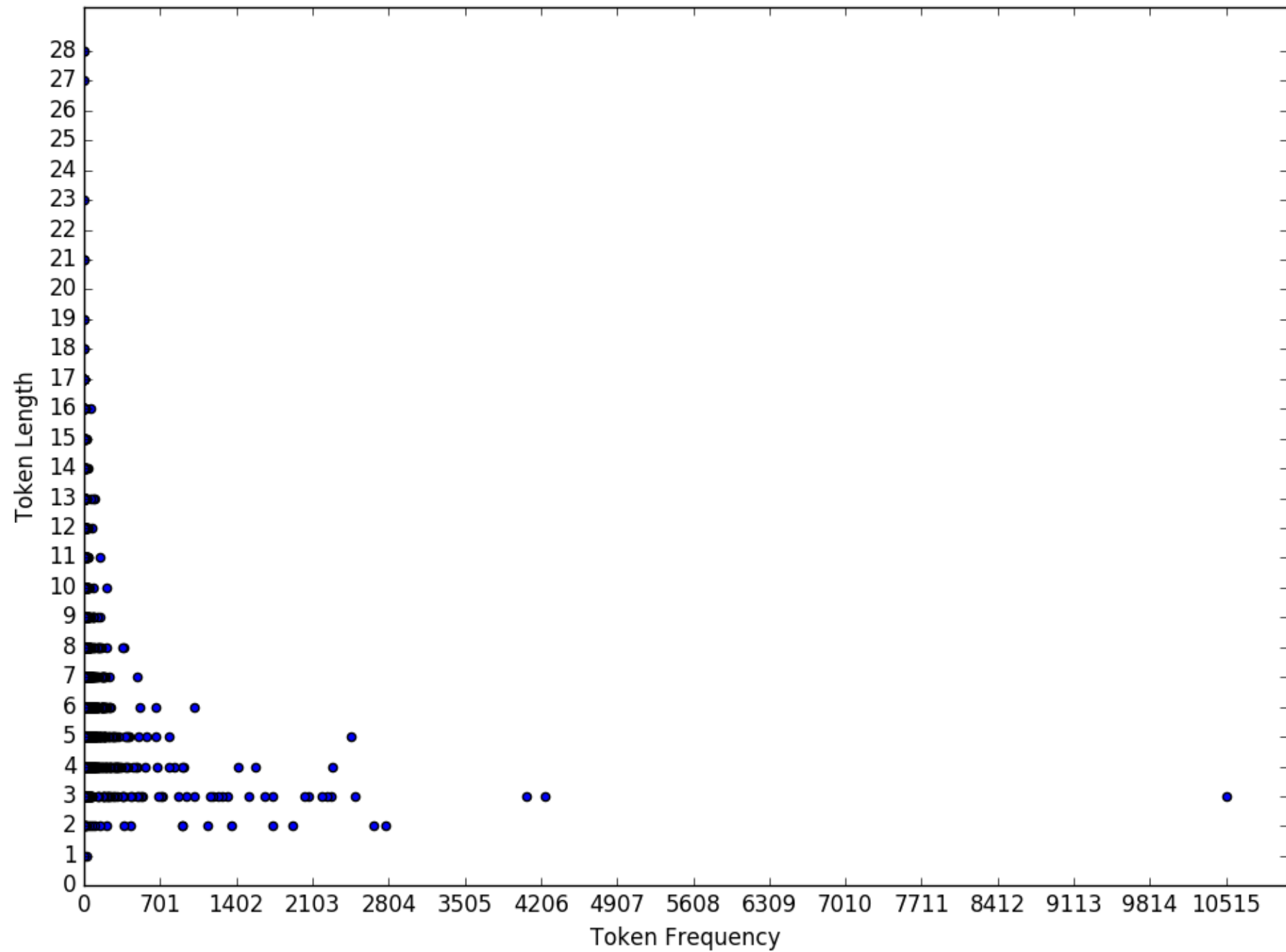
# French



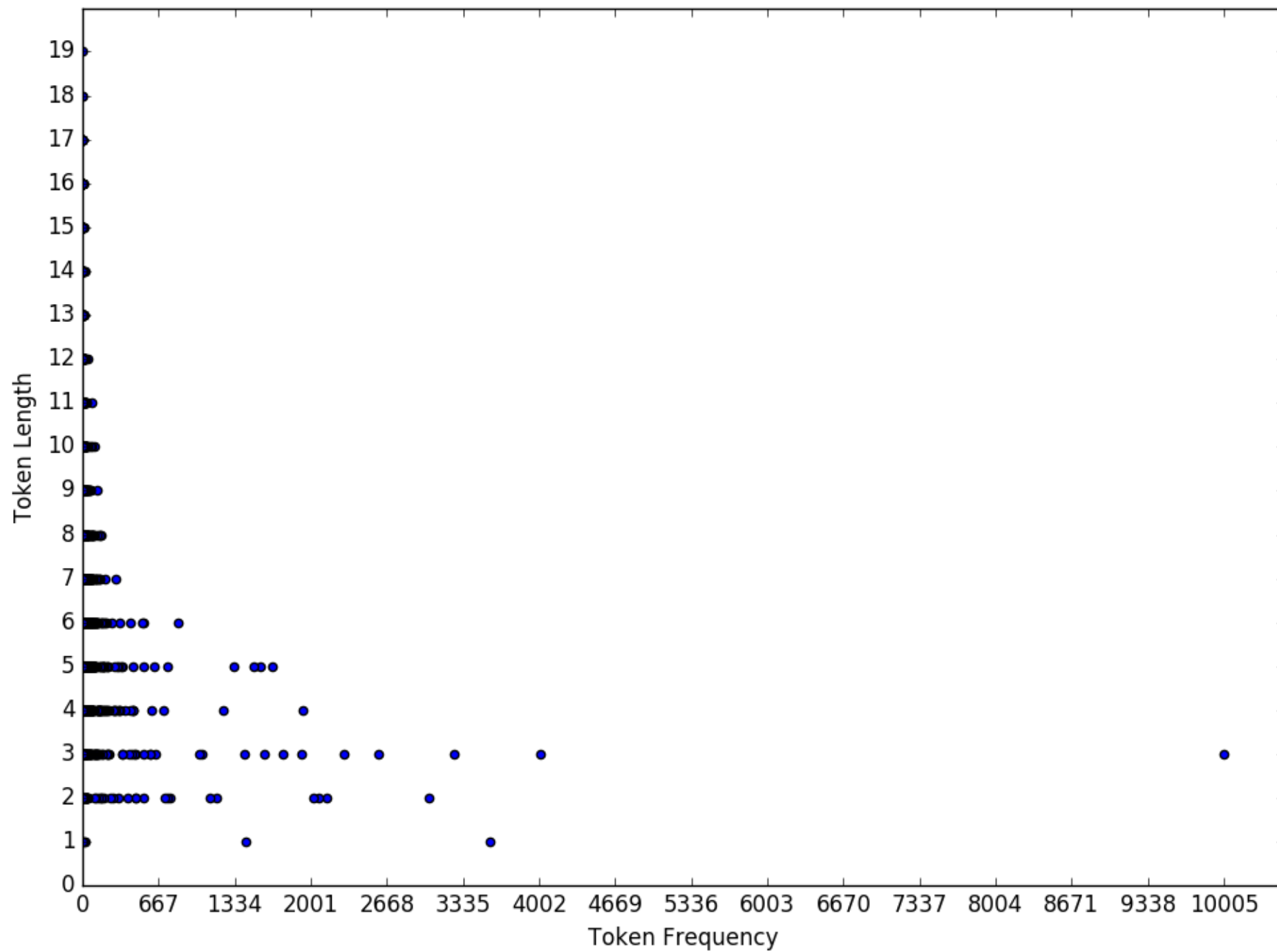
Galela



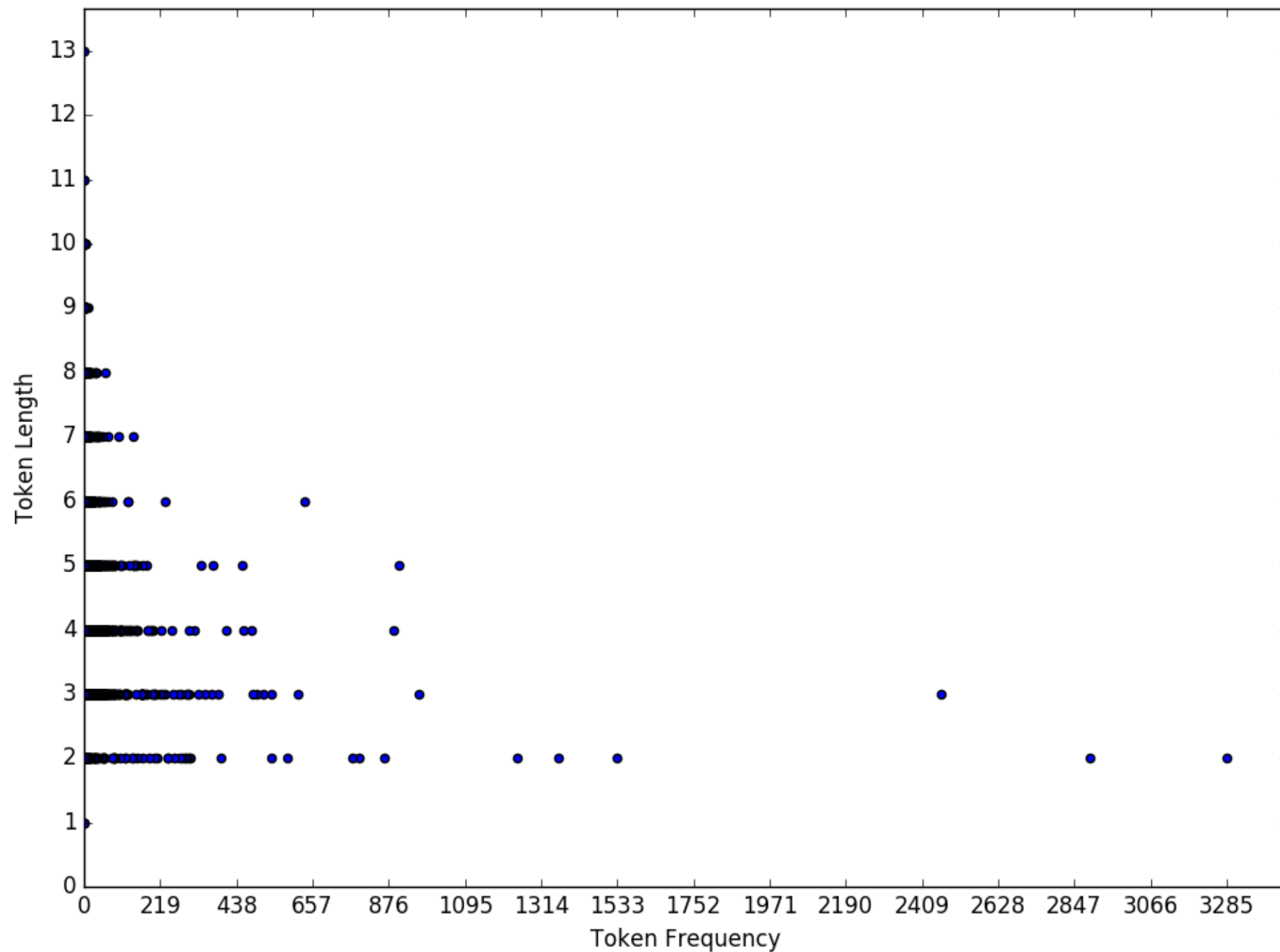
German



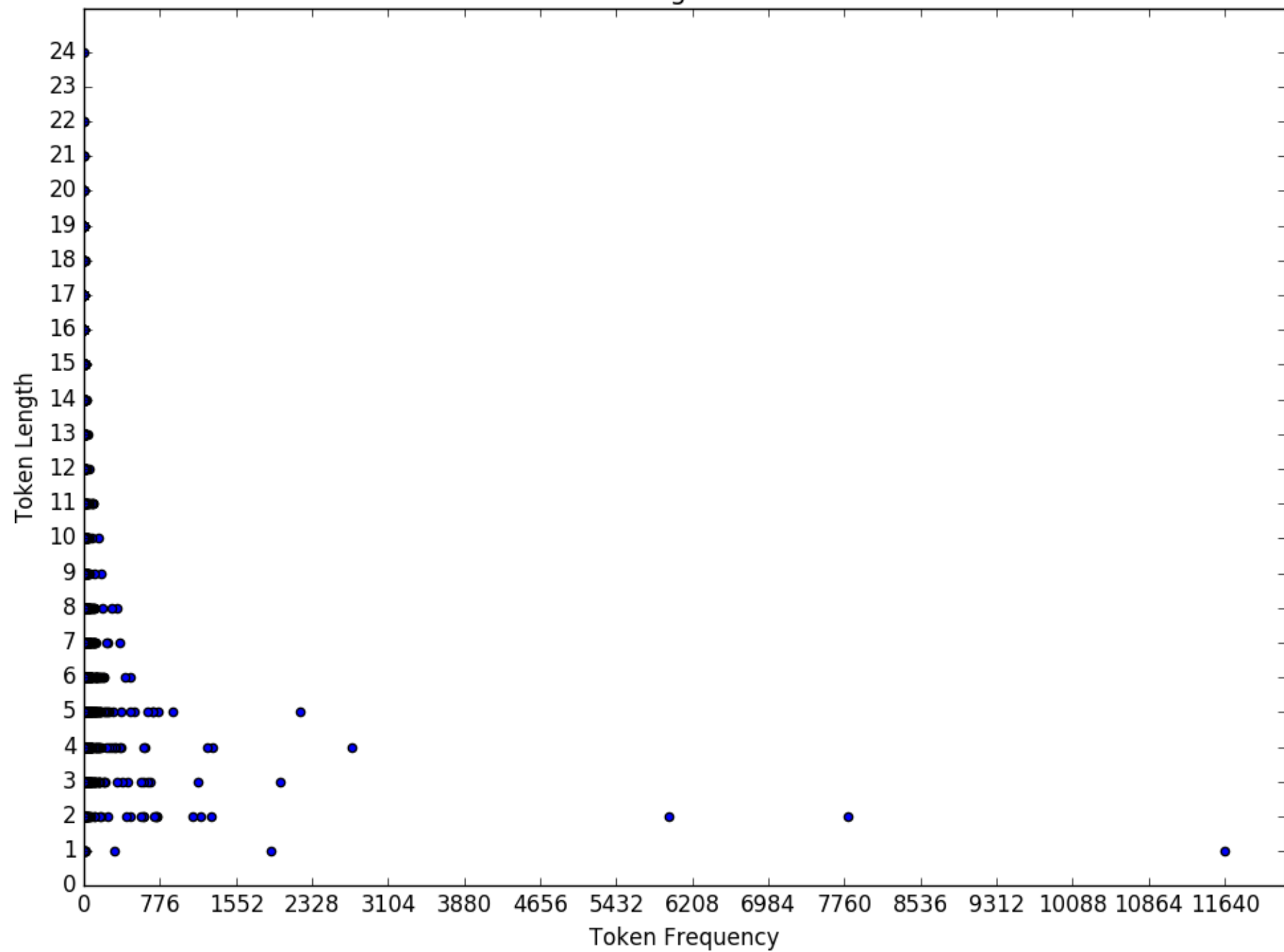
Greek



# Hebrew

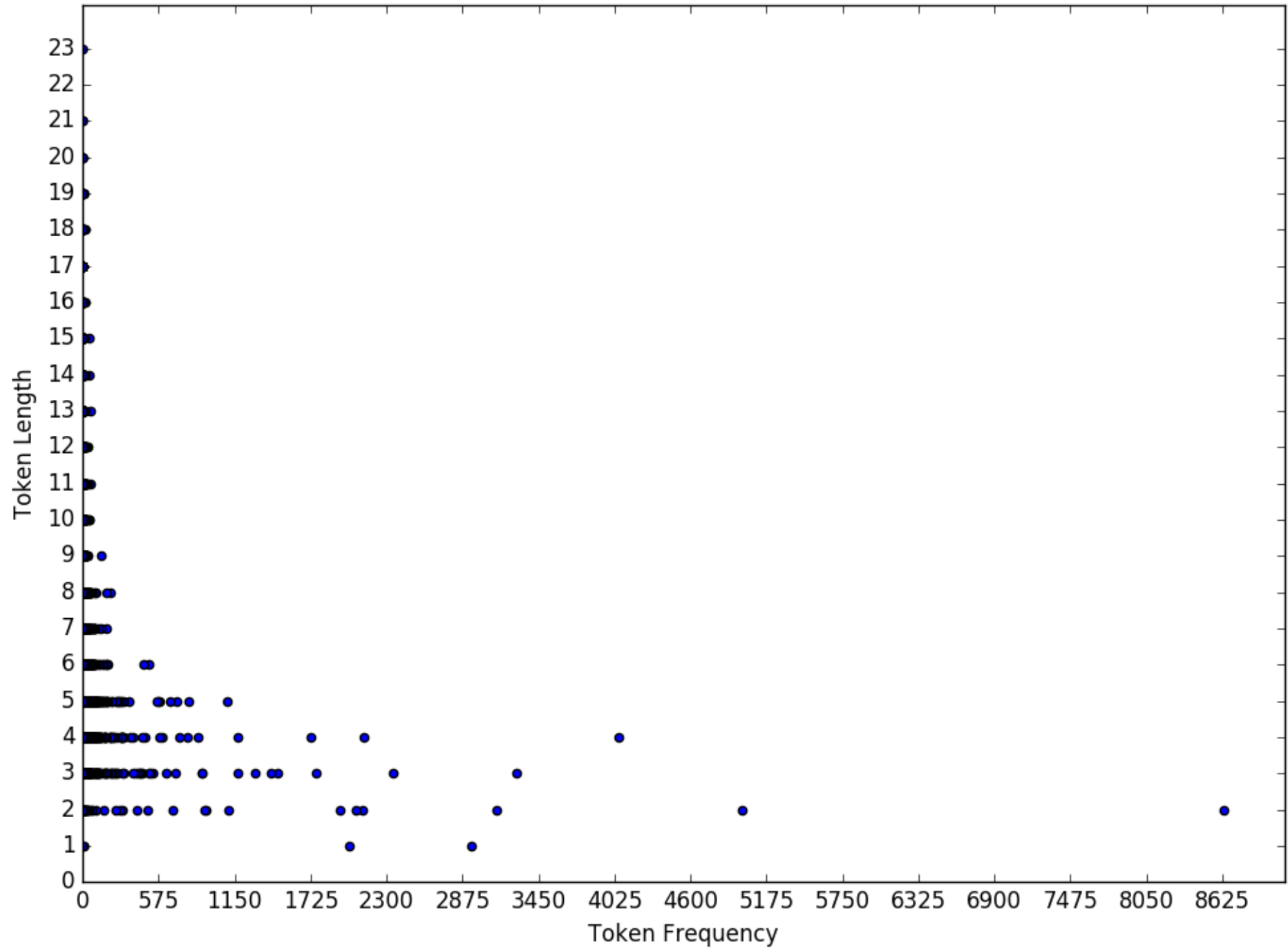


# Hungarian

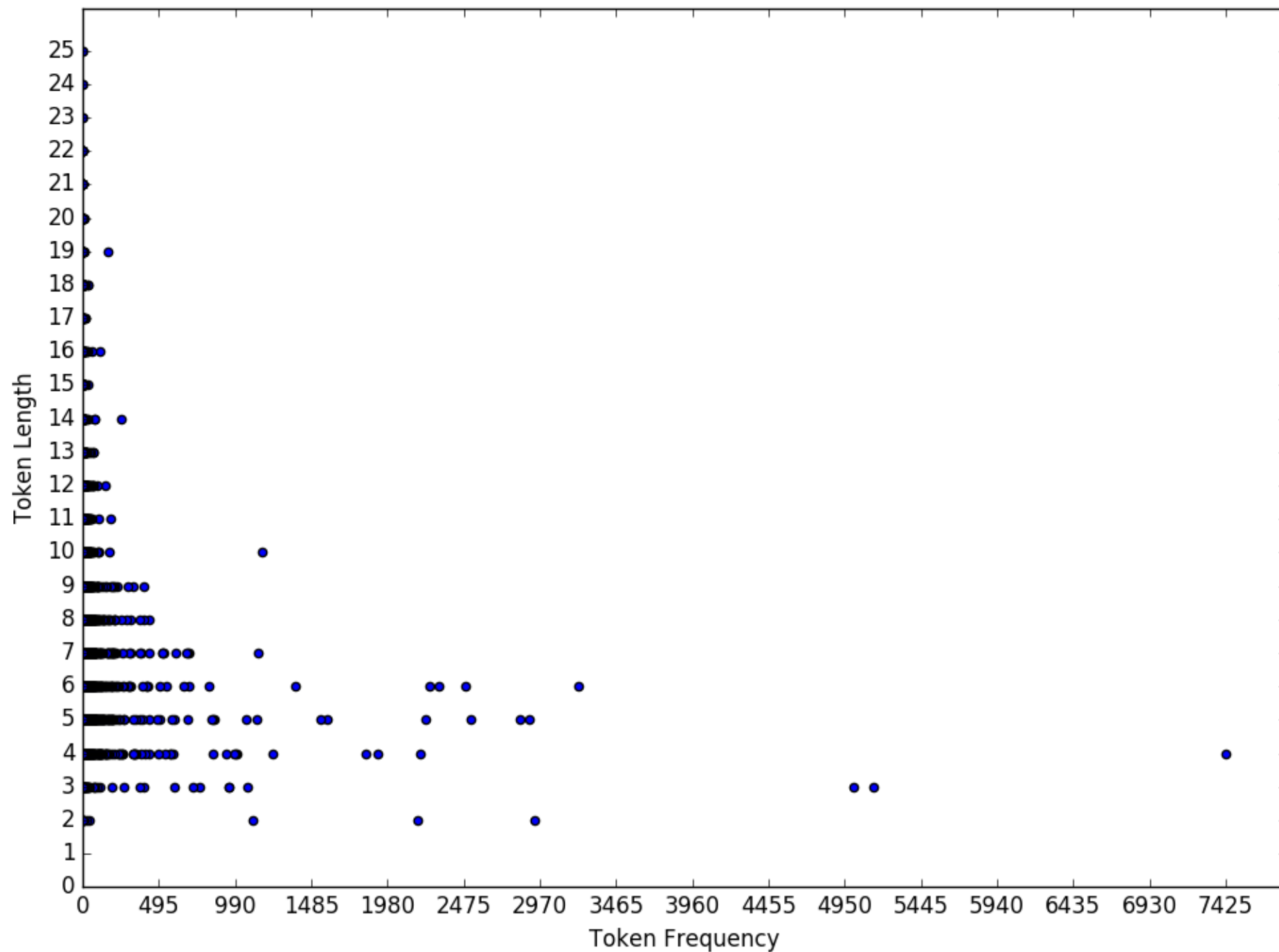




