

Introduction à la textométrie avec le logiciel IRaMuTeQ

Natacha Souillard

▶ To cite this version:

Natacha Souillard. Introduction à la textométrie avec le logiciel IRaMuTeQ. Semaine Data-SHS, Dec 2022, Aix en Provence, France. hal-03942366

HAL Id: hal-03942366 https://amu.hal.science/hal-03942366

Submitted on 17 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Introduction à la textométrie avec le logiciel IRaMuTeQ

Natacha Souillard, MCF en Sciences de l'Information et de la communication

IMSIC, Aix-Marseille Université





Introduction à la textométrie avec le logiciel IRaMuTeQ

- 1.La textométrie, origines et épistémologie
- 2. Termes, définitions de base
- 3. Opérations initiales
- 4. Iramuteq, présentation générale
- 5. La Classification Hiérarchique Descendante
- 6.Quelques exemples



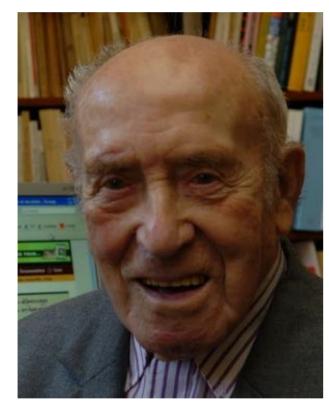
- Une approche interdisciplinaire du discours (Analyse du discours), plutôt comme manifestation de phénomènes sociaux que comme pratique sociale.
- Une certaine matérialité textuelle :
 - Focalisation sur des marqueurs textuels et lexicaux spécifiques, en étudier le déploiement dans des contextes discursifs distincts.
 - « Ce qui est dit et la façon de le dire » (Guilbert, 2011, p.24) les mots ne sont pas choisis au hasard.

Rechercher « dans le dit tel qu'il est dit, des traces des opérations cognitives effectuées par le sujet communicant avec autrui » (Ghiglione, 1989).

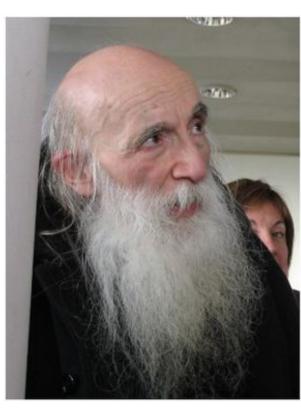
Comment « le langage se rapporte au monde par l'intermédiaire de l'esprit » (Johnson-Laird, 1988, p. 66).



« L'école française d'Analyse de discours » (St-Cloud, ENS Lyon)



Charles Muller, linguiste => TAL



Jean-Paul Benzécri, statisticien



Antoine Prost, historien

Publication emblématique: Demonet M., Geffroy A., Gouaze J., Lafon P., Mouillaud M., Tournier M. (1975) - *Des tracts en Mai 68. Mesures de vocabulaire et de contenu*, Armand Colin et Presses de la Fondation Nat. des Sc. Pol., Paris.



Des approches statistiques, plurielles et complémentaires :

• Approches plutôt inductives :

Etudier la fréquence, la présence ou l'absence de termes, les régularités/irrégularités, les rapprochements et les éloignements (distance), les convergences et divergences, l'environnement syntagmatique, l'originalité de textes, d'auteurs,...

...en fonction du lexique pris dans son environnement textuel et syntagmatique, entre linéarité et réticularité.

Approches plutôt déductives :

Repérer dans un corpus les traces de catégories prédéfinies, observer leur manifestation, les facteurs d'opposition ou de rapprochement, l'inattendu, etc.

Approches plutôt lexicales ou plutôt syntagmatiques



Le corpus, un "observable" et un "observatoire" (D.Mayaffre)

Observable:

- Construit à partir de données invoquées pour avoir une prise directe sur une partie de la réalité de l'objet étudié.
- Hypothèses, caractéristiques, contextualisation, critères d'homogénéité (signifiance, acceptabilité, complétude, exploitabilité, Bommier-Pincemin, 1999).
- Clôture ?

```
**** *var1_mod1 *var2_mod3 *.....
```

texte texte

texte texte ...



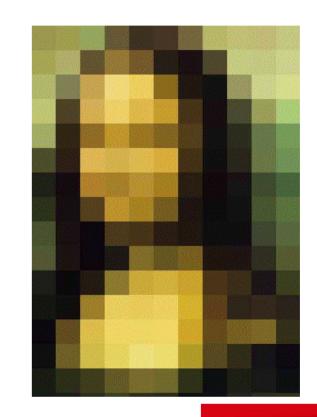
Le corpus, un "observable" et un "observatoire" (D.Mayaffre)

Observatoire:

- Objectivation : aborder les objets discursifs comme des phénomènes, sociaux
- Approche Statistique :
 - Limiter la part d'échantillonnage a priori : corpus massifs, exploration, effets massifs de structure (invisibles à « l'œil nu »), sérendipité
 - Déplacement du travail interprétatif.
 - Triangulation
 - Réticularité : texte, co-texte, contexte (intra/inter-corpus, intra/intertextualité)

Ce que la textométrie ne fait pas

- Le rapport aux données ne restitue pas la réalité, ni de l'objet, ni de la réception.
- Le choix d'une description formelle qui permet de reconnaître l'objet par rapport aux autres.
- Identification par des caractéristiques, informations, jugées structurantes *via* le lexique.





Violence des Gilets Jaunes?





2. Termes & Définitions

« Les questions que se donne la statistique lexicale sont les suivantes : « quels sont **les textes les plus semblables** en ce qui concerne le vocabulaire et la fréquence des formes utilisées ? Quelles sont **les formes qui caractérisent** chaque texte, par leur présence ou leur absence ? »

(Lebart & Salem, 1994, p.135).

