

Logiciel IRaMuTeQ :

EXPLORATION D'UN CORPUS TEXTUEL À L'AIDE DE LA LEXICOMÉTRIE

ANF TDM 2022 - Exploration documentaire et extraction d'information - 3-5 oct. 2022



Stéphéline Ginguené,
Pôle de Psychologie Sociale
de l'Unité INSERM U1296
Université Lumière Lyon 2



Unité U1296 : Radiations
Défense, Santé, Environnement
Lyon & Brétigny-sur-Orge, France

IRaMuTeQ : Qu'est-ce que c'est ?

- Interface de **R** pour les **Analyses Multidimensionnelles de Textes et de Questionnaires**
- Développé par Pierre Ratinaud (Loubère & Ratinaud, 2014 ; Ratinaud, 2009)
- Logiciel libre
- Fonctionnement avec Python et R
- Lexicométrie

Et la lexicométrie ?

- « Analyse informatisée du discours et du lexique »
(Souti, 2015, p.123)
 - Étude du lexique via des méthodes statistiques
- Visualisation, compréhension de la structure et des mondes lexicaux au sein d'un corpus
- Quelques notions :
 - Occurrences
 - Formes
 - Lemmatisation (≠ stemming)
 - Formes actives

Représentation de la
maladie mentale ?

« Selon les statistiques, il y a une personne sur cinq qui est déséquilibrée. S'il y a quatre personnes autour de toi et qu'elles te semblent normales, c'est pas bon. »
→ 32 occurrences

Selon les statistiques **il y a** une personne sur cinq qui **est** déséquilibrée s **il y a** quatre personnes autour de toi et qu'elles te semblent normales c **est** pas bon → 28 formes

Selon les statistique **il y avoir** une **personne** sur cinq qui **être** déséquilibré s **il y avoir** quatre **personne** autour de toi et qu'elles te sembler normal c **être** pas bon → 27 formes

statistique **personne** déséquilibré **personne** sembler normal bon → 6 formes actives

QUELLES ANALYSES ?

- **Nuage de mots**
- **CHD**
- **Analyse de Similitudes**
- Analyses de spécificités
- Analyse Factorielle de Correspondance
- ...

Deuxième analyse :
LA CLASSIFICATION
HIÉRARCHIQUE
DESCENDANTE (CHD)

Au commencement, il y a le
BIG BANG...

DEUXIÈME ANALYSE : LA CLASSIFICATION HIÉRARCHIQUE DESCENDANTE (CHD)

- Méthode Reinert (Reinert, 1999)
- Segmentation du texte
- Création d'une matrice
- Éclatement
- Ré-éclatement, ...

- Génère des classes de mots
- Mondes lexicaux qui renvoient à des appartenances sociales (Marchand, 2007)

Exemple (si phrase = segment) :

Segment 1 : Selon les **statistiques**, il y a une **personne** sur cinq qui est **déséquilibrée**.

Segment 2 : S'il y a quatre **personnes** autour de toi et qu'elles te **semblent normales**, c'est pas **bon**.

	Statistique	Personne	Déséquilibré	Sembler	Normal	Bon
Segment 1	1	1	1	0	0	0
Segment 2	0	1	0	1	1	1

LA CHD EN APPLICATION ... SUR LA PRESSE



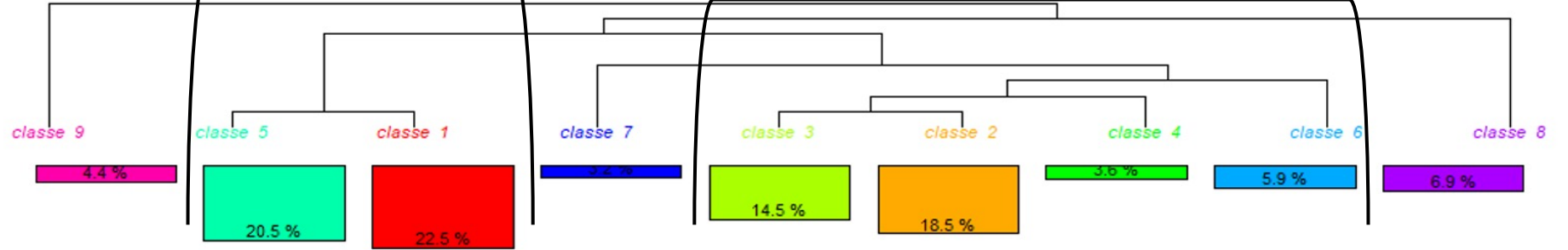
Cluster 5 : Débordements liés au démantèlement et émotions liées

