

Frequences Lexicales du Français

Christophe Pallier

5/16/2019

Nous allons afficher les distributions des fréquences lexicales des mots listés dans la base lexique3 (voir <http://www.lexique.org>)

Nous chargeons d'abord la table lexique382

```
source('https://raw.githubusercontent.com/chrplr/openlexicon/master/datasets-info/fetch_datasets.R')
```

```
## Loading required package: rjson
```

```
## Loading required package: tools
```

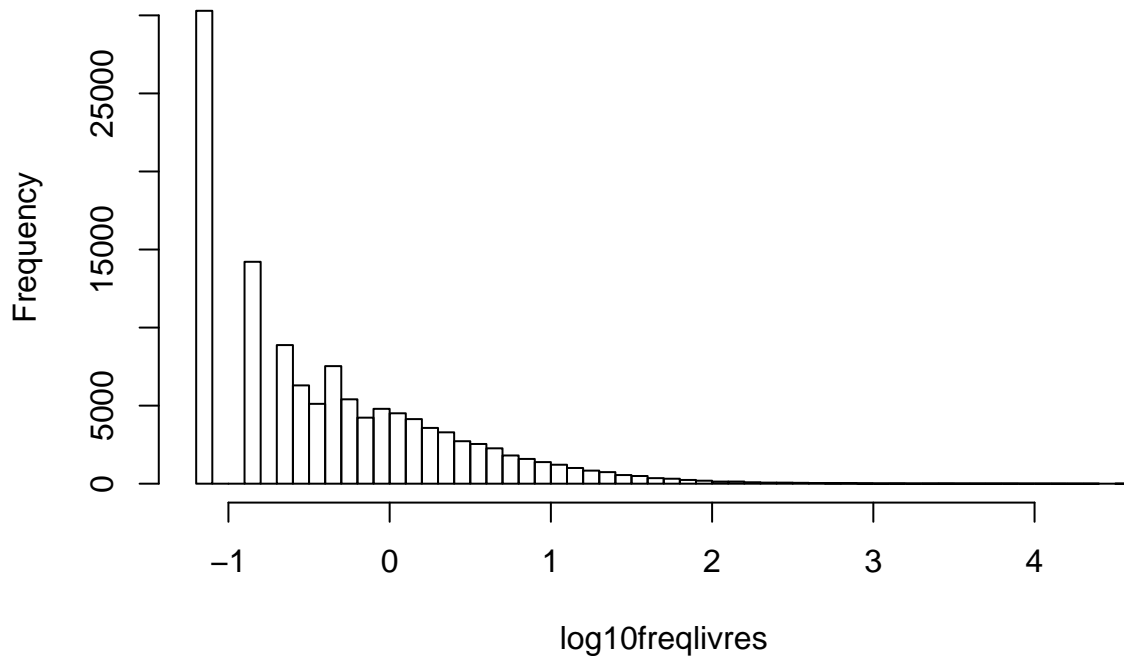
```
lexique <- get_lexique382()
```

Comme l'échelle des fréquences est extrêmement large, et dominée par les mots de très basse fréquence, nous compressons en passant par logarithme (ainsi, sur l'axe des x sur la figure:

- 0 = fréquence de 1 par million
- 1 = 10 par million
- 2 = 100 par million ...

```
s = min(lexique$freqlivres[lexique$freqlivres>0])
log10freqlivres = log10(lexique$freqlivres + s / 2)
hist(log10freqlivres, nclass = 50)
```

Histogram of log10freqlivres



```
s = min(lexique$freqfilms2[lexique$freqlivres2>0])
log10freqfilms2 = log10(lexique$freqfilms2 + s / 2)
```

```
hist(log10freqfilms2, nclass = 50)
```

Histogram of log10freqfilms2

