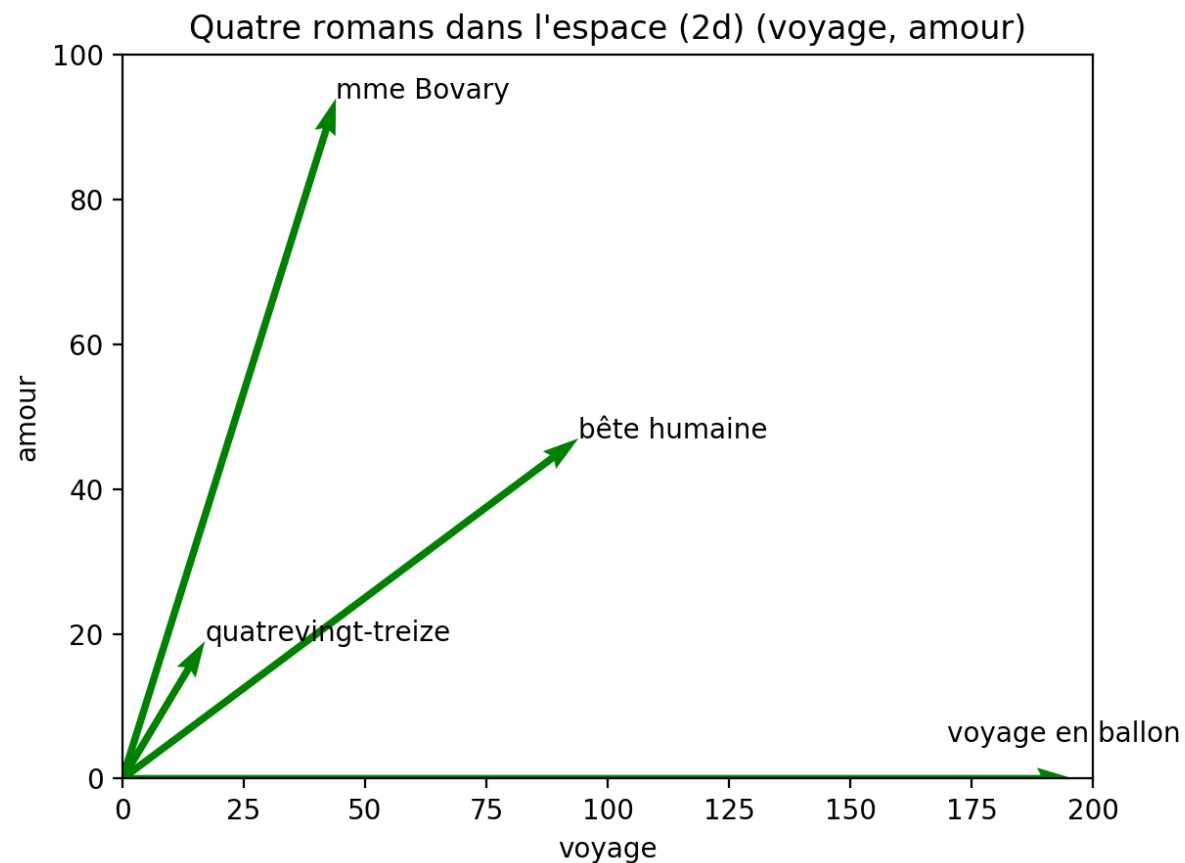


- *If A and B have almost identical environments we say that they are synonyms. (Harris, 1954)*
- *You shall know a word by the company it keeps. (Firth, 1957)*
- *The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear. (Lenci, 2008)*

	QuatreVT 119 Kw	Voyage Bal 82 kw	Bête Hum. 128 kw	Mme Bovary 117 kw
bataille	35	4	6	2
clair	105	26	96	52
facile	12	19	6	10
politique	11	0	9	5
voyage	17	196	94	44
idiot	2	1	2	6
amour	19	0	47	94

