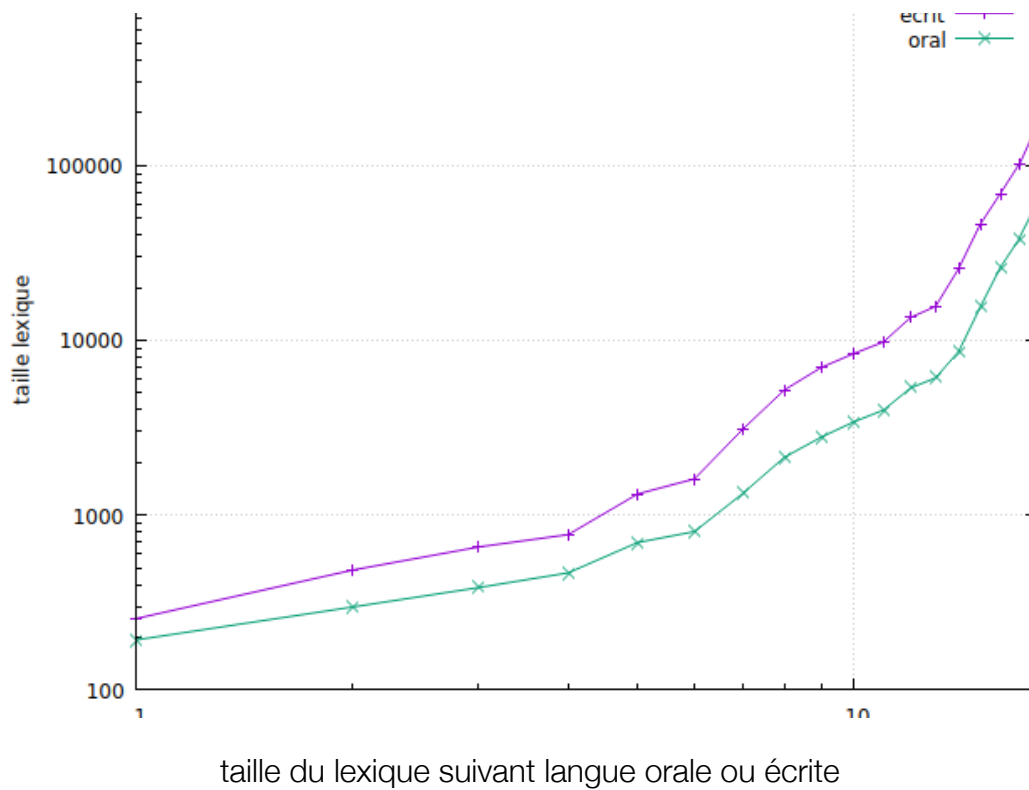


## TP1 : Intro au TAL

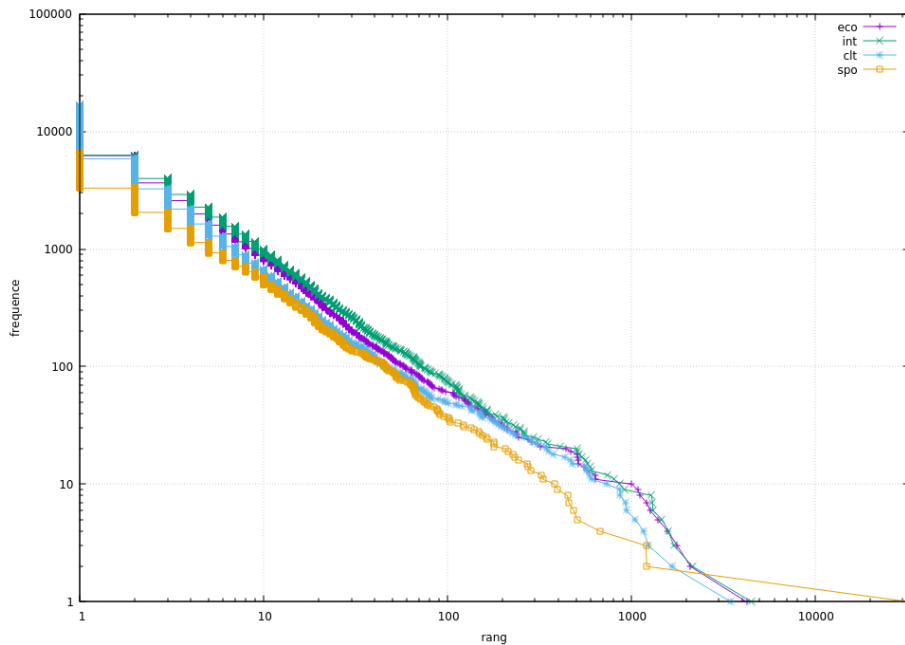
### Exercice 1 :

Comme on peut le voir les courbes croissent de manière exponentielle mais arrivent à une certaine taille de corpus élevée la taille du lexique n'augmente plus et quant aux caractéristiques des langues orales ou écrites elles ont exactement le même comportement.



