Aviv shisman, 206558157
Itay Hassid, 209127596

NLP - Exercise 3 - Report results(Cont)

(The comparison is included)

2) added a file named 'WordLists - part 2 (Second Order) ' with the tables requested , as for the conclusions they are similar to the conclusions of 'by window approach' and 'by dependency' (that bag of words with k=5 has a chance of missing some words and etc...)

In addition i can also conclude that because we used word2Vec method we increased the percentage of 'good words', **its visible that this list of words are better** and like we talked in the class one of the reasons for this is the hyper parameters they used for their model, another reason may be that there corpus was more informed or that they extracted from it better.

3) added a file named 'att- part 2 (First Order) ' with the tables requested, as for the conclusions there are 2 things that come to my mind:
1. The features in the bag of words are more similar to the targets then the features in the other methods(probably for reasons similar to what i mentioned above)
2. It seems that yoav and omer model of word2Vec with the dependency method is using the Tag of the feature(of some unknown TagSet) including to the feature it self. (Tag_feature or Tag-feature)

4)

We've calculated the Average Precision Evaluation for every type for the words **car** and **piano**.

We set N variable in the AP function to be 20. It's should be the number of all relevant documents (words in our case), but we can't know how many exist in such tremendous corpus. It's also possible to set N to be the number of all relevant words found from all the co-occurrences types but then AP would never be 1 because the numerator smaller (or equal) than to 20 and the denominator will be around 35-40 so setting N to 20 seems like the best choice for us.

In this part, all the words similar to the target word so AP would be 1 and MAP be also 1.

	By bag of word k=5	By dependency
car	1	1
piano	1	1
MAP	1	1

We tried other approach, we considered words as related if they could replace the target word.

	By bag of word k=5	By dependency
car	14.783/20 = 0.74	19.85/20 = 0.992
piano	17.26/20 = 0.863	20/20 = 1
MAP	0.8015	0.996