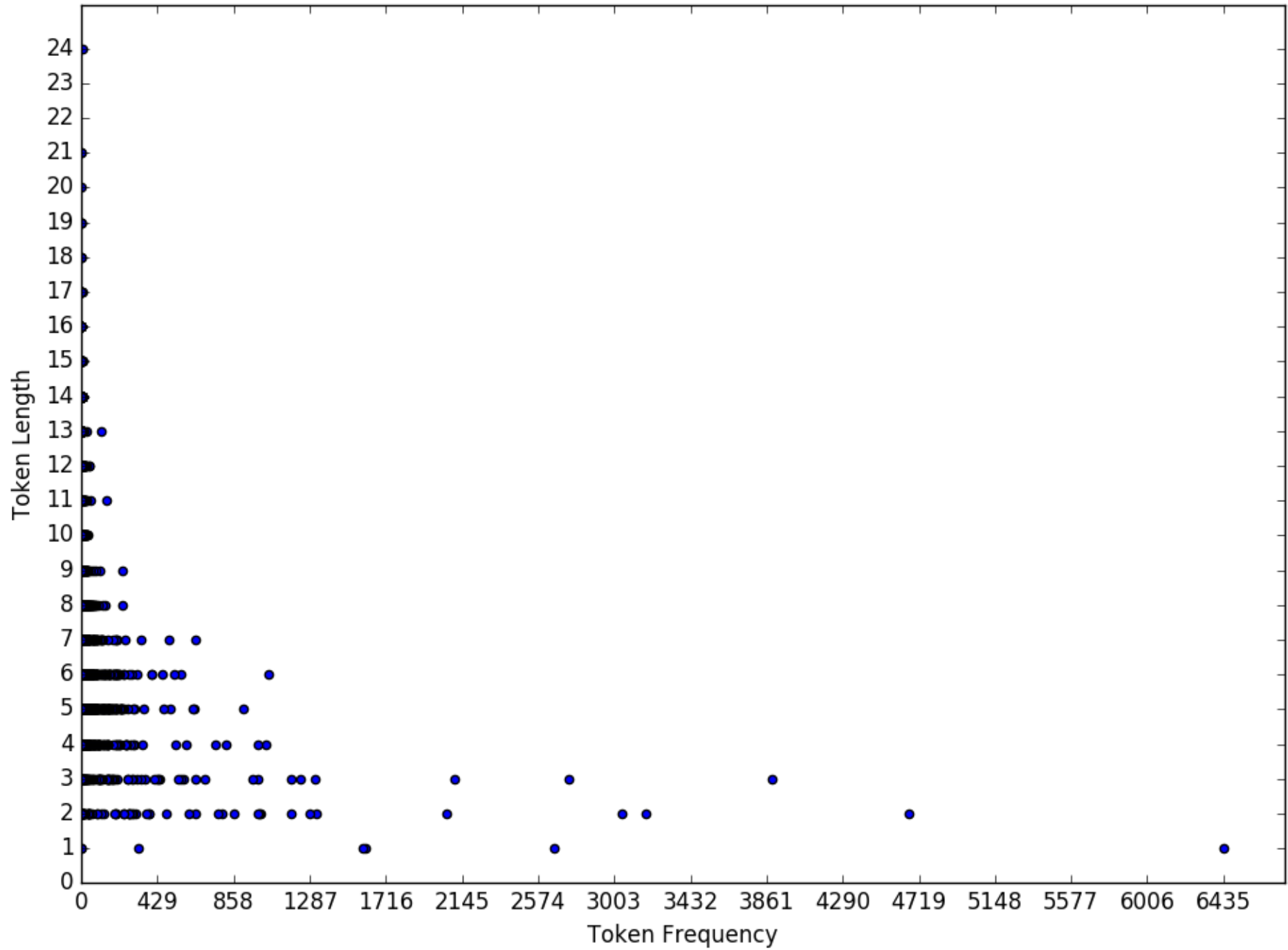
## Token Frequency



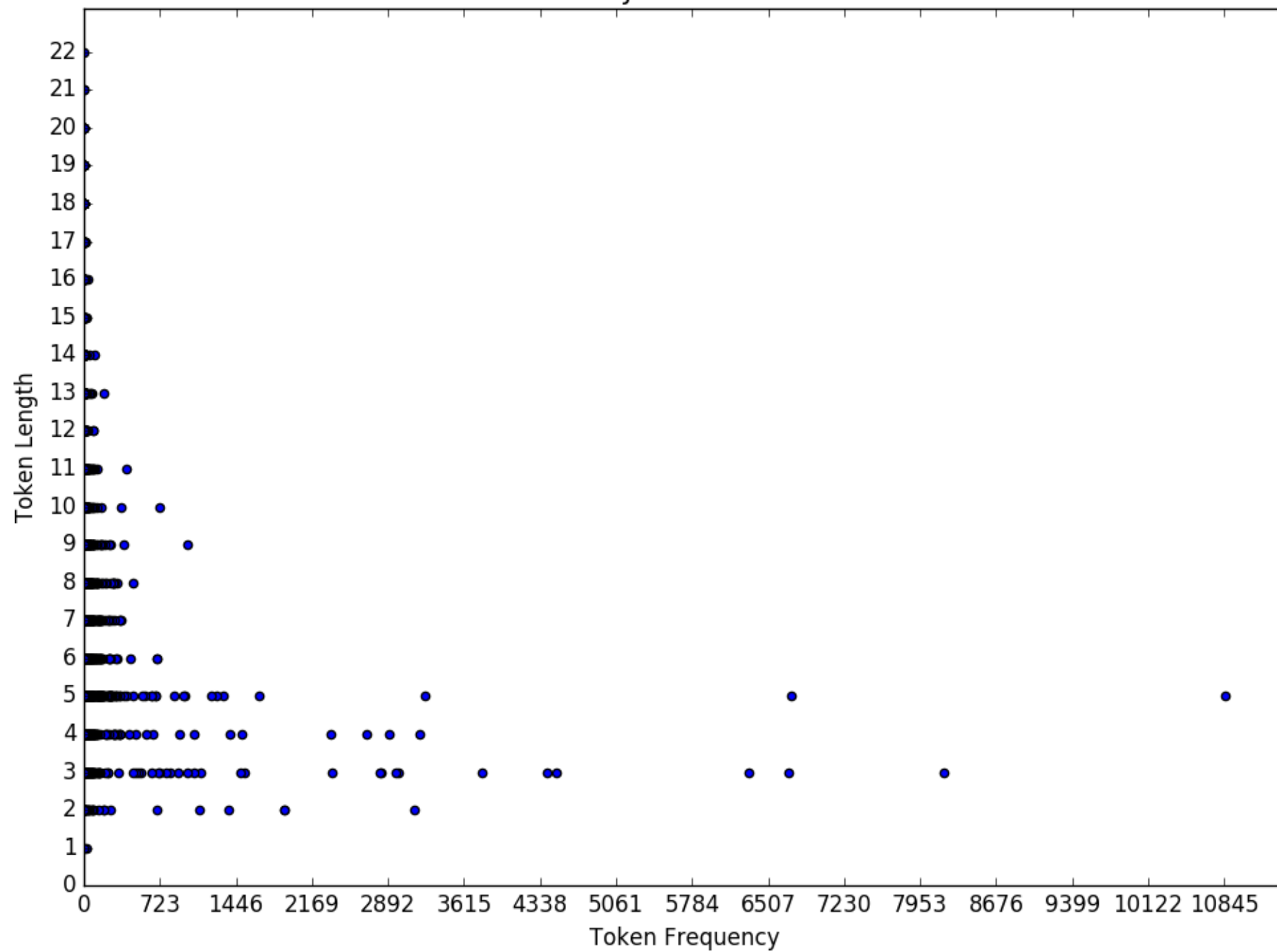
# Indonesian



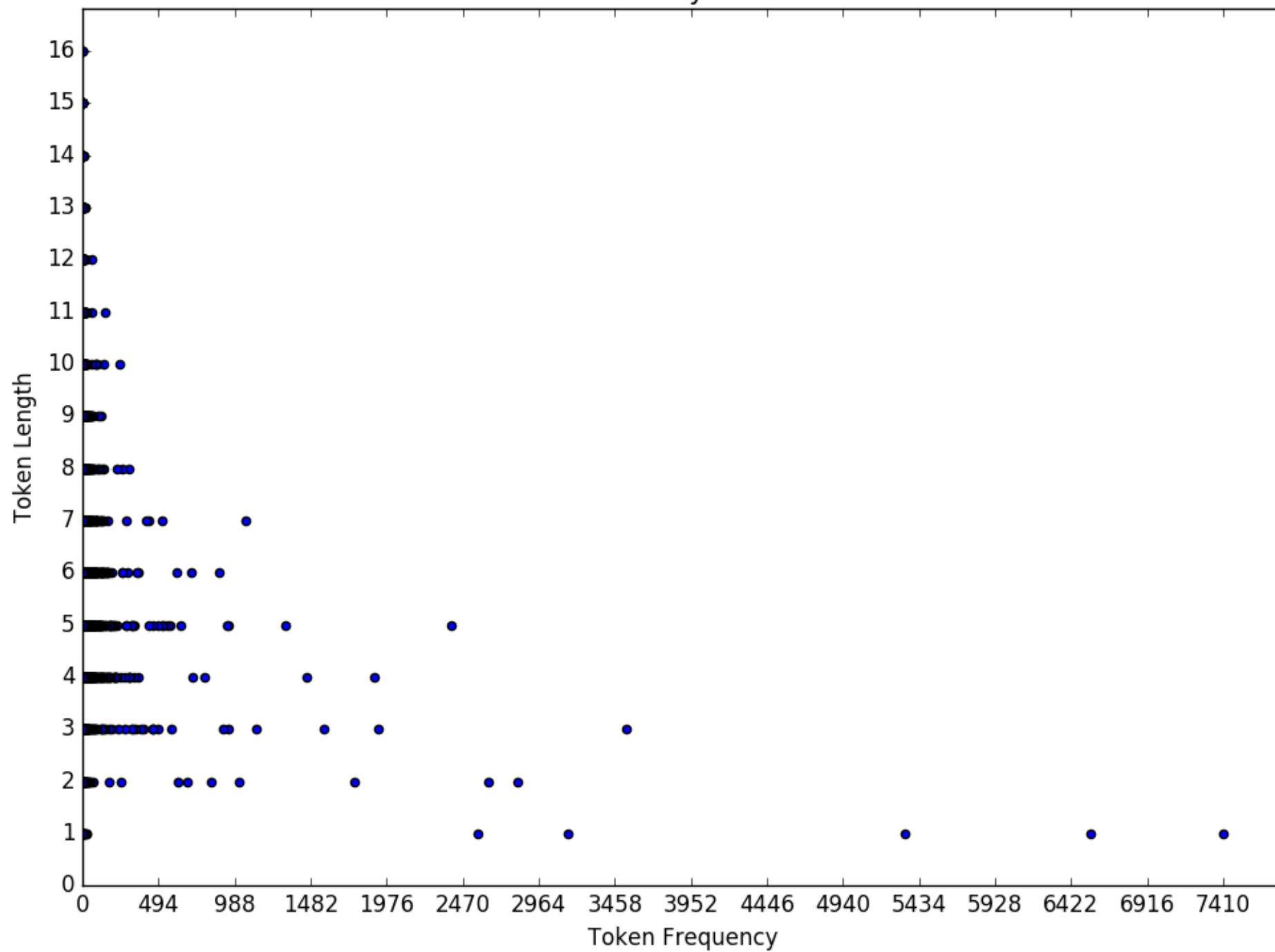
## Token Frequency



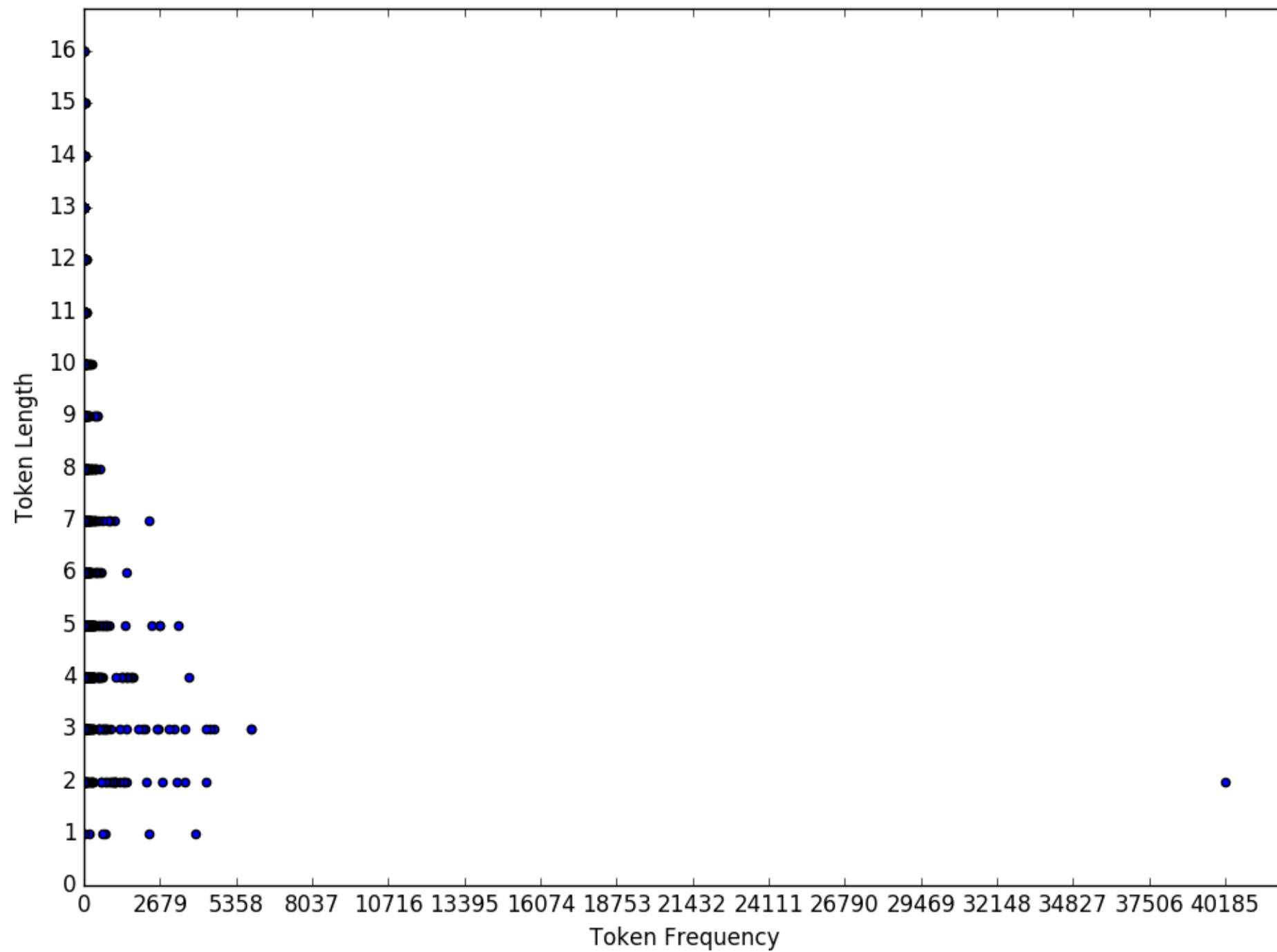
# Jakalteko



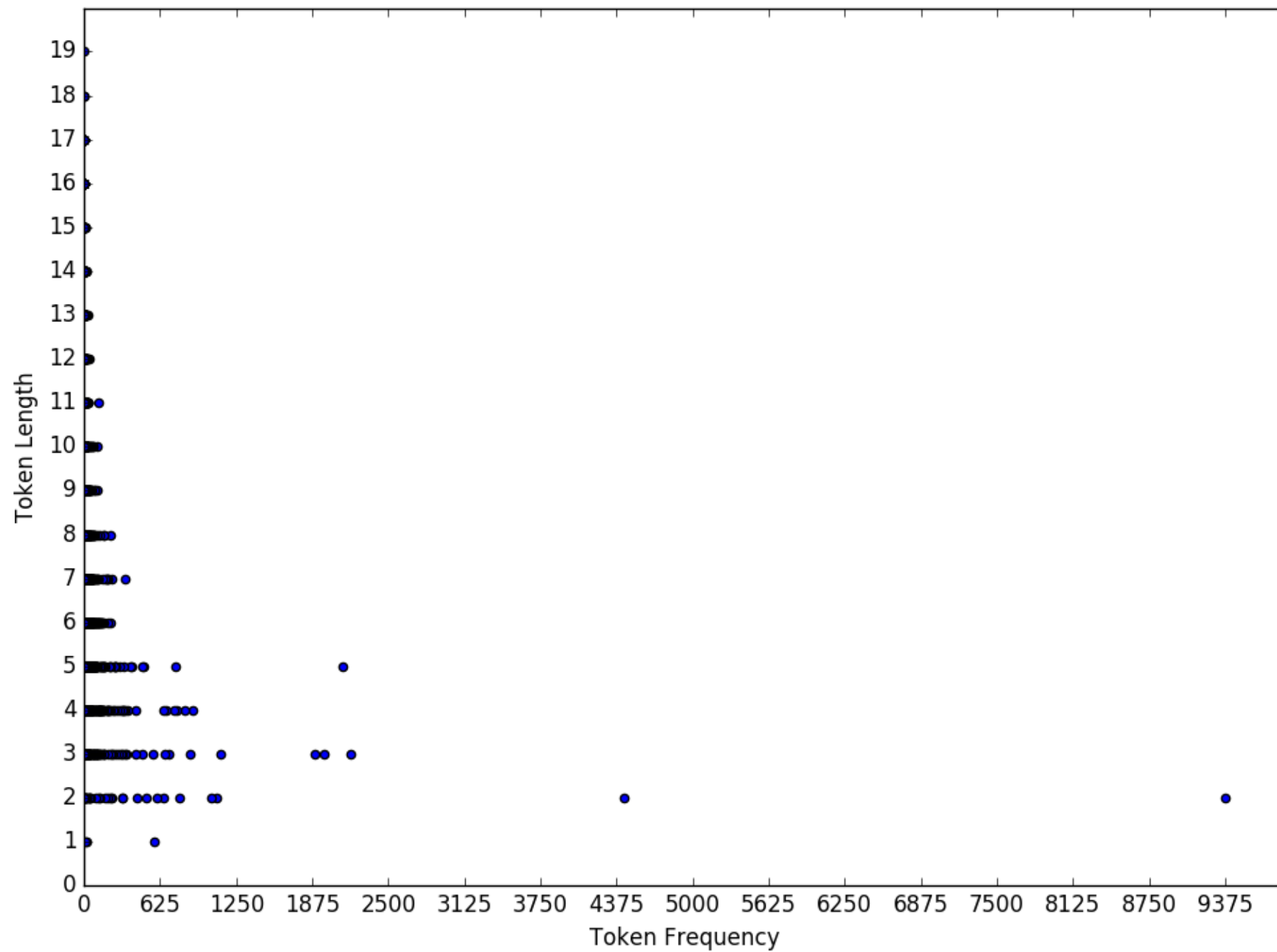
## Token Frequency



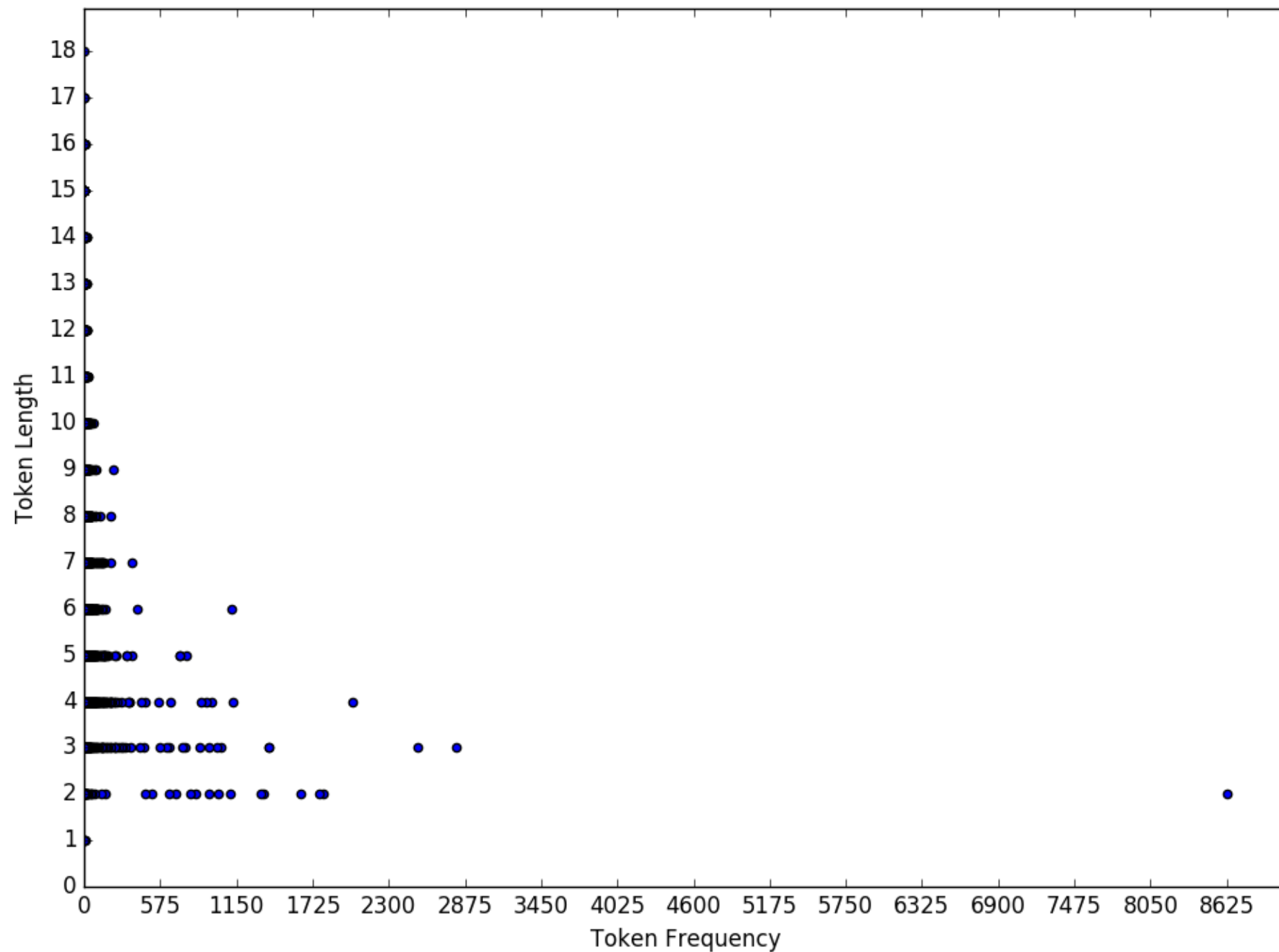
# K'iche'



