## Percentage of appearance of each token in the training data

