



HAL
open science

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.



Maison Méditerranéenne
des Sciences de l'Homme
Aix-Marseille Université

Introduction à la textométrie avec le logiciel IRaMuTeQ

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

IMSIC, Aix-Marseille Université



IRaMuTeQ



Institut Universitaire
de Technologie
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



1. La textométrie, origines et épistémologie

- 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 étudiant 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).

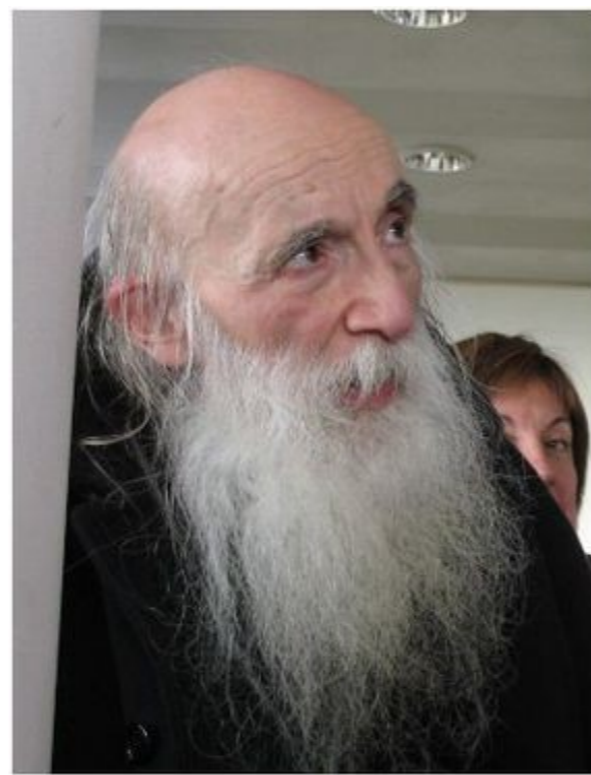
Comment « *le langage se rapporte au monde par l'intermédiaire de l'esprit* » (Johnson-Laird, 1988, p. 66).

1. La textométrie, origines et épistémologie

« 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.

1. La textométrie, origines et épistémologie

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**

1. La textométrie, origines et épistémologie

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 (signifiante, acceptabilité, complétude, exploitabilité, Bommier-Pincemin, 1999).
- Clôture ?

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

texte texte texte texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte textetexte texte texte texte texte texte texte texte
texte texte ...

1. La textométrie, origines et épistémologie

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

Observatoire :

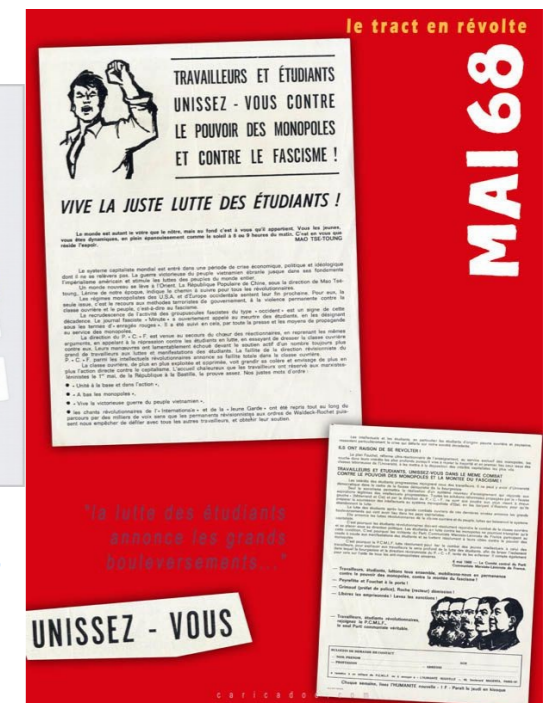
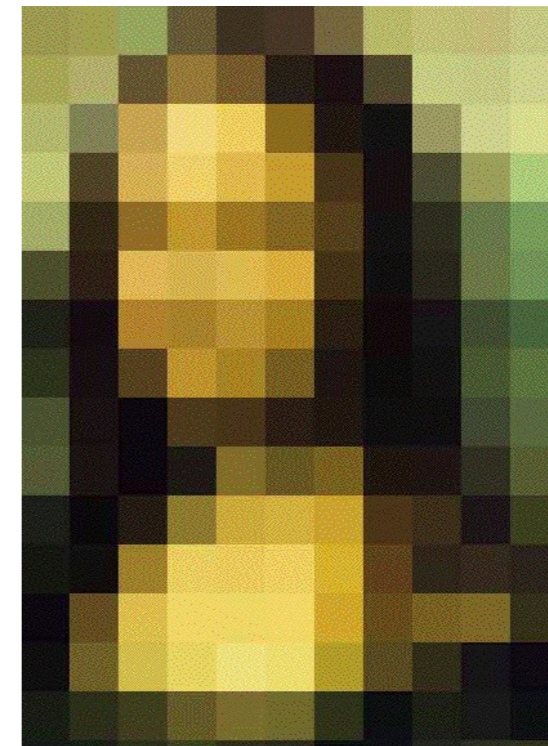
- Objectivation : aborder les objets discursifs comme des phénomènes, sociaux
- **Approche Statistique :**
 - **Limiter la part d'échantillonnage a priori** : corpus massifs, exploration, effets massifs de structure (invisibles à « l'œil nu »), sérendipité
 - Déplacement du travail interprétatif.
 -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