Latin

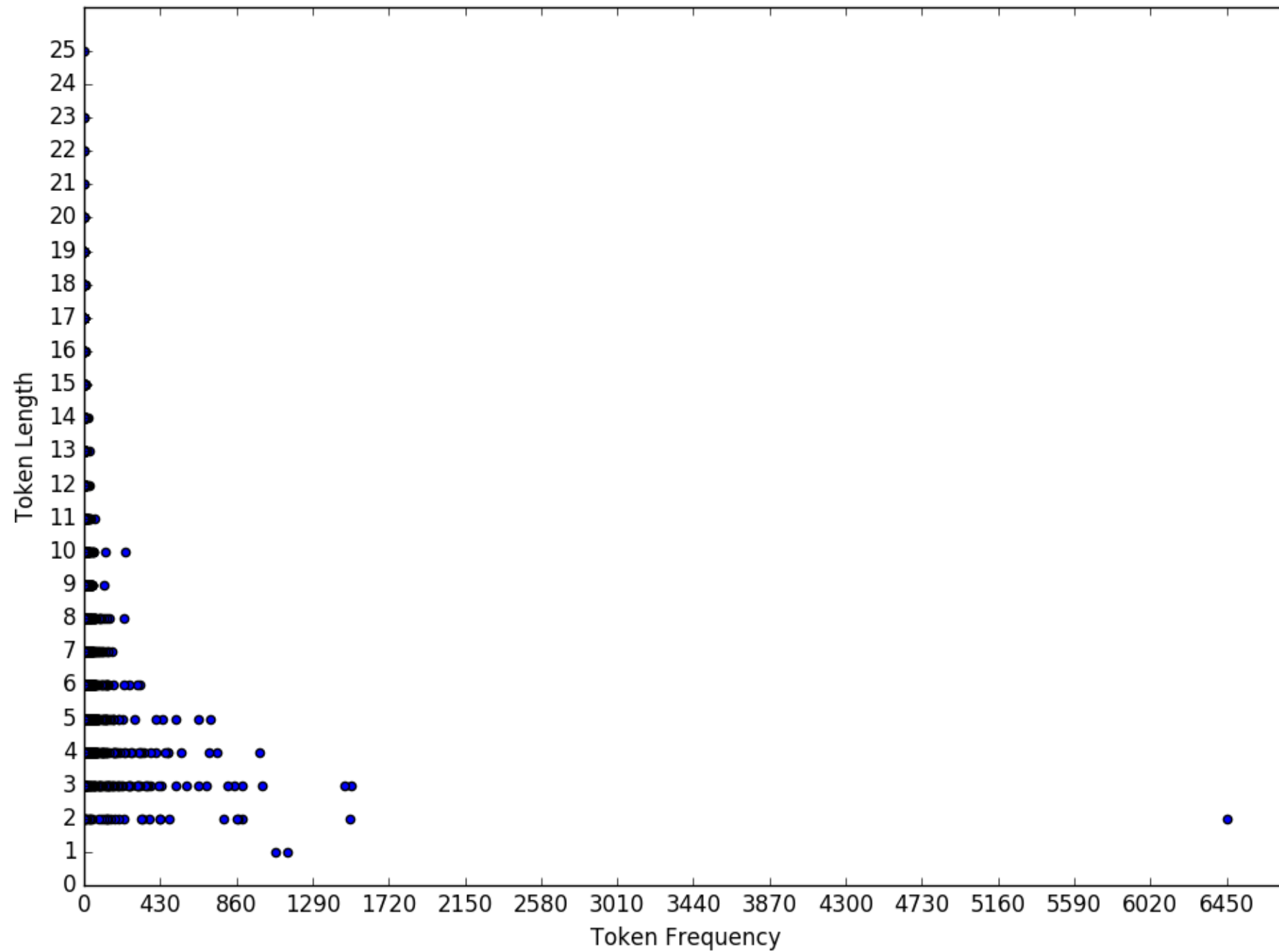


Latvian

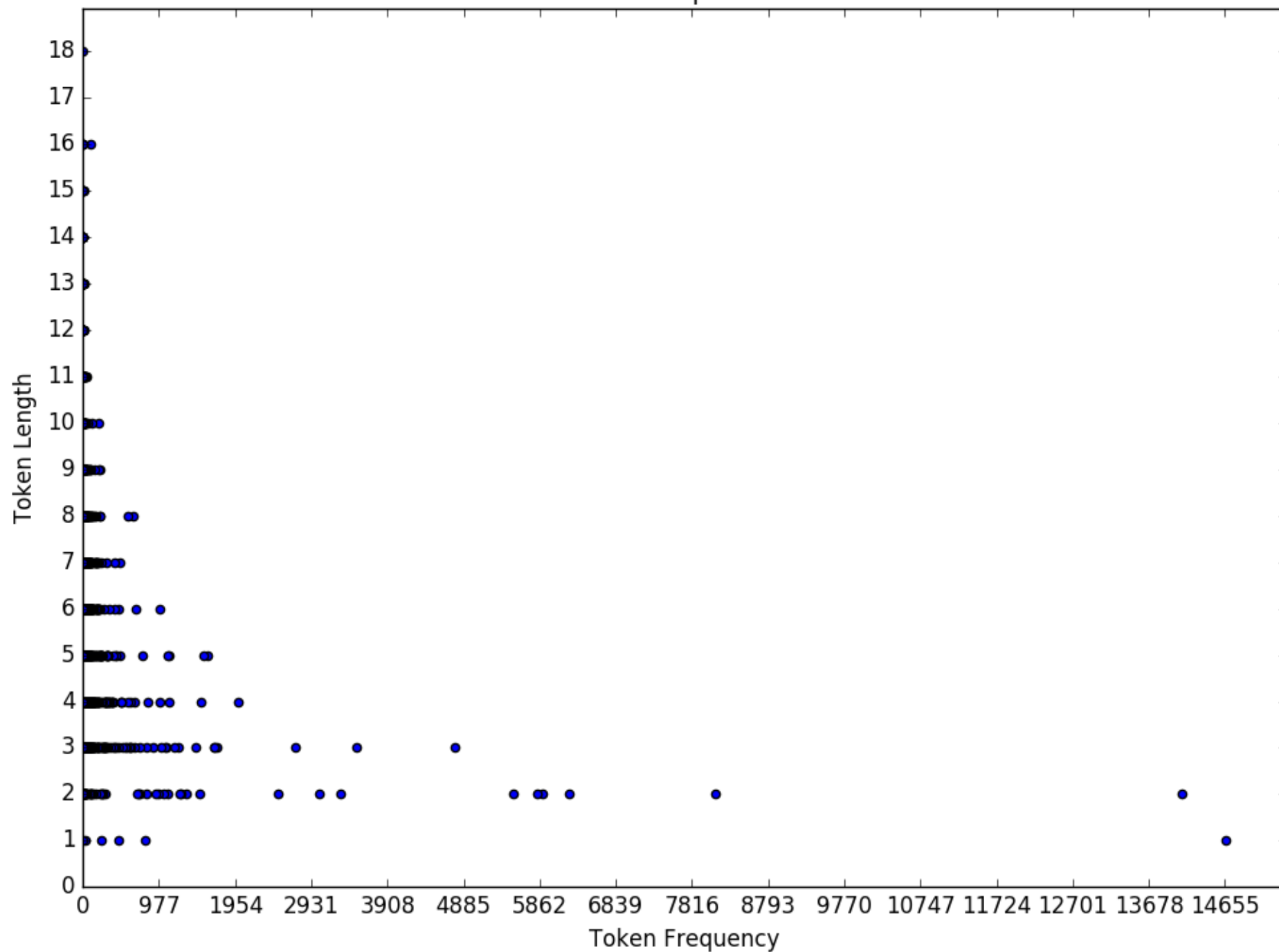




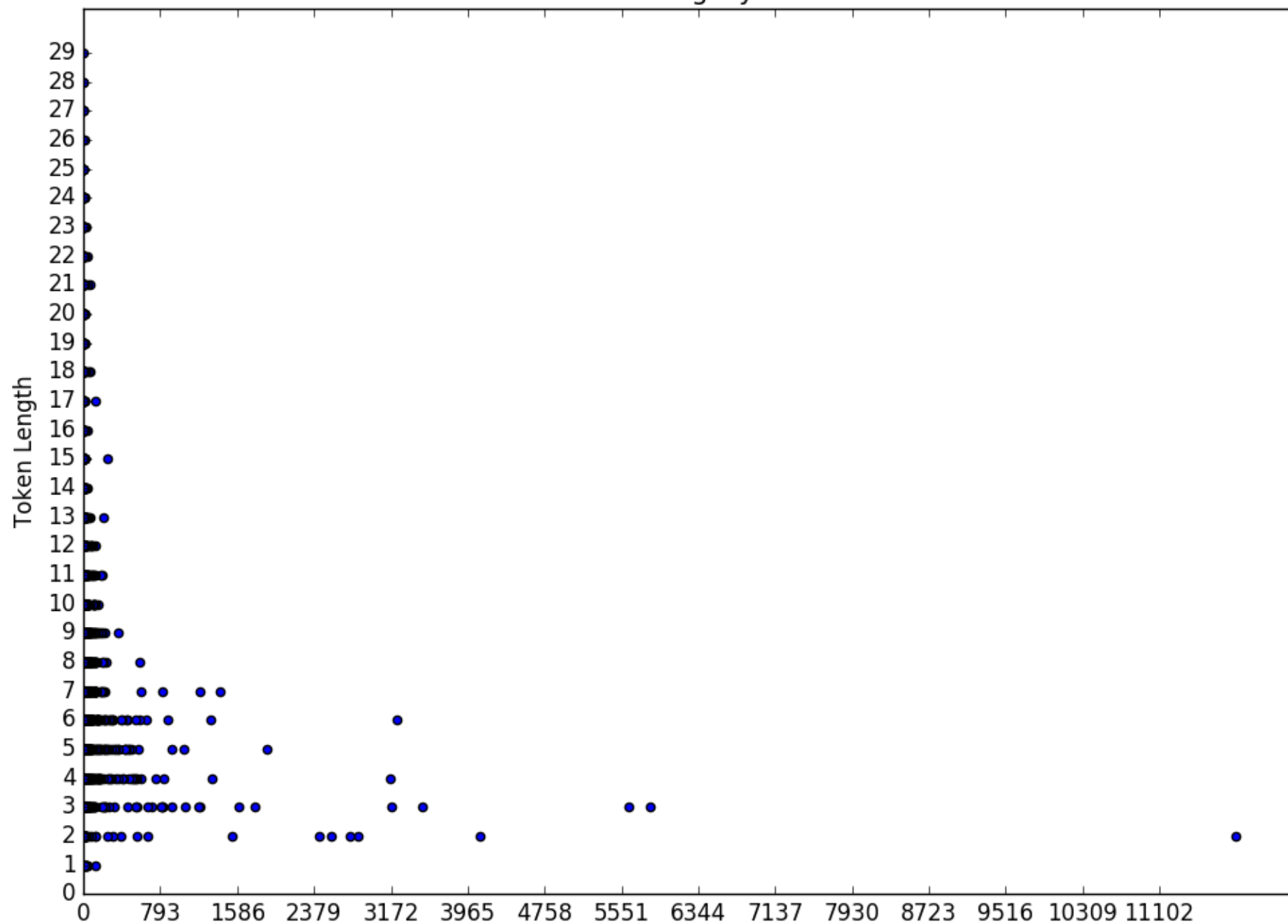
# Lithuanian



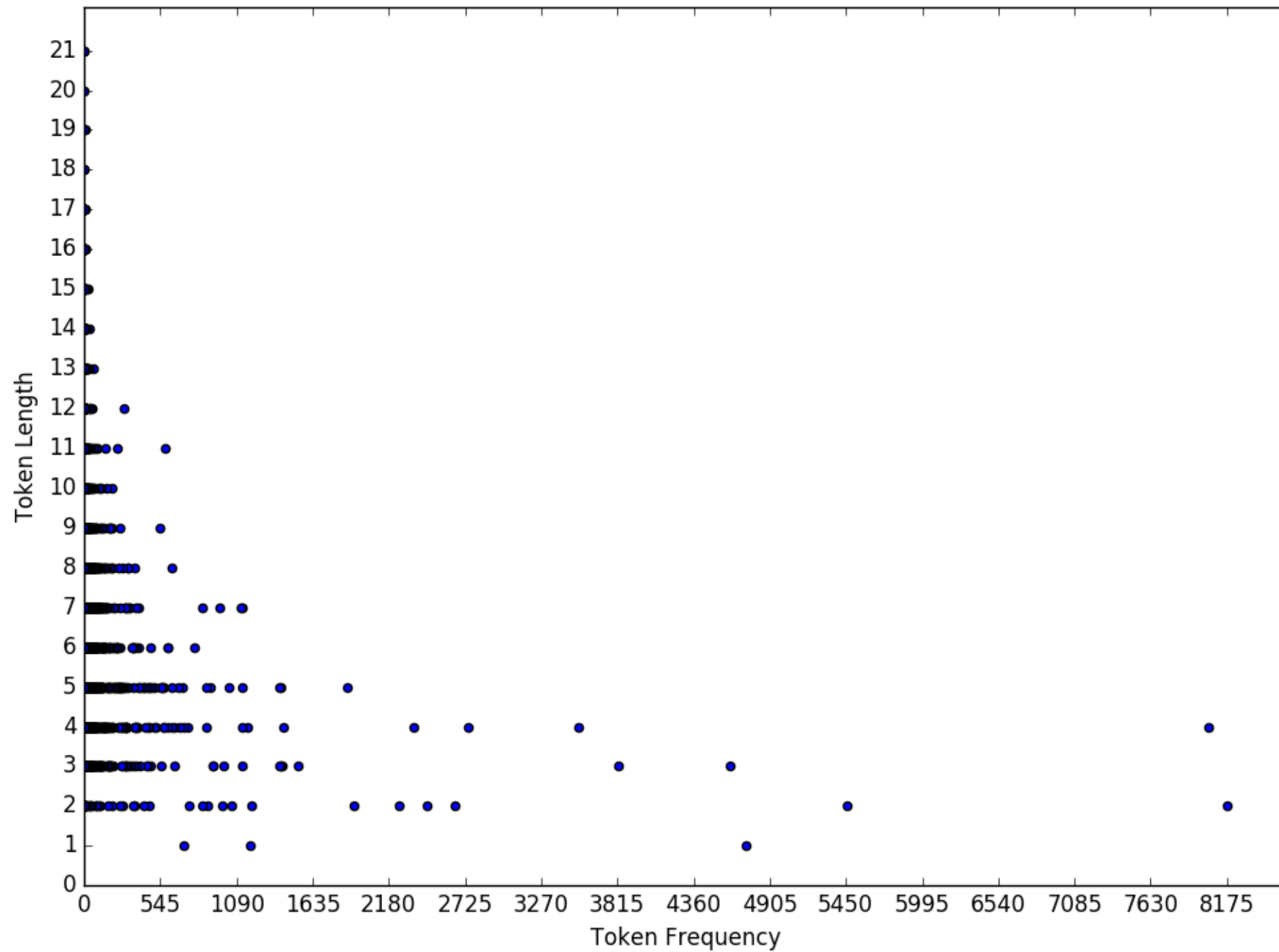
# Lukpa



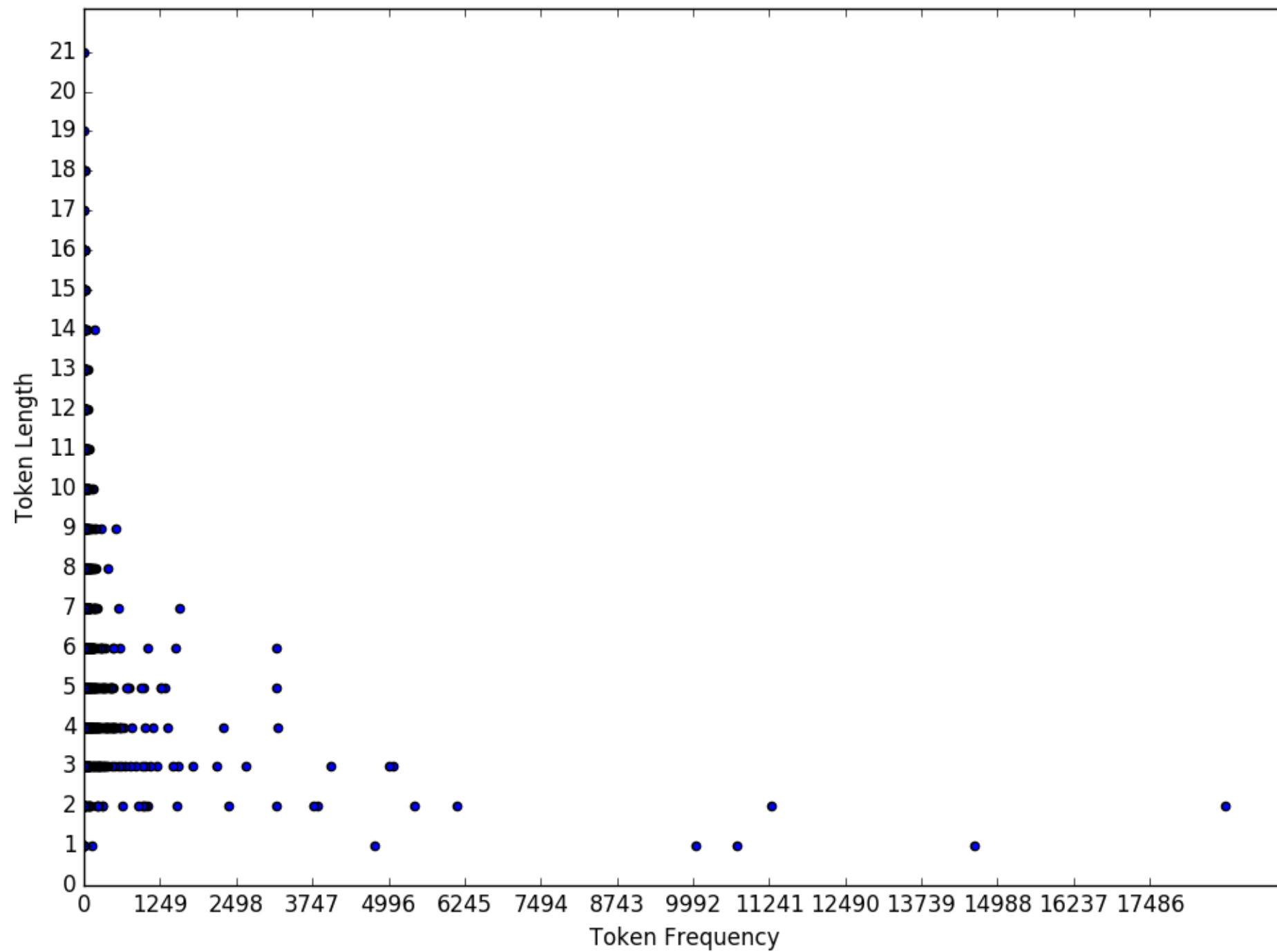
## Token Frequency



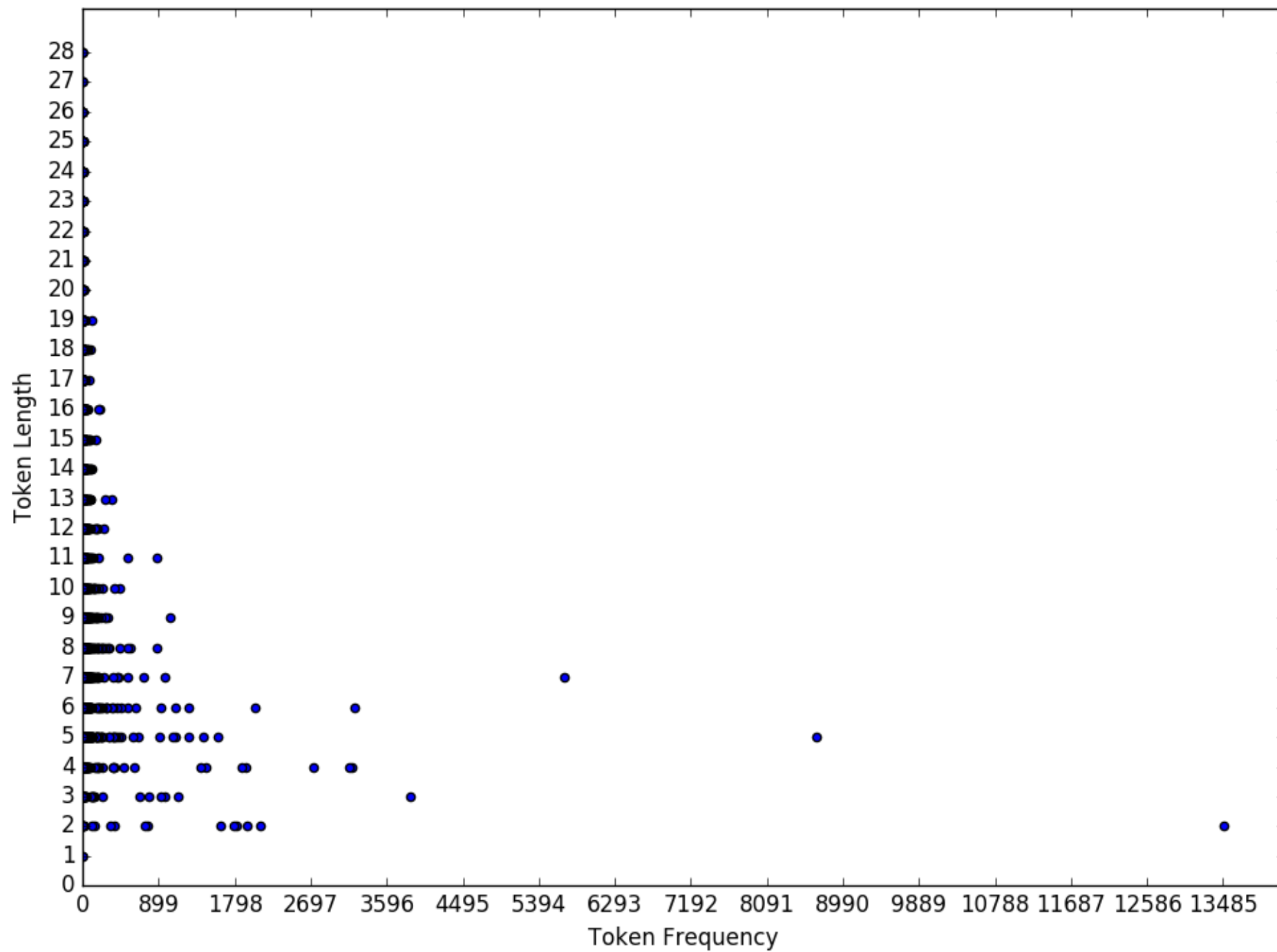
Mam



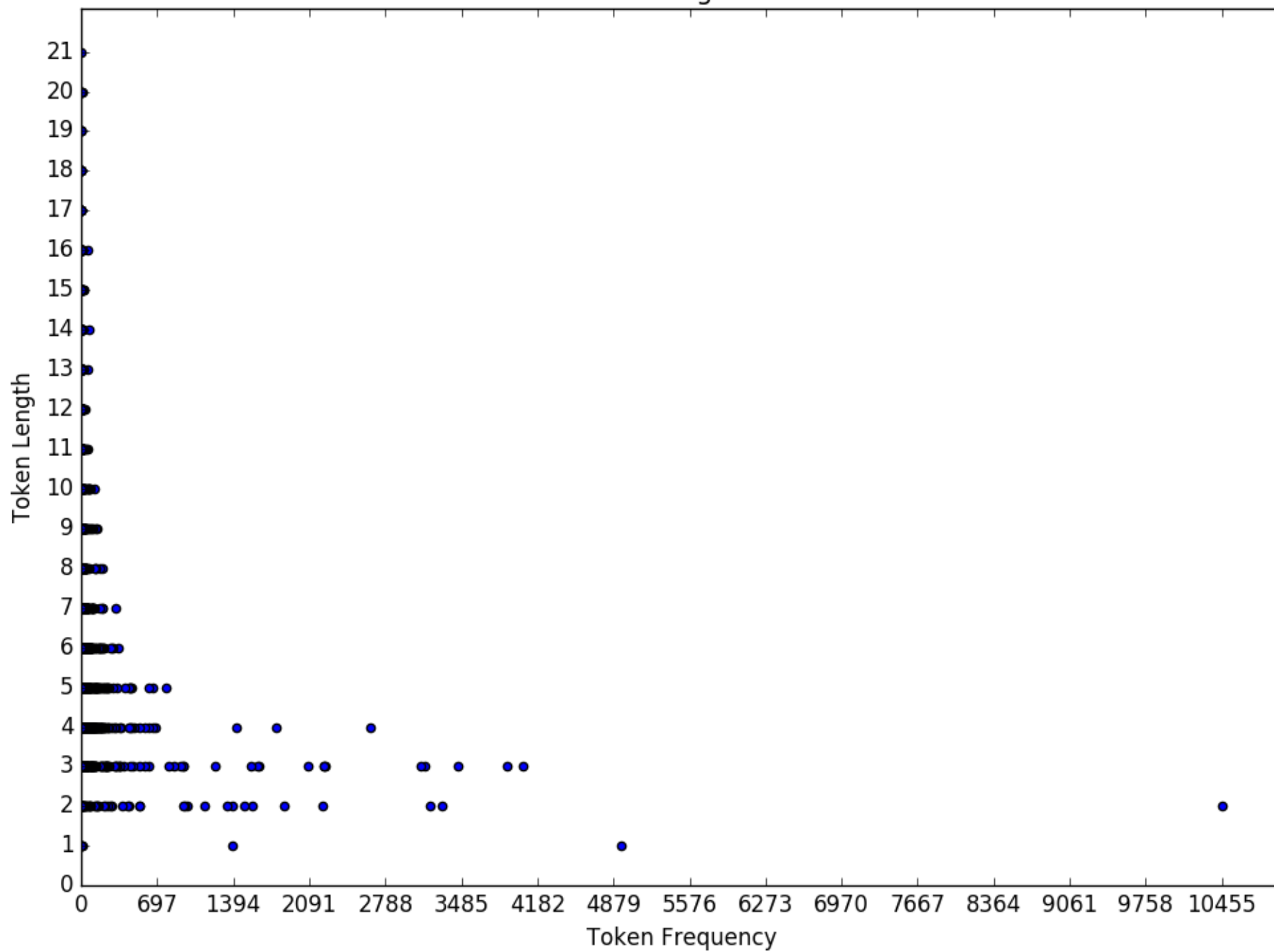
