Weekly report

Yevgen

Amir

* General Tasks done:

1. Words occurring only once were removed from the English Data
2. Words occurring only once were removed from the Dutch data
3. Orthographic Links were added. (measured using levenstein distance)

The levenstein distance for every pair of ENG-DUTCH was calculated, however it was added only if the value was 0.5 or greater

1. Code for activation propagation on the graphs was added.

* Experiments:

1. The results were evaluated using T-tests.

Metrics used: - Total variation distance

* Rank biased distance
* Average precision

So in total we had three T-test results

1. Two sets of comparisons:

\* TEST vs Bilingual model

↑

|

| T-test

|

↓

\* TEST vs Monolingual model

TEST data are obtained from the experiments done by JANE G VAN HELL.

The monolingual data the Big Dutch and English graphs.

The Bilingual data are the same graphs connected using the three links mentioned below.

3. Three sets of links:

\*Associative links (obtained from the free association norms)

\*Translation equivalence links (obtained from the merged dictionaries)

\*Orthographic links (obtained using levenstein distance)

* Results Obtained:

When using only association links and Translation equivalence links, the results were not optimal.

When added the orthographic links for all the links, the improvement was not very significant.

* Things to be discussed:

Assigning link weights.

Non-symmetric weight assignment.

Build the network using word frequency (remove the less frequent words).

End goals of the project:

Trying to construct the bilingual graph using the monolingual graphs modified by adding different sets of edges.