1. La textométrie, origines et épistémologie

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

2. Opérations initiales

Le corpus

```
**** *var1_mod1 *var2_mod3 *.....  
texte texte texte texte texte texte texte texte texte texte texte texte texte texte texte  
texte texte texte texte texte texte texte texte texte texte texte texte texte texte texte  
texte texte  
**** *var1_mod2 *var2_mod1  
texte texte texte texte texte texte texte texte texte texte texte texte texte texte texte  
texte texte texte texte texte texte texte texte texte texte
```

Pour Iramuteq :

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

3. Opérations initiales

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
àÀâÂäÄáÁâÃäÉèÈêÊëËìÎîíÍíóÓòÒôÔöÖõÕøØùÙûÛüÜúÚçÇβœ
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

3. Opérations initiales

Index des formes

12528 de	1195 c	530 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 y	290 aujourd	226 temps
4631 à	985 qu	462 j	290 lui	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 a	533 tous	342 sa	235 président	195 croissance

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

3. Opérations initiales

Index des formes

12528 de	1195 c	530 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 y	290 aujourd	226 temps
4631 à	985 qu	462 j	290 lui	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 a	533 tous	342 sa	235 président	195 croissance

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

2. Opérations initiales

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)

2. Opérations initiales

Distinction formes actives et formes supplémentaires (mots-outils)

Clés d'analyse
Choix des clés d'analyse
0=éliminé ; 1=active ; 2=supplémentaire

Adjectif	1	voir liste	Conjonction	2	voir liste
Adjectif démonstratif	2	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	voir liste
Adverbe	1	voir liste	Pronom personnel	2	voir liste
Adverbe supplémentaire	2	voir liste	Pronom possessif	2	voir liste
Article défini	2	voir liste	Pronom relatif	2	voir liste
Article indéfini	2	voir liste	Préposition	2	voir liste
Auxiliaire	2	voir liste	Verbe	1	voir liste
Chiffre	2	voir liste	Verbe supplémentaire	2	voir liste

✓ Valider

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).



3. Opérations initiales

Étiquetage (imparfait)

de	426235	pre	au	48626	art def	département	2218	nom
la	239114	art def	par	47445	pre	salarié	2217	adj
l	184719	art def	pas	40592	adv sup	borne	2216	nom
le	176677	art def	ce	35715	pro dem	concerner	2215	ver
être	163412	aux	plus	34780	adv sup	montrer	2214	ver
et	158730	con	s	31445	pre per	atteindre	2213	ver
les	158324	art def	faire	28506	ver sup	adjoint	2202	nom
à	151137	pre	avec	27056	pre	juillet	2199	nom
des	148395	art ind	qu	25265	con	poursuivre	2194	ver
d	131664	pre	se	24540	pro per	modèle	2189	nom
avoir	123244	aux	nous	23789	pro per	responsable	2175	adj
en	112780	pre	aux	23411	art def	déclarer	2173	ver
du	93212	art def	on	23376	pro per	particulier	2172	adj
un	83924	art ind	pouvoir	22443	ver sup	fonds	2171	nom
pour	79628	pre	mais	22264	con	continuer	2167	ver
une	75276	art ind	écologique	20989	adj	direction	2166	nom
dans	56382	pre	o	20642	pro dem	représenter	2162	ver
qui	56298	pro rel	transition	20589	nom	retraite	2159	nom
sur	54240	pre	cette	18987	adj dem	mener	2159	ver
que	53898	pro rel	son	18546	adj pos	international	2155	adj
il	53808	pro per	ou	17780	con	dispositif	2143	nom
ne	52150	adv sup	devoir	1731	ver sup	surtout	2136	adv sup

3. Opérations initiales

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) :

1. ".", "?", "!"
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
- ...