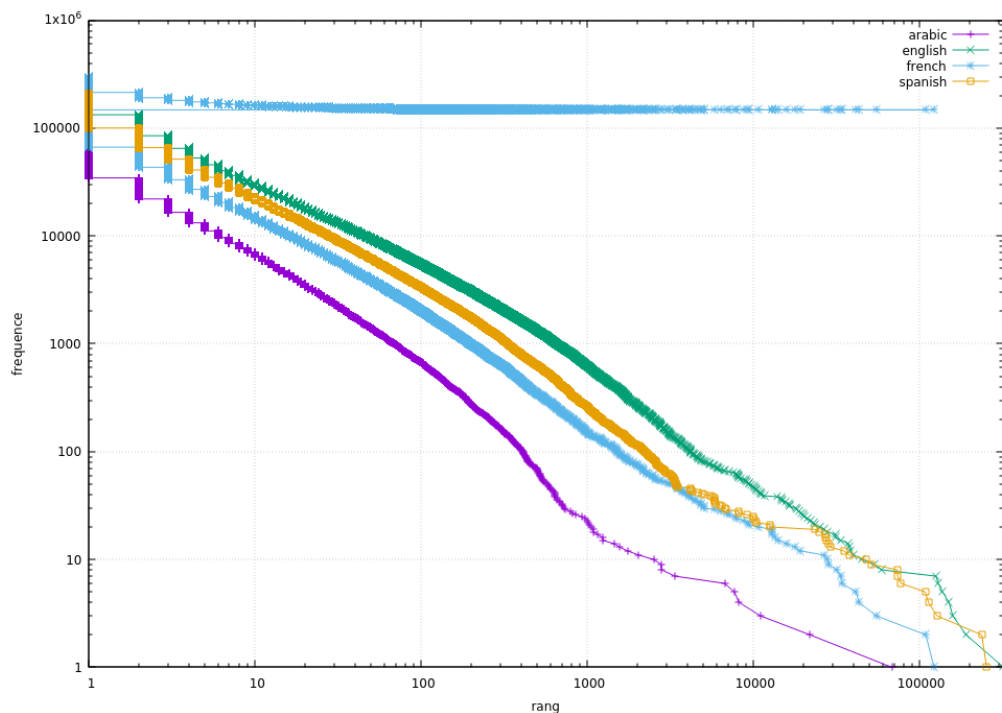
## Exercice 2 :

- 1) On peut en déduire que les courbes ont le mêmes comportements, elles semblent être plus ou moins linéaires décroissantes et la loi de Zipf semble être respectée hormis quelques cas particulier comme le range 1.



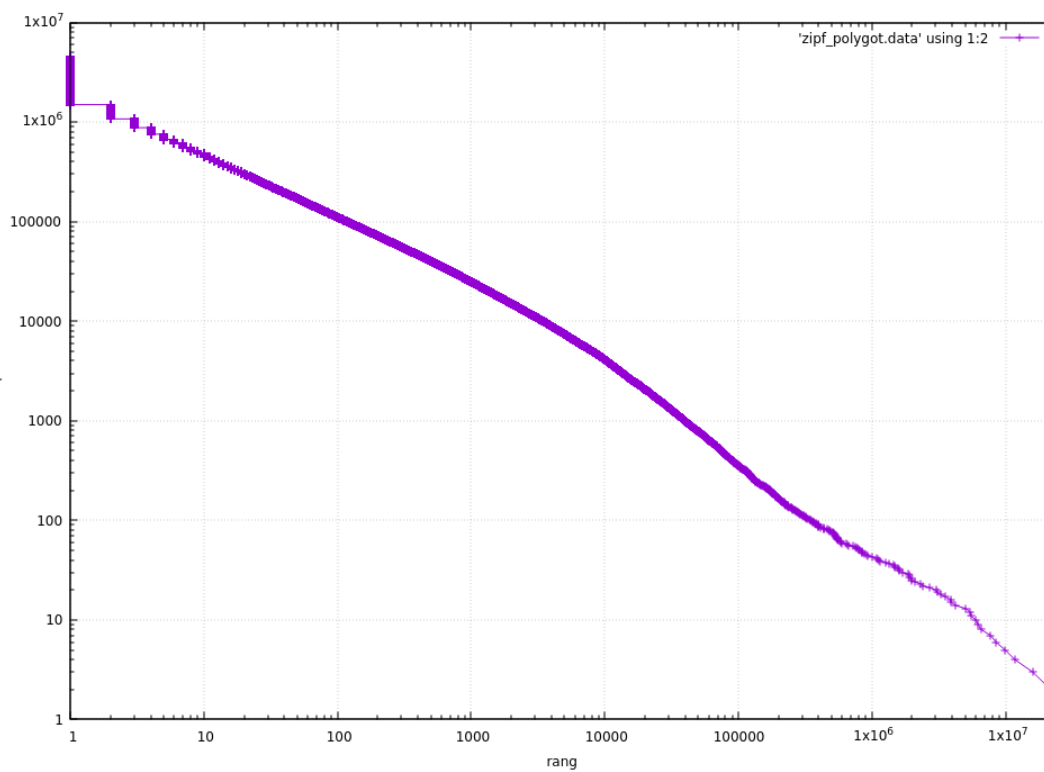
courbes rang / fréquence des différents topics

- 2) Pour les corpus multilingues ont remarques que les courbes sont similaires à celles du dessus et ont donc le même comportement.



Courbes rang / fréquence des différentes langues

- 3) Pour le fichier frwiki-from-polyglot.txt on retrouve encore les mêmes caractéristiques que dessus on peut donc en déduire que la loi de Zipf est respectée.



courbe rang / fréquence sur le fichier frwiki-from-polyglot