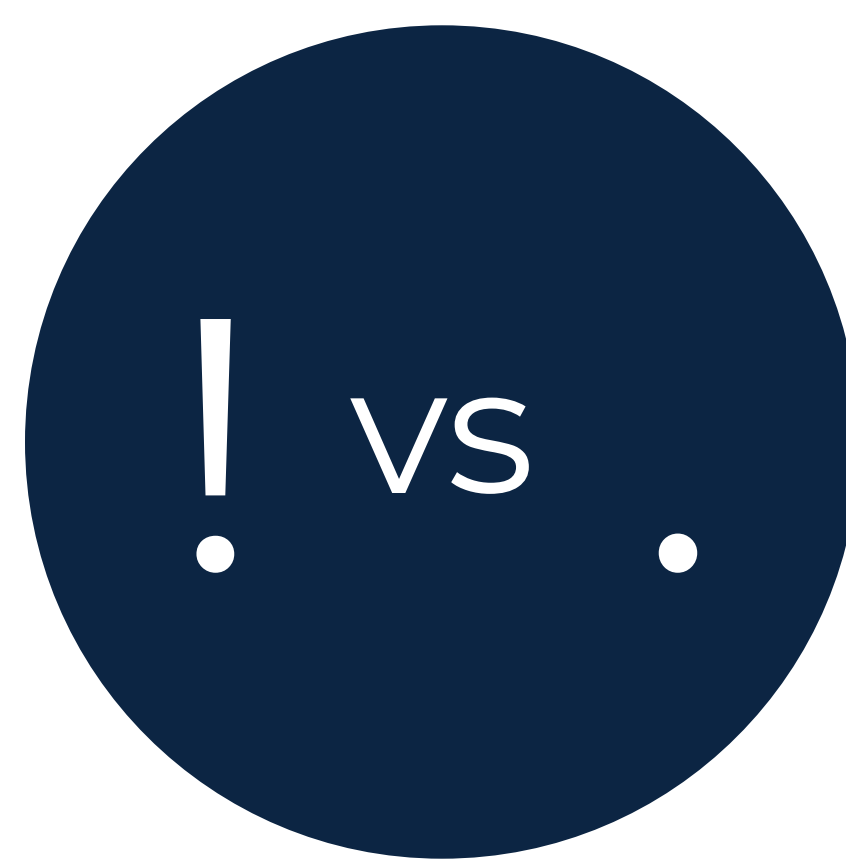


BOURG PALETTE

HACKATHON 2018



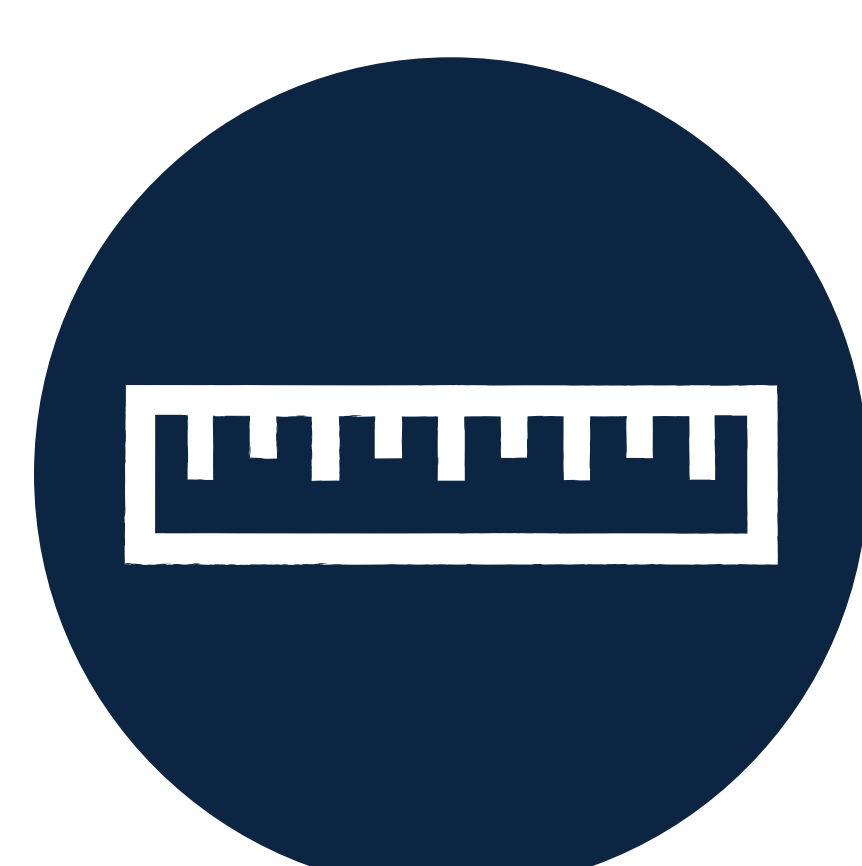
Casse



Ponctuation



Nombre de mots/phrases