3. Opérations initiales

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

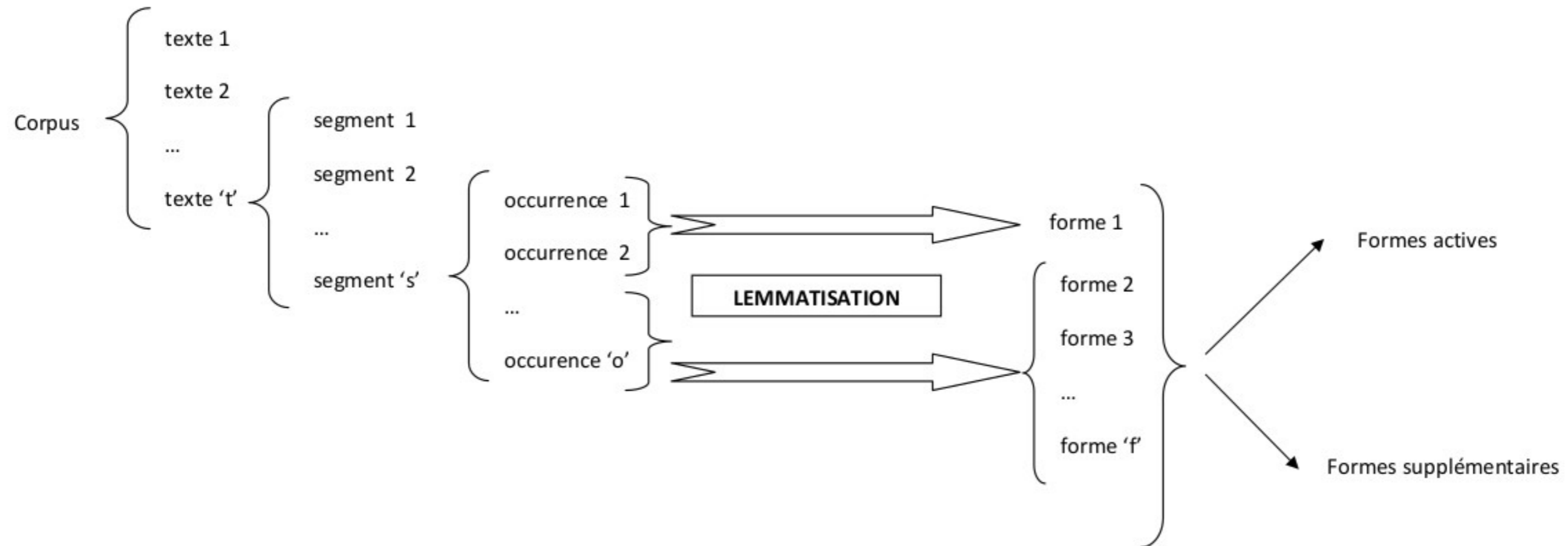
Num	Partie	Occurenc	Formes	Hapax	Fmax	Forme
✓ 1	159debré	8810	2017	1188	442	de
✓ 2	162pompidou	4395	1333	875	204	de
✓ 3	168couve	4182	1176	706	181	de
✓ 4	169chaban	1804	742	517	95	de
✓ 5	172messmer	7545	2061	1344	424	de
✓ 6	174chirac	8519	1997	1224	509	de
✓ 7	176barre	7013	1802	1135	388	de
✓ 8	177barre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

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



3. Opérations initiales

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>

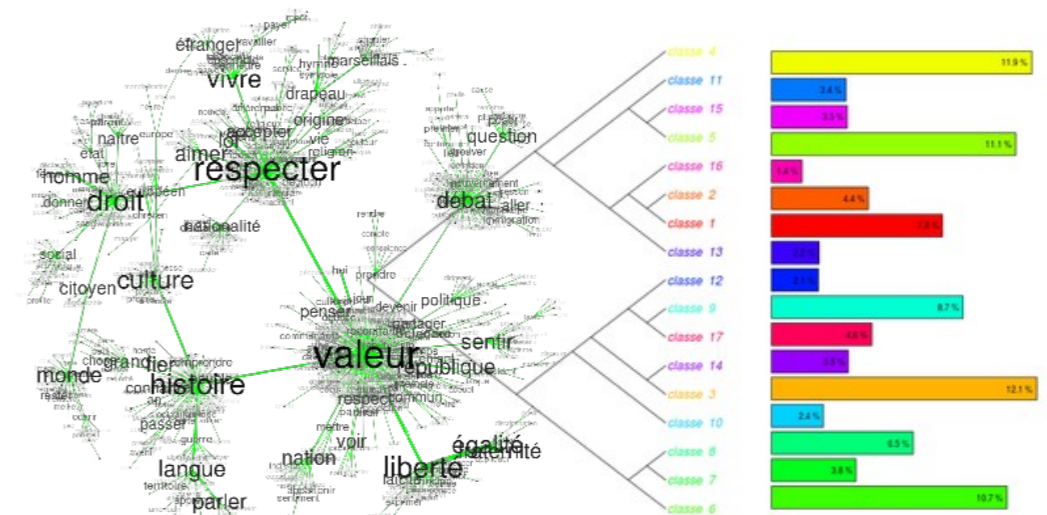
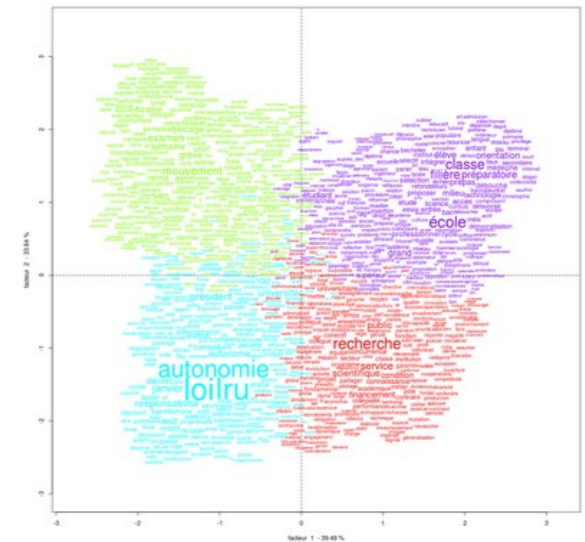
4. IRAMuTeQ, présentation générale

