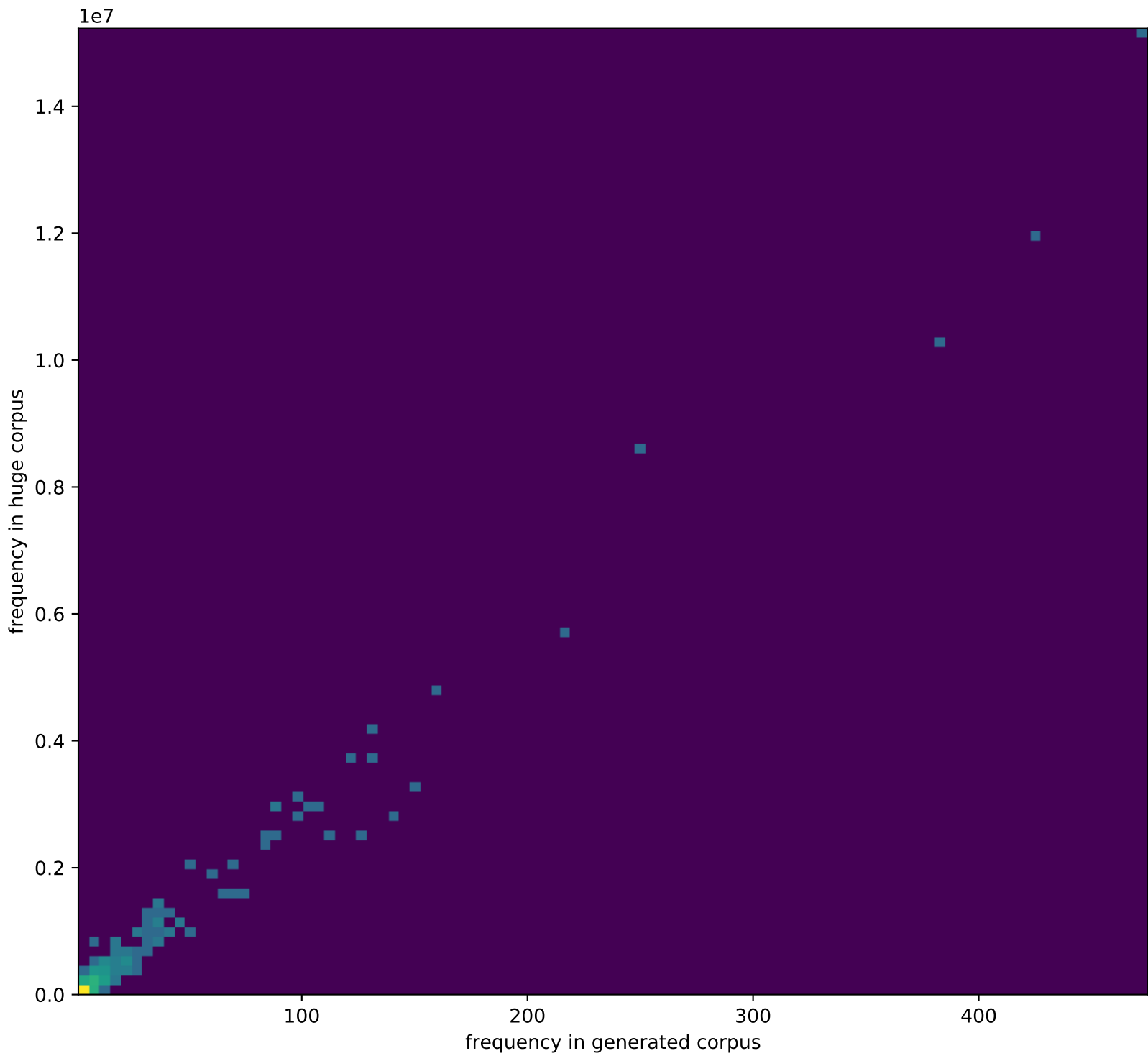
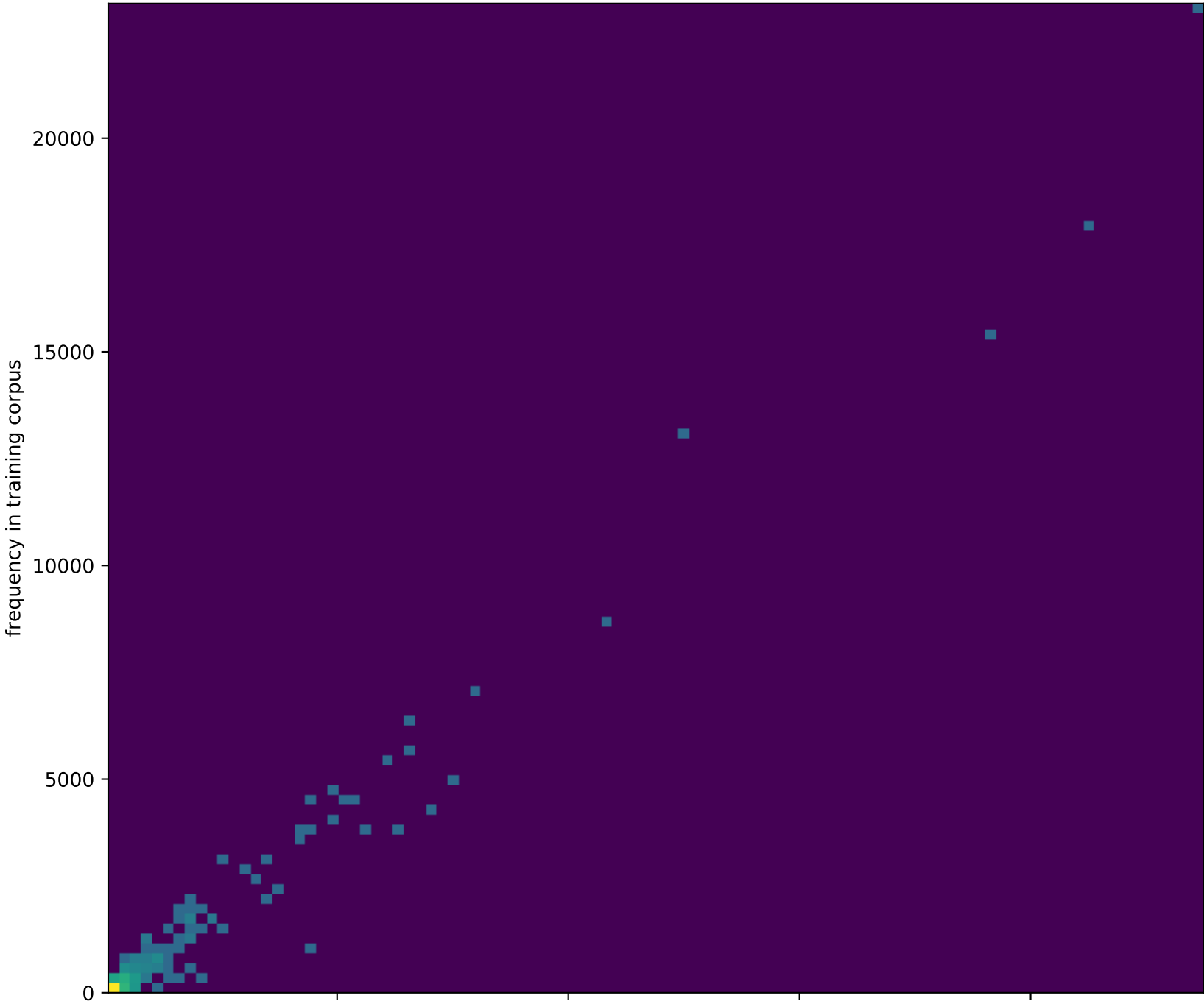


Generated words with frequency in training and huge corpus



percentage in train, token: 70.29 % , type: 37.44 %
percentage in huge, token: 76.33 % , type: 49.93 %
Percentage of new words present in huge: token: 20.40 % , type: 20.04 %