# Maori



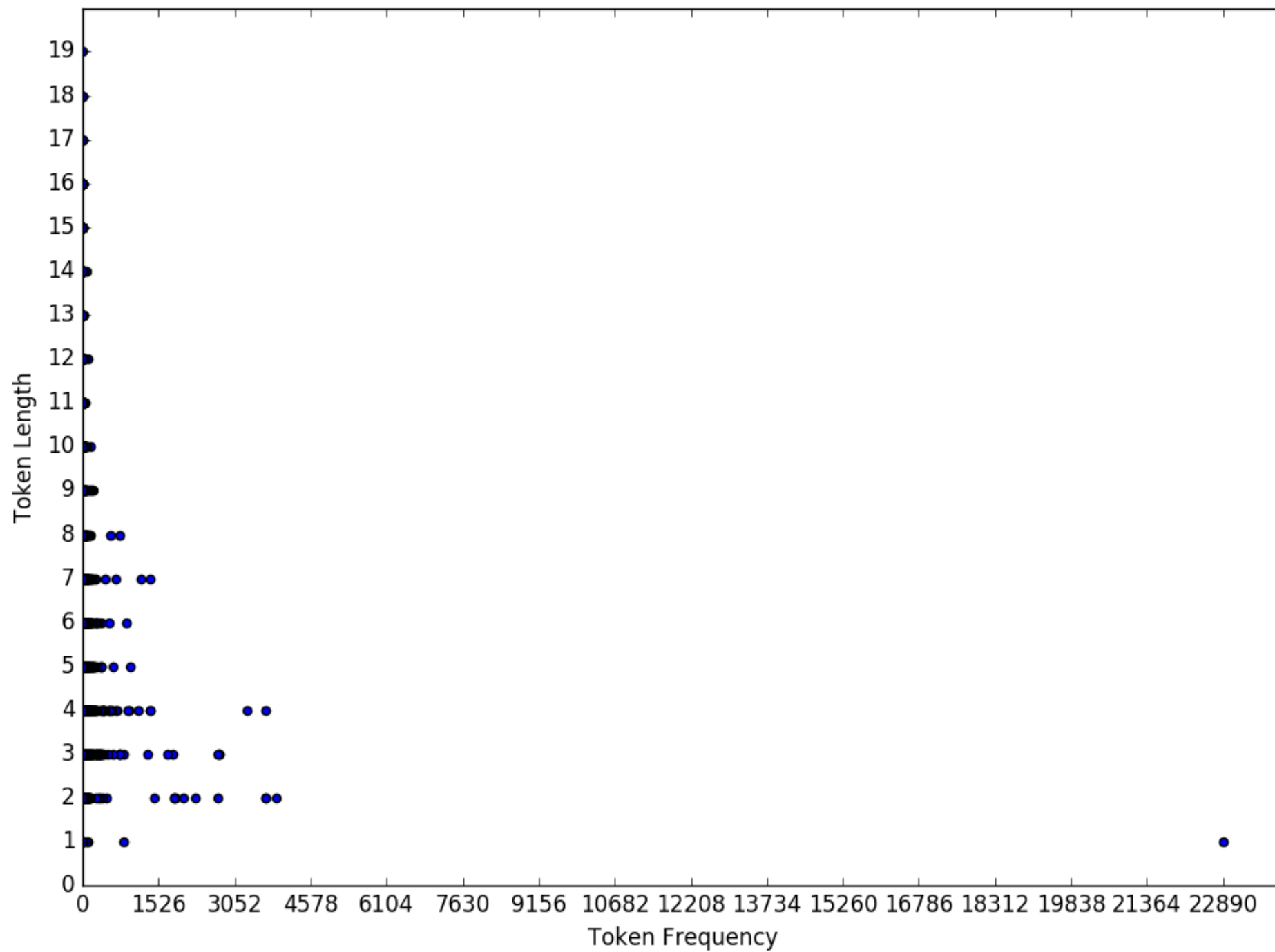
# Nahuatl



## Token Frequency

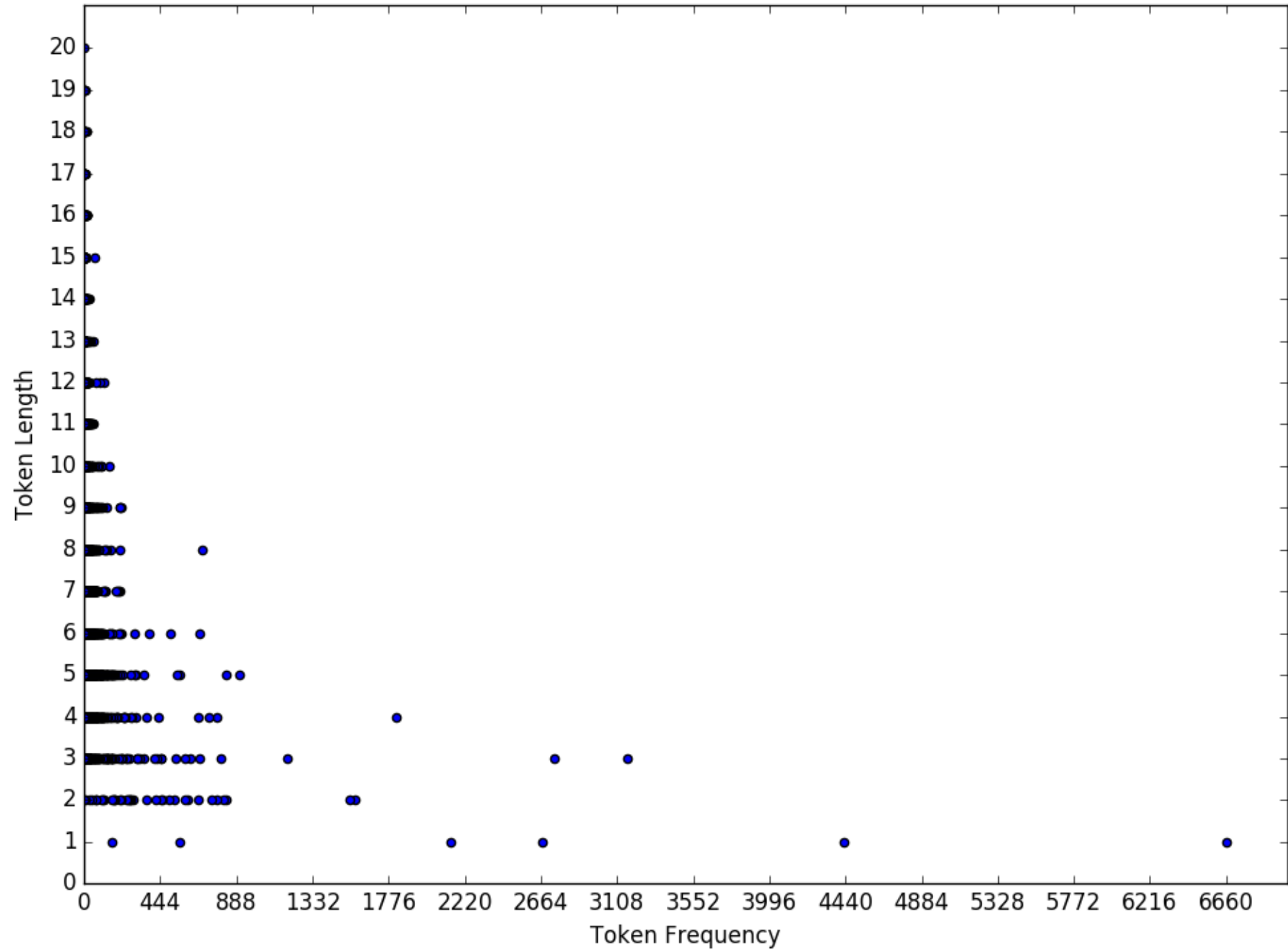


Paite

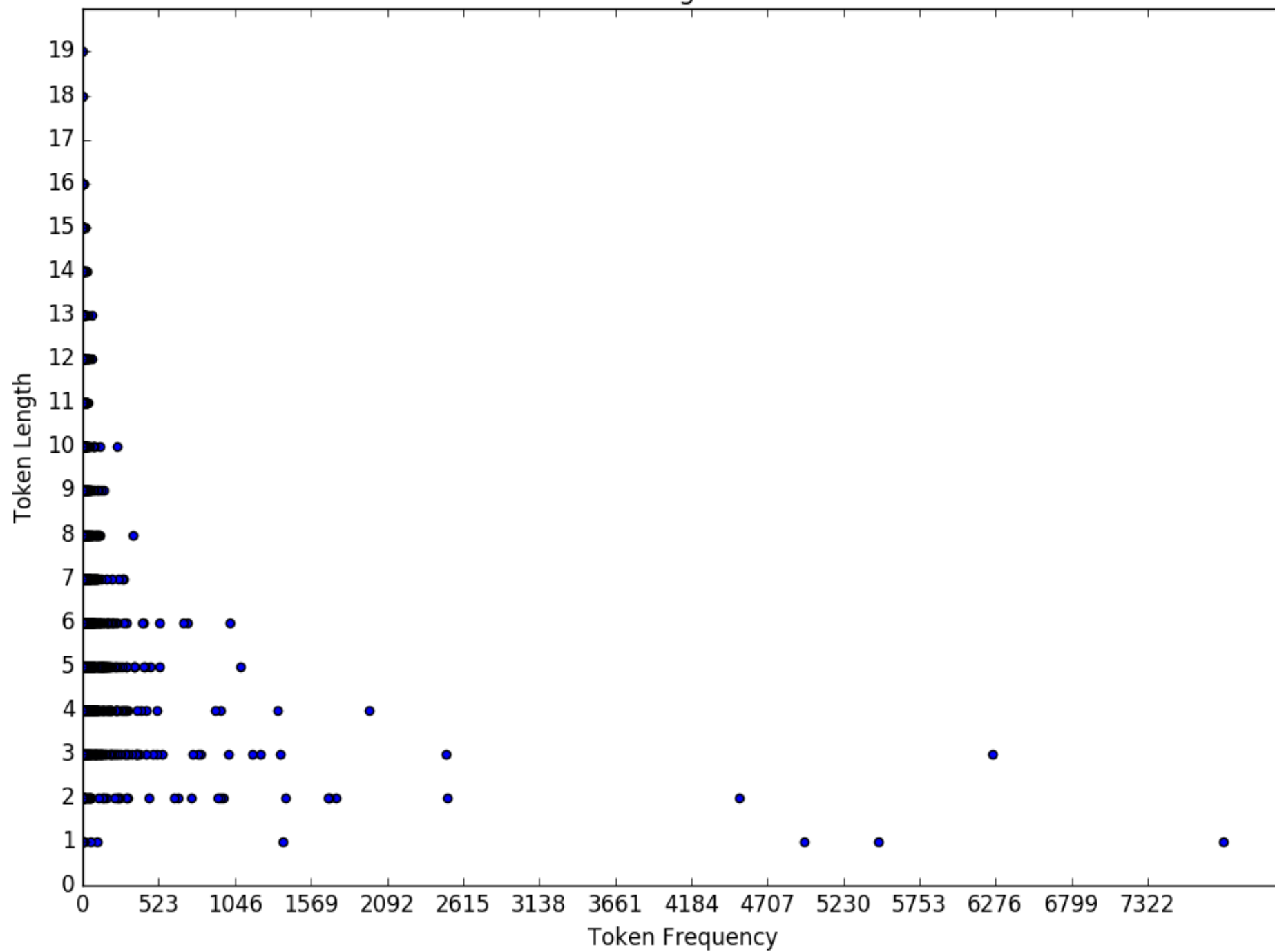




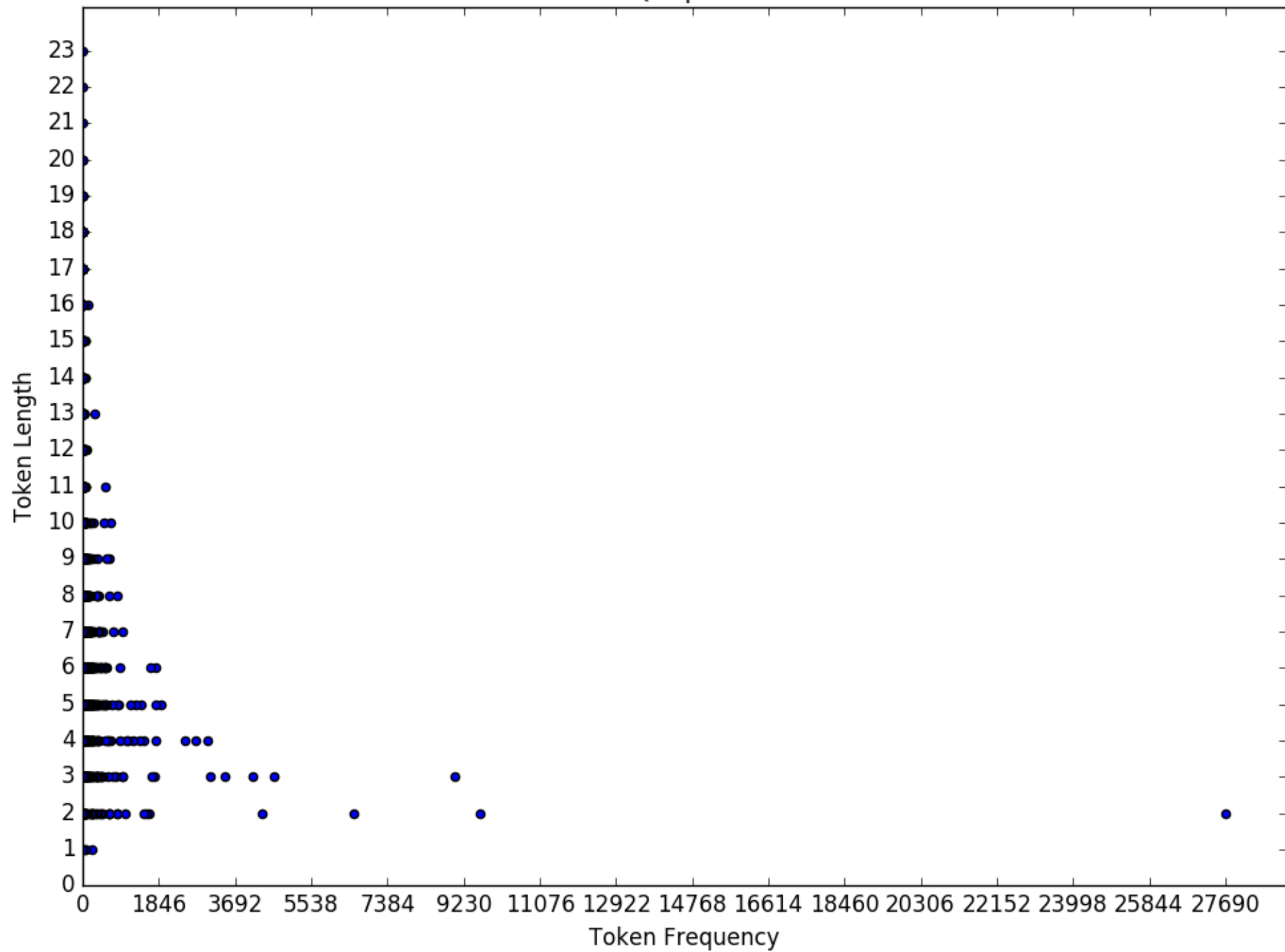
Polish



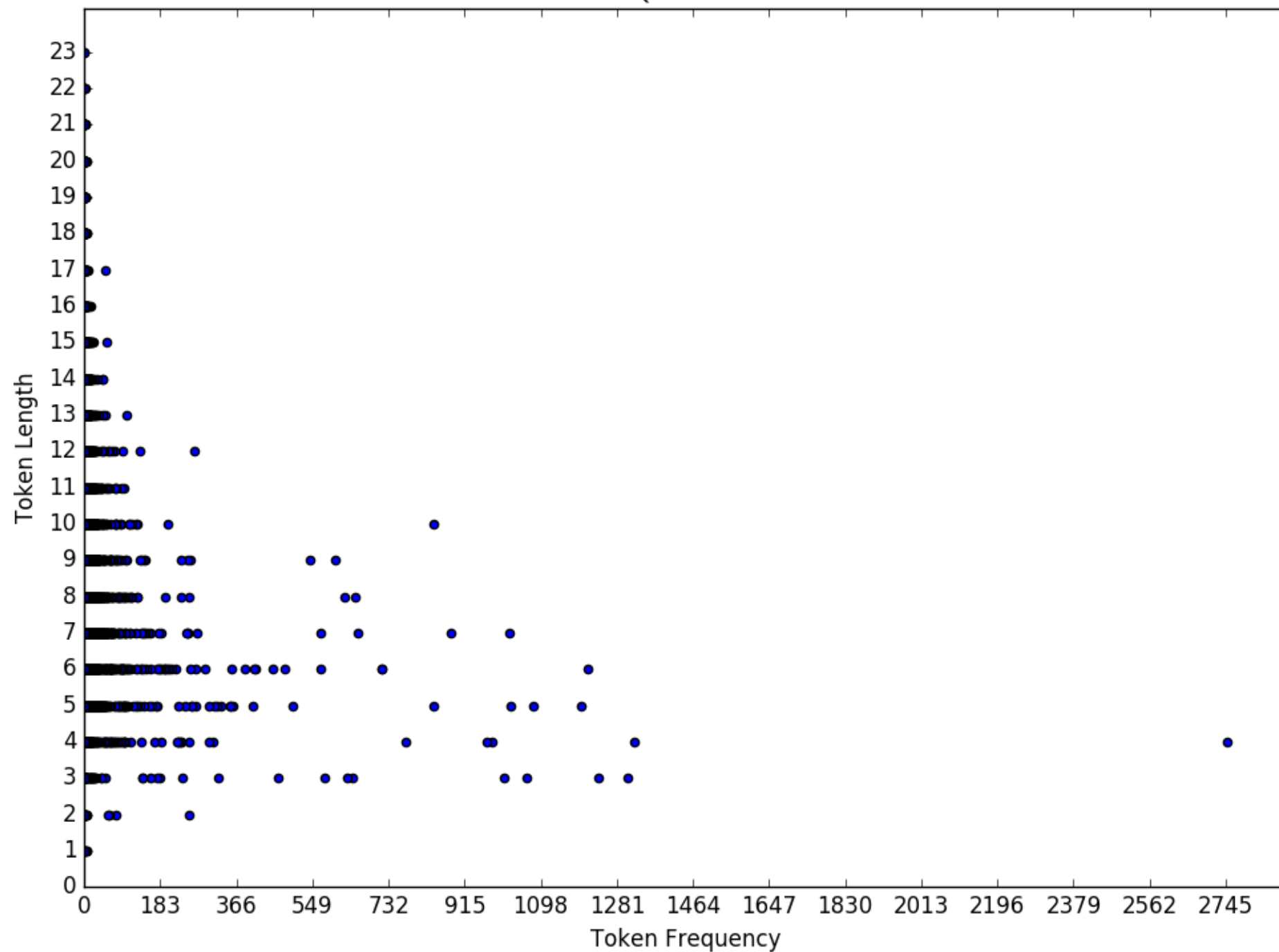
# Portuguese



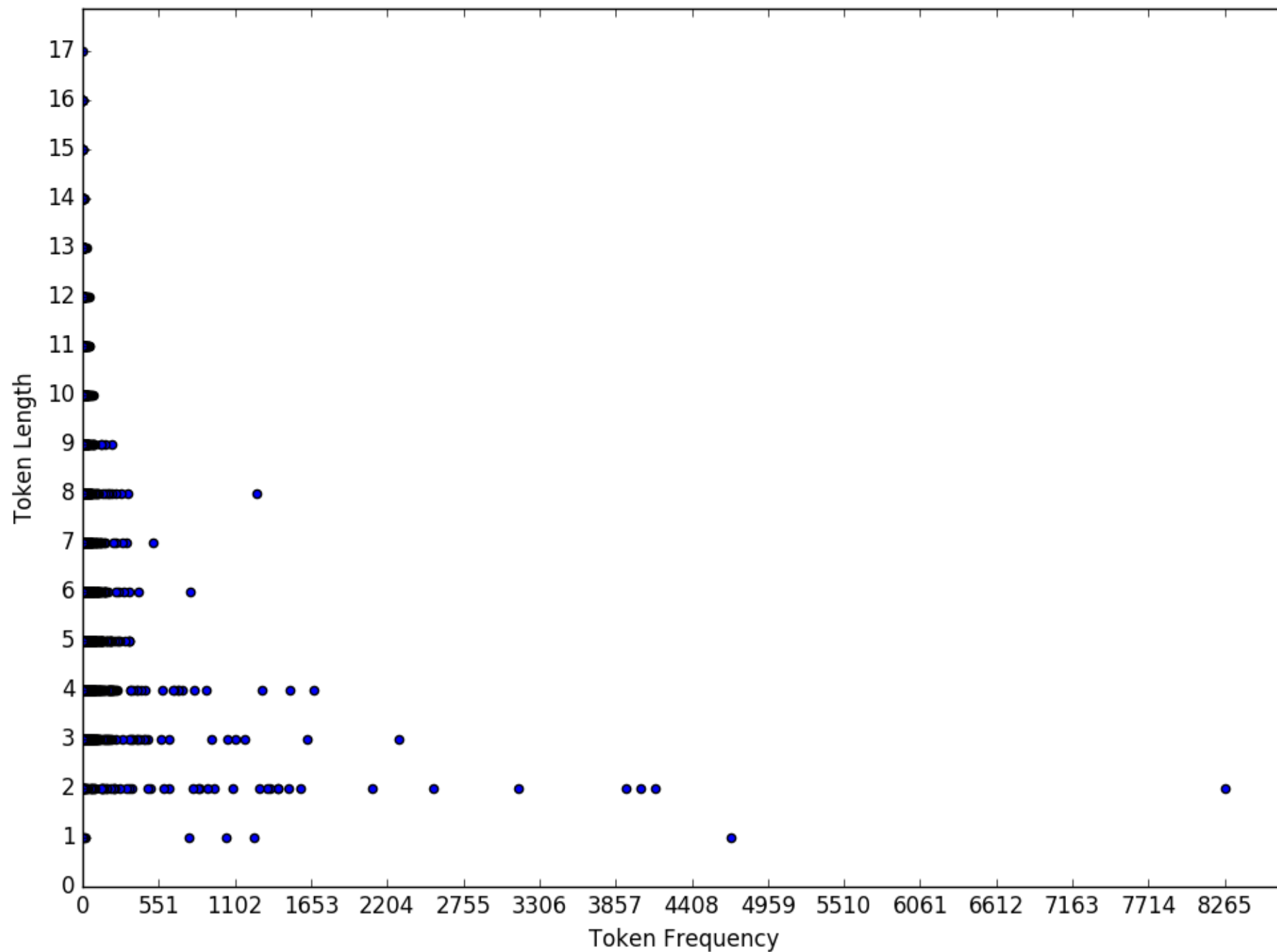
Q'eqchi'