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/iramuteq/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)

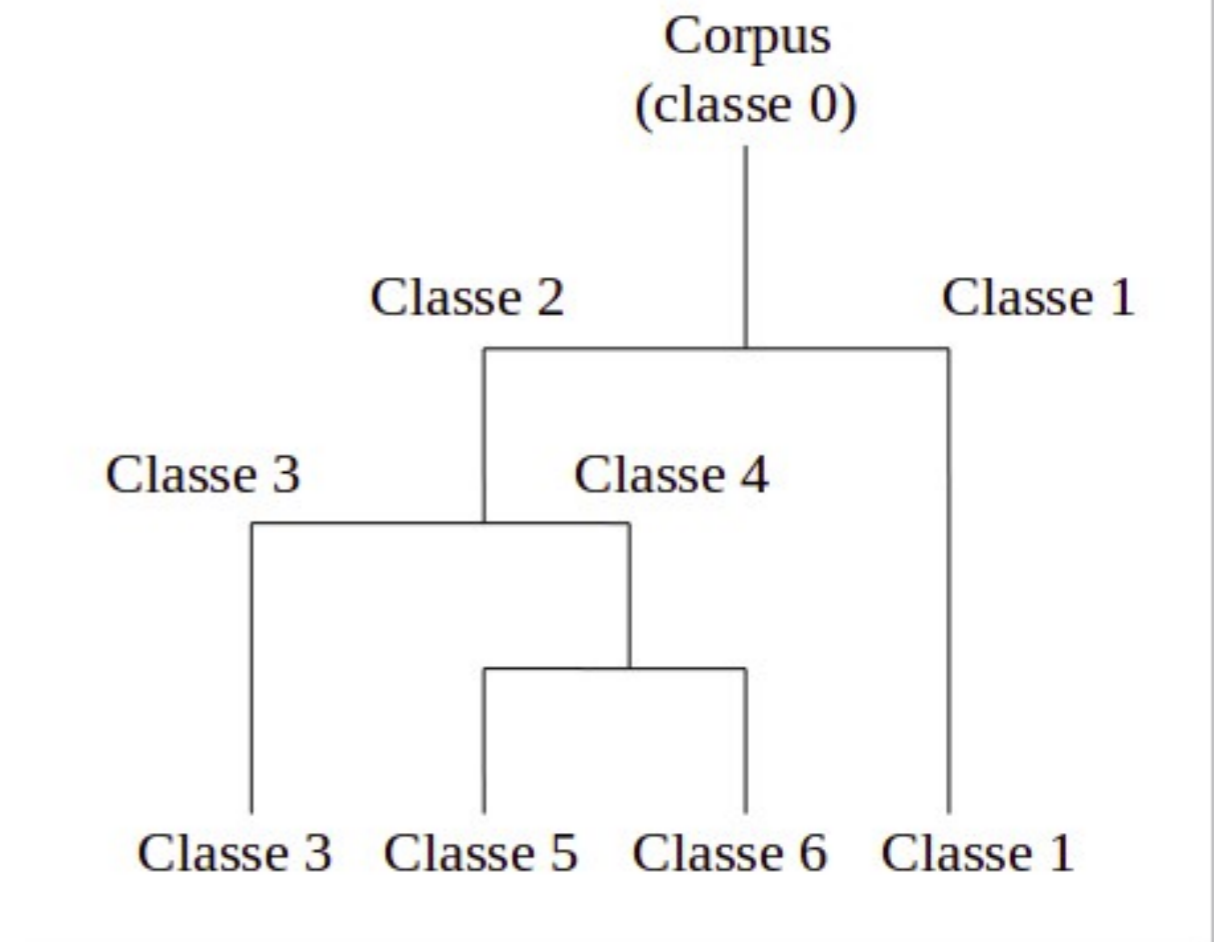


Partition (contributions ou paragraphes)

Lexique

- Tokenisation
- Reconnaissance
- Lemmatisation
- statut

0	0	1	1	0	1
0	0	1	1	0	1
0	0	1	1	0	1
1	1	1	1	1	0
0	0	1	1	0	1
1	1	0	0	1	0
1	1	0	0	1	0
0	0	1	1	0	1
1	1	0	0	1	0
0	0	1	1	0	1
1	1	0	0	0	1
0	0	1	1	0	1



- 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 (PQR)	
La Dépêche 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 (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