Cluster 4 : Classes descriptives du démantèlement

Cluster 3 : Buts et obstacles de la migration

Cluster 2 : Expérience de la personne en situation de migration

Cluster 1 : Situation des mineurs et question du regroupement familial



pompier
incendie
maîtrisé
habitation
blesser
gaz
maisonnette
communauté
tympaan
simultanément
orange
feu
caillassés
blessure
éteindre

déclencher
sweat
anxiété
bonbonne
tradition
ors
sale
extincteur
intervenir
panique
esthétique
animé
prévention
04h00
brûler
hôpital
combattre

évacuation
caler
jungle
commencer
débuter
démantèlement
lundi
matin
déblaiement
mardi
préfète
opération
total

bidonville
mercredi
afp
début
démarrer
oat
orchestre
migrant
calme
destruction
poursuivre
clair
rythme
millier
deuxième
nord
achever
construire

centre
accueil
orientation
provisoire
hangar
filer
cao
bus
répartir
valise
adulte
attente
majeur
autocar
attendre
vulnérable
orienter
sas
mineur
cap
transit
région
conteneur
groupe
bord
08h00
acheminer
situer
grille
prévoir
seul

soudan
afghanistan
erythrée
manche
gagner
traverser
sporadique
rassemblement
grenade
heur
venir
habiter
naître
femme
déterminer
devenir
lacrymogène
rêve
enfant
boueux
confiance
côte
homme
année
érythrée
produit
préserver
mois

squat
ordre
ville
maison
force

caravane
démanteler
poser
soupirer
ami
voiture
proche
content
ancien
mosquée
jeune

commerçant

bas
église
perdre
coup
retrouver
mot
immatriculer
empêcher
multiplier

asile
demander
droit
pays
association
politique
solution
européen
cas
social
état
demandeur
comprendre
condition
humain
papier
fuir
trouver
retourner
guerre
traiter
risquer
migration
doute
travailleur
expulser
allemagne
éligible
économique
interroger
empreinte

barreau
bénéficiaire
samedi
infirmier
syndicat
conseil
maire
accord
initial
valeur

perplexe
avocat

garantie
adjoint
franco_britannique
touquet
national
introduire
sgp
satisfaction
outre_manche
calais
difficulté

auberge
gens
partir

vice_président
représentant
ndlr
tenter
rumeur
revenir

maraudeur
infondé
exhorter
arrivant

difficile
inquiet

tentant
ramadan
fête
compter
pakistanaï
persister
irréductible

royaume
unir
ministre
attache

familial
isolé
établir
londres
accueillir
intérieur
pomme
discord
maîtrise

mineur

dossier
percée
bretagne
logement
britannique
exaspération
aboès
annoncer
euro
dérouler
insécurité
débat
empoisonner