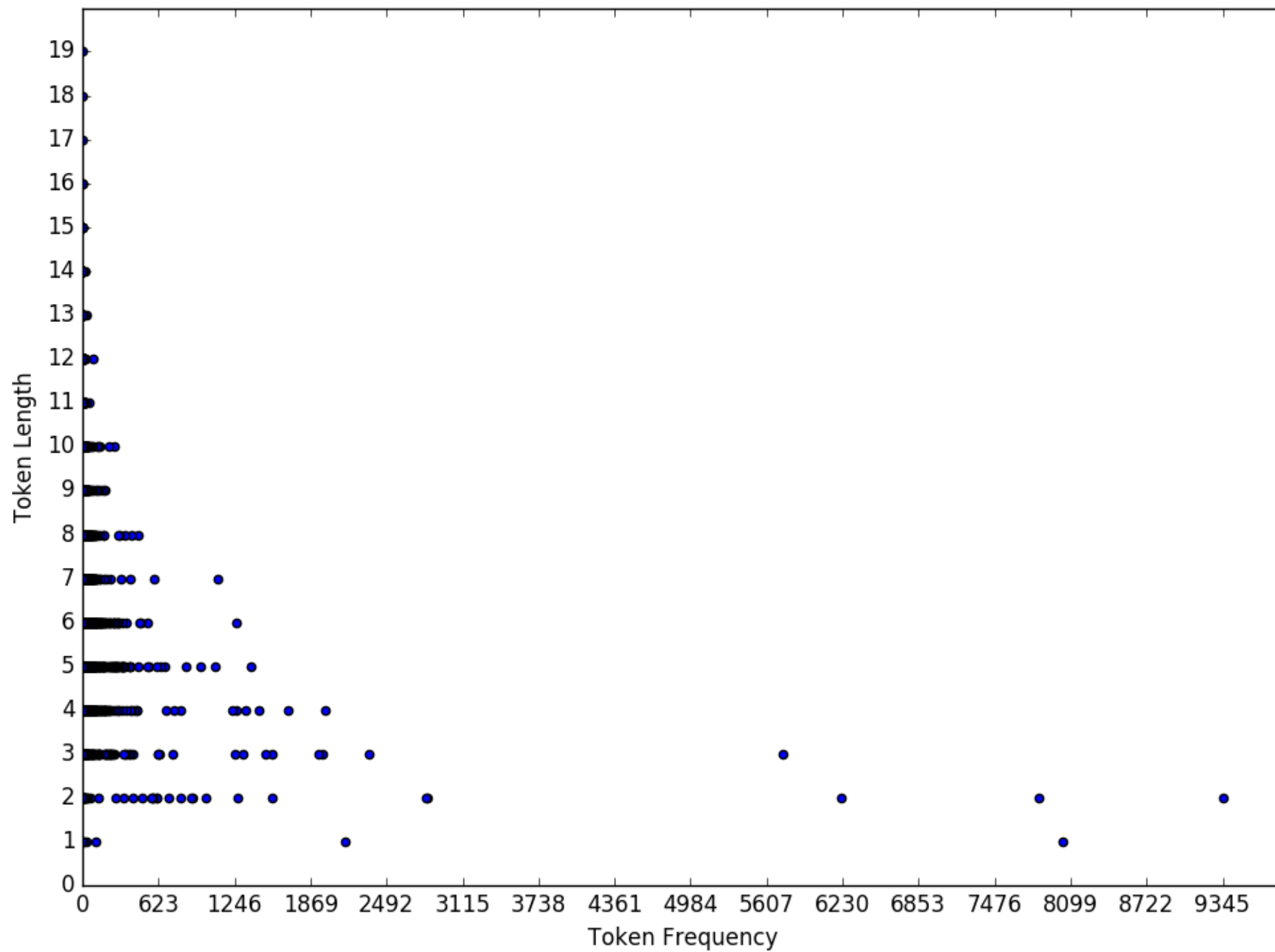
# Quichua



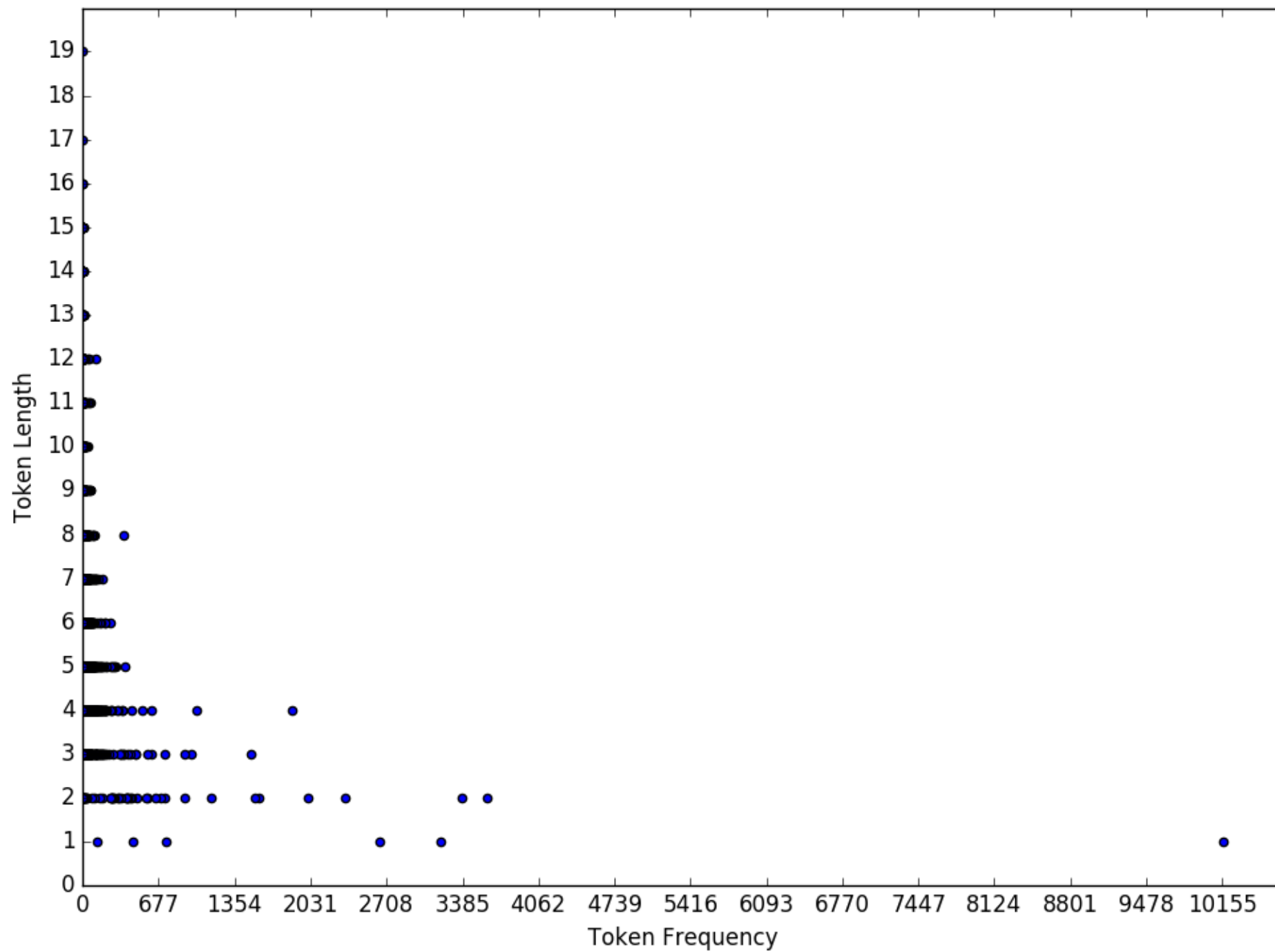
# Romanian



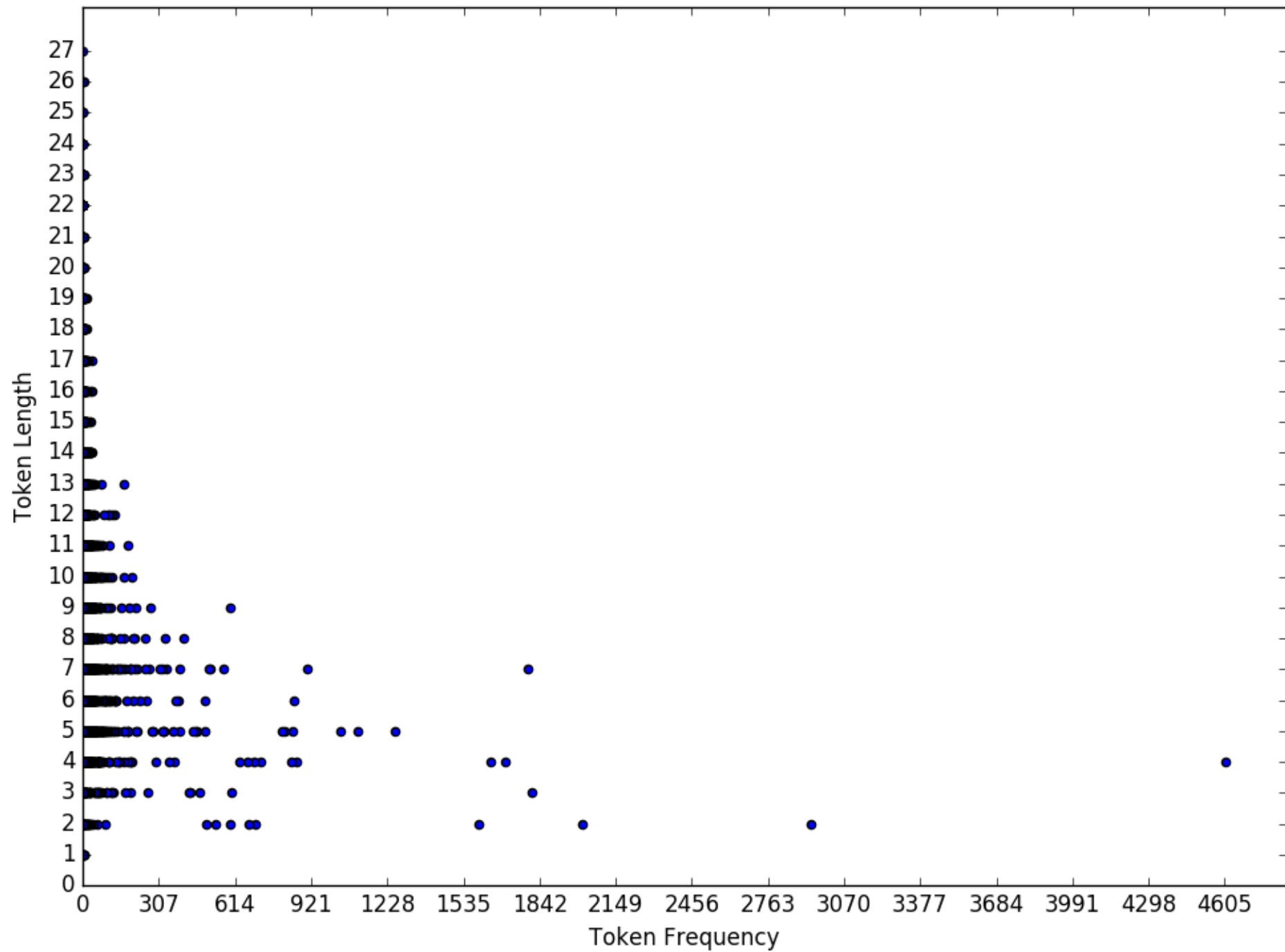
## Romani



Serbian

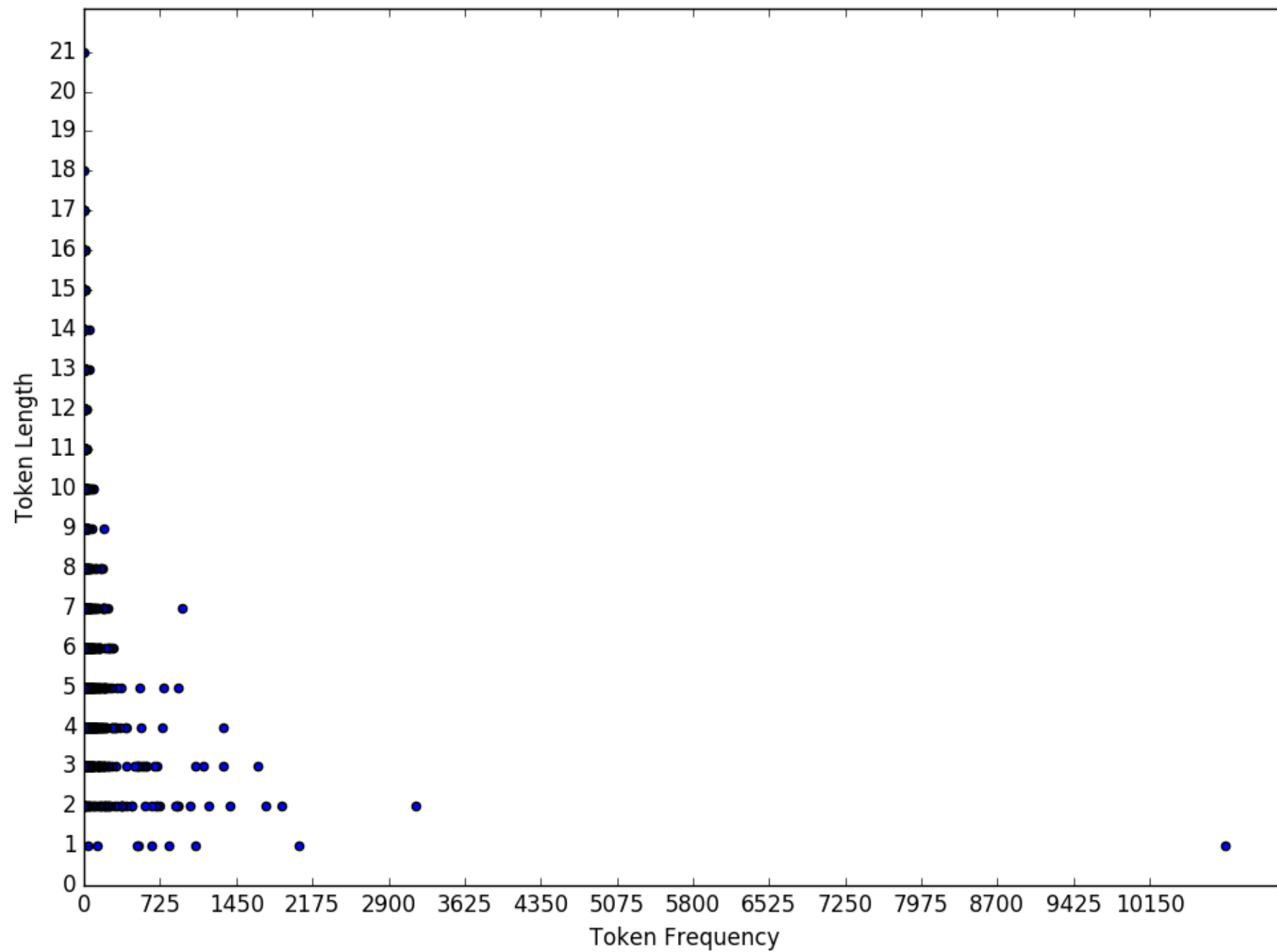


Shuar

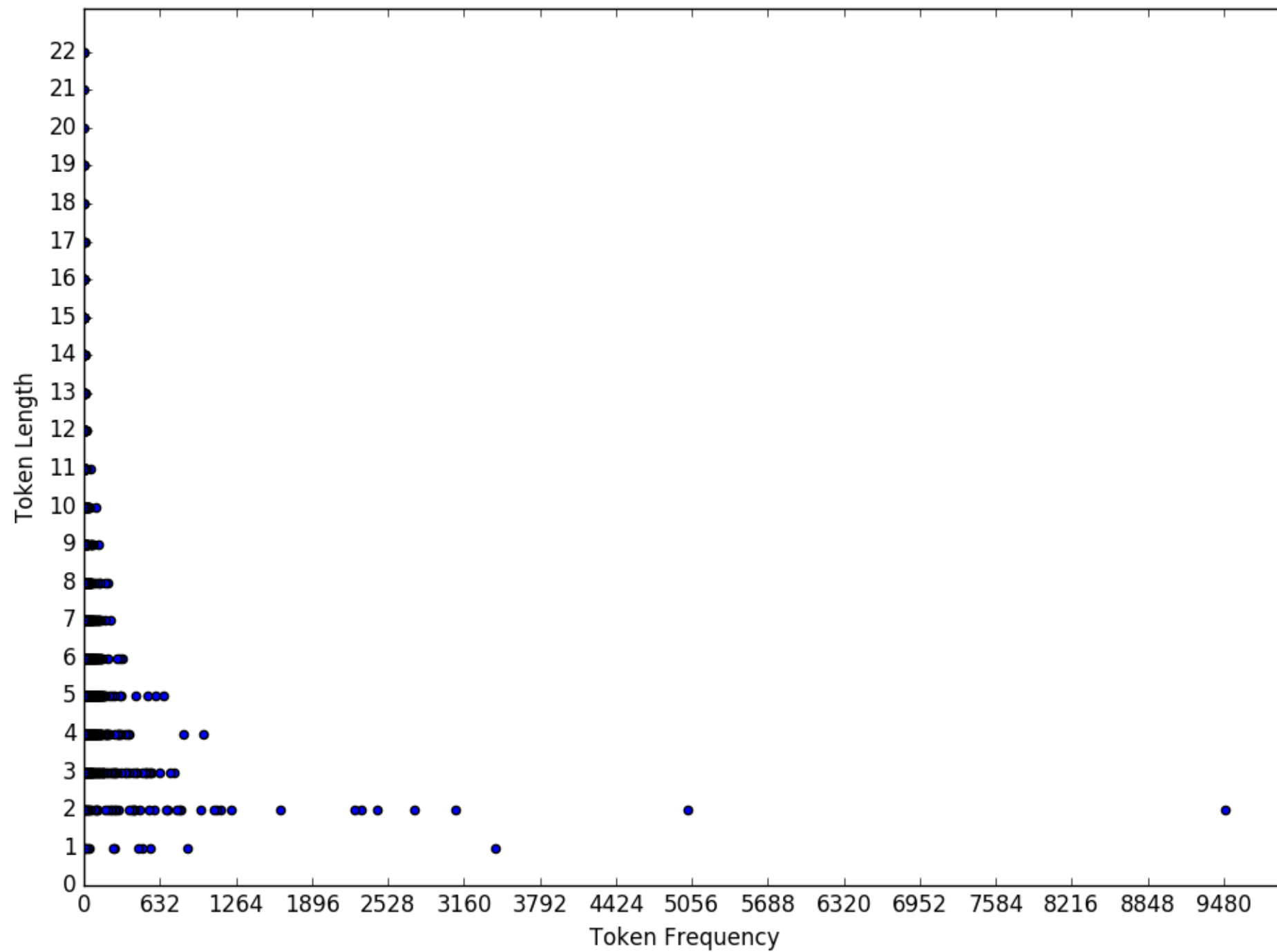




