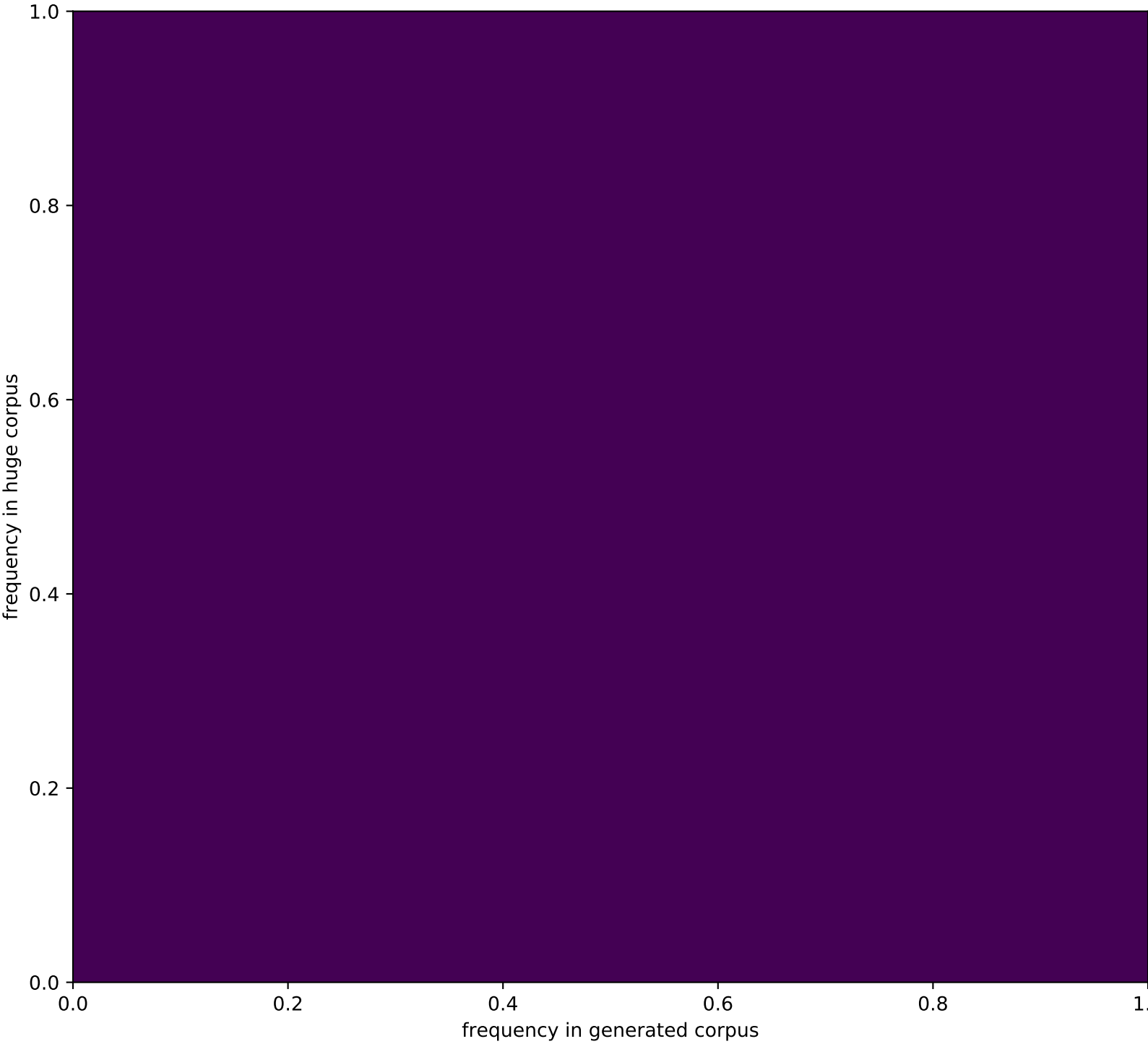


Generated words with frequency in training and huge corpus



percentage in train nan %  
percentage in huge nan %  
Percentage of words not present in training corpus that are present in the bigger corpus: nan %