Longueur des mots/phrases



Entités nommées

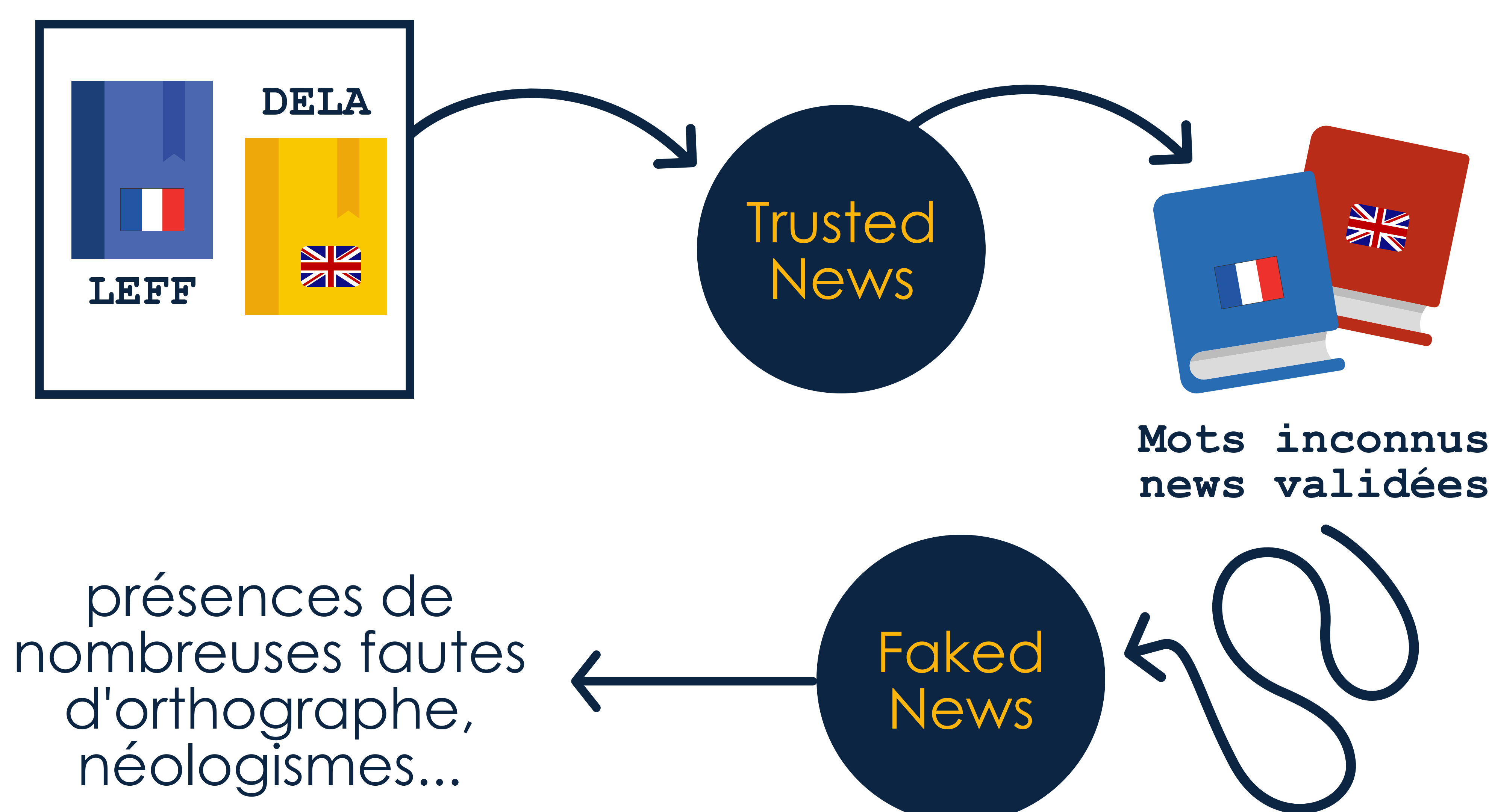


Stratégies éditoriales

Features

Richesse du vocabulaire ?

