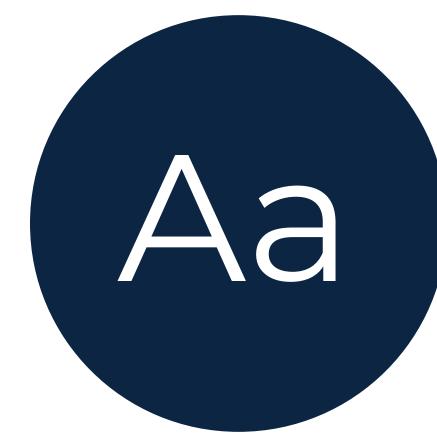
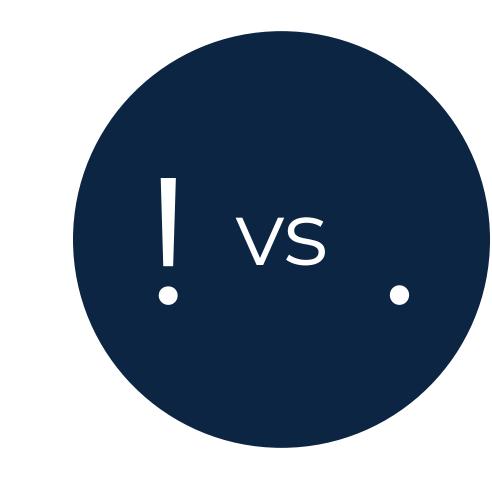
### BOURG PALETTE

