1，所有的单词数：32492

2，单词中所有的bigram数：224864

3，所有的bigram去重之后，不重复的bigram数量为：2719

4，统计的频率如下