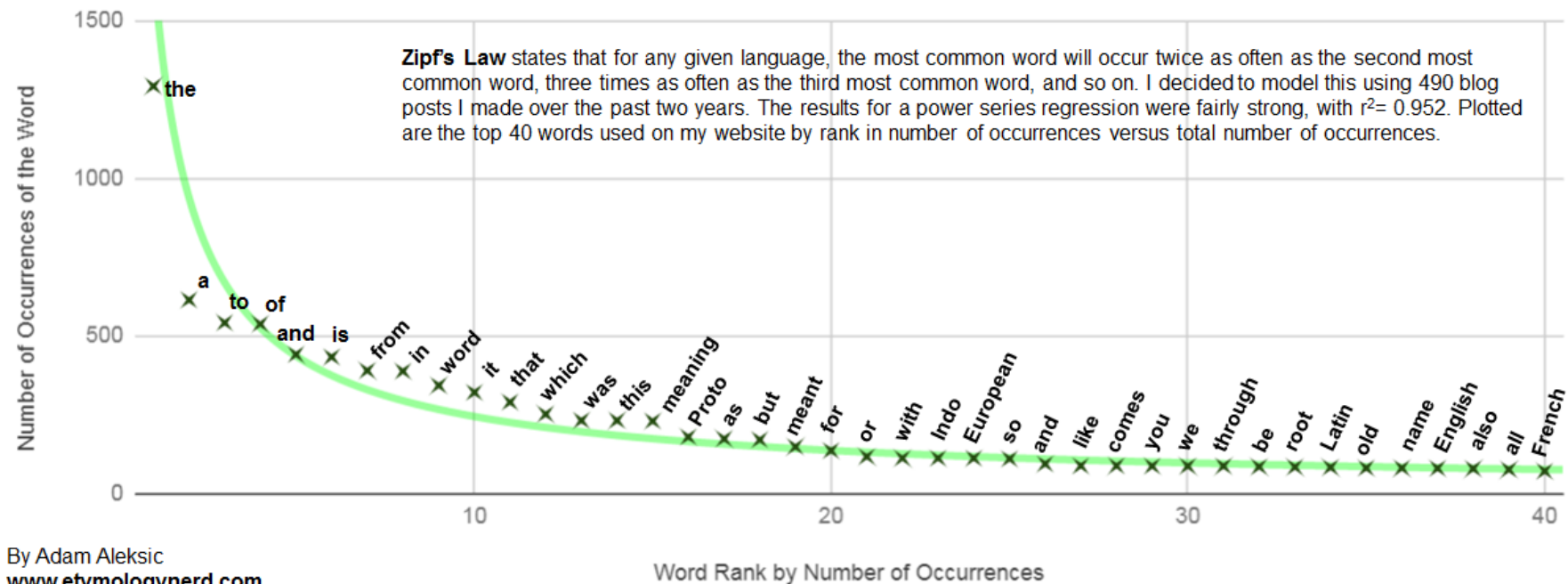
Table – Matrice terme-documents pour quelques mots et 4 romans (Quatrevingt-treize (Hugo) ; Le voyage en ballon (Verne) ; La bête humaine (Zola) ; Mme Bovary (Flaubert)).

	QuatreVT	Voyage Bal	Bête Hum.	Mme Bovary
voyage	17	196	94	44
amour	19	0	47	94



Loi de Zipf

Zipf's Law and My Blog on Language



<i>Pondération</i>	<i>Formule du tf</i>
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>	
$td-idf_{t,d} = tf_{t,d} \times idf_t$	

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)