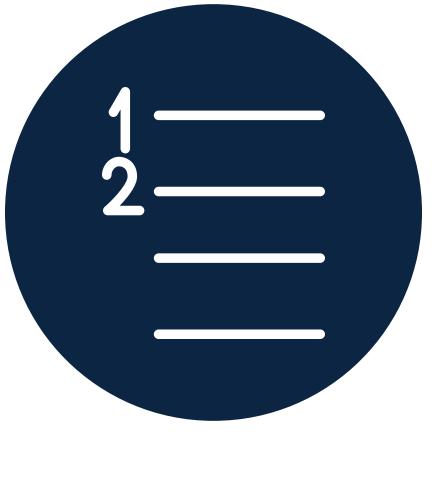
HACKATHON 2018



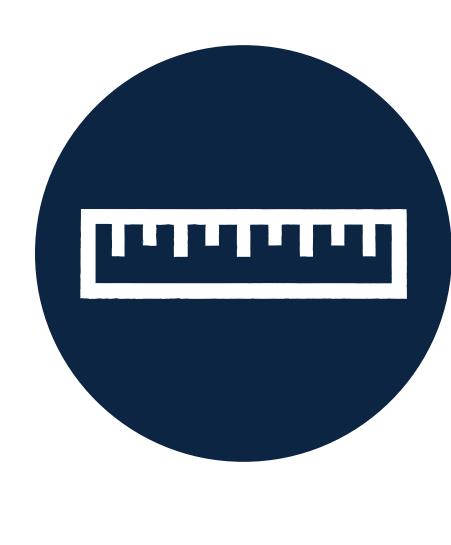
Casse



Ponctuation



Nombre de mots/phrases



Longueur des mots/phrases



Entités nommées

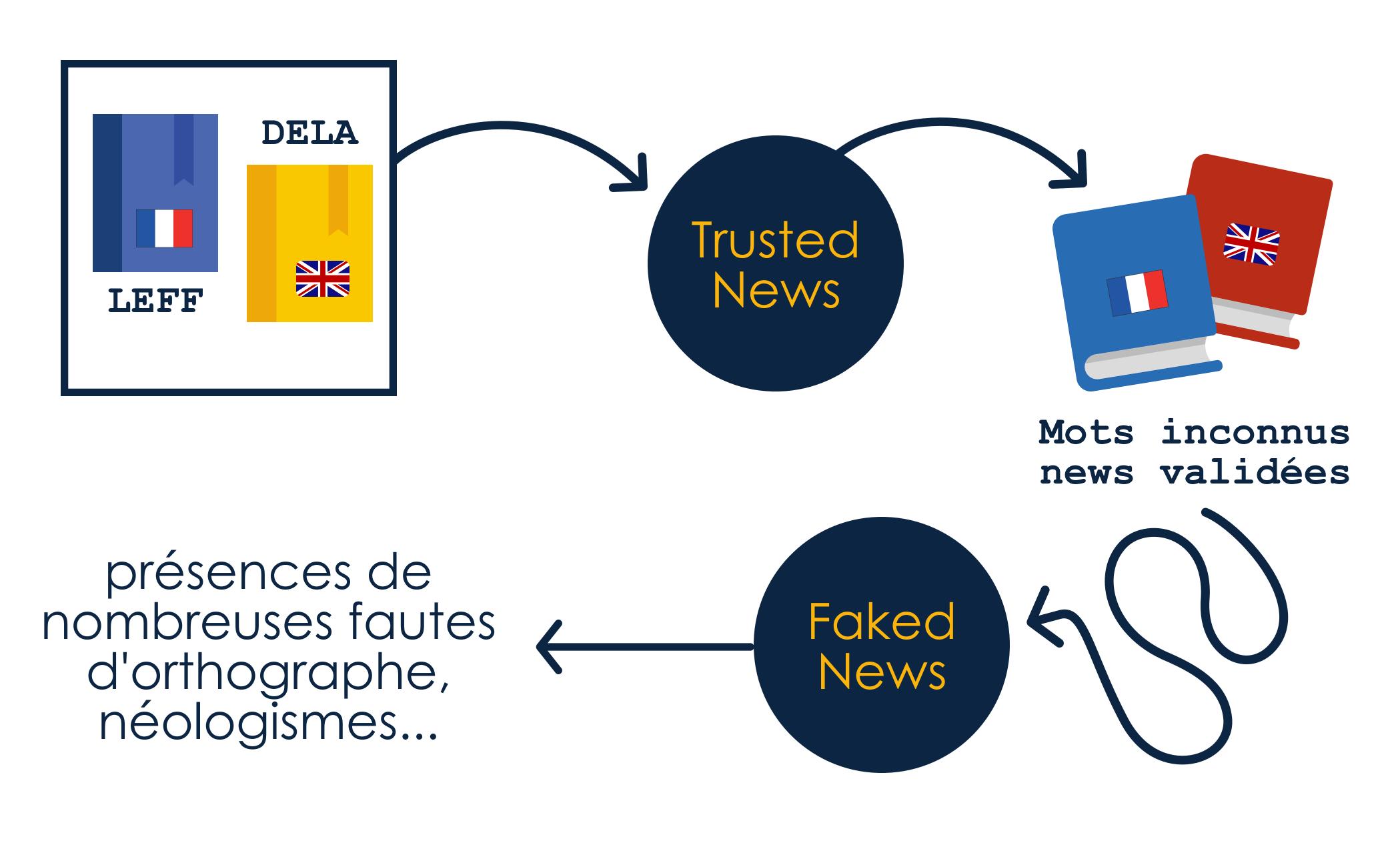


Stratégies éditoriales

## Features

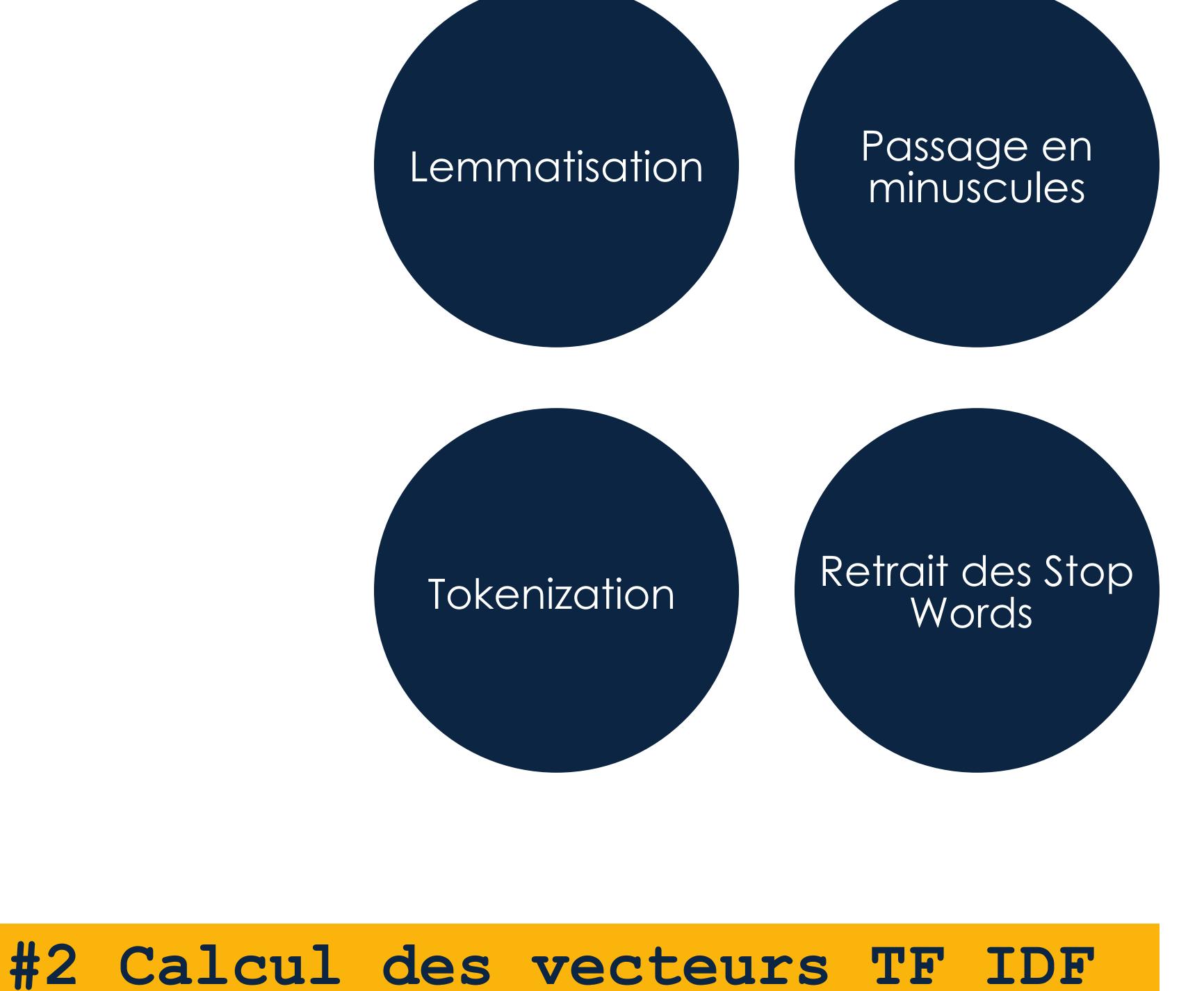
Richesse du vocabulaire ?

Influence de l'orthographe ?



