Dauphine | PSL
UNIVERSITÉ PARIS

# Translation using word embeddings

CASTRO ROS Alejandro - EL MANSOURI Hassan - Massoulié ELOI

# Problem Definition

Source language embedding

$$X$$

Target language embedding

$$Z$$

**Objective**

Finding a linear mapping *W* such that

$$Z = WX$$

- We utilize pretrained embeddings from fastText based on skip-gram methods

# Supervised methods

- Aligning vectors by maximising the sum of cosine distance sum($(Wx_i)^T z_{i)}$;
- Standard gradient descent, minimising the sum quadratic distance sum($||Wx_i - z_i||^2$);
- Closed-form *Procrustes* solution: minimise a Frobenius distance by bringing the singular values of $ZX^T$ to 1.

# Supervised methods results

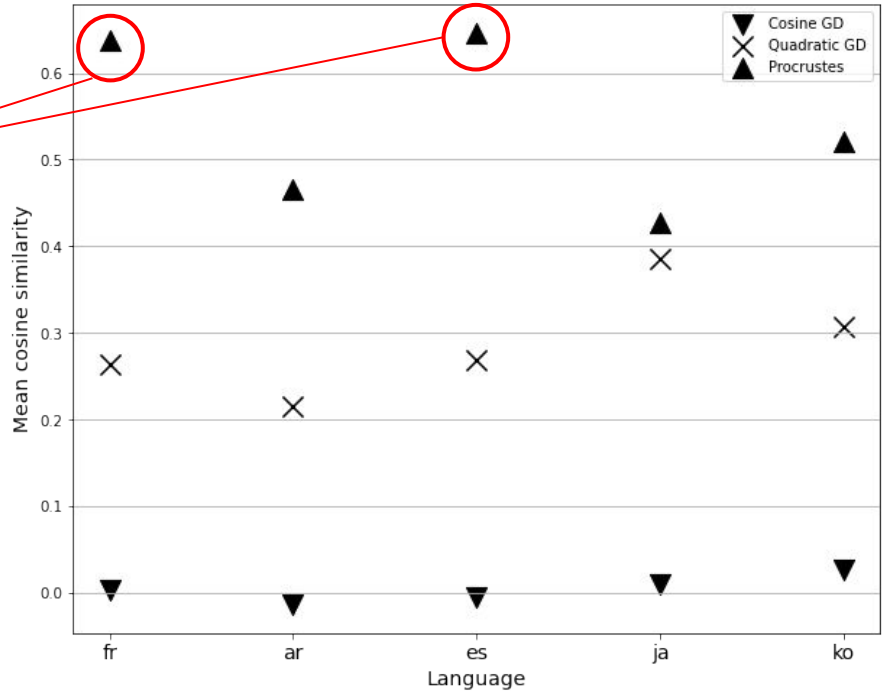
- Best results for **Procrustes** method.
- **Cosine distance minimization** didn't give good results.

| | Linear Loss | Quadratic Loss | Procrustes |
|---|---|---|---|
| 1 | remit (0.185) | pivoter (0.320) | **asseoir** (0.502) |
| 2 | aïeux (0.180) | glisser (0.307) | dormir (0.499) |
| 3 | battit (0.179) | hall (0.305) | **assis** (0.492) |
| 4 | retourna (0.176) | **asseoir** (0.305) | reposer (0.485) |

