Pun Detection

Team 25: Venkat,Ramachandra,Arun,Ram Manoj

Word Sense Disambiguation

Word sense disambiguation is the task of computationally determining which sense of a

polysemous term is the one intended when that term is used in a particular communicative act.

Pun Detection

A pun is a form of wordplay in which one signifier (*e.g.,* a word or phrase) suggests two or more meanings by exploiting polysemy, or phonological similarity to another signifier, for an intended humorous or rhetorical effect.

- I used to be a banker but I lost interest.
- When the church bought gas for their annual barbecue, proceeds went from the sacred to the propane.

For example, the first of the following two punning jokes exploits contrasting meanings of the word "interest", while the second exploits the sound similarity between the surface form "propane" and the latent target "profane":

Puns where the two meanings share the same pronunciation are known as *homophonic* or *perfect*, while those relying on similar- but not identical-sounding signs are known as *heterophonic* or *imperfect*.

lexical semantics and phonology – is an essential prerequisite for the production and interpretation of puns.

Heterophonic puns need Phonology and more.So we consider only Homophonic puns in our study.

precision

# of true positives ÷ (# of true positives + # of false positives)

recall

# of true positives ÷ (# of true positives + # of false negatives)

accuracy

(# of true positives + # of true negatives) ÷ (# of true positives + # of true negatives + # of false positives + # of false negatives)

F1

(2 × precision × recall) ÷ (precision + recall)

http://alt.qcri.org/semeval2017/task7/index.php?id=results

Method using co-occurrence matrix

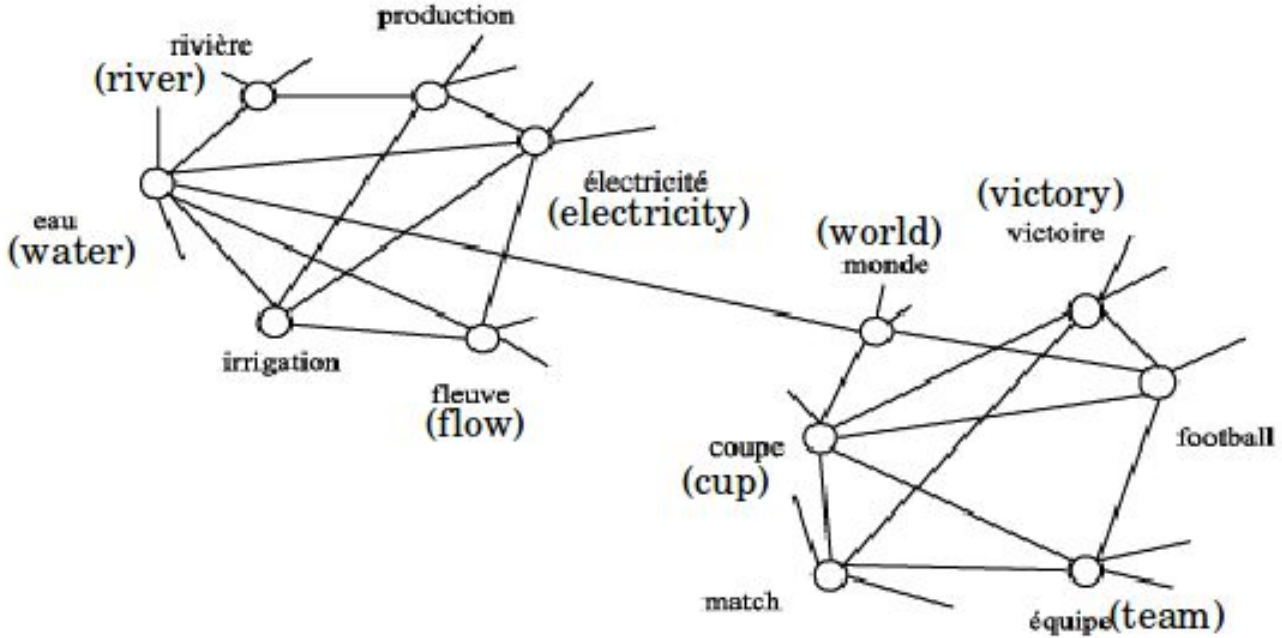


Fig. 4. Graph of the cooccurrents of the French word *barrage*.

Co-occurrence matrix

First we tried to build a co-occurrence matrix to see how the world are related in the given pun sentences.

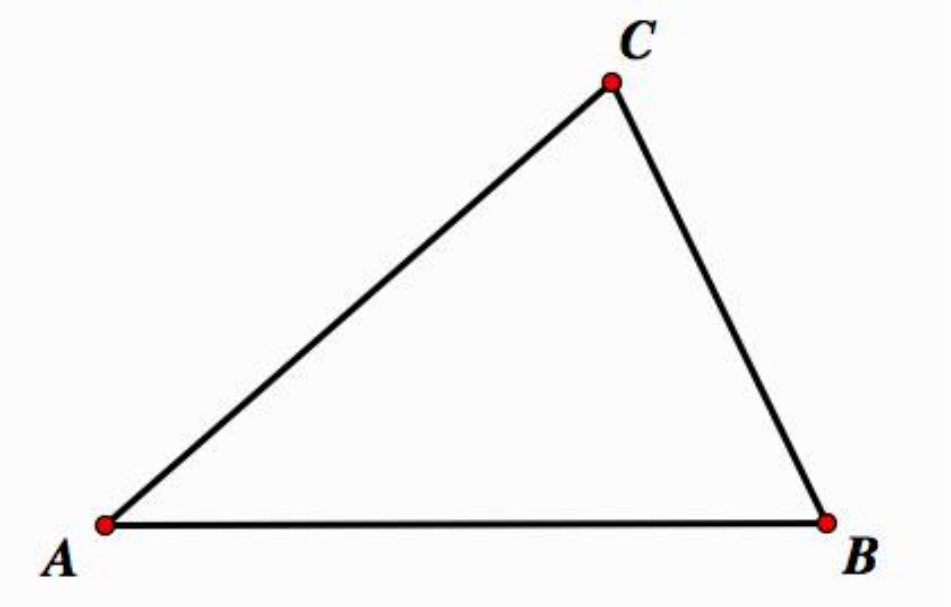
For the given set of words from pun sentences we went through wiki-dump to extract their co-occurrence counts. Using these we used to calculate different edge weights.,

Once we the edge weights we went through each and every line to find whether the transitivity property holds or not.

Working example.

He used to work near the bank of river but he lost interest .

Transitivity property



A = River
B = Bank
C = Interest

Here AB, BC, CA are the weights caliculated from co-occurence matrix.

So how does it work ?? Since words river and bank related with each we get more edge weight.

Similarly same with the case of bank and interest.

But river and interest are not related with each other they get less edge weight so here transitivity property failed.

So this sentence will be detected as a pun sentence.

Finding non-transitive edge

Sort the edges using edge weights.

Method1: Take largest and smallest edge now consider mean of the edge. If the difference between mean and the middle edge is lower than some threshold then transitivity property fails.

Method2: Take top two largests edge’s now consider mean of the edge. If the difference between mean and the smallest edge is lower than some threshold then transitivity property fails.

Edge weight calculation.

Method 1:

We simple Co-occurrence count. And got accuracy around 39%.

Recall-0.35

Precession-0.61

Method 2:

Instead of using co-occurrence count's we tried to tried normalise each and every edge. Here the edge weight is max(P(A/B), P(B/A)) and got accuracy around 51%.

Recall-0.61

Pressession-0.69