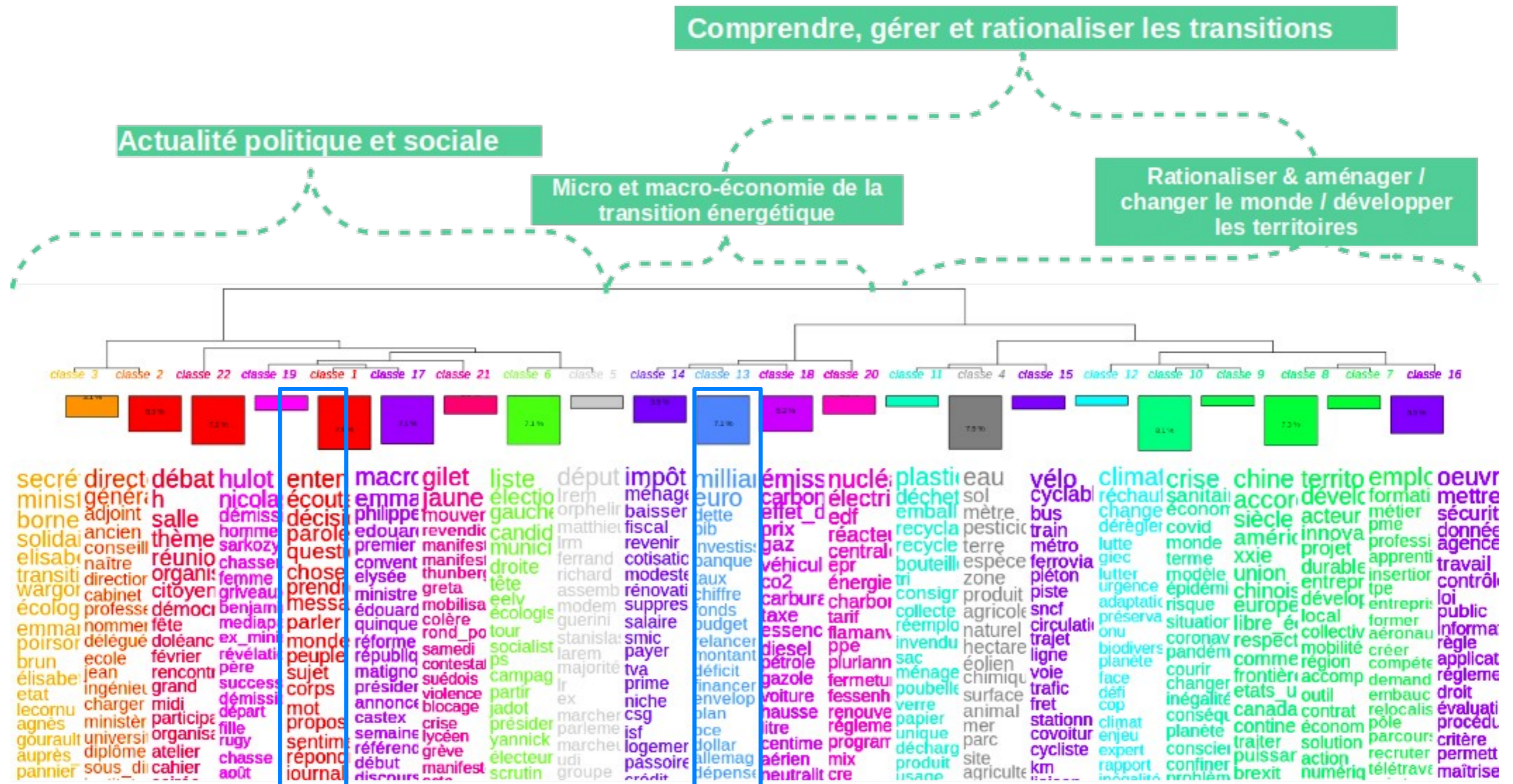
Moyenne de formes par segment: 35.937154

**** *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)



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

Historique

- Corpus textuel
 - TGENnéolib3
 - TGENnéolib2
 - Tgennéolib
 - SS CORPUS CROISSA
 - SSCORPUSVALEUR
 - SSCORPUSMARCHÉ
 - CLASSE 8_global_TE
 - CLASSE 10_global_T
 - CLASSE 4_global_TE
 - Subcorpus_global_T
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_T
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE
 - corpus_global_TE

