

Zipf's Law and My Blog on Language

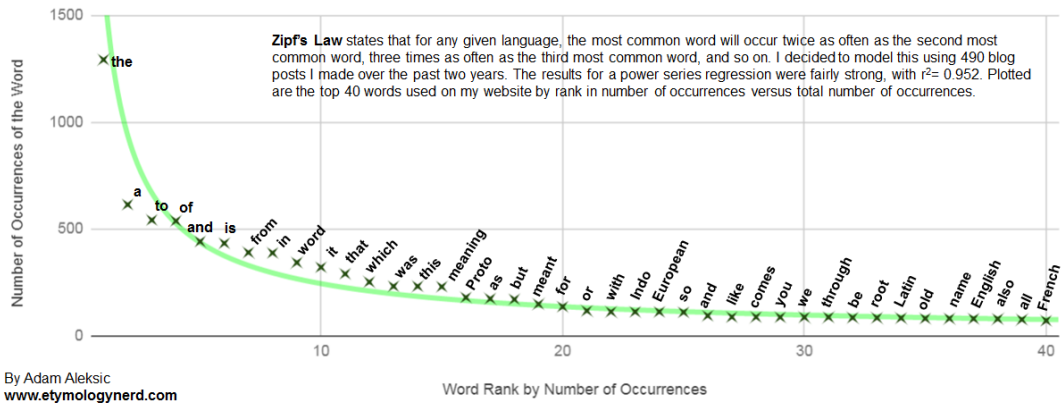


FIGURE K.10 – Une illustration de la loi de Zipf (Source : Adam Aleksic, www.etymologynerd.com)

Pondération	Formule du tf
binaire	$\begin{cases} 1 & \text{si } f_{t,d} > 0 \\ 0 & \text{sinon} \end{cases}$
fréquence brute	$f_{t,d}$
fréquence (normalisée)	$\frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$
normalisation logarithmique	$\log(1 + f_{t,d})$
normalisation par le max	$K + (1 - K) \frac{f_{t,d}}{\max_{t' \in d} f_{t',d}}$
<i>Formule pour idf</i>	
$idf_t = \log\left(\frac{ D }{df_t}\right)$	
<i>Formule complète</i>	
$td-idf_{t,d} = tf_{t,d} \times idf_t$	

TABLE K.2 – Formules pour le calcul du tf-idf, avec les variantes les plus courantes pour le calcul du tf.

- Notations :
- t terme
 - D ensemble de documents, $d \in D$
 - $f_{t,d}$ fréquence de t dans d
 - df_t $|\{d \in D / f_{t,d} \neq 0\}|$

bataille (35,2)
 politique (11,5)
 amour (19,94)
 voyage (17,44)

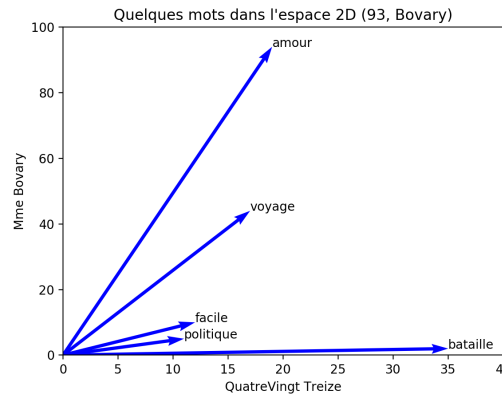


FIGURE K.11 – représentation graphique des mots dans le plan (93, Bovary)

je connaissais le	voyagiste	qui les amène
voyages : Différent du	voyagiste	ou tour-opérateur . En
compagnie à un	voyagiste	. Animaux domestiques : Lorsqu'
que disent certains	voyagistes	, légalement , aucune n'
Transport / Hébergement Le	voyagiste	FRAM propose les
plus vite votre	voyagiste	. Le Comité scientifique
le monde ... 6ème	voyagiste	européen , Kuoni est
autres , les grands	voyagistes	essayent de prendre
fév 2008 Ces	voyagistes	, qui opèrent principalement
parmis les meilleurs	voyagistes	en ligne . Service
les 15 principaux	voyagistes	en ligne en
ses amies des	voyagistes	locaux qui fêtent
foyers ...) et des	voyagistes	. La commercialisation Le
lancé vers des	voyagistes	susceptibles d' offrir
clients chez un	voyagiste	- Développement d' un
des plus grands	voyagistes	européens dont ils
voyage à cheval .	Voyagiste	spécialiste du voyage
clients chez un	voyagiste	- Développement d' un
sur cette destination	Voyagiste	Catalogue destination du
hôtel , à votre	voyagiste	, à votre ambassade

TABLE K.3 – extrait kwic de la concordance [-3,3] autour de *voyagiste* dans frWaC, consultation 2021-08-11.