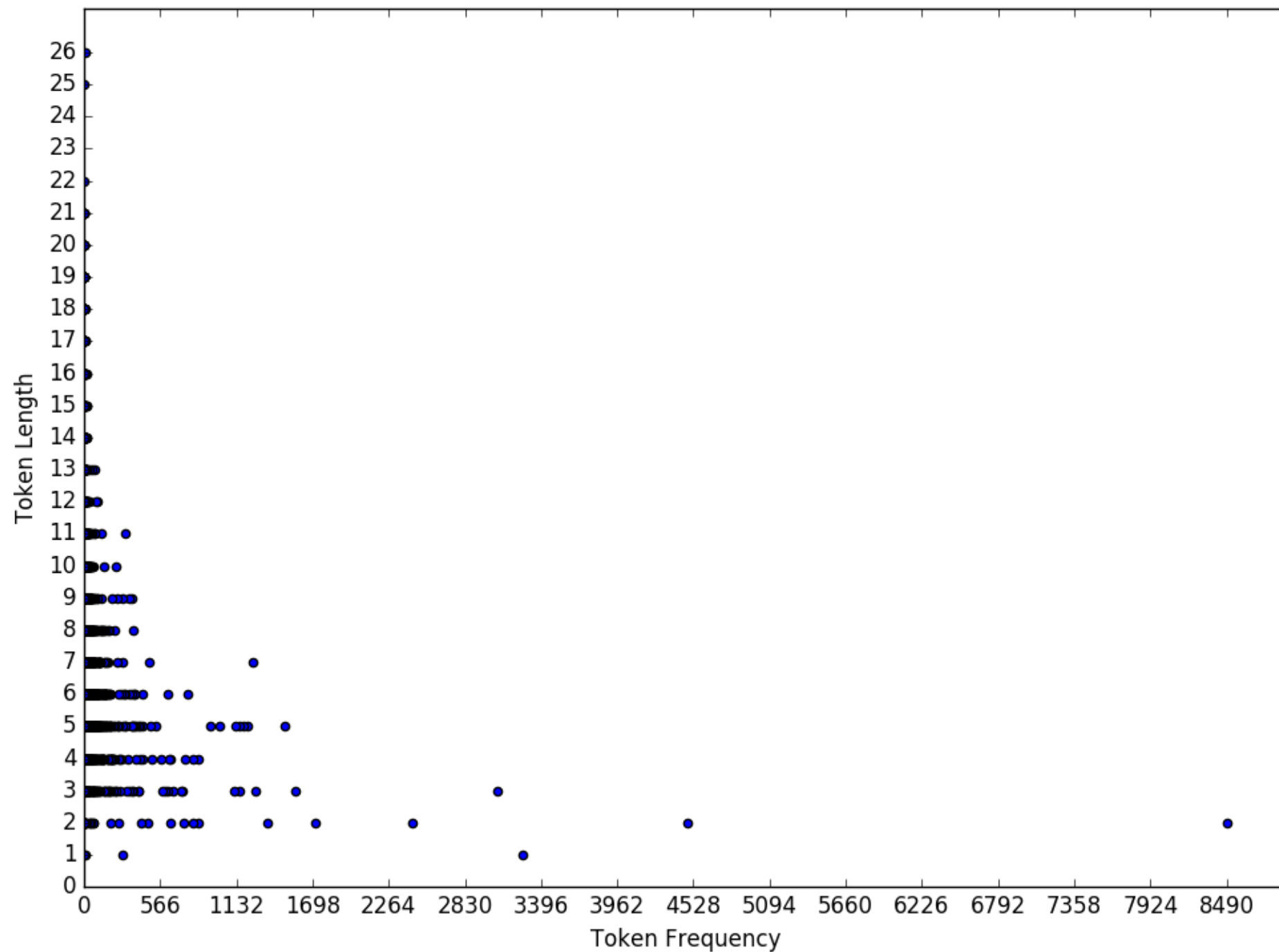
# Slovak



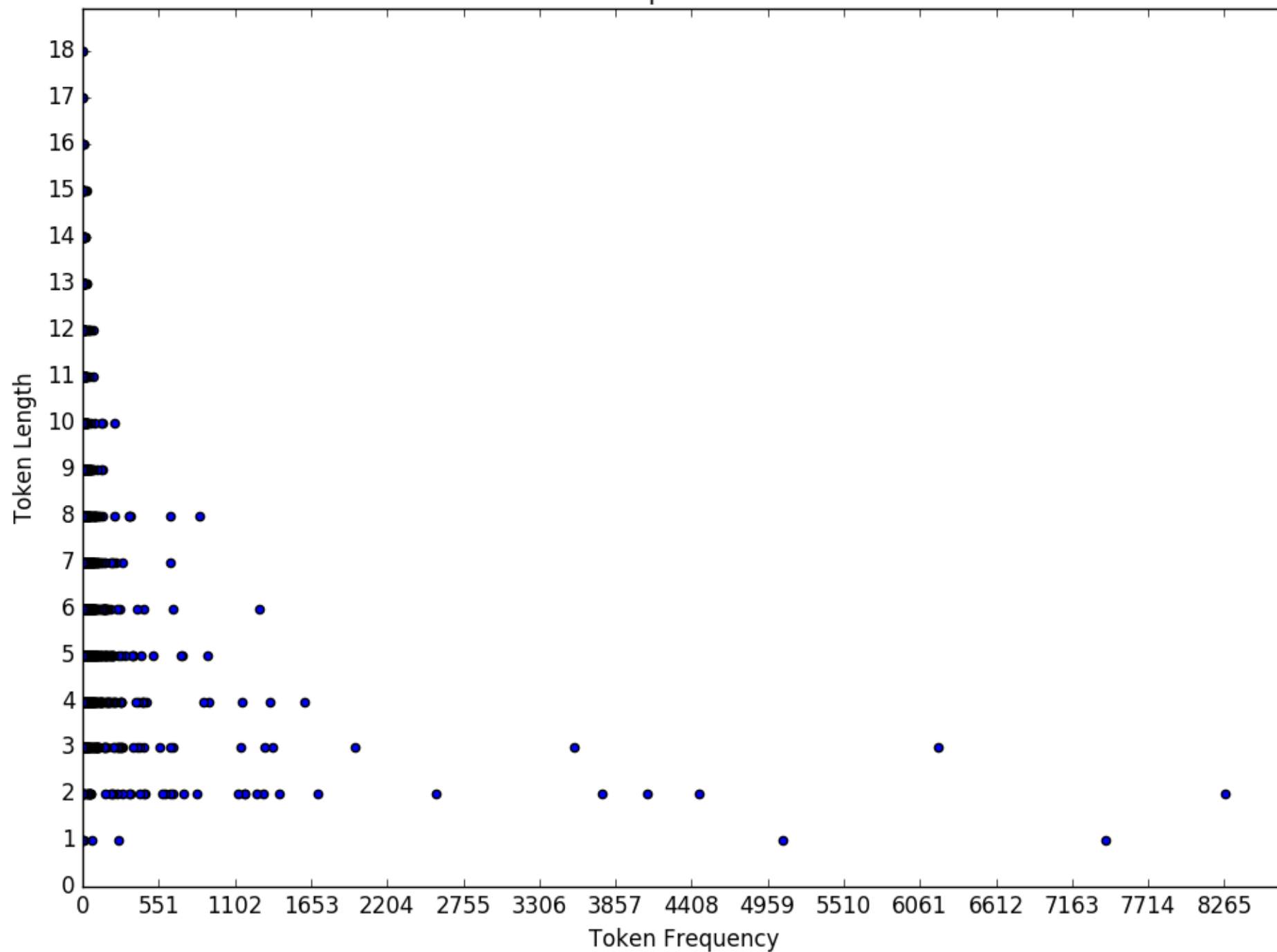
Slovene



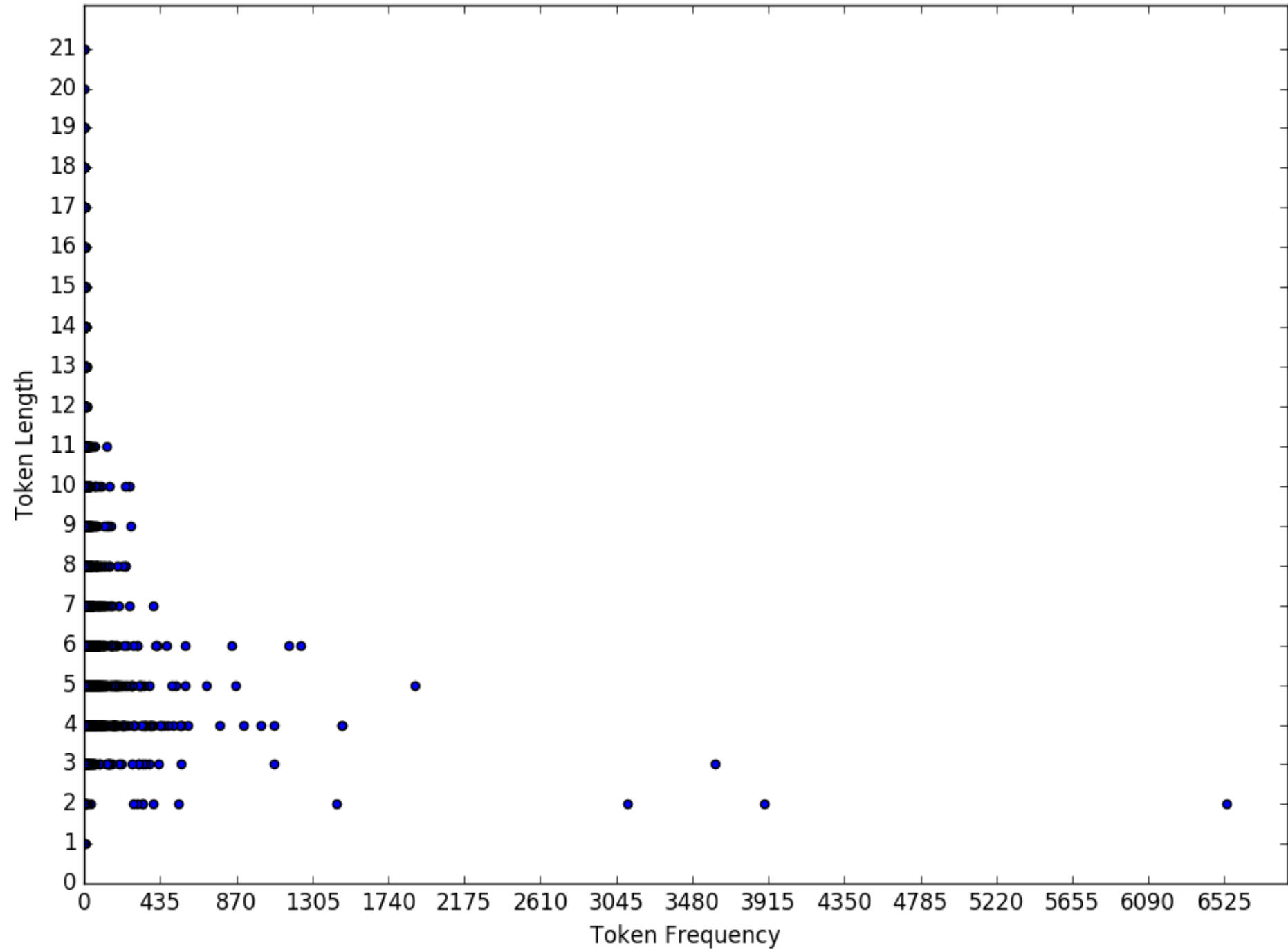
Somali



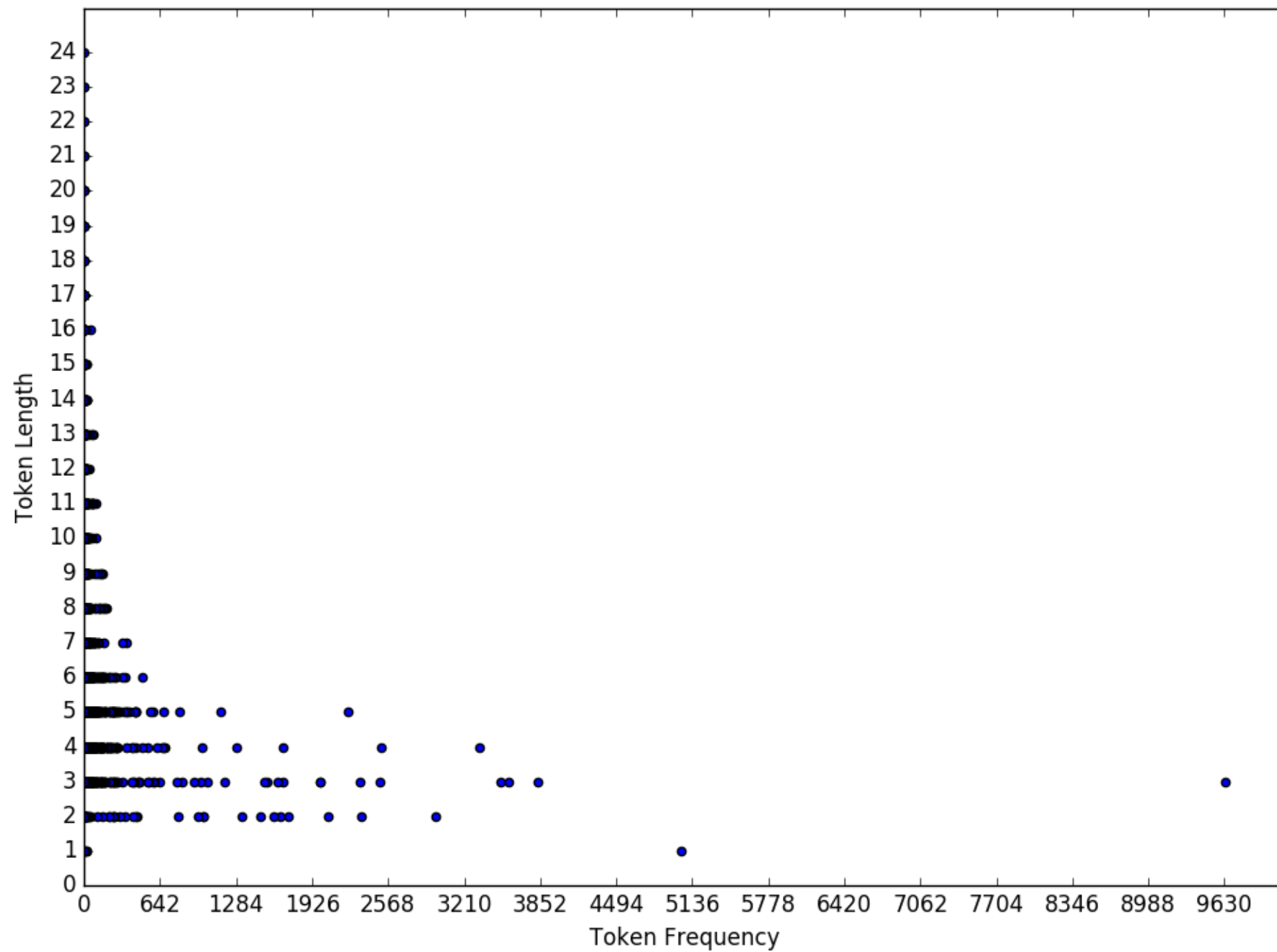
Spanish



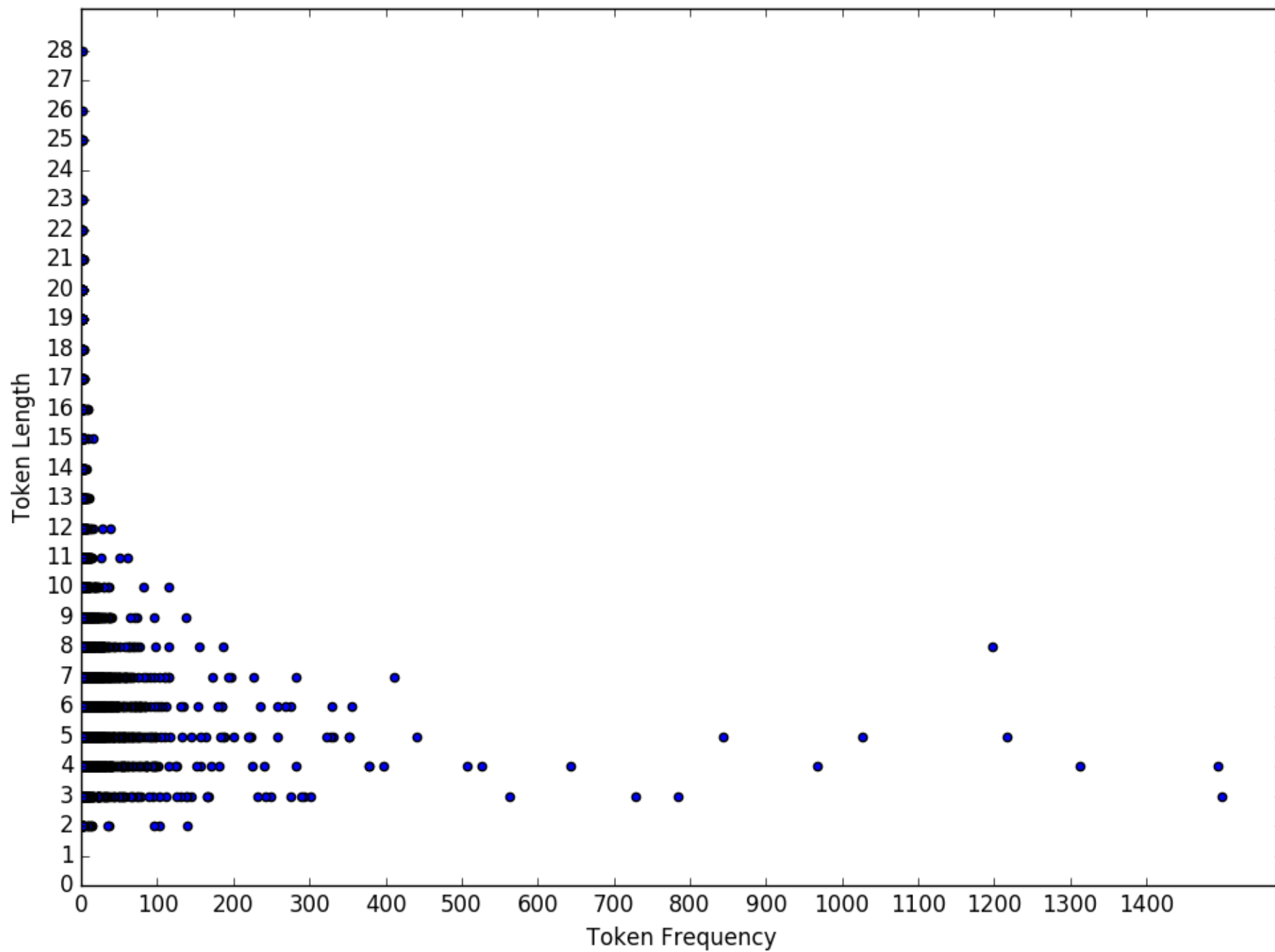