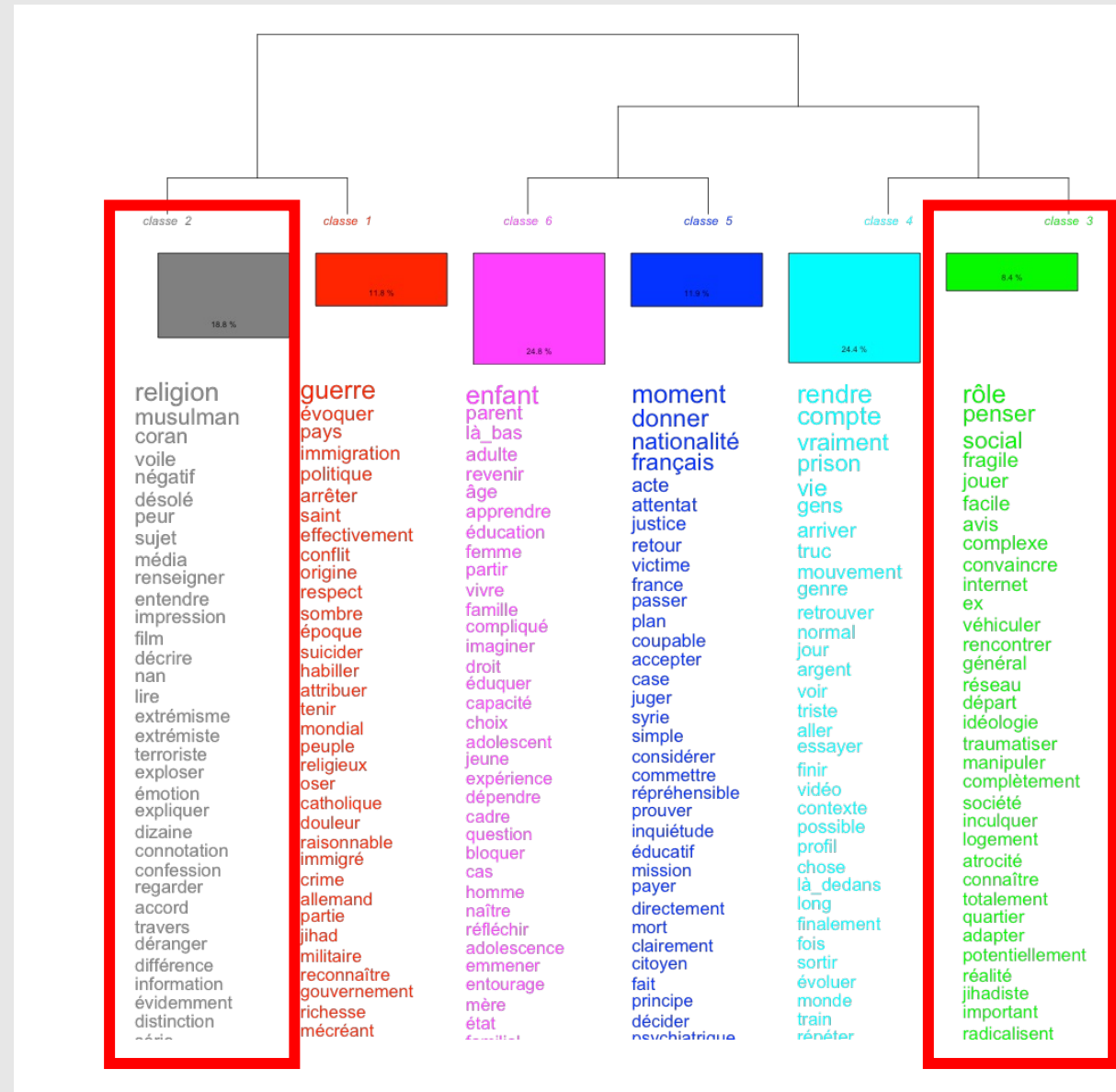


LA CHD EN APPLICATION... SUR DES ENTRETIENS

- Religion et individus de confession musulmane dans le lexique employé
(Classe 2 - 18.81%)
 - Associé au voile, à la négativité, à la peur, aux médias...
 - Et à l'extrémisme/terrorisme
- Traits stéréotypiques :
 - Fragiles/endocrinés/manipulés...
(Classe 3 - 8.4%)