La lexicométrie regroupe « toute une série de méthodes qui permettent d'opérer des **ré-organisations formelles de la séquence textuelle** et des analyses statistiques portant sur le vocabulaire à partir d'une **segmentation** » (Salem, 1986)

=> Segmentation du lexique, segmentation des textes ; Tableau lexical (formes * textes) ; Réticularité/Linéarité



2. Termes & Définitions

- Forme (graphique) =/≠ « mot » (erreurs de graphie d'un même mot ; mise à distance du sens => « individu »)
- Occurrence: apparition d'une forme.
- => Le nombre total d'occurrences = taille du corpus
- Hapax : une forme qui n'apparaît qu'une seule fois dans le corpus
- => Nombre hapax/nombre occurrences = indice de richesse du vocabulaire



Le corpus

```
**** *var1_mod1 *var2_mod3 *.....
```

texte texte

```
**** *var1_mod2 *var2_mod1
```

texte texte

toyto toyto toyto toyto toyto toyto toyto toyto toyto

Pour Iramuteq:

- Format texte brut
- UTF-8 de préférence
- pas d'accents, d'espaces, de caractères spéciaux (+ / * etc.)
- On peut intégrer des thématiques



La segmentation du lexique (Tokenisation)

- Une suite de caractères bornée par deux caractères délimiteurs est une occurrence :
- Délimiteurs (Lexico): espace, retour à la ligne, [(« ,.;?:!'/-_ »)]
- Caractères pleins (Iramuteq):

```
a-z;A-Z;0-9
àÀâÂäÄáÁåÅãéÉèÈêÊëËìÌîÎïÏíÍóÓòÒôÔöÖõÕøØùÙûÛüÜúÚçÇßœ
Œ'ñÑ[.:,;!?'-]
```

- Deux suites identiques constituent deux occurrences d'une même forme graphique
- L'ensemble est regroupé dans un index des formes et de leurs occurrences



Index des formes

12528	de	1195	С		530	sera	341	ai		développement
8324	la	1188	je		528	doit	323	travail	231	économie
6211	I	1183	ne		527	aussi	310	entre	229	deux
5815	et	1127	par		509	ont	306	si	227	enfin
5217	les	1117	ce		494	français	297	économique	226	encore
4908	le	1074	sur		479	У	290	aujourd	226	temps
4631	à	985	qu		462		290	hui	222	ensemble
4435	des	908	france		453	etat	288	dont	221	vie
3832	d	855	S		447	sans	283	sociale	220	société
3051	est	838	aux		434	ou	282	on	219	depuis
2982	en	838	n		425	comme	280	seront	216	ceux
2799	que	816	nos		422	ces	278	monde	215	donc
2441	une	810	gouvernem	ent	422	tout	278	république	210	toutes
2425	nous	803	avec		421	son	266	fait	209	soit
2273	qui	744	mais		413	avons	265	loi	208	droit
2142	un	711	elle		410	ses	265	où	208	sécurité
2060	pour	697	cette		409	même	264	contre	207	ainsi
2024	du	695	vous		406	été	263	leurs	206	elles
1977	dans	693	politique		400	faire	262	action	206	moyens
1809	il	667	se		390	ils	256	europe	203	cet
1410	au	651	être		386	faut	243	effort	202	autres
1393	notre	647	sont		375	entreprises	241	peut	202	cela
1368	plus	633	leur		362	emploi	236	nationale	199	mesures
1275	pas	603	pays		346	bien	235	avenir	197	jeunes
1214	a	533	tous		342	sa	235	président	195	croissance
·										

P. MARCHAND, corpus discours d'investiture premiers ministres Vème République



Index des formes

_		 _							
12528	de	1195	С		sera	341	ai	233	développement
8324	la	1188	je	528	doit	323	travail	231	économie
6211	l	1183	ne	527	aussi	310	entre	229	deux
5815	et	1127	par	509	ont	306	Si	227	enfin
5217	les	1117	ce	494	français	297	économique	226	encore
4908	le	1074	sur	479	У	290	aujourd	226	temps
4631	à	985	qu	462		290	hui	222	ensemble
4435	des	908	france	453	etat	288	dont	221	vie
3832	d	855	S	447	sans	283	sociale	220	société
3051	est	838	aux	434	ou	282	on	219	depuis
2982	en	838	n	425	comme	280	seront	216	ceux
2799	que	816	nos	422	ces	278	monde	215	donc
2441	une	810	gouvernement	422	tout	278	république	210	toutes
2425	nous	803	avec	421	son	266	fait	209	soit
2273	qui	744	mais	413	avons	265	loi	208	droit
2142	un	711	elle	410	ses	265	où	208	sécurité
2060	pour	697	cette	409	même	264	contre	207	ainsi
2024	du	695	vous	406	été	263	leurs	206	elles
1977	dans	693	politique	400	faire	262	action	206	moyens
1809	il	667	se	390	ils	256	europe	203	cet
1410	au	651	être	386	faut	243	effort	202	autres
1393	notre	647	sont	375	entreprises	241	peut	202	cela
1368	plus	633	leur	362	emploi	236	nationale	199	mesures
1275	pas	603	pays	346	bien	235	avenir	197	jeunes
1214	а		tous	342	sa	235	président	195	croissance

P. MARCHAND, corpus discours d'investiture premiers ministres Vème République



La lemmatisation

« l'opération de regroupement qui rassemble les formes différentes appartenant au même vocable - ce que font précisément les auteurs de dictionnaires quand ils établissent leur nomenclature » Etienne Brunet :

http://lexicometrica.univ-paris3.fr/article/numero2/brunet2000.html

- Des formes initiales aux formes réduites :
 - Noms et adjectifs => racine
 - Verbes => infinitif
- Étiquetage grammatical (plus ou moins bon) en fonction de l'outil que l'on utilise et du dictionnaire dont on dispose (https://cental.uclouvain.be/treetagger/)
- Reconnaissance des locutions, expressions (dictionnaire d'expressions)
- Une intervention manuelle toujours possible sur les dictionnaires
- Ex : « Commune » ; « Paris » (désambiguïsation manuelle des homographes)



Distinction formes actives et formes supplémentaires (mots-outils)

▼			Clé	ės d'analyse			
				lés d'analyse			
			0-elimine ; 1-ac	tive ; 2=supplémentaire			
Adjectif	1	\$	voir liste	Conjonction	2	÷	voir liste
Adjectif démonstratif	2	ĵ.	voir liste	Formes non reconnues	1	\$	voir liste
Adjectif indéfini	2	÷	voir liste	Nom commun	1	‡	voir liste
Adjectif interrogatif	2	÷	voir liste	Nom supplémentaire	2	\$	voir liste
Adjectif numérique	2	\$	voir liste	Onomatopée	2	-	voir liste
Adjectif possessif	2	\$	voir liste	Pronom démonstratif	2	^	voir liste
Adjectif supplémentaire	2	÷	voir liste	Pronom indéfini	2	1	voir liste
Adverbe		\$	voir liste	Pronom personnel	2	1	voir liste
dverbe supplémentaire	2	÷	voir liste				
Article défini	2	‡	voir liste	Pronom possessif	2		voir liste
Article indéfini	2	‡	voir liste	Pronom relatif	2	Ŷ.	voir liste
Auxiliaire	2	‡	voir liste	Préposition	2	÷	voir liste
Chiffre	2	‡	voir liste	Verbe	1	‡	voir liste
				Verbe supplémentaire	2	÷	voir liste

adv_sup : environ,
encore, ensuite,
autour...
nom_sup : an, ans,
alentours, bien,
biens...
ver_sup : faire,
pouvoir, aller, etc

Là encore, des focales d'observation, des hypothèses, etc : R. Ghiglione, psychosociologue (« un foetus » /« l' » enfant) ; E. Brunet, linguiste (la + « ion » = indicateur d'abstraction) (TXM).

