

**IRAMUTEQ**

**[iramuteq.org](http://iramuteq.org)**

# Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires

Développé par Pierre Ratinaud @ LERASS, Toulouse

Un logiciel libre construit avec des logiciels libres

- Python
- R
- Lexique 3 ([www.lexique.org](http://www.lexique.org))

Open source

# Entrée

Nécessite pré-traitement avant import

## Exemple d'entrée :

0001 \*an\_2000 \*sexe\_2 \*age\_6 \*cspc\_3 \*nivet\_5 \*sympeco\_1  
JE CROIS QUE C'EST LE RECHAUFFEMENT DE L'ATMOSPHERE DUE A LA POLLUTION

0003 \*an\_2000 \*sexe\_2 \*age\_3 \*cspc\_3 \*nivet\_3 \*sympeco\_2  
COUCHE D OZONE POLLUTION

0005 \*an\_2000 \*sexe\_1 \*age\_2 \*cspc\_3 \*nivet\_3 \*sympeco\_2  
RECHAUFFEMENT DE LA PLANETE

0006 \*an\_2000 \*sexe\_2 \*age\_4 \*cspc\_1 \*nivet\_4 \*sympeco\_2  
ÇA NE ME DIT RIEN NON

0007 \*an\_2000 \*sexe\_1 \*age\_5 \*cspc\_1 \*nivet\_5 \*sympeco\_1  
C'EST LE RECHAUFFEMENT DE LA PLANETE DU AU DEVELOPPEMENT DES INDUSTRIES

[...]

(Exemple : résultat d'enquête ADEME réalisée par Daniel Boy)

## Format des variables associées au segment de texte :

```
0003 *an_2000 *sexe_2 *age_3 *cspc_3 *nivet_3 *sympeco_2  
COUCHE D OZONE POLLUTION
```

```
N° de segment *variable1_valeur1 *variable2_valeur2 [...]  
segment de texte
```

## Paramètres d'import de l'exemple

- Langue : français
- Marqueur de texte = 0000
- Segment de textes = paragraphes

# Fonction n°1 : Statistiques

Statistiques textuelles descriptives basiques

## Résumé de l'import

- Nombre de textes
- Nombre d'occurrences
- Nombre de formes
- Nombre d'hapax
- Moyenne d'occurrences par texte
- Visualisation fréquence/rang : loi de Zipf

Forme	Freq. 	Types
rechauffement	3118	nr
pollution	2365	nom
coucher	2240	ver
ozone	2093	nom
gaz	1993	nom
planete	1810	nr
terre	1168	nom
climatique	1100	adj
effet	1034	nom
serrer	965	ver

Fréquences de termes lemmatisés et catégorisés grammaticalement

# **Fonction n°2 :**

## **Nuage de mots**

## **Paramètres du nuage de mots**

- Lemmatisation
- Formes actives ou supplémentaires
- Nombre de mots dans le nuage
- Taille maximum et minimum des mots
- Liste des mots à inclure dans le nuage



# **Fonction n°3 :**

# **Analyse de similitudes**

Réseau de cooccurrences ou de similarités de mots

## Paramètres de l'analyse de similitudes

- Indice de similitude : **cooccurrences**, Jaccard, Dice, Chi-2 etc...
- Spécialisation du graphe : random, cercle, **Fruchterman-Reingold**
- Totalité des liens, poids minimum des liens ou **arbre maximum**
- Clustering en communautés : 8 algos dispo (**betweenness**)
- Liste des mots à inclure dans le graphe



# Fonction n°4 : Spécificités et AFC

Produit une **analyse factorielle des correspondances** sur un tableau de contingence qui croise formes actives et les variables choisies

## **Paramètres des spécificités**

- Formes actives et/ou supplémentaires
- Variables ou modalités choisies
- Indice : loi hypergéométrique ou chi-2
- Fréquence minimale des termes considérés