Swahili



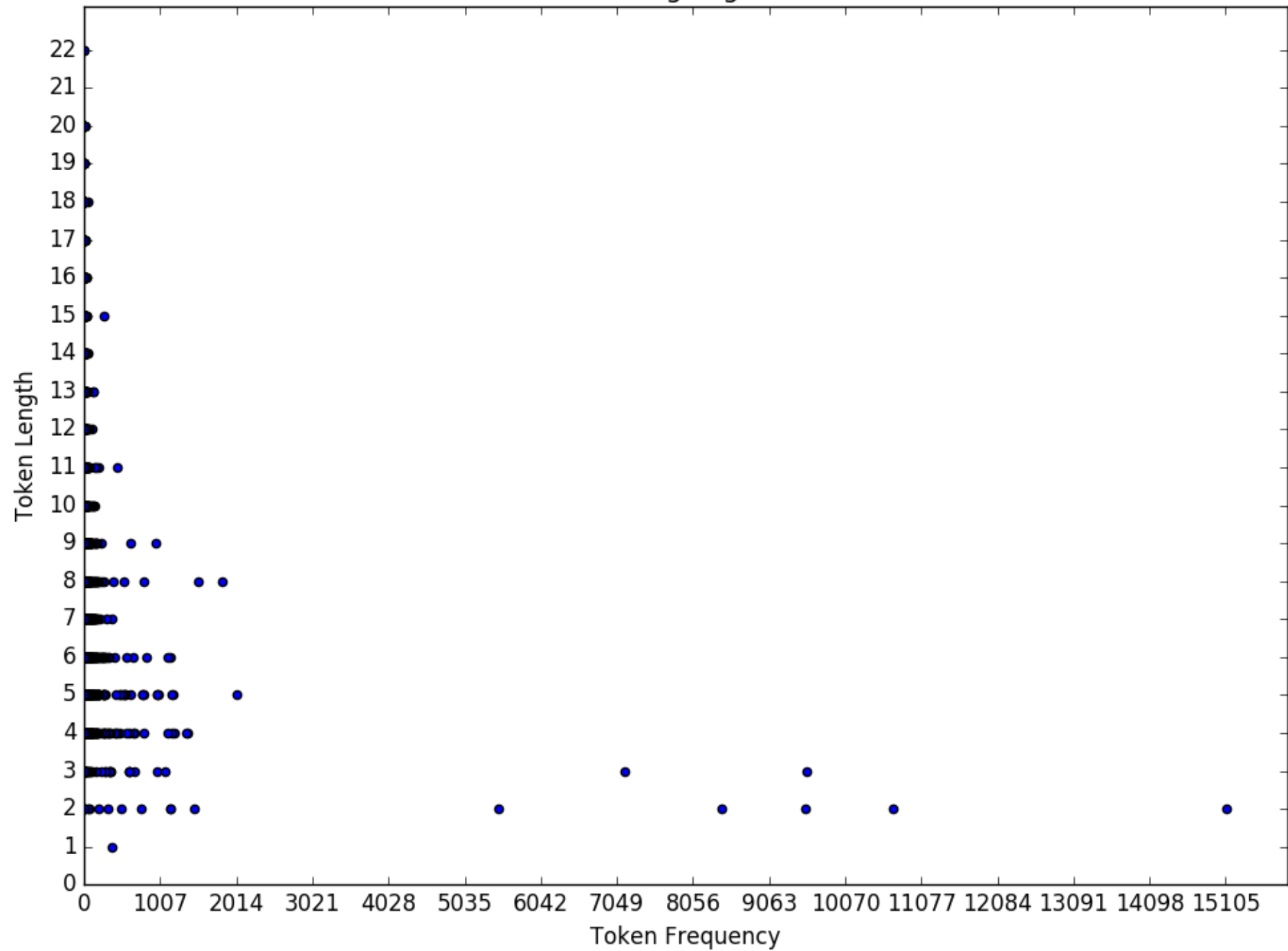
Swedish



## Token Frequency

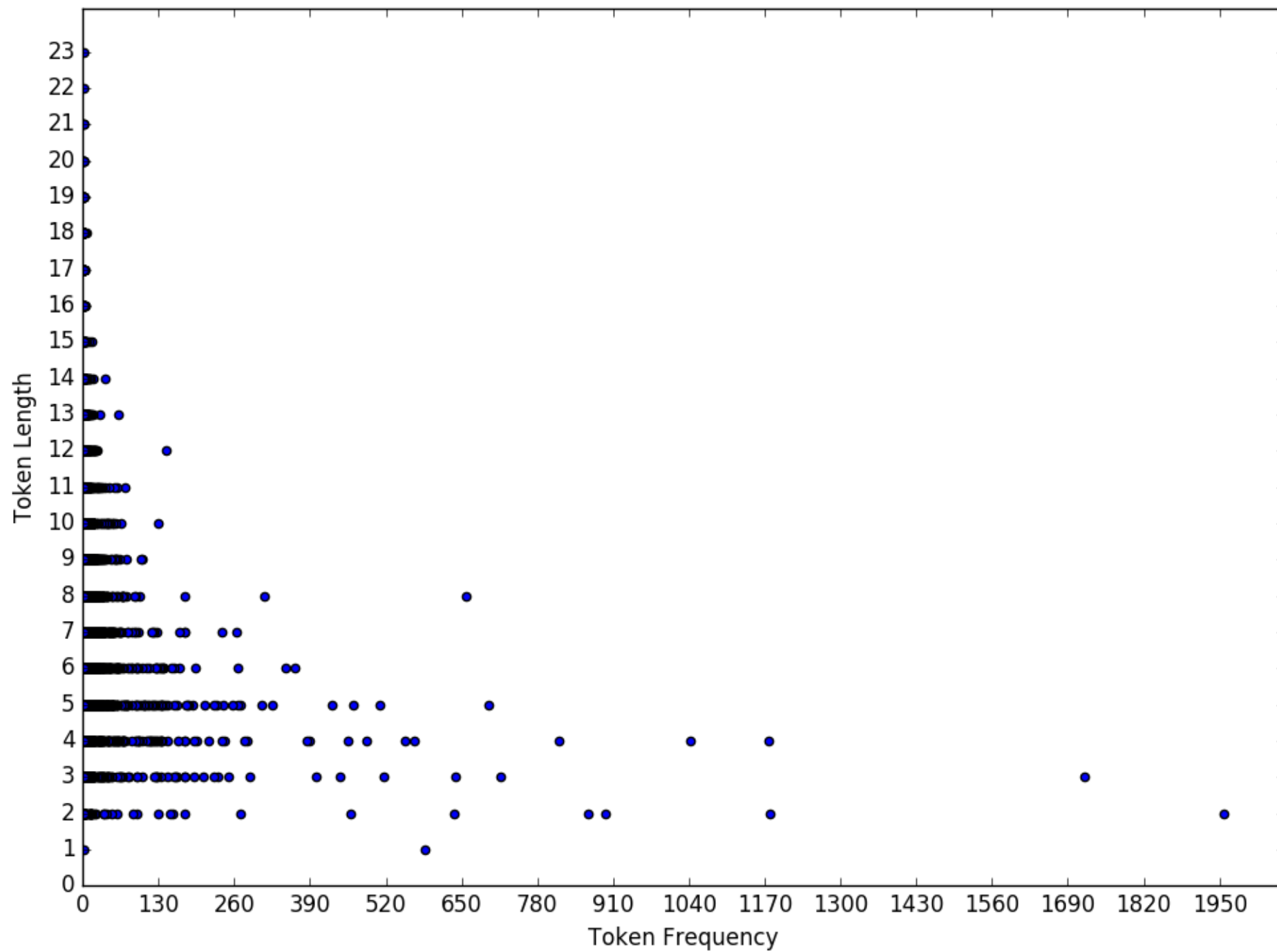


Tagalog

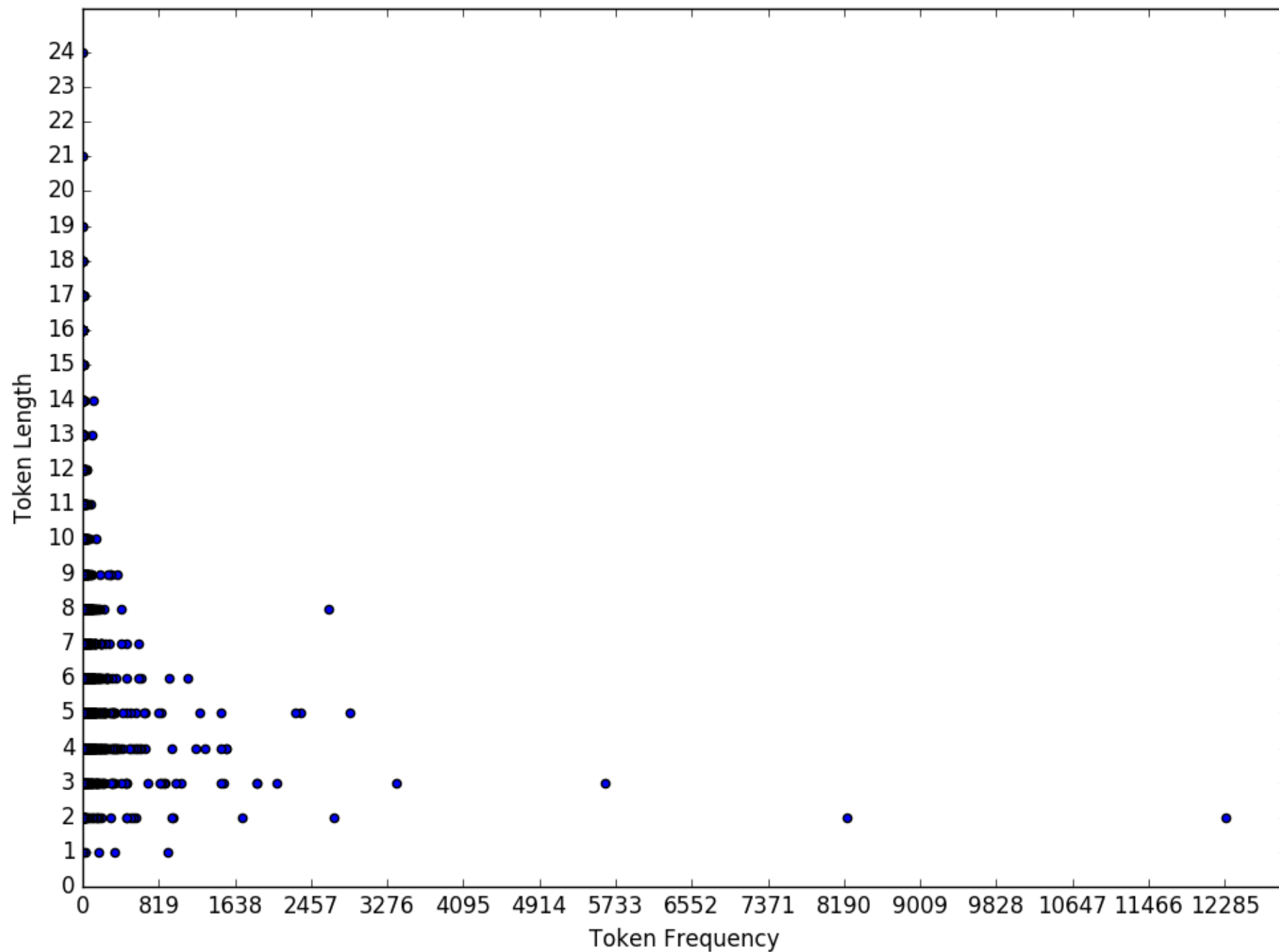




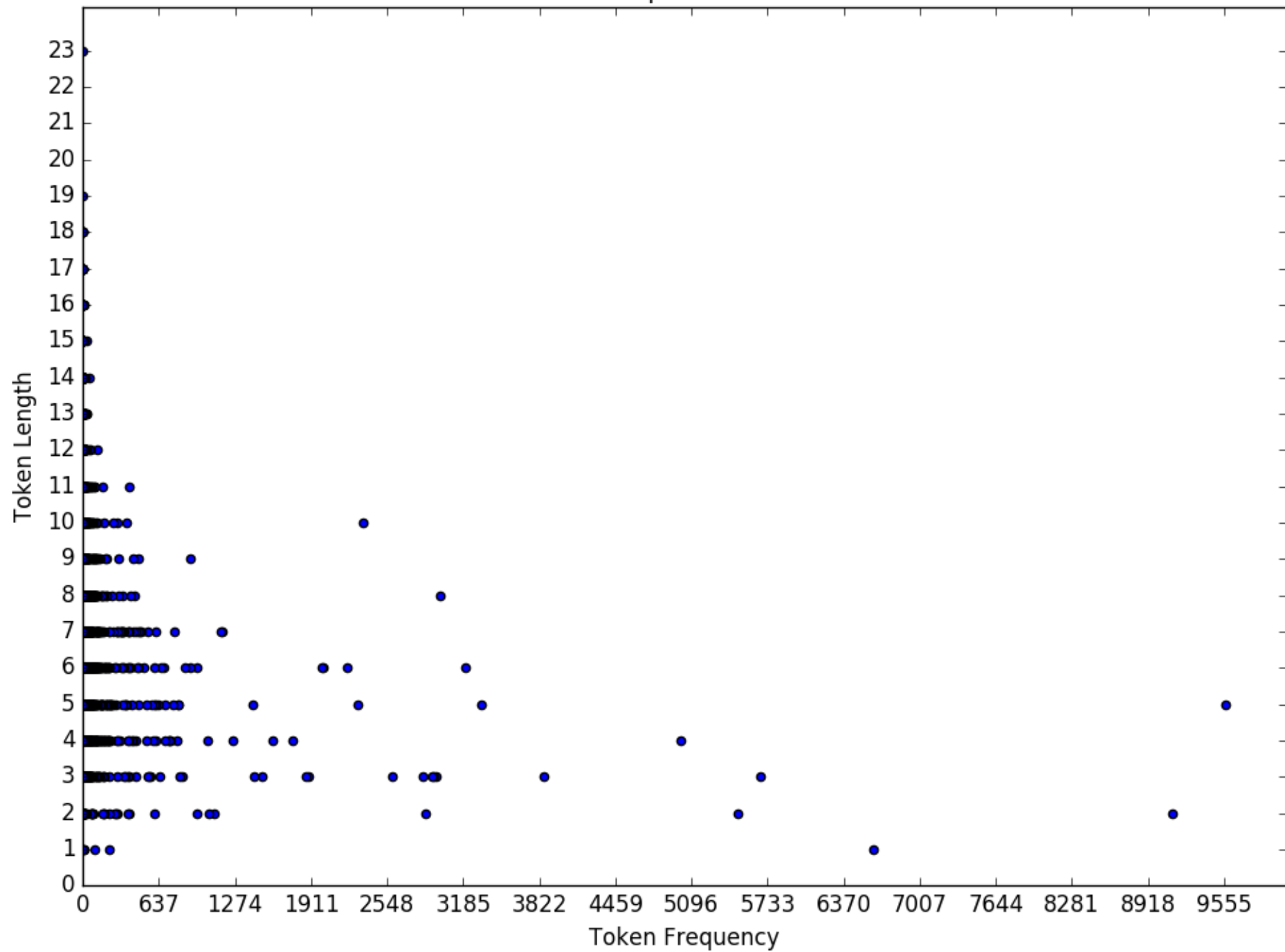
# Turkish



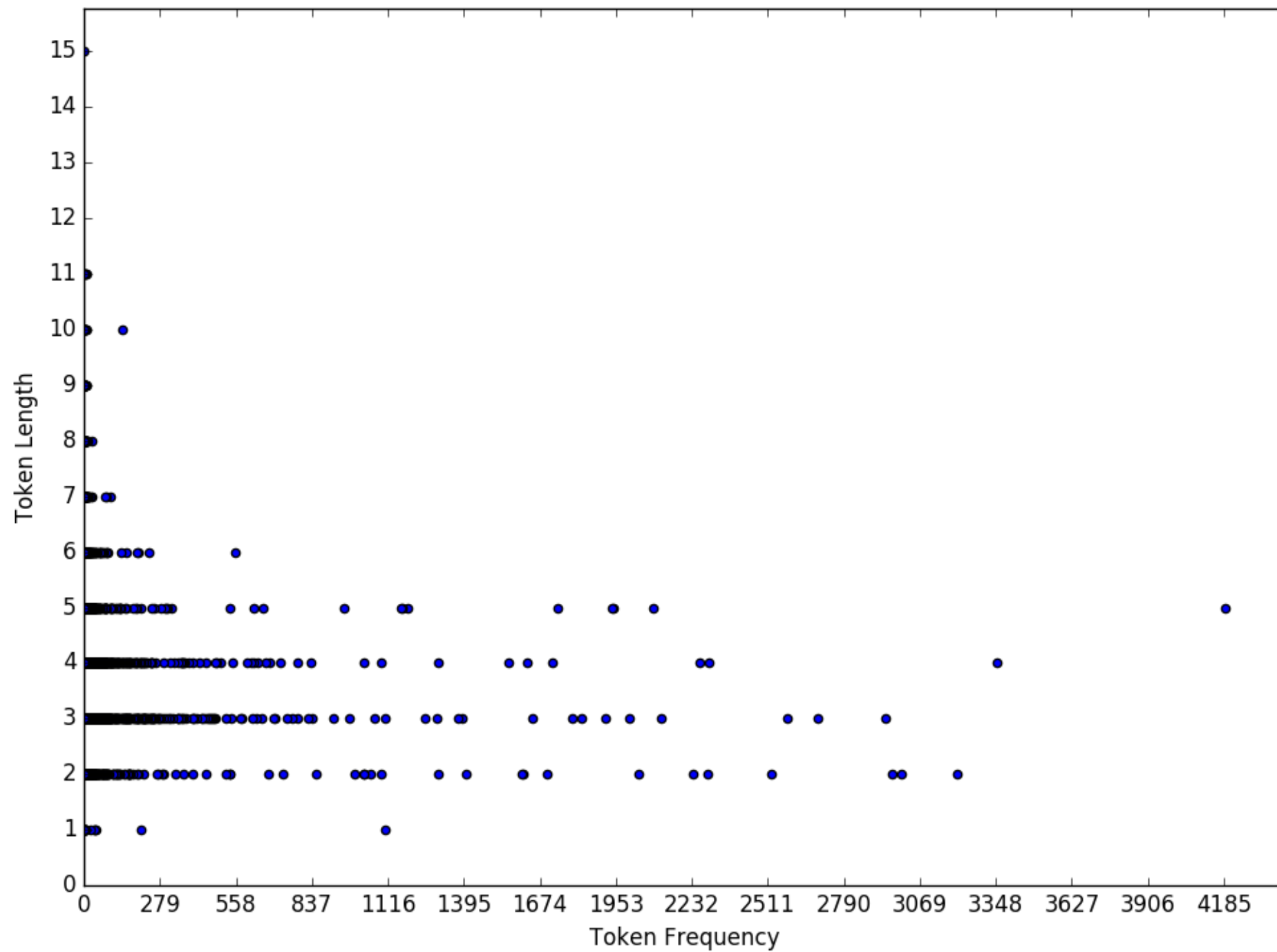