Étiquetage (imparfait)

de	426235 pre	au	48626art_def	département	2218 nom
la	259114art def	par	47445 pre	salarië	2217 ad
	184719art def	pas	40592 adv_sup	borne	2216 nom
le	176677 art def	ce	35715pro dem	concerner	2215ver
ëtre	163412 aux	plus	34780 adv sup	montrer	2214ver
et	158730 con	5	31445psa_psg	atteindre	2213ver
les	158324art def	faire	285 (6)ver sup	adjoint	2202 nom
a	151137 pre	avec	27050pre	juillet	2192 muni
des	148395 art ind	qu	25265 con	poursulvre	2.94 ver
d	131664pre	se	24540pro per	moděle	2189 nom
avoir	123244 aux	nous	23789 pro per	responsable	2175 ad
en	112780 pre	aux	23411 art def	déclarer	2173 ver
du	93212 art def	on	23376pro per	particulier	2172 ad
un	83924art ind	pouvoir	22443ver_sup	fonds	2171 nom
pour	79628pre	mals	22264con	continuer	2167 ver
une	75276 art ind	écologique	20989ad	direction	2166 nom
dans	56382 pre	О	20542 pro_dem	représenter	2162 ver
qui	56298 pro_rel	transition	20589nom	retraite	2159nom
sur	54240 gre	cette	18987 adj dem	mener	2159ver
que	53898 pro_rel	son	18546 adj pos	International	2155 adj
ii e	53808 pro_per	ou	17780 con	dispositif	2143 nom
ne	52150 adv_sup	devoir	1731 ver_sup	urtout	2136 adv_sup



Tableau lexical, partition

Textes (parties, variables, segments...)

Lexique (formes)

- Segmentation
- Reconnaissance
- Lemmatisation
- Statut :
 (actif, supplémentaire, éliminé)

« Les segments de texte : critère de taille et de ponctuation (ratio taille/ponctuation) :

2. ";" ":"
3., et espace

L'objectif : des segments de tailles homogènes en respectant le plus possible la structure du langage (analyse propositionnelle, R.Ghiglione). Présence / absence Nombre d'occurrences

Statistiques:

- Factorielles
- Classificatoires
- Arborées
- ..



Tableau lexical, partition

• nombre de parties : 37

• nombre d'occurrences : 239802

• nombre de formes : 7455

 moyenne d'occurrences par forme : 32.17

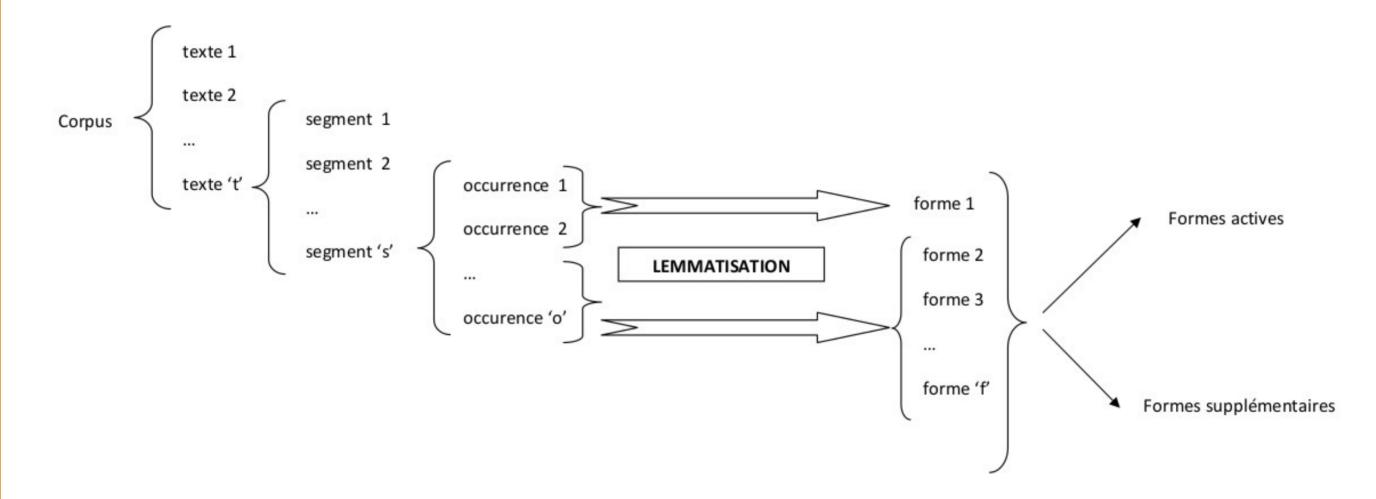
• nombre d'hapax : 2669 (1.11% des occurrences - 35.80% des formes)

moyenne d'occurrences par partie:
6481.14

P. MARCHAND, corpus discours d'investiture premiers ministres Vème République

	Num	Partie	Occurenc	Formes	Нарах	Fmax	Forme
1	1	159debré	8810	2017	1188	442	de
1	2	162pompidou	4395	1333	875	204	de
1	3	168couve	4182	1176	706	181	de
1	4	169chaban	1804	742	517	95	de
1	5	172messmer	7545	2061	1344	424	de
1	6	174chirac	8519	1997	1224	509	de
1	7	176barre	7013	1802	1135	388	de
1	8	177barre	4121	1229	821	183	de
1	9	178barre	6048	1636	1007	354	de
1	10	181 amauroy	11652	2648	1590	662	de
1	11	181bmauroy	8711	2168	1333	477	de
1	12	183mauroy	9277	2112	1229	563	de
1	13	184amauroy	5425	1507	916	287	de
1	14	184bfabius	5351	1477	953	224	de
1	15	186chirac	9416	2415	1511	569	de
1	16	187achirac	7856	1840	1109	415	de
1	17	187bchirac	6687	1837	1160	357	de
1	18	188rocard	7610	2091	1366	407	de
1	19	191 arocard	1794	631	427	71	la
1	20	191cresson	5265	1532	983	291	de
1	21	192acresson	4266	1381	925	226	de
1	22	192bbérégovoy	5484	1556	1006	227	de
1	23	192cbérégovoy	2570	801	521	122	de
1	24	193aballadur	11531	2411	1406	658	de
1	25	193bballadur	7011	1736	1053	356	de
1	26	195ajuppé	13063	2689	1495	787	de
1	27	195bjuppé	4769	1324	840	322	de
1	28	196juppé	5058	1303	787	243	de
1	29	197jospin	6715	1794	1105	368	de
1	30	202raffarin	9217	2087	1216	511	de
1	31	203raffarin	3247	971	608	171	de
1	32	204raffarin	3717	1040	640	212	de
1	33	205villepin	6595	1656	971	394	de
1	34	206avillepin	4833	1366	871	247	de
1	35	206bvillepin	3527	1027	673	175	de
1	36	207fillon	8148	2294	1546	409	de

