

# Éthique : introduction aux *embeddings* (plongements lexicaux)

Karën Fort

karen.fort@loria.fr / <https://members.loria.fr/KFort>

## Quelques sources d'inspiration

- ▶ Présentation de D. Seddah à la journée ATALA RobusTAL, avec son accord :  
[http://pauillac.inria.fr/~seddah/Talk\\_Robustal.zip](http://pauillac.inria.fr/~seddah/Talk_Robustal.zip)
- ▶ Présentation de Y. Goldberg : *Demystifying Neural Word Embeddings* (2015) [https://www.rocq.inria.fr/alpage-wiki/tiki-download\\_file.php?fileId=25](https://www.rocq.inria.fr/alpage-wiki/tiki-download_file.php?fileId=25)

# Le problème : les mots et l'ordinateur

- ▶ La soupe était mauvaise
- ▶ La soupe était infâme
- ▶ Le gâteau était infâme
- ▶ Le gâteau était affreux

→ Pour l'ordinateur, les mots ne sont que des ensembles de symboles, tous les mêmes.

→ Mais pour nous, certains se ressemblent plus que d'autres

→ Nous avons besoin d'une représentation des mots qui capture cela

# L'hypothèse

L'hypothèse distributionnelle (Harris 1954) :

*Word in similar contexts tend to have similar meanings*

Linguistique de corpus (Firth, 1957) :

*You should know a word by the company it keeps*

Le Jabberwocky (L.Carroll, 1871) :

*Il était grilheure ; les slictueux toves Gyraient sur l'al-  
loinde et vriblaient : Tout ivoreux allaient les borogoves ;  
Les verchons fourgus bourniflaient.*

# Collectionner les contextes

he curtains open and the moon shining in on the barely  
ars and the cold , close moon " . And neither of the w  
rough the night with the moon shining so brightly , it  
made in the light of the moon . It all boils down , wr  
surely under a crescent moon , thrilled by ice-white  
sun , the seasons of the moon ? Home , alone , Jay pla  
m is dazzling snow , the moon has risen full and cold  
un and the temple of the moon , driving out of the hug  
in the dark and now the moon rises , full and amber a  
bird on the shape of the moon over the trees in front  
But I could n't see the moon or the stars , only the  
rning , with a sliver of moon hanging among the stars  
they love the sun , the moon and the stars . None of  
the light of an enormous moon . The plash of flowing w  
man 's first step on the moon ; various exhibits , aer  
the inevitable piece of moon rock . Housing The Airsh  
oud obscured part of the moon . The Allied guns behind

→ Les mots sont similaires s'ils sont utilisés dans des contextes similaires

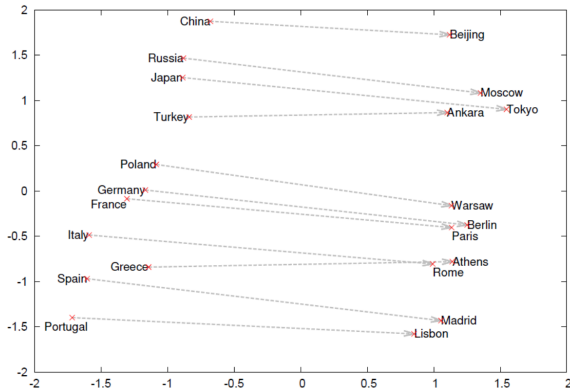
## Les plongements lexicaux : représenter les mots comme des vecteurs

moon = (the :324, shining :4, cold :1, brightly :2, stars :12,  
elephant :0, ...)

→ Les mots sont similaires si leurs vecteurs sont similaires

# Les plongements lexicaux : calculer des distances

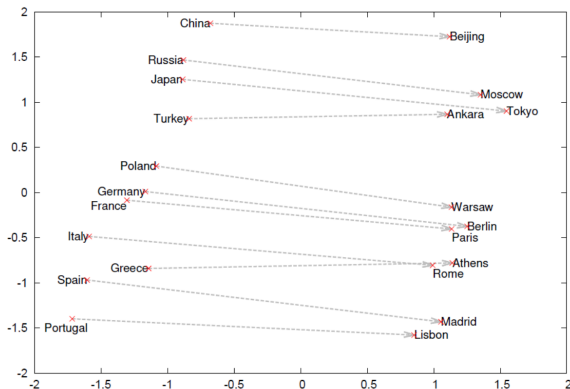
ex : Word2Vec [Mikolov et al., 2013]





# Les plongements lexicaux : calculer des distances

ex : Word2Vec [Mikolov et al., 2013]



king - man + woman = queen  
Tokyo - Japan + France = Paris  
best - good + strong = strongest



Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013).  
Efficient estimation of word representations in vector space.  
CoRR, abs/1301.3781.