AML Assignment #3

Talal Zahid – tz01815

1- The code (Scrapper.py) and the corpus (your_file.txt) is provided separately.
2- The code is explained in the comments. After the Corpus is saved. The 'your_file.txt' is fed into the code in 'skip_gram.ipynb'.
3- Below are the results from 3(a) and 3(b)

K = 5, ED = 100, Epoch = 5, lr = 0.001

```
Beginning epoch 0
```
```
c:\users\talal\appdata\local\programs\python\python36\lib\site-packages\ipykernel_launcher.py:17: UserWarning: Implicit dimensi
on choice for log_softmax has been deprecated. Change the call to include dim=X as an argument.
100% |####################################################################|
  0% |                                                                    |
```
```
Epoch 0 Loss: 704794.00000
Beginning epoch 1
```
```
100% |####################################################################|
  0% |                                                                    |
```
```
Epoch 1 Loss: 656219.43750
Beginning epoch 2
```
```
100% |####################################################################|
  0% |                                                                    |
```
```
Epoch 2 Loss: 633359.12500
Beginning epoch 3
```
```
100% |####################################################################|
  0% |                                                                    |
```
```
Epoch 3 Loss: 619207.50000
Beginning epoch 4
```
```
100% |####################################################################|
```
```
Epoch 4 Loss: 609138.18750
```

K=10, ED = 200, Epoch = 5, lr = 0.001

```
Beginning epoch 0
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:17: UserWarning: Implicit dimension choice for log_softm

Epoch 0 Loss: 1362448.37500
Beginning epoch 1

Epoch 1 Loss: 1258619.62500
Beginning epoch 2

Epoch 2 Loss: 1217635.00000
Beginning epoch 3

Epoch 3 Loss: 1192218.75000
Beginning epoch 4
Epoch 4 Loss: 1173661.75000
```
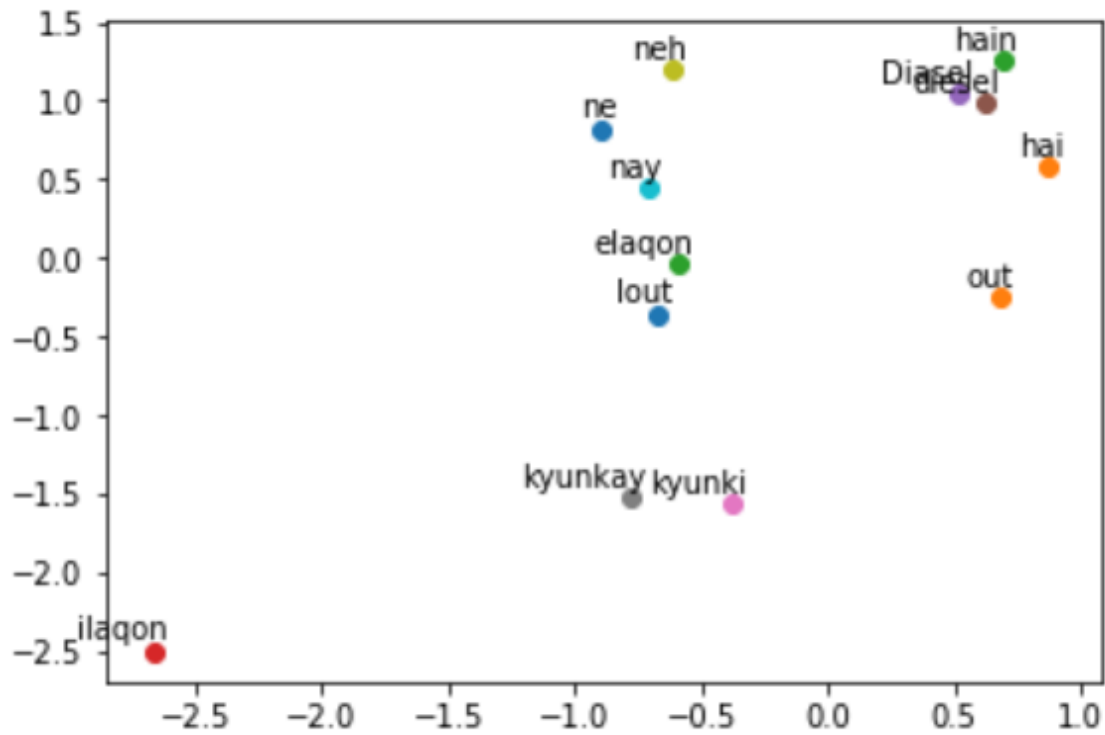
K=15, ED = 200, Epoch = 5, lr = 0.001

```
             Beginning epoch 0
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:17: UserWarning: Implicit dimension choice for log_softmax has k

Epoch 0 Loss: 1996397.75000
Beginning epoch 1

Epoch 1 Loss: 1845255.37500
Beginning epoch 2

Epoch 2 Loss: 1788186.50000
Beginning epoch 3

Epoch 3 Loss: 1724256.48500
Beginning epoch 4

Epoch 4 Loss: 1664212.82750
```

