**A. Own problem definition for problem solving**

**This will be done with deciding the intensity-modifying words in my own, find some synonym pairs in the text, and matching them with those intensity-modifying words.** This is the whole scheme I chose to find such pairs. First, find some candidate words in the text corpora which have probability to be included in such pairs, and also decide the intensity-modifying words that will be used in these pairs. Then, find the synonym of the word in the text corpora. Finally, in such pairs, find which word is in higher intensity between two synonyms, and return the pair with intensity-modifying words.

To achieve this, there were several methods to identify. First, how to select the words which can be candidates of those pairs? Since the words such as I, is, and that will definitely not included in such pairs, I needed a way to select the probable words. Also, deciding the intensity modifying words will affect the quality of the output, too. Then, finding synonyms is also an important problem. Although synsets() and lemma\_names() in wordnet gives some synonyms, I needed a way to make a better quality synonym sets. Finally, I had to find out which is in higher intensity, and make them in a pair.

**C. Quality of output and further improvement**

**1) Overall accuracy:** Intensity-modifying words are listed in ‘boosters’ in the code. It is originally from nltk.sentiment.vader booster\_dict and chose only incrementing words then modified. ‘more’ and ‘most’ were deleted due to the use of comparative and superlative, and ‘so’ was deleted due to the use of the meaning ‘therefore’. Also, intensity-modifying adverbs appearing more than 50 times were added.

In overall, the accuracy was satisfying. Among 50 outputs, 40 outputs were accurate, due to the Collins Cobuild Advanced Learner’s English Dictionary. In the definition, they were often described as synonyms with an intensity modifying word. For example, there is (‘very simple’, ‘elementary’). When finding ‘elementary’ in dictionary, it says that something that is elementary is very simple and basic. Thus, these word pairs can be regarded as correct outputs.Six of them were synonyms with similar intensity. For example, when looking for (‘very like’, ‘alike’), since alike is a synonym for like in similar intensity, but not used with such intensity-modifying words, those word pairs appeared in output. However, it cannot be regarded as a correct output since they are similar in intensity. Three of them has reversed matching. For instance, (‘totally different’, ‘dissimilar’) is in the output. However, different is in higher intensity than dissimilar. Thus, the intensity-modifying words make the intensity difference bigger. Finally, one of them are not really synonyms. In (‘very handsome’, ‘giving’) ‘giving’ can sometimes be used as similar meaning with handsome, but not really regarded as a ‘good synonym’.

**2) Further improvement:** If one synonym appears significantly with intensity-modifying words and one does not, then I picked those pairs for the output. This approach has a huge drawback that word pairs such as (very good, great) will not appear. If the higher-intensity synonym is not superlative, then it can appear with some intensity-modifying words and suddenly be removed from the outputs. Also, it can’t detect the words like (totally different, dissimilar). Dissimilar has lower intensity, but since it does not appear with intensity-modifying words, it is shown as superlative of ‘different’. Therefore another approach is needed to recognize these pairs. Using the definition of the word would be good. If a definition of w1 is very w2, then (very w2) is same as (w1).

When looking into my approach, lots of possible outputs have been deleted for further accuracy. For example, for the outputs, (adv+adj), (adv+v), (v+adv) are possible but only (adv+adj) were used since the other pairs didn’t give significant amount of accurate outputs. Also, in finding synonyms, words end with ‘er’ and ‘st’ were regarded as comparative and superlative. However, it is also sure that some cases like fast and queer exists, thus deleted. Also, words with too many synonyms was deleted in the dictionary since it tended to have not-very-similar words. For the intensity-modifying words, words such as ‘so’ were deleted. In this perspective, accuracy and the number of word pairs is a tradeoff. Thus, if there are better criteria of removing those words, there will be more outputs. In finding synonyms, it will be better if I search the words used in similar contexts. If the words appear in synsets-lemma and also used in similar contexts, then it can be regarded as ‘synonym’. In addition, sometimes, synonym pairs such as (‘higher’, ‘high’) appears. If I further filter the words with same ‘stem’, it will boost the accuracy of the synonym dictionary.