Influence de l'orthographe ?



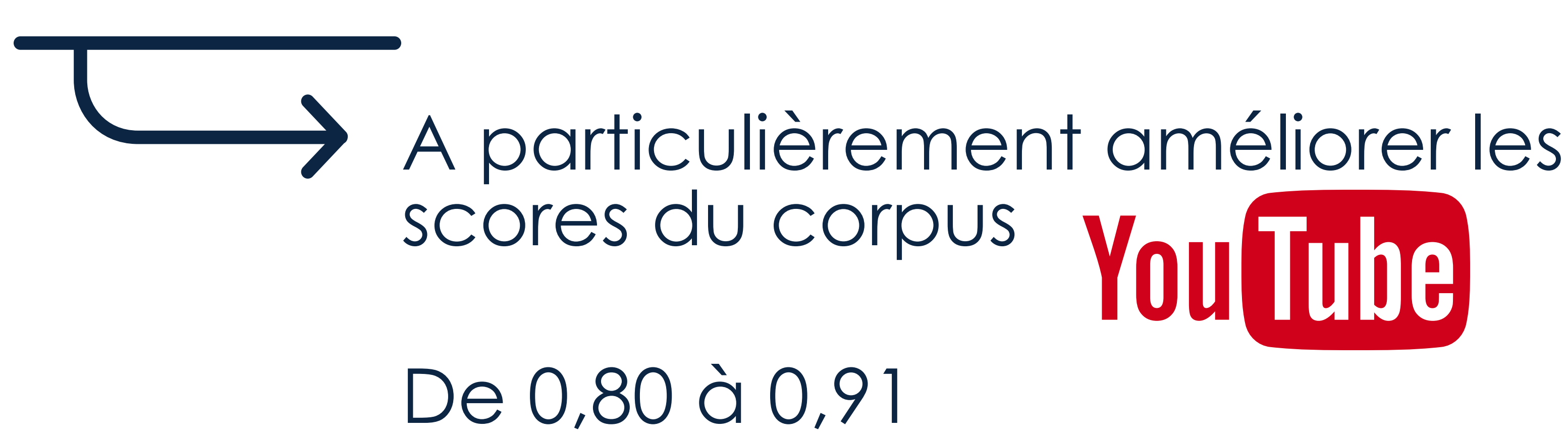
Concrètement...