K=20, ED = 300, Epoch = 5, lr = 0.01

```
             Beginning epoch 0
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:17: UserWarning: Implicit dimension choice for log_softmax has

Epoch 0 Loss: 509114.50240
Beginning epoch 1

Epoch 1 Loss: 498455.37500
Beginning epoch 2

Epoch 2 Loss: 449390.34234
Beginning epoch 3

Epoch 3 Loss: 418304.51940
Beginning epoch 4

Epoch 4 Loss: 384175.77740
```

4- Part C (there wasn't many misspellings as it was from dawn.com and auto generated)

misspelled = [('lout','out'),('elaqon','ilaqon'),('Diasel','diesel'),('kyunki','kyunkay'),
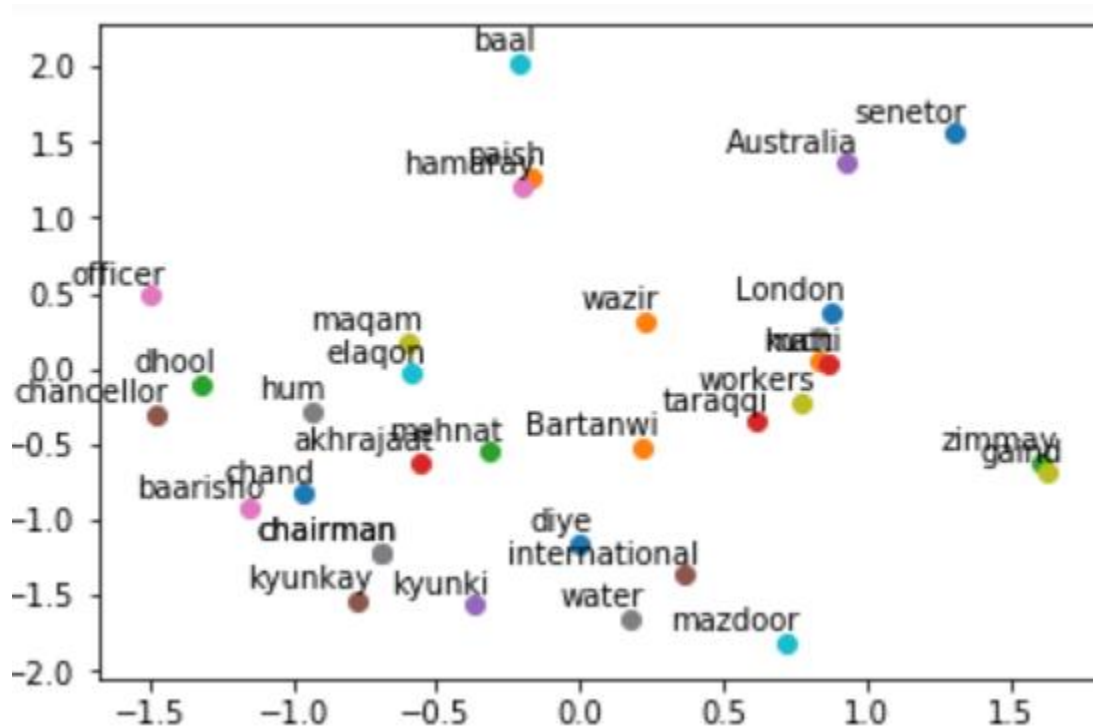('neh','nay','ne'),('hai','hain')]

ne, neh and nay have strong correlation. Also Diasel and diesel, Kyunkay and kyunki.