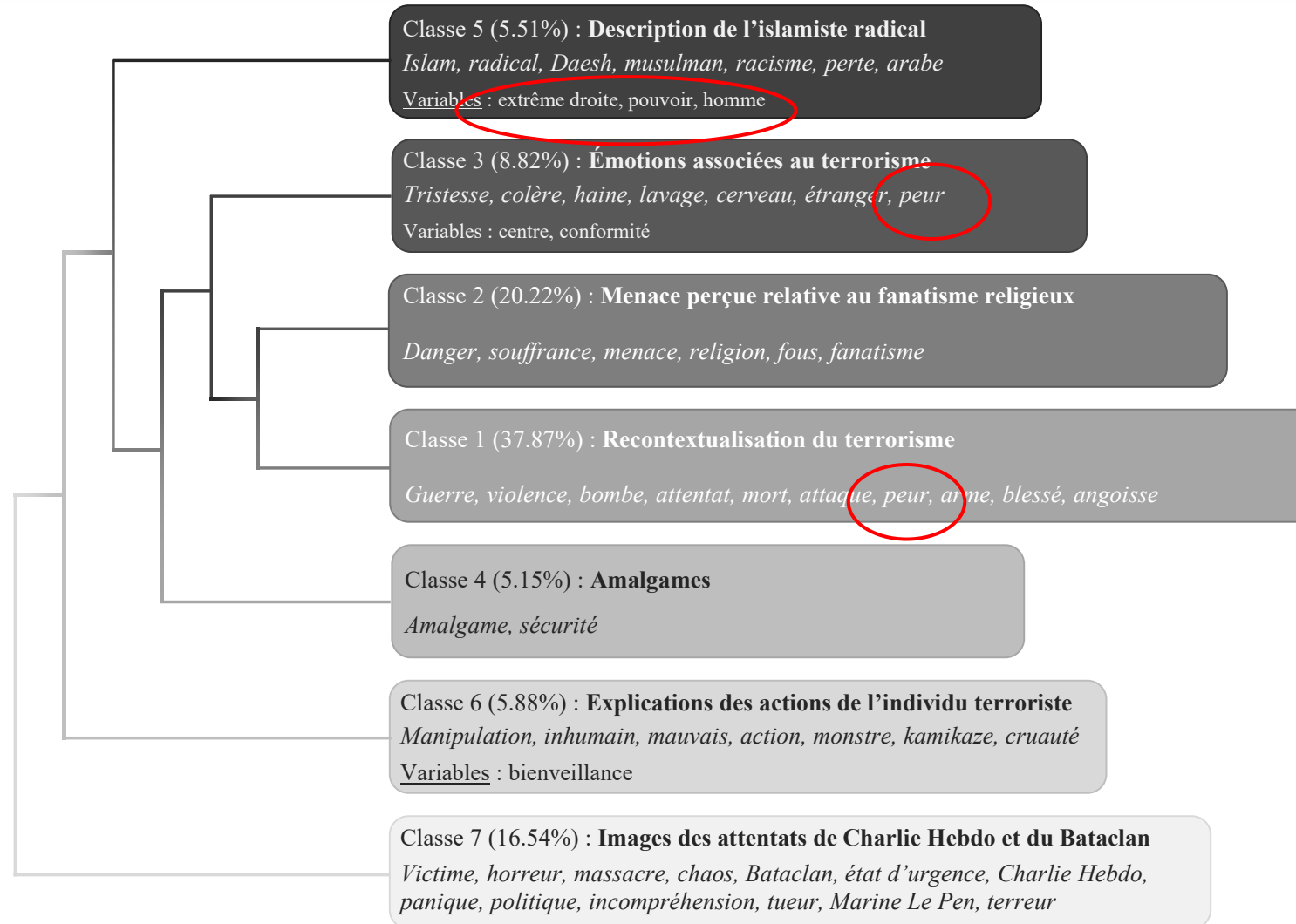
Ex de verbatim :

« ça dépend de l'intention, ça dépend de si c'est des personnes fragiles qui ont été manipulées, là elles sont un peu plus excusables »
(F30 – Sans enfant)



85.46% du corpus classé

LA CHD EN APPLICATION... SUR DES ASSOCIATIONS LIBRES



CE QU'ON EN CONCLUT...

- Applications pratiques diverses
- Pour toute personne qui souhaite explorer les champs lexicaux au sein d'un corpus textuel
 - Quelques exemples :
Presse, tweets, entretiens, études de cas, questions ouvertes, discours, textes de loi, littérature, forums de discussion, sites internet, méta-analyse...
- Complémentaire des analyses thématiques et sémantiques

CE QU'ON EN CONCLUT...