En résumé



Daniel Pélissier, "Comment préparer l'analyse de textes de sites Web grâce à la lexicométrie et au logiciel Iramuteq?," dans Présence numérique des organisations, 14/04/2016, https://presnumorg.hypotheses.org/187



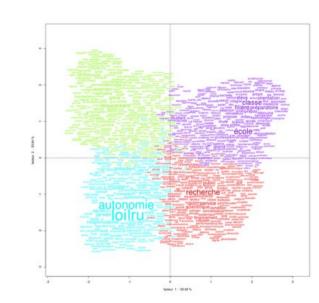
4.IRAMuTeQ, présentation générale

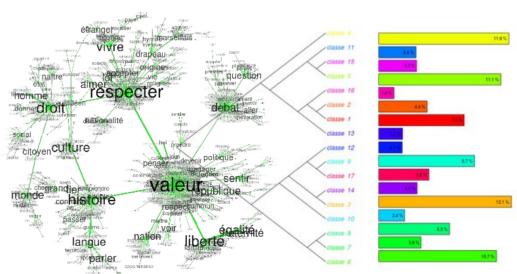
Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires

- Python, compatible Microsoft, Mac, Linux
- Licence GNU-GPL, (libre), Pierre Ratinaud, LERASS (python)
- Site internet : http://www.iramuteq.org/
- Forum: https://sourceforge.net/p/iramuteg/discussion/1068065/thread/9d256b60/

Analyses statistiques

- Formes, fréquences, caractéristiques (profil)
- Chi2 (coeff.liaison)
- Méthodes factorielles (AFC)
- Classification automatique + segments de textes caractéristiques
- Concordanciers
- Calculs de distance (connexion lexicale) et représentation arborée
- Cooccurrences et similitudes
- Statistiques, tableurs, bases de données, représentations graphiques...





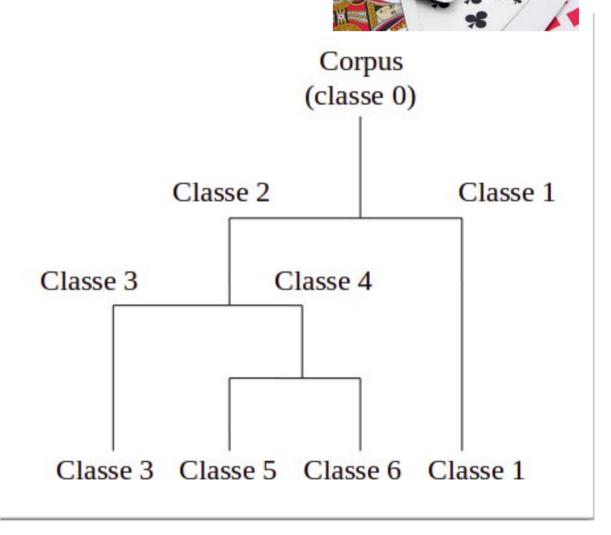


5.La Classification Hiérarchique Descendante (Reinert, Alceste)

Partition (contributions ou paragraphes)

Lexique • Tokenisation • Reconnaissance • Lemmatisation • statut	0 0 0 1 0 1 0 0 1 0	0 0 0 1 0 1 0 0 1 0	1 1 1 1 0 0 1 1 0 0	1 1 1 1 0 0 1 1 0 0	0 0 0 1 0 1 1 0 0 1 1 0	1 1 0 1 0 0 1 1 0 0

- L'arborescence n'est pas accessoire
- Équilibre nombre de classes : effet de structure/finesse du niveau d'interprétation





6.Quelques cas: La transition écologique PQN PQR

Critères d'homogénéité : Du 1^{er} août 2018 au 1^{er} octobre 2022 – Presse Quotidienne

- « transition écologique »

Presse Quotidienne Régionale (l	PQR)
La Dépéche du Midi	2433
Midi Libre	1287
L'Indépendant	646
Centre presse Aveyron	306
La Nouvelle république des Pyrénées	239
Sous-total PQR	4911
Presse Quotidienne Nationale (F	PQN)
Le Monde	1240
Le Figaro	1006
L'Humanité	468
Les Echos	1066
La Correspondance économique	1012
La Tribune	945
Libération	550
L'Agéfi Quotidien	87
La Croix	452
Aujourd'hui en France	593
Sous-total (PQN)	7419
Total articles collectés	12330