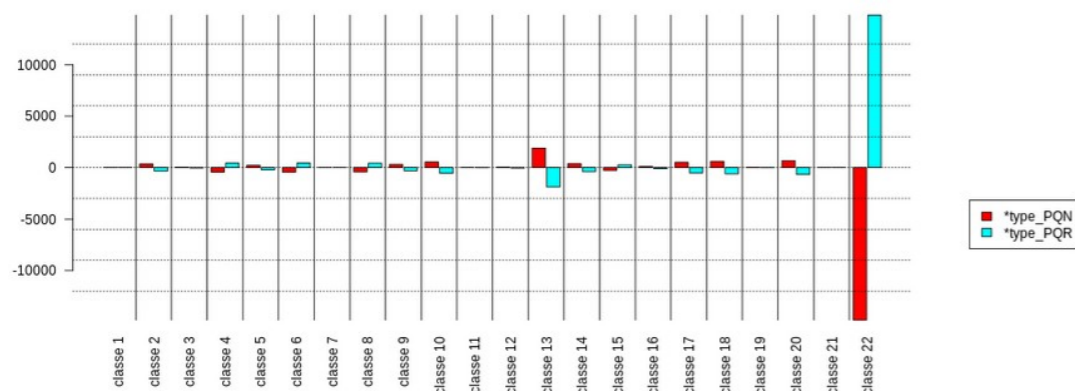
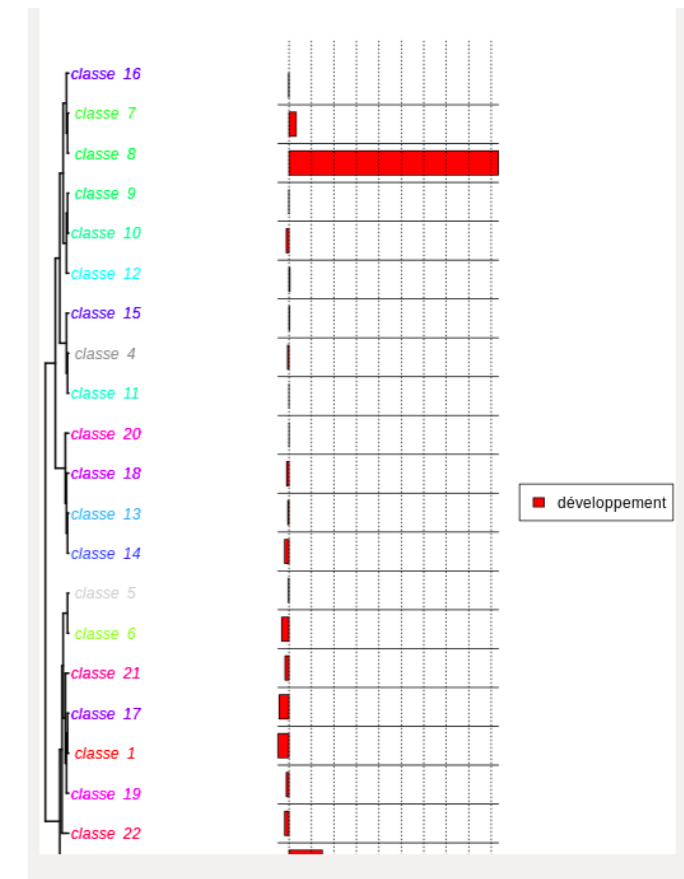
Classification - Subcorpus_global_TE_011020_avec_idroyaume_unis_corpus_2

CHD Profils AFC Spécificités des TGens

	2 Classe 2	3 Classe 3	4 Classe 4	5 Classe 5	6 Classe 6	7 Classe 7	8 Classe 8	9 Classe 9
	11125/210518	6510/210518	16545/210518	3870/210518	15015/210518	3932/210518	15343/210518	3083/210518
	5.28%	3.09%	7.86%	1.84%	7.13%	1.87%	7.29%	1.46%

num	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p
0	1885	5965	31.6	5370.41	nom	territoire	< 0,0001
1	1524	4592	33.19	4660.38	nom	développement	< 0,0001
2	923	2489	37.08	3309.17	nom		< 0,0001
3	602	1300	46.31	2947.41	nom		< 0,0001
4	2138	10680	20.02	2698.47	nom		< 0,0001
5	835	2493	33.49	2564.05	adj		< 0,0001
6	1820	8782	20.72	2448.4	nom		< 0,0001
7	666	1911	34.84	1272.38	nom		< 0,0001
8	1155	4911	23.54	1379.02	nom		< 0,0001
9	729	2811	25.93	1414.61	nom		< 0,0001
10	454	1338	33.93	1414.61	nom		< 0,0001
11	850	3689	23.04	1379.02	nom		< 0,0001
12	565	1992	28.36	1321.93	ver	ac	< 0,0001
13	403	1177	34.24	1272.38	nom		< 0,0001
14	421	1309	32.16	1206.08	nom		< 0,0001
15	1077	5624	19.15	1203.25	nom		< 0,0001

Prêt Bienvenue



**** *id_8760 *type_PQN *source_la_tribune_france *date_2019-09-06 *am_2019-09 *annee_2019
score : 25505.88

comment réveiller et redynamiser les **économies locales** 1 **construire** des **projets de territoires** avec l **ensemble** des **acteurs locaux** **élaborer** un **contrat de développement local** avec la **participation** de tous les **acteurs** citoyennes et citoyens **entreprises secteur associatif** et en **veillant** à ce que les **moyens** associés soient assurés