# Uma



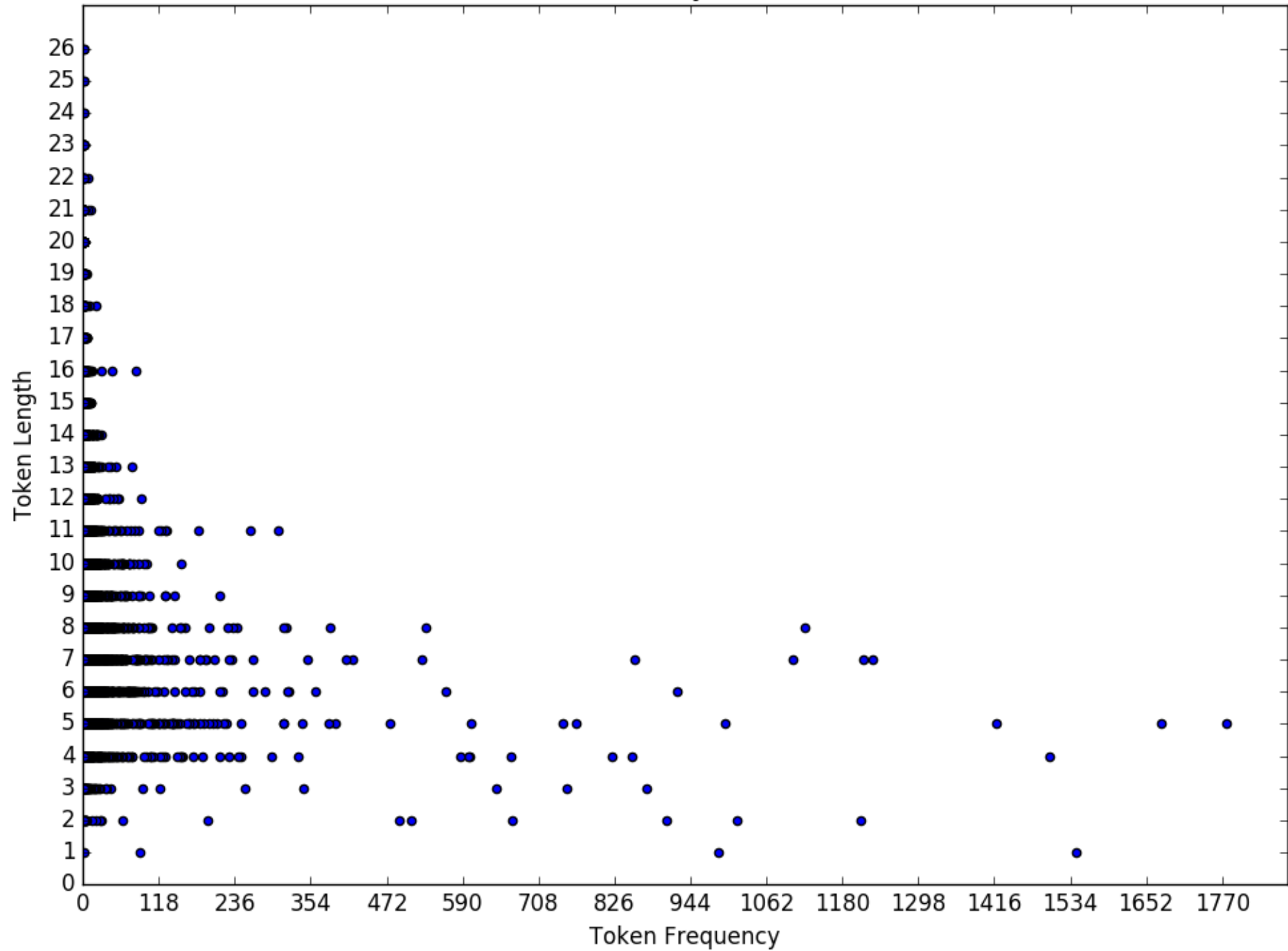
Uspanteco



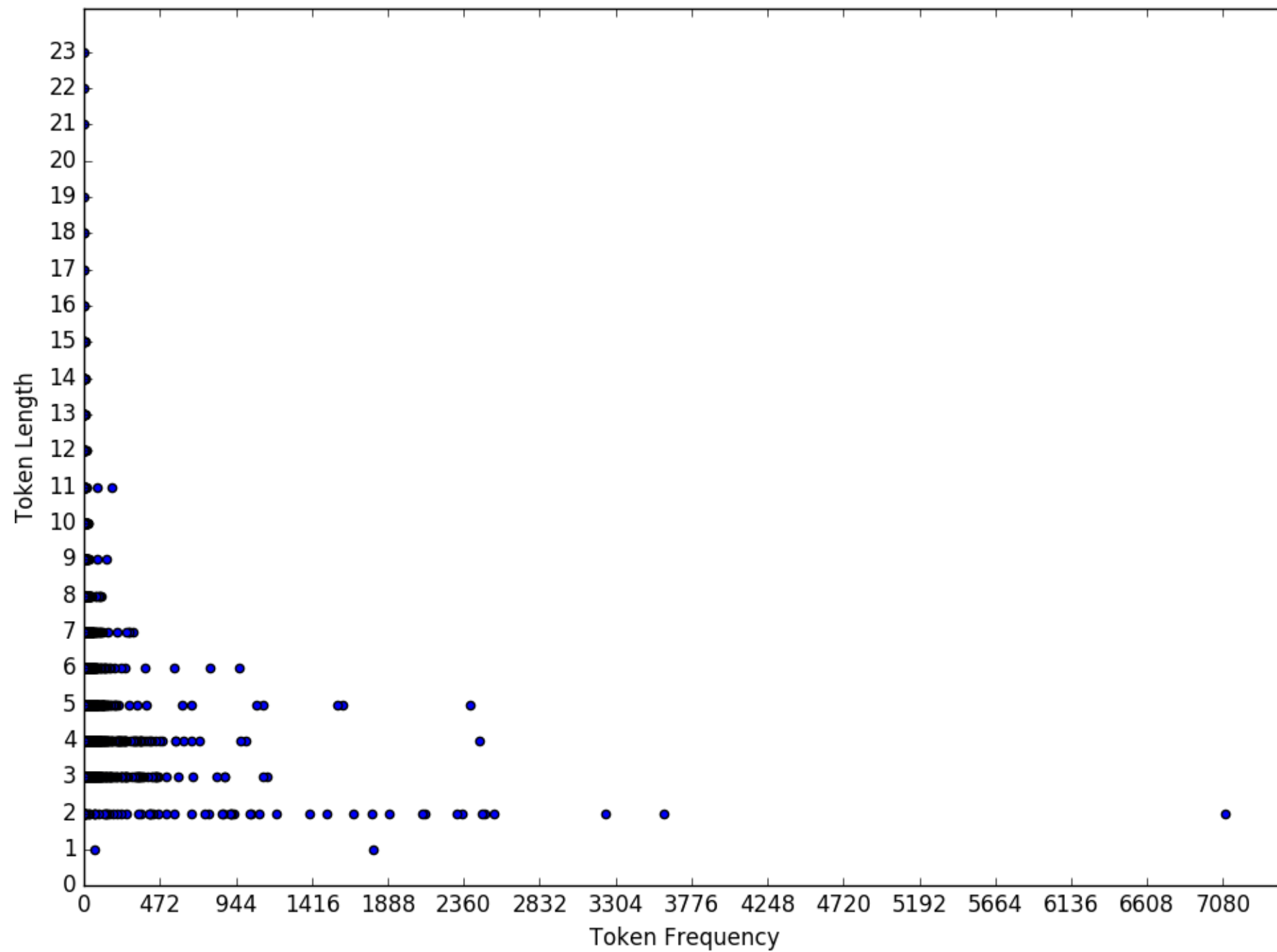
Vietnamese



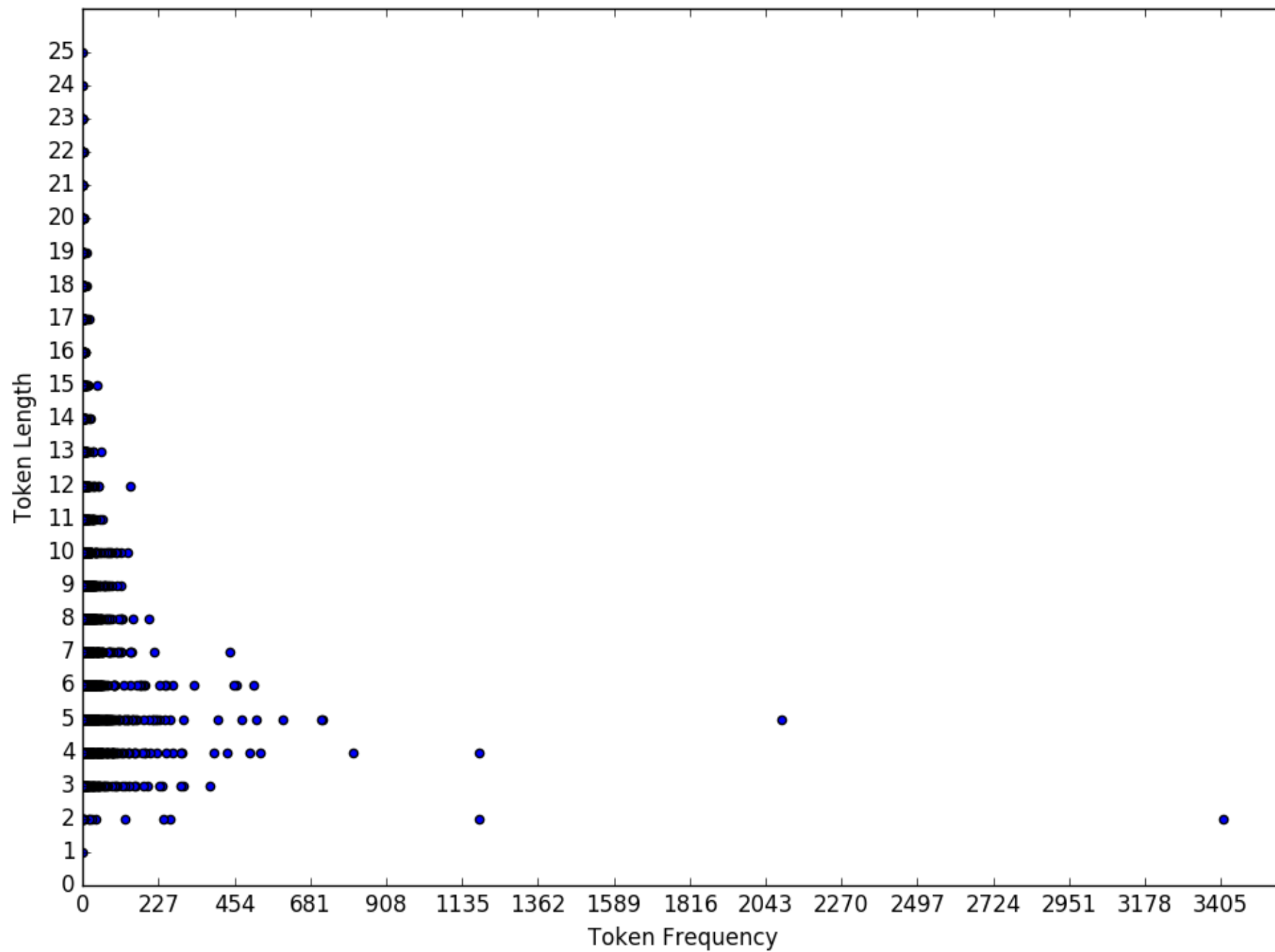
# Wolaytta



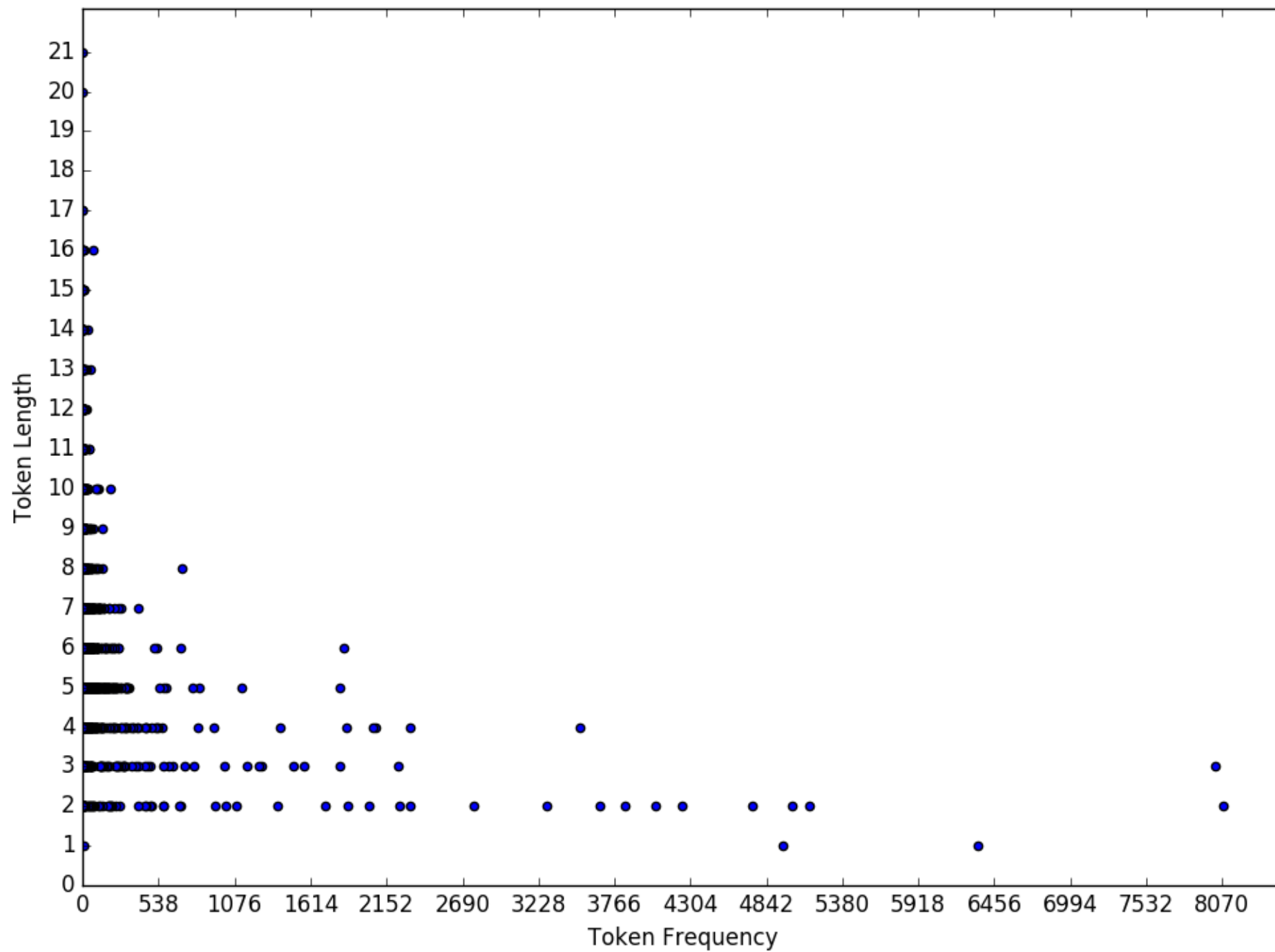
Wolof



# Xhosa



Zarma





Zulu

