Generated words with frequency in training and huge corpus
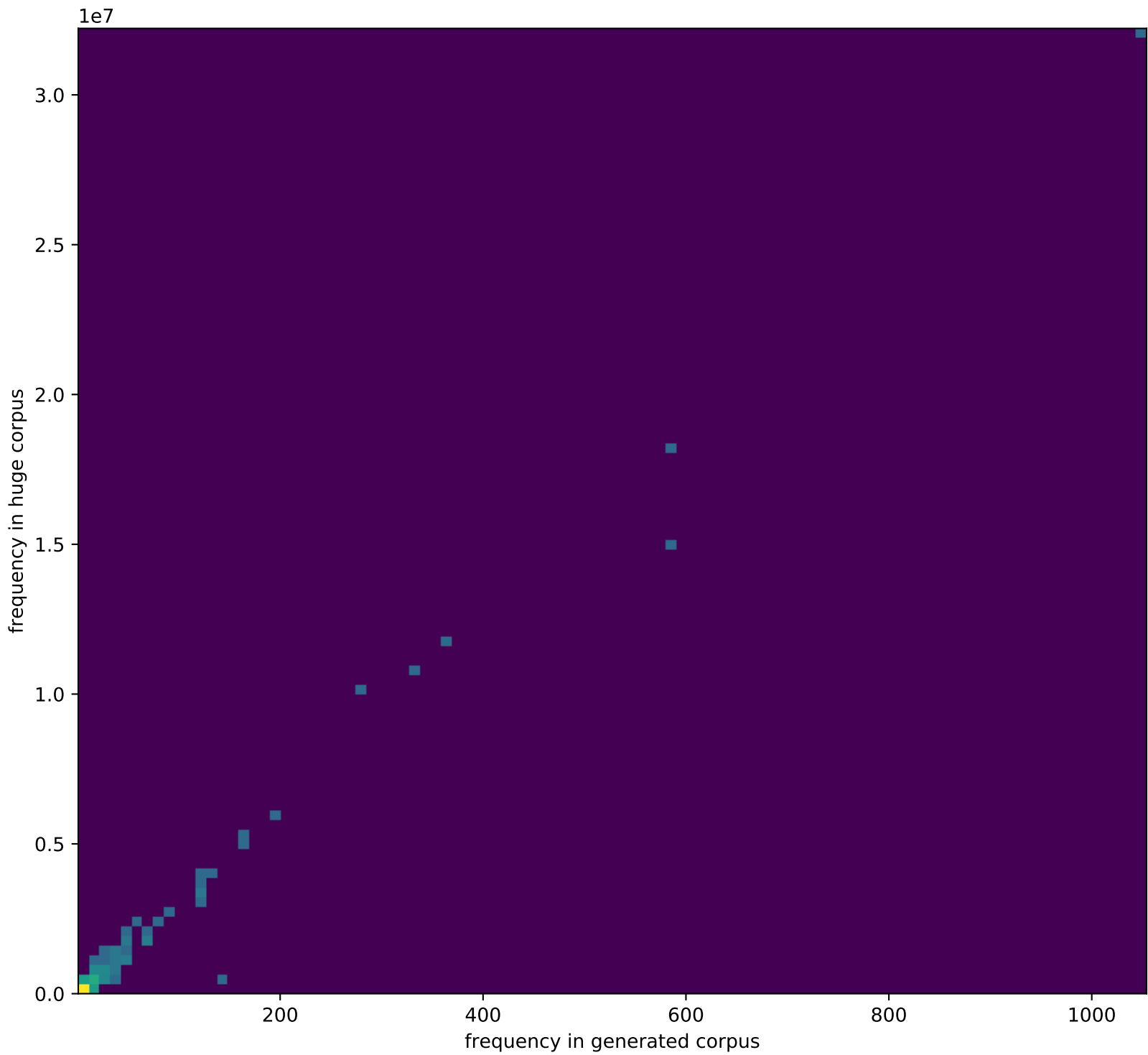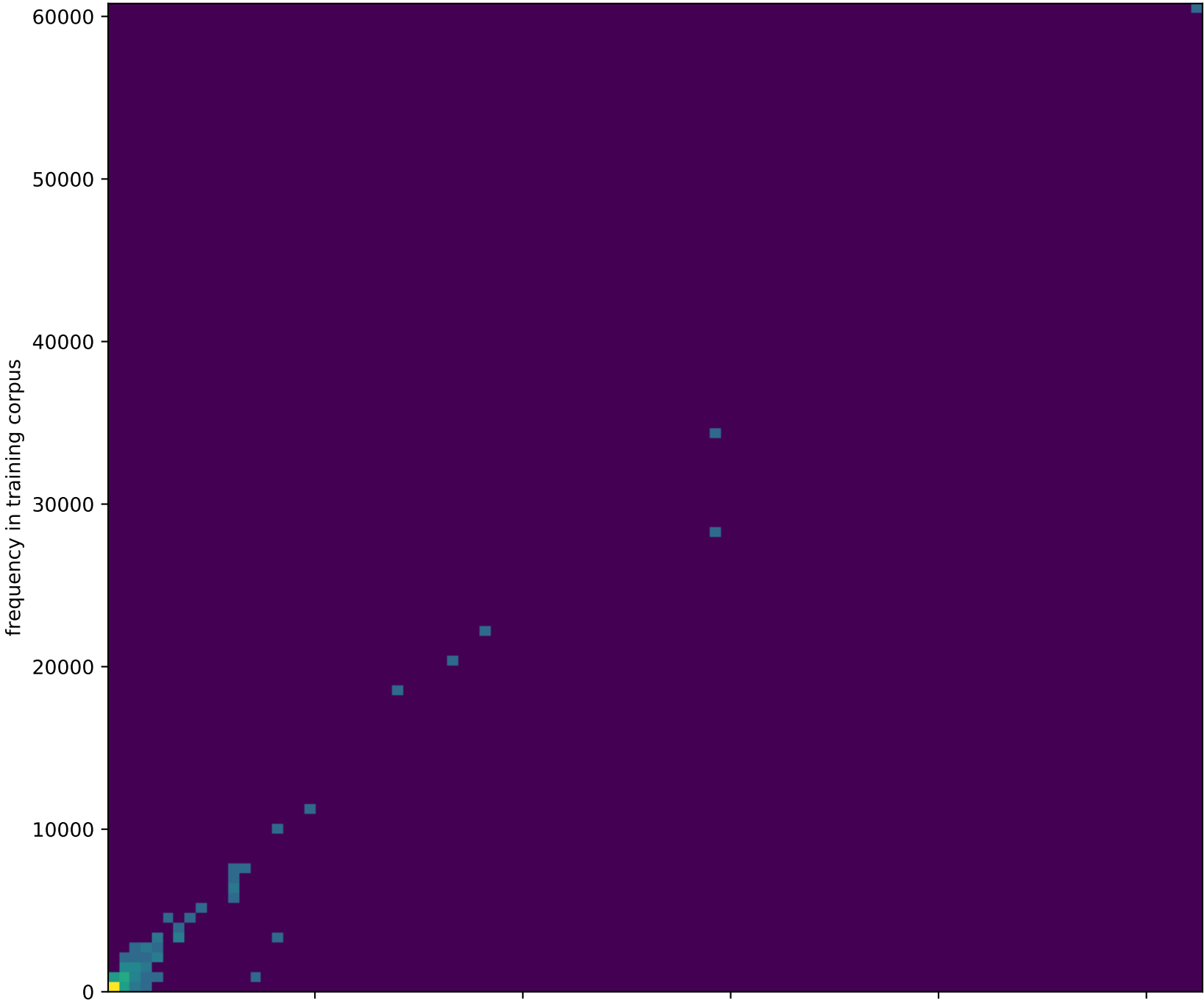

frequency in training corpus

frequency in huge corpus

frequency in generated corpus

percentage in train, token: 78.44 % , type: 47.98 %
percentage in huge, token: 83.32 % , type: 59.56 %
Percentage of new words present in huge: token: 22.66 % , type: 22.28 %