Report

Thresholds:

The first threshold was filtering the lemmas which doesn’t reach the 100 occurrences we also clean up the entire sentence that means, we took context which appeared at least 100 times.

While we count context for each word no matter which variation we execute, we took only 100 most common contexts for each word.

**Word2Vec**

6-2- We received lists of similar words for our target words according to the dependencies (Type 1) and according to word to bag k=5 (type 2).

We can notice that we receive very good results.

Through a subjective judgement, we can really see the similarity between the words. For example, the word ‘car’ we receive different type of cars (limosine), different type of means of transport (motorcycle), or what’s within the car (rear-engined). There is not a big difference between BOW and Deps. Indeed, we receive almost the same words but in a different order.

6-4- To judge the similarity, we need to do it semanticaly and topically for type 1 and for type 2.

Semantics : MAP\_car = 0.9615

Topically : MAP\_CAR = 0.924

We do the same for piano

Semantics : MAP\_PIANO = 0.975

Topically : MAP\_PIANO = 0.41