Nombre de textes: 12331

Nombre de segments de texte: 216099

Nombre de formes: 99814

Nombre d'occurrences: 7765983

Nombre de lemmes: 73003

Nombre de formes actives: 70074

Nombre de formes supplémentaires: 2929

Nombre de formes actives avec une

fréquence >= 174: 2989

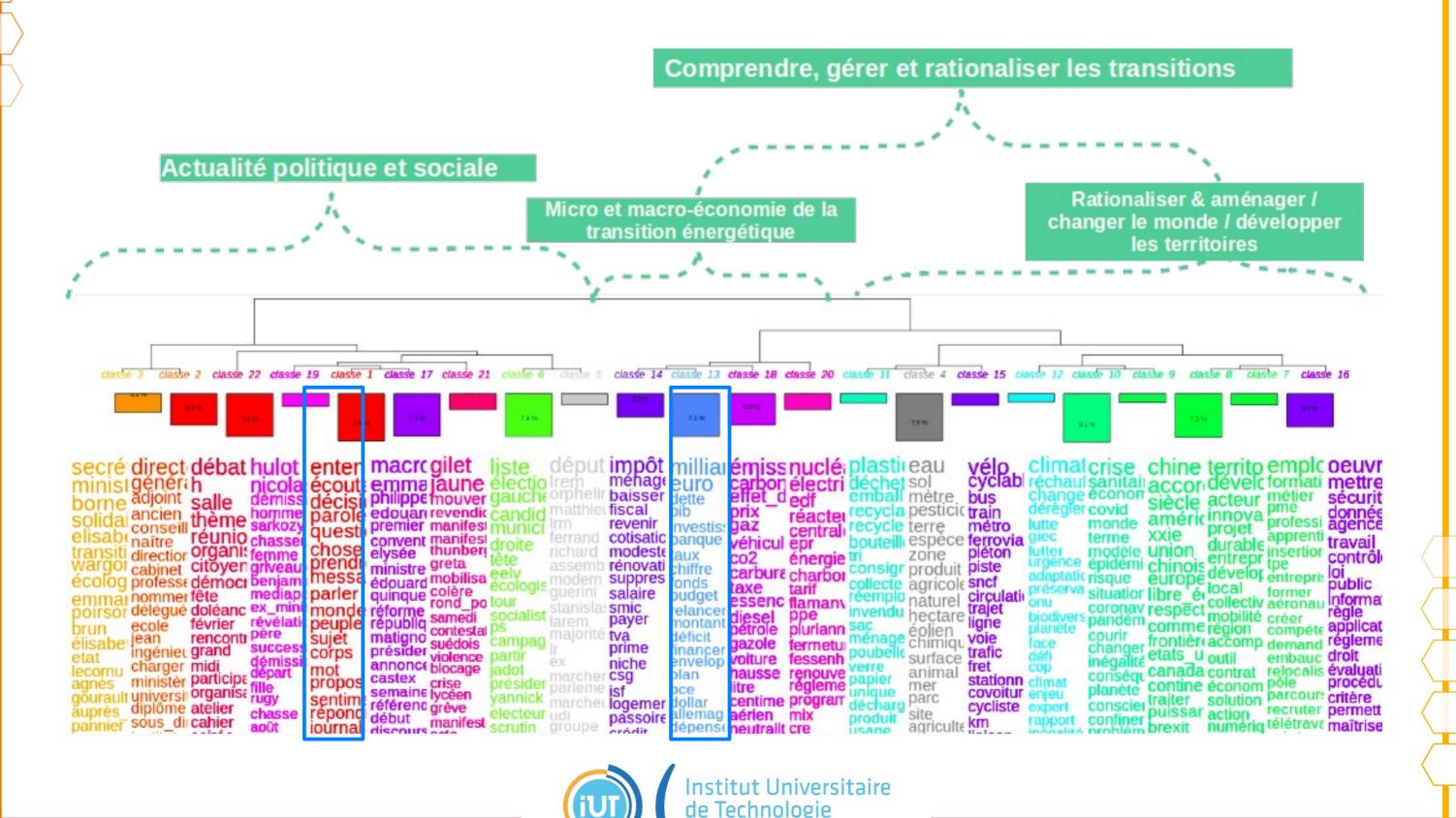
Moyenne de formes par segment: 35.937154

**** *id_1 *type_PQR *source_la_depeche_du_midi *date_2018-08-01 *am_2018-08 *annee_2018

Bonne nouvelle pour les usagers de la SNCF. La compagnie ferroviaire a indiqué qu'elle prévoyait trois trains sur quatre en gare Montparnasse mercredi. Elle confirme ainsi un « retour progressif à la normale », après que le gestionnaire du réseau haute tension RTE a pu rétablir, lundi soir, l'alimentation électrique de cette gare parisienne. [...]