- De nombreux intérêts :
 - Richesse du dictionnaire (français, anglais, espagnol, portugais)
 - Facilité d'utilisation
 - Tarif
 - ...
- Limites de l'outil : pas de prise en compte de la ponctuation par exemple, ni de la sémantique, ...
- Pas de maîtrise du code...
 - Interface logicielle
- ... mais Open source !

EN PRATIQUE...

- Corpus des déclarations des 12 candidat-es aux présidentielles
- Variables illustratives :
 - Sexe,
 - Parti politique (nom du/de la candidat-e),
 - Hauts de France (E. Macron, Y. Jadot, F. Roussel) vs les autres



EN PRATIQUE...

EXTRACTION D'UN CORPUS

Le camp des travailleurs

Élection présidentielle 2022

Travailleuses, travailleurs,

En avril 2020, pendant le premier confinement, Macron expliquait qu'il faudrait se rappeler que le pays avait tenu grâce à celles et ceux « que nos économies reconnaissent et rémunèrent si mal ». Deux ans plus tard, le personnel des hôpitaux et des Ehpad, les aides à la personne, les ouvriers de l'agroalimentaire et de la logistique, les travailleurs des transports, les agents de nettoyage et ceux du gardiennage, l'armée des livreurs et des coursiers, les travailleurs de l'Éducation nationale, continuent d'être sous-payés et méprisés. Des travailleurs de la santé sont licenciés de fait et privés de salaire, en particulier en Guadeloupe, en Martinique et en Guyane. Tous ceux dont les conditions d'existence se dégradent doivent faire entendre leurs intérêts.

Profits en hausse et pouvoir d'achat en baisse : ça suffit !

Suppressions d'emplois, cadences infernales, salaires insuffisants : plus les grandes entreprises sont rentables, plus leurs actionnaires sont rapaces et s'attaquent aux travailleurs.

Alors qu'une minorité très riche s'offre des voyages dans l'espace, il y a 300 000 sans-domicile-fixe et plusieurs millions de mal-logés. Des étudiants et des retraités ne survivent que grâce à l'aide alimentaire.

Les profits se sont envolés, pour le plus grand bonheur des riches, des grands actionnaires, c'est-à-dire de ceux qui ne font rien de leurs dix doigts, quand ceux qui se tuent à la tâche s'appauvrissent et sont précarisés.

La flambée des prix des carburants, du gaz, de l'électricité et de nombre de produits de première nécessité ne gêne pas les grandes fortunes. Mais, pour des millions de femmes et d'hommes, ce sont de nouvelles privations insupportables.

CODAGE DU CORPUS

```
**** *candidat_nathart *genre_f *hdf_non
```

Le camp des travailleurs

Élection présidentielle 2022

Travailleuses, travailleurs,

En avril 2020, pendant le premier confinement, Macron expliquait qu'il faudrait se rappeler que le pays avait tenu grâce à celles et ceux « que nos économies reconnaissent et rémunèrent si mal ». Deux ans plus tard, le personnel des hôpitaux et des Ehpad, les aides à la personne, les ouvriers de l'agroalimentaire et de la logistique, les travailleurs des transports, les agents de nettoyage et ceux du gardiennage, l'armée des livreurs et des coursiers, les travailleurs de l'Éducation nationale, continuent d'être sous-payés et méprisés. Des travailleurs de la santé sont licenciés de fait et privés de salaire, en particulier en Guadeloupe, en Martinique et en Guyane. Tous ceux dont les conditions d'existence se dégradent doivent faire entendre leurs intérêts.

Profits en hausse et pouvoir d'achat en baisse : ça suffit !

Suppressions d'emplois, cadences infernales, salaires insuffisants : plus les grandes entreprises sont rentables, plus leurs actionnaires sont rapaces et s'attaquent aux travailleurs.