**Traduction of *sit* to French for different methods.**

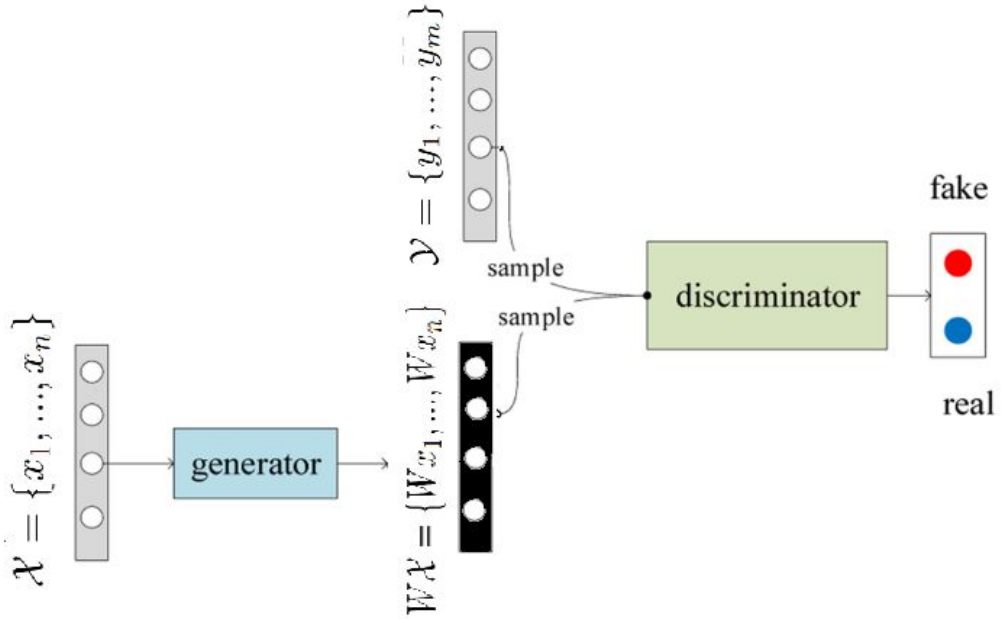
# Results and Language Comparison

- Best metric for **Indo-European** languages.
- **Procrustes** method outperforms iterative optimization.



# Unsupervised method

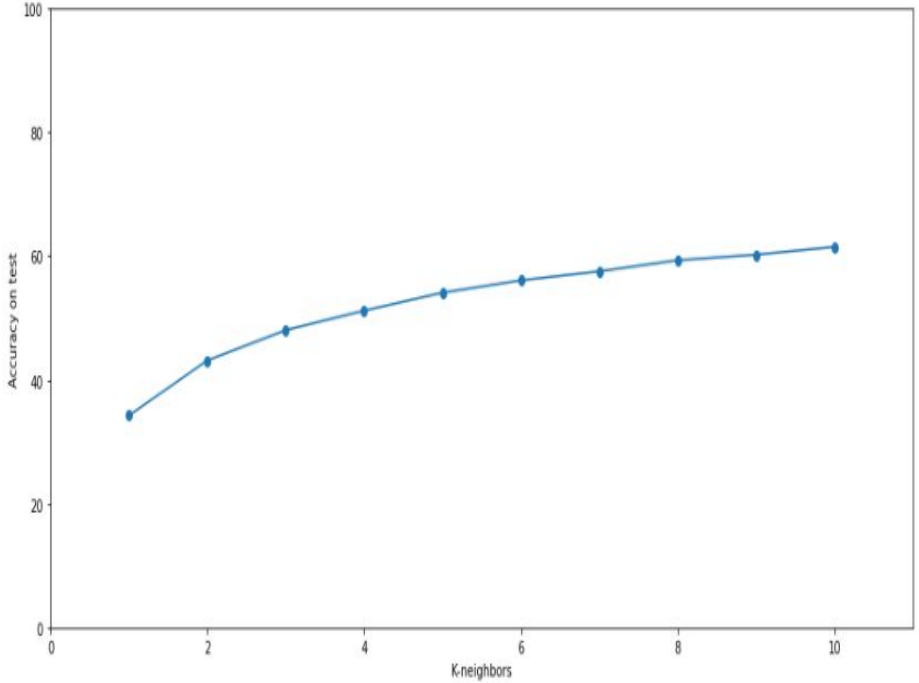
- No cross-lingual supervision.
- We alternate between training the generator (mapping) and the discriminator.
- The generator gets better at fooling the discriminator and vice-versa.



# Results of GAN and limitations

| Clusters | accuracy (K=1) | mean cosine similarity |
|---|---|---|
| Unsupervised GAN | 34.36 % | 0.47 |
| Procrustes | **69.56 %** | **0.63** |

Figure 4: Evaluation and comparison with best supervised method




⇨ Supervised methods apparently better than unsupervised counterparts, but …

- More refinement methods could be used.
- Limited by computing power overhead
- Understandable due to no supervision

# Merci pour votre attention