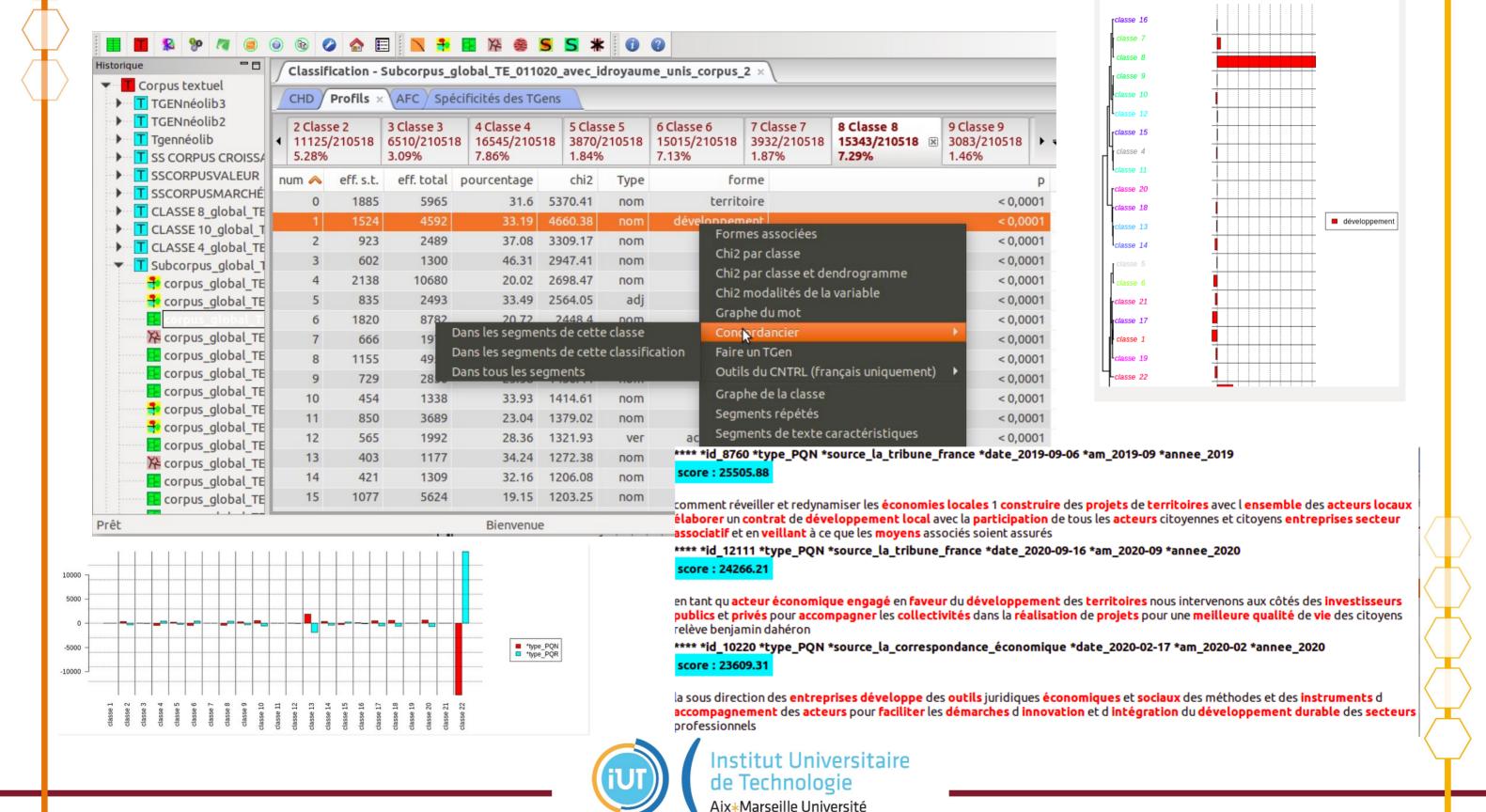


6. La transition écologique PQN PQR, CHD (Reinert)

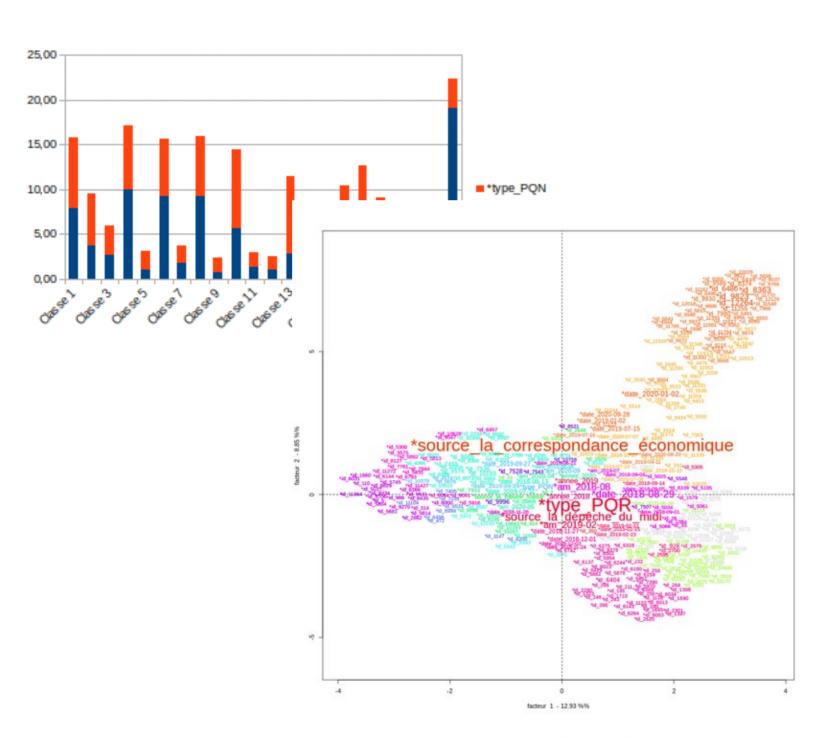


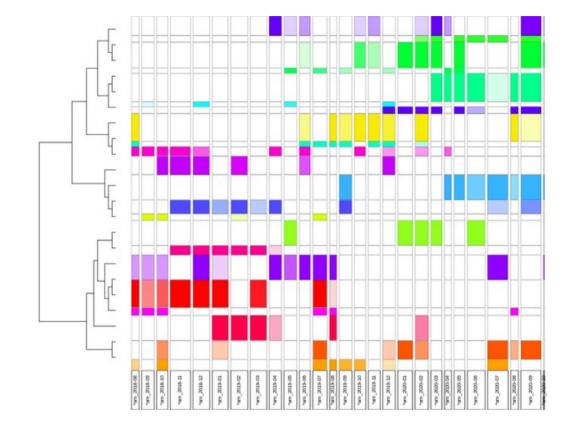
Aix*Marseille Université

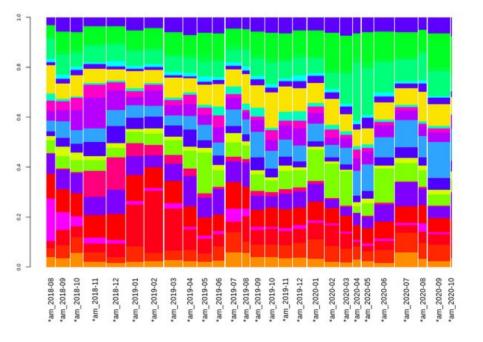
6. La transition écologique PQN PQR, CHD (Reinert)



6. La transition écologique PQN PQR CHD (Reinert) et analyses complémentaires

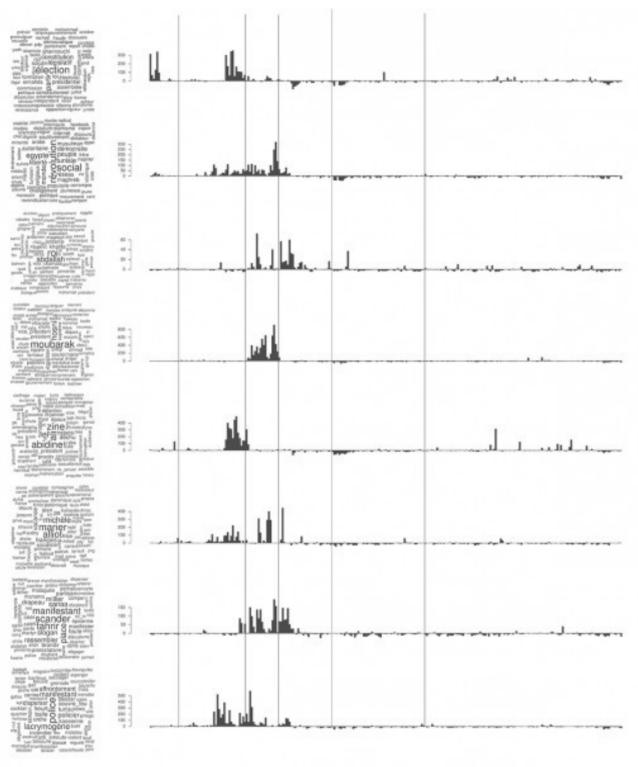








6. Focus lexico-chronologique



Les Révolutions Arabes d'après les Dépêches de l'AFP

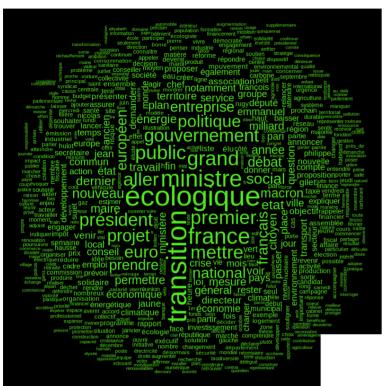
- Analyse lexico-chronologique des dépêches AFP, de décembre 2010 à septembre 2011
- Mots-clés choisis : « Tunisie »,
 « Égypte », « Libye » et « Syrie »
- 7 562 733 occurrences;
 208 187 dépêches
- Une réadaptation du cadrage médiatique : émeute, révolution, guerre