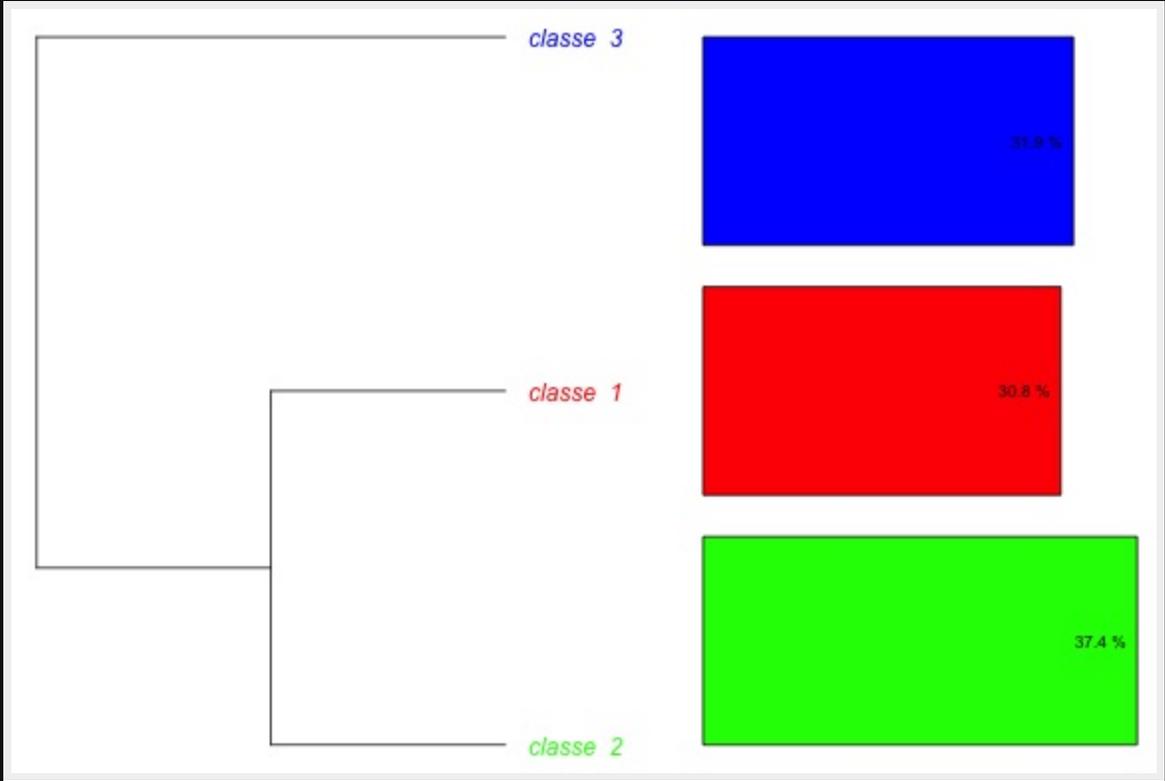


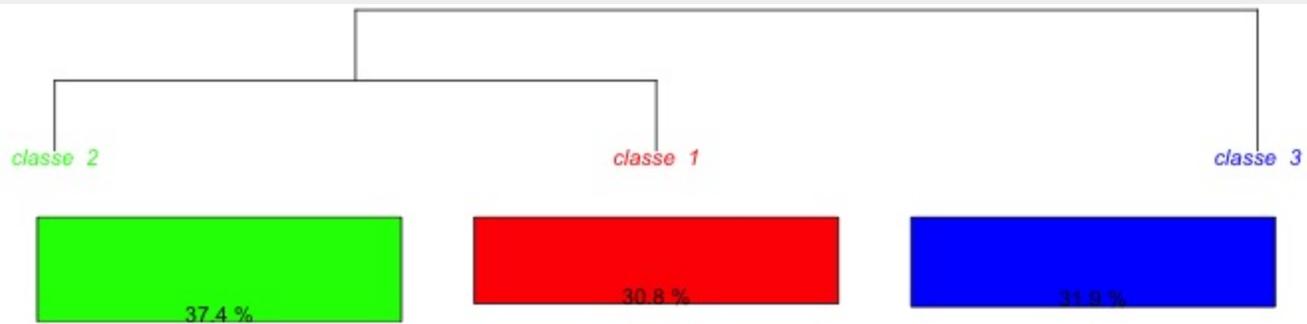
# Fonction n°5 : Clustering

Cette analyse propose une classification hiérarchique descendante selon la méthode décrite par Reinert

## Paramètres de la classification

- Trois modalités : double sur RST (méthode héritée d'Alceste), simple sur segments ou sur texte
- Nombre de classes terminales
- Nombre de formes (mots) considérées dans le corpus
- **MODE PATATE** (optimisation pour gros corpus)





coucher  
 ozone  
 rayon  
 soleil  
 trou  
 atmosphere  
 passer  
 chaleur  
 solaire  
 rayonnement  
 empeche  
 protege  
 uv  
 co2  
 dozone  
 filtre  
 terre  
 bloquer  
 accumulation  
 repartir  
 ultra  
 violet  
 retenir  
 laisser  
 accumuler  
 protection  
 carbonique  
 gaz  
 rester  
 detruit  
 traverser  
 proteger

rechauffement  
 planete  
 fonte  
 climatique  
 climat  
 glacier  
 glace  
 changement  
 eau  
 saison  
 faune  
 mer  
 flore  
 montee  
 niveau  
 dereglement  
 fondre  
 rechauffement  
 nord  
 froid  
 banquise  
 fonder  
 chaud  
 modification  
 temperature  
 planete  
 inondation  
 catastrophe  
 pole  
 oceans  
 neige  
 temperatures

voiture  
 usine  
 pollution  
 polluer  
 produit  
 air  
 dechets  
 industrie  
 echappement  
 chose  
 vehicules  
 automobile  
 pot  
 chauffage  
 chimique  
 gens  
 echappements  
 transport  
 rejeter  
 industriel  
 respirer  
 bombe  
 utilisation  
 fumees  
 nucleaire  
 utiliser  
 entreprise  
 toxique  
 centrale  
 echappement  
 nature  
 ville



# Sorties

Iramuteq met à disposition de nombreuses sorties en format .csv dans le dossier local correspondant au corpus courant

# Autres outils d'analyses textuelles

- **TXM** : agrège beaucoup des fonctions d'analyse et de statistiques textuelles disponibles ailleurs
- **CorText** : représentation de réseaux de cooccurrences, extraction de concepts saillants
- **Gargantext** : directement branché sur les API de sites tels que wikipédia ou europresse et factiva
- **Voyant Tools** : facile et rapide à prendre en main

# Merci

Un groupe MATE-SHS s'est formé lors de l'ANF de Fréjus autour des données textuelles - contact : [diego.antolinosbasso@sciencespo.fr](mailto:diego.antolinosbasso@sciencespo.fr)