Alors qu'une minorité très riche s'offre des voyages dans l'espace, il y a 300 000 sans-domicile-fixe et plusieurs millions de mal-logés. Des étudiants et des retraités ne survivent que grâce à l'aide alimentaire. Les profits se sont envolés, pour le plus grand bonheur des riches, des grands actionnaires, c'est-à-dire de ceux qui ne font rien de leurs dix doigts, quand ceux qui se tuent à la tâche s'appauvrissent et sont précarisés.

La flambée des prix des carburants, du gaz, de l'électricité et de nombre de produits de première nécessité ne gêne pas les grandes fortunes. Mais, pour des millions de femmes et d'hommes, ce sont de nouvelles privations insupportables.



À SAVOIR (principes de base)

- Lignes étoilées
- Introduction d'un texte : ****
- Variables :
 - *nomdelavariabile_modalité
 - ex : *genre_f
- Sous-corpus (questions, paragraphes, thématiques, ...) :
 - *nomdusouscorpus_modalité
 - ex : -*question_1

**** *candidat_nathart *genre_f *hdf_non

Le camp des travailleurs
Élection présidentielle 2022
Travailleuses, travailleurs,
En avril 2020, pendant le premier confinement, Macron expliquait qu'il faudrait se rappeler que le pays avait tenu grâce à celles et ceux « que nos économies reconnaissent et rémunèrent si mal ». Deux ans plus tard, le personnel des hôpitaux et des Ehpad, les aides à la personne, les ouvriers de l'agroalimentaire et de la logistique, les travailleurs des transports, les agents de nettoyage et ceux du gardiennage, l'armée des livreurs et des coursiers, les travailleurs de l'Éducation nationale, continuent d'être sous-payés et méprisés. Des travailleurs de la santé sont licenciés de fait et privés de salaire, en particulier en Guadeloupe, en Martinique et en Guyane. Tous ceux dont les conditions d'existence se dégradent doivent faire entendre leurs intérêts.

**** *candidat_yanjad *genre_h *hdf_oui

Faire face avec Yannick Jadot
Une présidence pour le climat
et le vivant, dans la paix, la dignité et la justice
Madame, Monsieur, chers compatriotes,
Les 10 et 24 avril, par votre vote, vous allez choisir le destin de la France et de l'Europe. Vous allez choisir dans quelle société vous voulez vivre et faire grandir vos enfants.
La guerre en Europe rappelle l'exigence de protéger nos libertés, notre démocratie, nos valeurs. L'épreuve de la pandémie l'a montré, nous devons reprendre le contrôle de notre destin pour protéger la vie. Les impacts du changement climatique, qui nous touchent déjà, exigent d'agir. Vite. Dès demain. Les énergies renouvelables sont les seules énergies de paix et de sécurité. Seule l'agriculture biologique construit notre souveraineté alimentaire.
Le quinquennat qui s'achève a été celui de l'inaction climatique. Il a aggravé les inégalités et la situation de notre jeunesse. Canicules, inondations, sécheresses, mégafeux, pollution de l'air, malbouffe, extinction des espèces, souffrance animale... Le statu quo, c'est le chaos.

RÉFÉRENCES

Loubère, L., & Ratinaud, P. (2014). *Documentation IRaMuTeQ 0.6 alpha 3 version 0.1. Iramuteq.*

http://www.iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf

Marchand, P. (2007). Concepts, méthodes, outils. Dans *Analyse statistique de données textuelles en sciences de gestion* (Gauzente, C., Peyrat-Guillard, D., p. 49-70). Caen : EMS Management et société.

Ratinaud, P. (2009). *Iramuteq : Interface de R pour les analyses multidimensionnelles de textes et de questionnaires.* <http://www.iramuteq.org>

Reinert, M. (1999). Quelques interrogations à propos de l'objet d'une analyse de discours de type statistique et de la réponse Alceste. *Langage et Société*, 90, 57–70.

Souti, C. (2015). La lexicométrie ou l'analyse du discours assistée par ordinateur: ce que l'informatique et les mathématiques peuvent apporter à la littérature et la linguistique, *Expressions*, 1, 122-130.