# #1 Préprocessing

Concrètement...

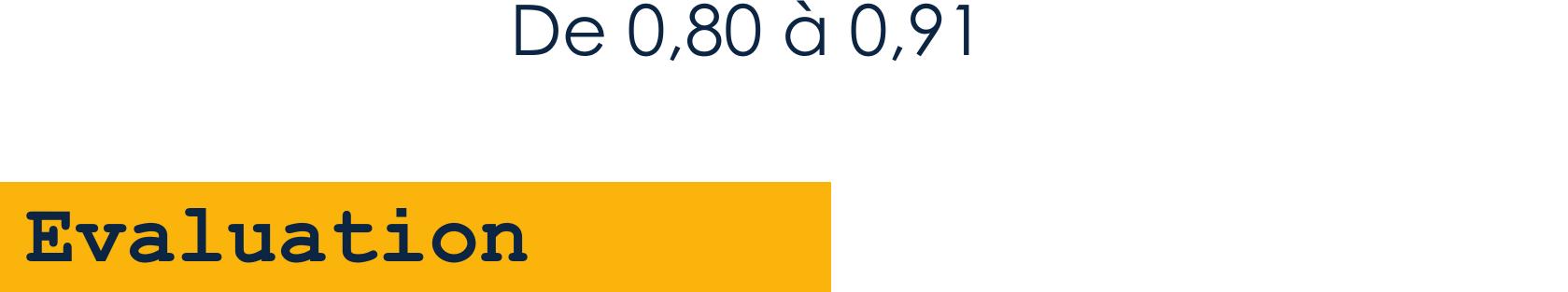


## #3 Réduction des dimensions avec LSA

A particulièrement améliorer les scores du corpus

0,91 pour l'anglais et YouTube

Avec Crossvalidation avec 10 paquets



F-Mesure: 0,85 pour le français

GILBERT DELAHAYE - MARCEL MARLIER



casterman

Résolution de problèmes majeurs