4- Part (A) Extracting synonyms manually proved to be a tedious task. Below are the ones I could manage.
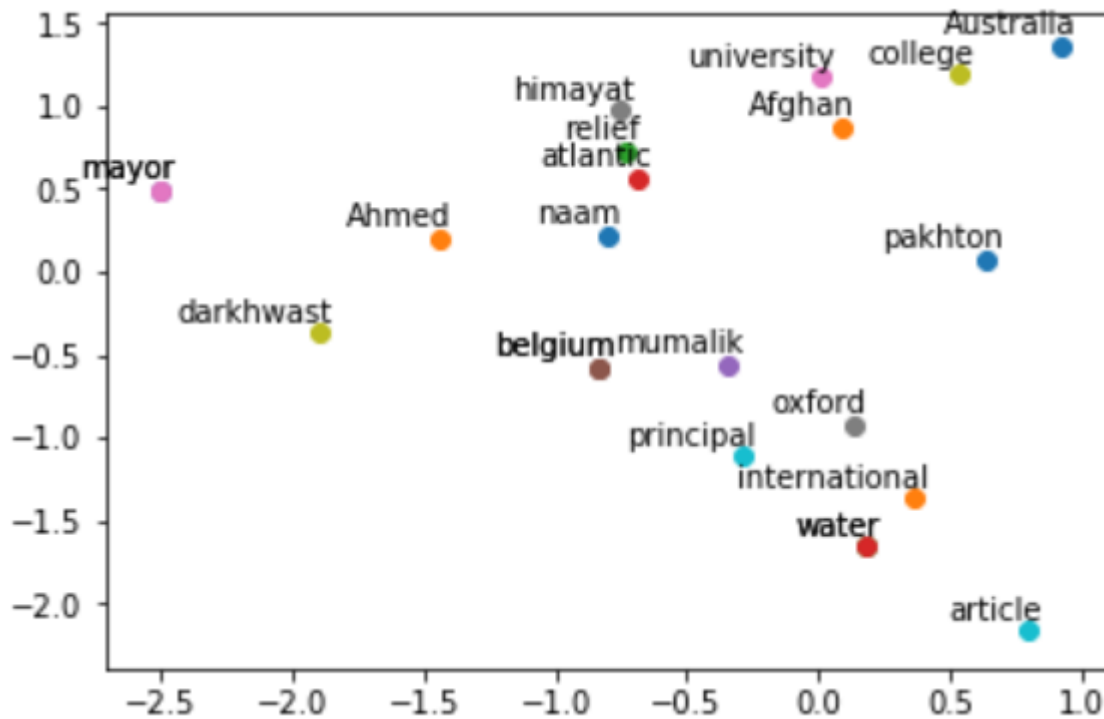
synonyms =
[('chand','kuch'),('dhool','matti'),('australia','international'),('baarisho','water'),('workers','mazdoor'),
('diye','paish'),('mehnat','taraqqi'),('chairman','chancellor'),('officer','chairman'),('maqam','elaqon'),('sen
etor','wazir'),('zimmay','akhrajaat'),('kyunki','kyunkay'),('hamaray','hum'),('gaind','baal'),('London','Barta
nwi')]

Office & Chancellor, Kyunkay & kyunki, maqam & elaqon, have strong correlation.

4- Part (B) Extracting Analogies manually proved to be a tedious task. Below are the ones I could manage.

analogies = [('australia','international'),('water','atlantic'),('mayor','belgium'),('university','oxford'), ('darkhwast','article'),('pakhton','Afghan'),('relief','water'),('mumalik','belgium'),('mayor','himayat'),('coll ege','principal'),('news','ghalat'),('naam','Ahmed'),('wet','water')]

Belgium & mumalik, university & college, Ahmed & naam, himayat & relief, international & oxford have strong correlation.