# 第二章

**17.编写一个函数，找出文本中最常出现的50个词，停用词除外。**

**import** nltk  
**from** nltk.book **import** \*  
**import** json  
**def** often(text):  
 stopwords = nltk.corpus.stopwords.words(**'english'**)  
 ttext = [word **for** word **in** text **if** word.lower() **not in** stopwords **and** word.isalpha()]  
 fdist1 = FreqDist(ttext)  
 **return** sorted(fdist1.items(),key=**lambda** item:item[1], reverse=True)[:50]  
a=json.dumps(often(text2)).decode(**"Unicode-escape"**)  
**print** a

[["Elinor", 684], ["could", 568], ["Marianne", 566], ["Mrs", 530], ["would", 507], ["said", 397], ["every", 361], ["one", 304], ["much", 287], ["sister", 282], ["must", 279], ["Edward", 262], ["mother", 258], ["Dashwood", 252], ["time", 237], ["Jennings", 230], ["know", 230], ["Willoughby", 215], ["might", 215], ["think", 209], ["Miss", 208], ["though", 204], ["well", 191], ["Lucy", 185], ["thing", 185], ["never", 184], ["soon", 180], ["Mr", 178], ["see", 173], ["Colonel", 173], ["without", 171], ["nothing", 170], ["ever", 169], ["may", 169], ["good", 166], ["John", 163], ["say", 160], ["first", 160], ["house", 159], ["little", 158], ["day", 150], ["great", 149], ["two", 145], ["however", 145], ["Brandon", 144], ["make", 140], ["made", 138], ["sure", 134], ["Ferrars", 129], ["heart", 127]]