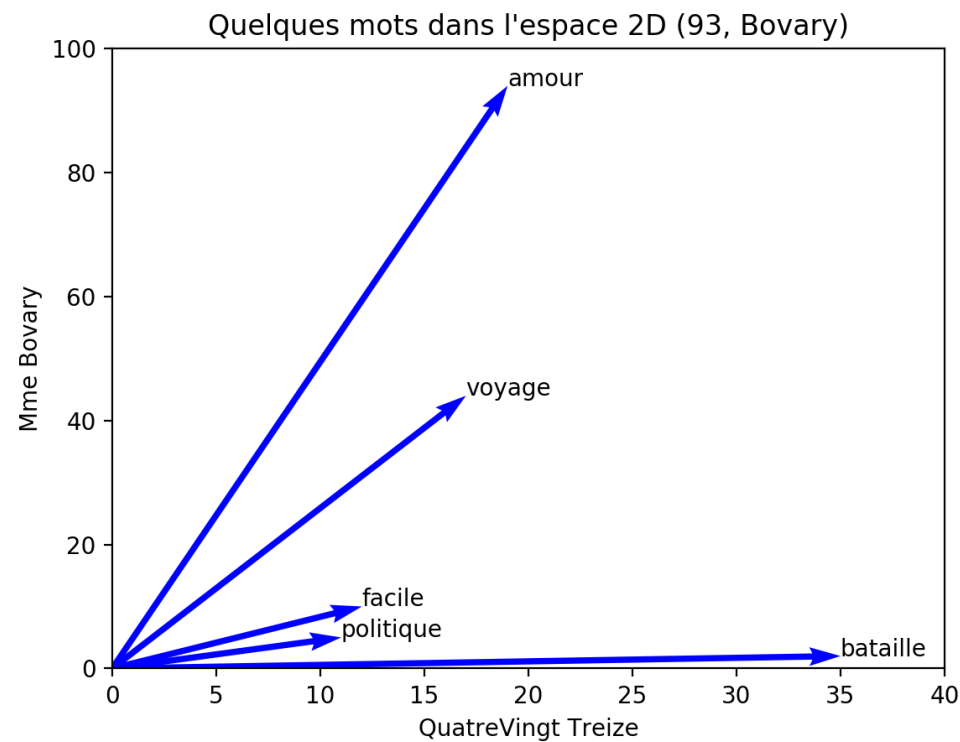


Figure – représentation graphique des mots dans le plan (93, Bovary)

KWIC voyageur

je connaissais le
voyages : Différent du
compagnie à un
que disent certains
Transport / Hébergement Le
plus vite votre
le monde ... 6ème
autres , les grands
fév 2008 Ces
parmi les meilleurs
les 15 principaux
ses amies des
foyers ...) et des
lancé vers des
clients chez un
des plus grands
voyage à cheval .
clients chez un
sur cette destination
hôtel , à votre

voyagiste
voyagiste
voyagiste
voyagistes
voyagiste
voyagiste
voyagistes
voyagistes
voyagistes
voyagistes
voyagistes
voyagistes
voyagistes
voyagistes
Voyagiste
voyagiste
Voyagiste
voyagiste

qui les amène
ou tour-opérateur . En
. Animaux domestiques : Lorsqu'
, légalement , aucune n'
FRAM propose les
. Le Comité scientifique
européen , Kuoni est
essayent de prendre
, qui opèrent principalement
en ligne . Service
en ligne en
locaux qui fêtent
. La commercialisation Le
susceptibles d' offrir
- Développement d' un
européens dont ils
spécialiste du voyage
- Développement d' un
Catalogue destination du
, à votre ambassade

frWaC, consultation 2021-08-11

	bataille	voyage	homme	femme
arriver	246	470	1 819	890
tomber	100	83	1 205	384
habiller	2	4	339	384
mourir	180	116	339	1 088
	55 331	208 520	668 289	346 093

Table – Matrice Terme-Terme obtenue dans frWaC. Contextes en ligne.

	arriver	tomber	habiller	mourir	
bataille	246	100	2	180	55 331
voyage	470	83	4	116	208 520
homme	1 819	1 205	339	1 499	668 289
femme	890	660	384	1 088	346 093

Table – Matrice Terme-Terme obtenue dans frWaC. Contextes en colonnes.