#1 Préprocessing

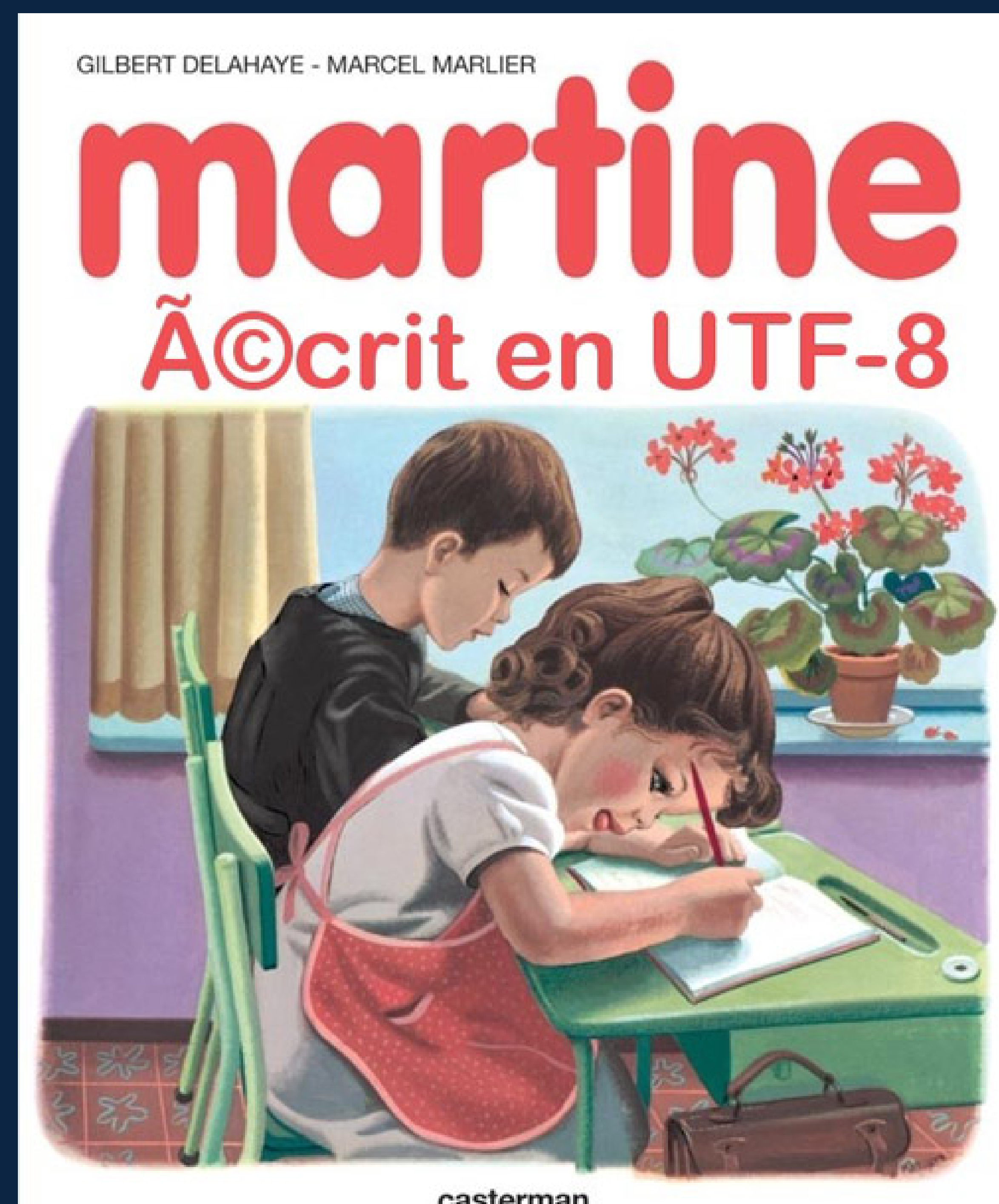
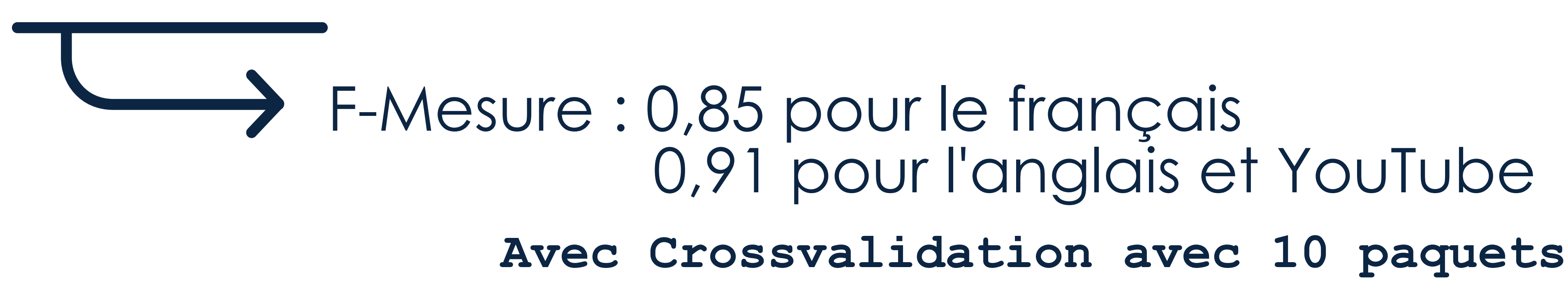


#2 Calcul des vecteurs TF IDF

#3 Réduction des dimensions avec LSA



#4 Evaluation



Résolution de problèmes majeurs