Souillard N, Ratinaud R., Marchand P., https://books.openedition.org/enseditions/7352

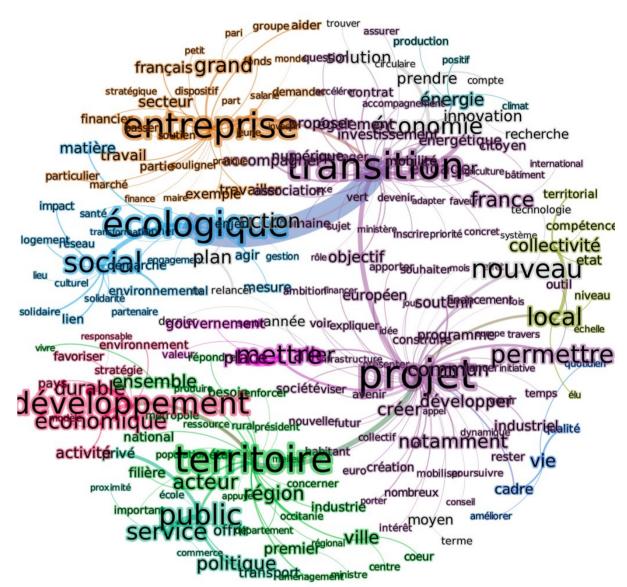
le 17 décembre 2010, immolation par le feu de Mohammed Bouazizi ; le 25 janvier 2011, début des mouvements de contestation au Caire ; le 15 février 2011, début des mouvements de contestation en Libye ; le 17 mars 2011, résolution 1973 de l'ONU création d'une zone d'exclusion aérienne au-dessus de la Libye pour protéger la population civile ; et le 9 mai 2011, mise en place d'un embargo par l'Europe contre le régime syrien.

de Technologie Aix*Marseille Université

6. Analyses de similitude ≠ nuage de mots



someday house crime trial thou touch allow goodbye believe daughter remember child kingdom return court understand guy show truththy show treat year and treat pray ready feel gonna allah fight deserve witctim show truththy show treat year and hate heart hope witch pain talk live gonna allah fight deserve victim show the pain talk live gonna allah fight deserve victim show the pain talk live gonna allah fight deserve victim show the pain talk live gonna show the pain talk live gonna gonna gonna allah fight deserve victim gonna allah fight deserve victim gonna allah fight deserve victim gonna show the pain talk live gonna gonna show the pain talk live gonna gonna show the pain to gonna show the pain talk live gonna sho



Graph retravaillé avec le logiciel Gephi



6. Analyses de similitude : Les figures de « l'habitant » et du « citoyen »

Tgen: Lamalle et Salem (2002)

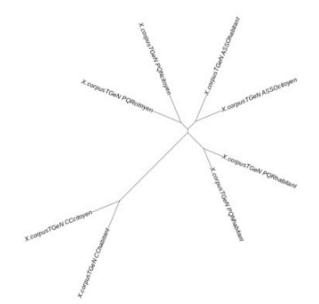
Un « ensemble d'occurrences sélectionnées parmi les occurrences du texte »

• **répertoire habitant** : « habitant » (1726 occ.), « habitants » (14 809 occ.), « habitante »

(605 occ.), « habitantes », (110 occ.), et « habitans » (2 occ.), « habitan » (1 occ.)

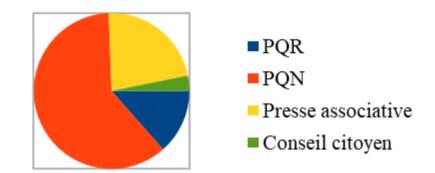
répertoire citoyen : « citoyen » (1080 occ.), « citoyens » (2042 occ.), « citoyenne » (556 occ.), « citoyennes » (165 occurrences), « citoyenneté » (743 occ.), « citoyennetés » (1 occ.) 6 , et « citoyennete » (3 occurrences), « citoyennété » (1

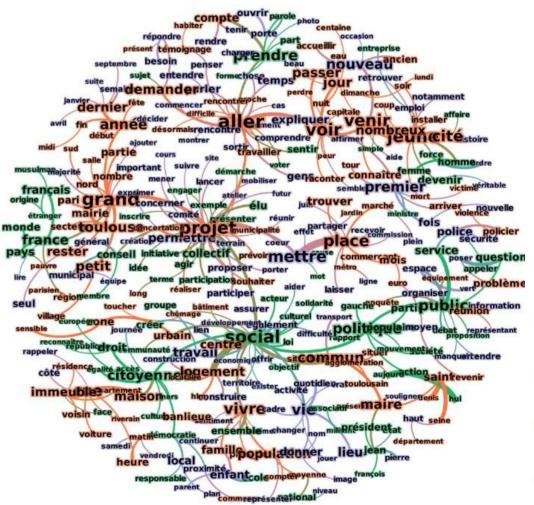
Métadonnées : *source_ ; *année_ ; *mois_*corpus_ ; *Tgen_ ; *TGeNcorpus_



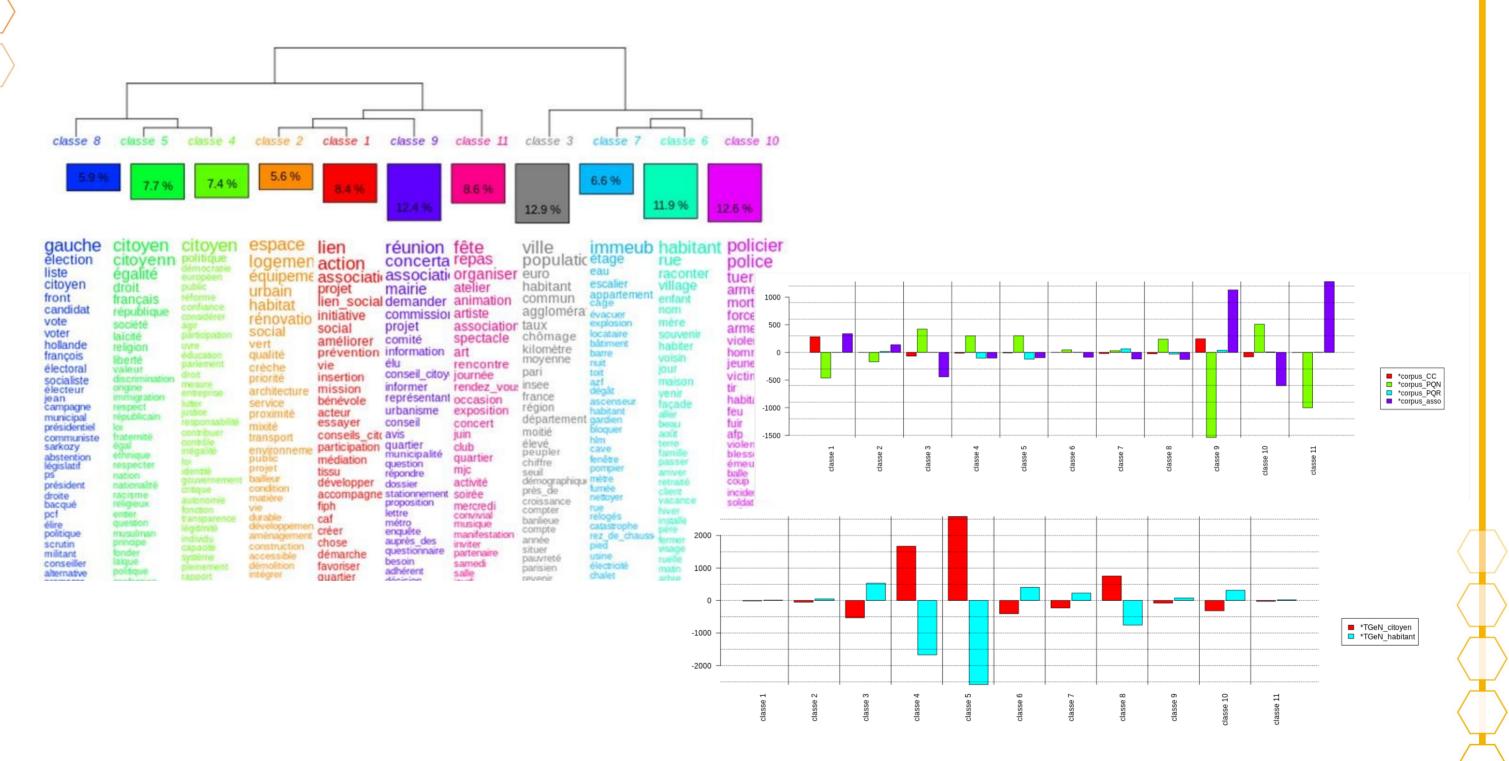








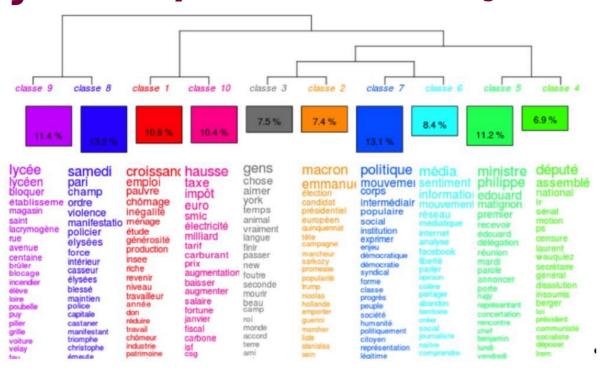
6. Analyses de similitude : Les figures de « l'habitant » et du « citoyen »

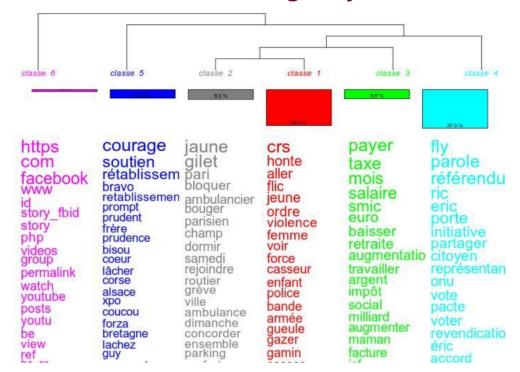




6. Corpus réseaux socio-numériques : comparer des « arènes » médiatiques d'un mouvement social (Gilets

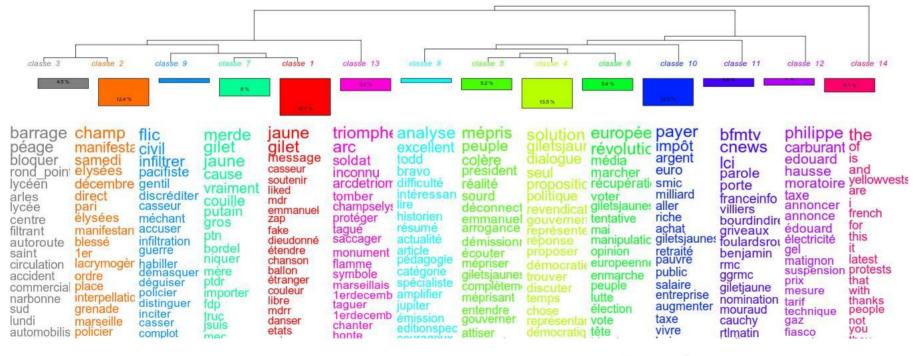
iaunes) Natacha Souillard, Brigitte Sebbah, Lucie Loubère, Laurent Thiong-Kay et Nikos Smyrnaios

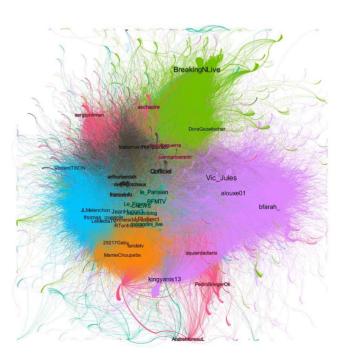




Dendrogramme, résultat de la CHD, corpus de PQN (Iramuteq)

Dendrogramme résultat de la CHD, Corpus Facebook (Iramuteq)





Dendrogramme résultat de la CHD, Corpus Twitter (Iranto



Graph de communauté (Iramuteq+ Gephi)

D'autres logiciels de textométrie

Alceste M. Reinert (http://www.image-zafar.com)

DtmVic L. Lebart (http://lebart.org)

Hyperbase E. Brunet

(http://ancilla.unice.fr/~brunet/pub/hyperbase.html)

Lexico 3 A. Salem (http://lexico3.no-ip.org)

SPAD Decisia / M. Bécue (http://www.spad.eu)

Sphinx Lexica Y. Baulac (http://www.lesphinx-developpement.fr)

Taltac

S. Bolasco (http://www.taltac.it/it/index.shtml)

ianac

TXM S. Heiden (http://textometrie.ens-lyon.fr/)

Une revue en ligne

Lexicometrica: http://www.cavi.univ-paris3.fr/lexicometrica/