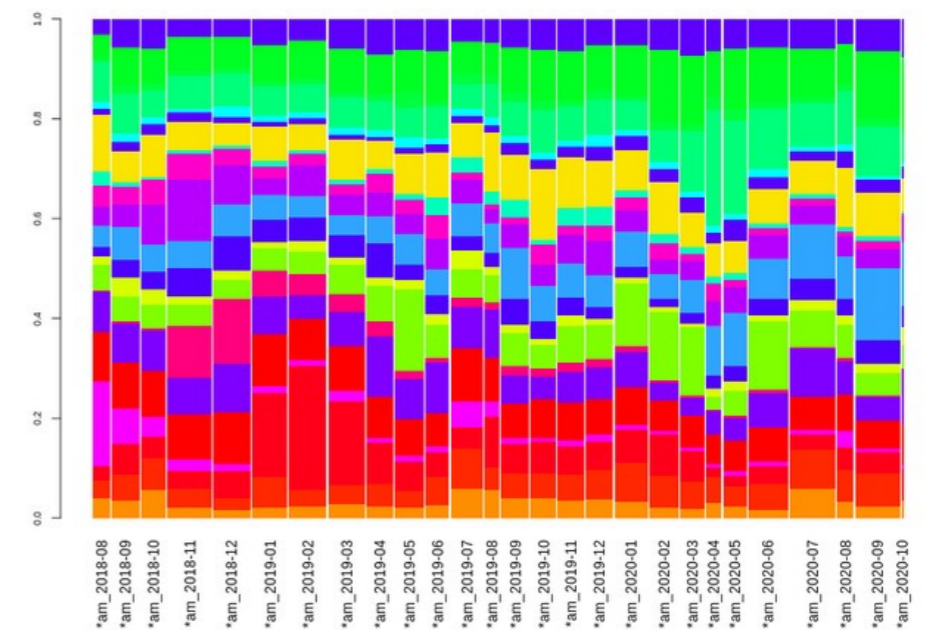
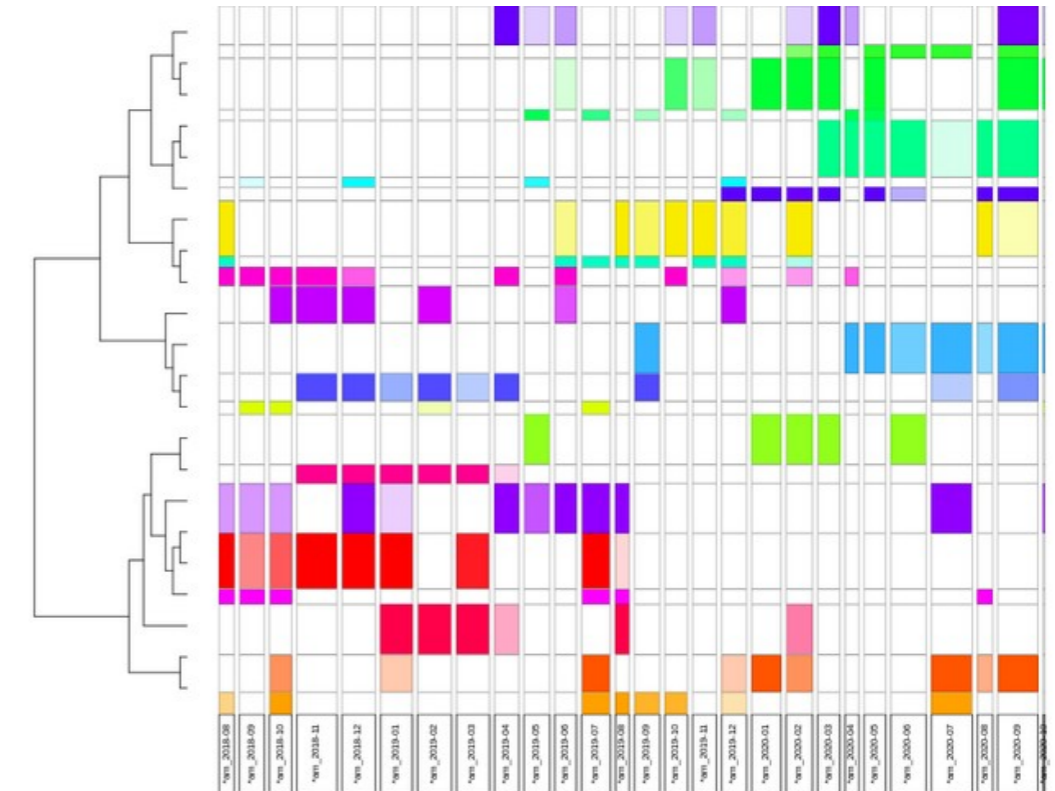
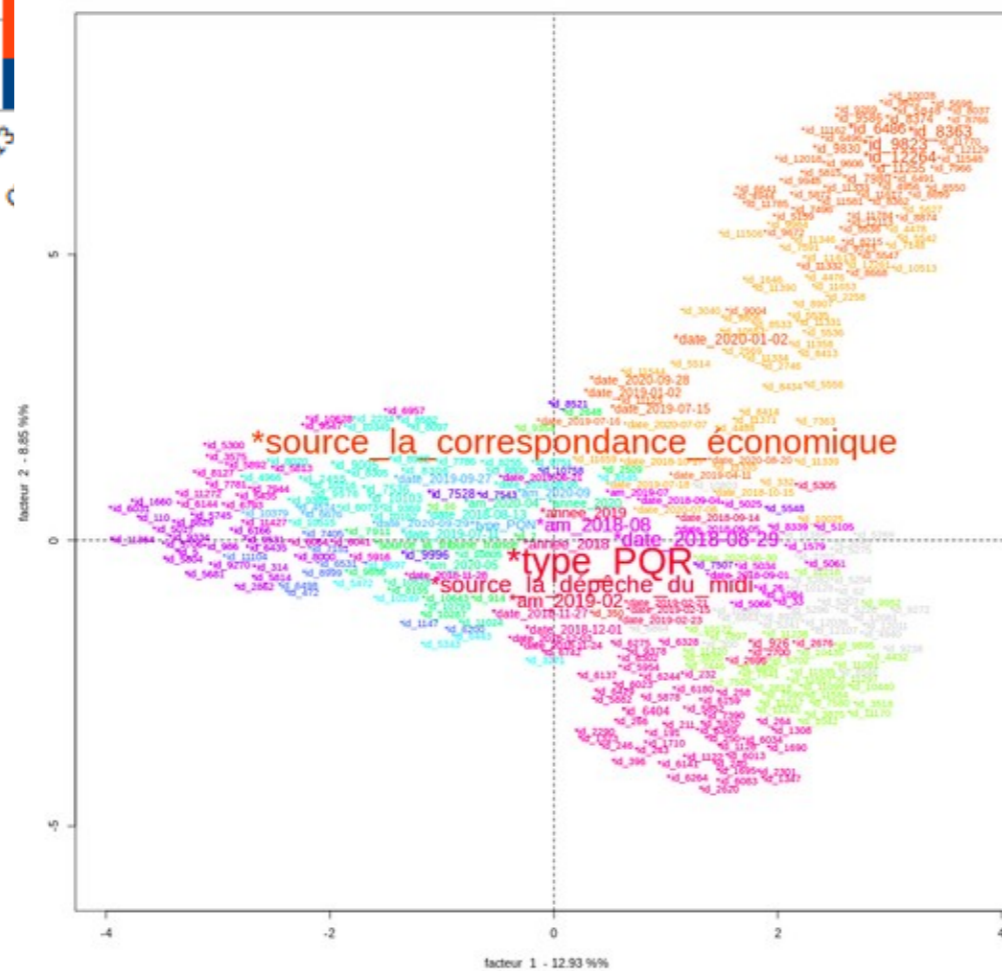
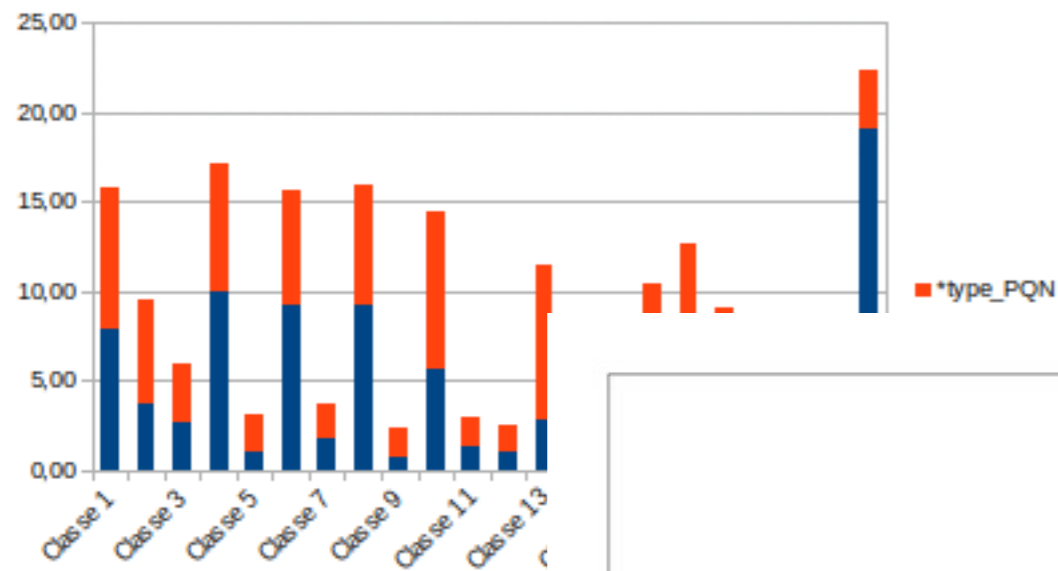
**** *id_12111 *type_PQN *source_la_tribune_france *date_2020-09-16 *am_2020-09 *annee_2020
score : 24266.21

en tant qu **acteur économique engagé** en **faveur** du **développement** des **territoires** nous intervenons aux côtés des **investisseurs publics** et **privés** pour **accompagner** les **collectivités** dans la **réalisation** de **projets** pour une **meilleure qualité** de **vie** des citoyens relève benjamin dahéron

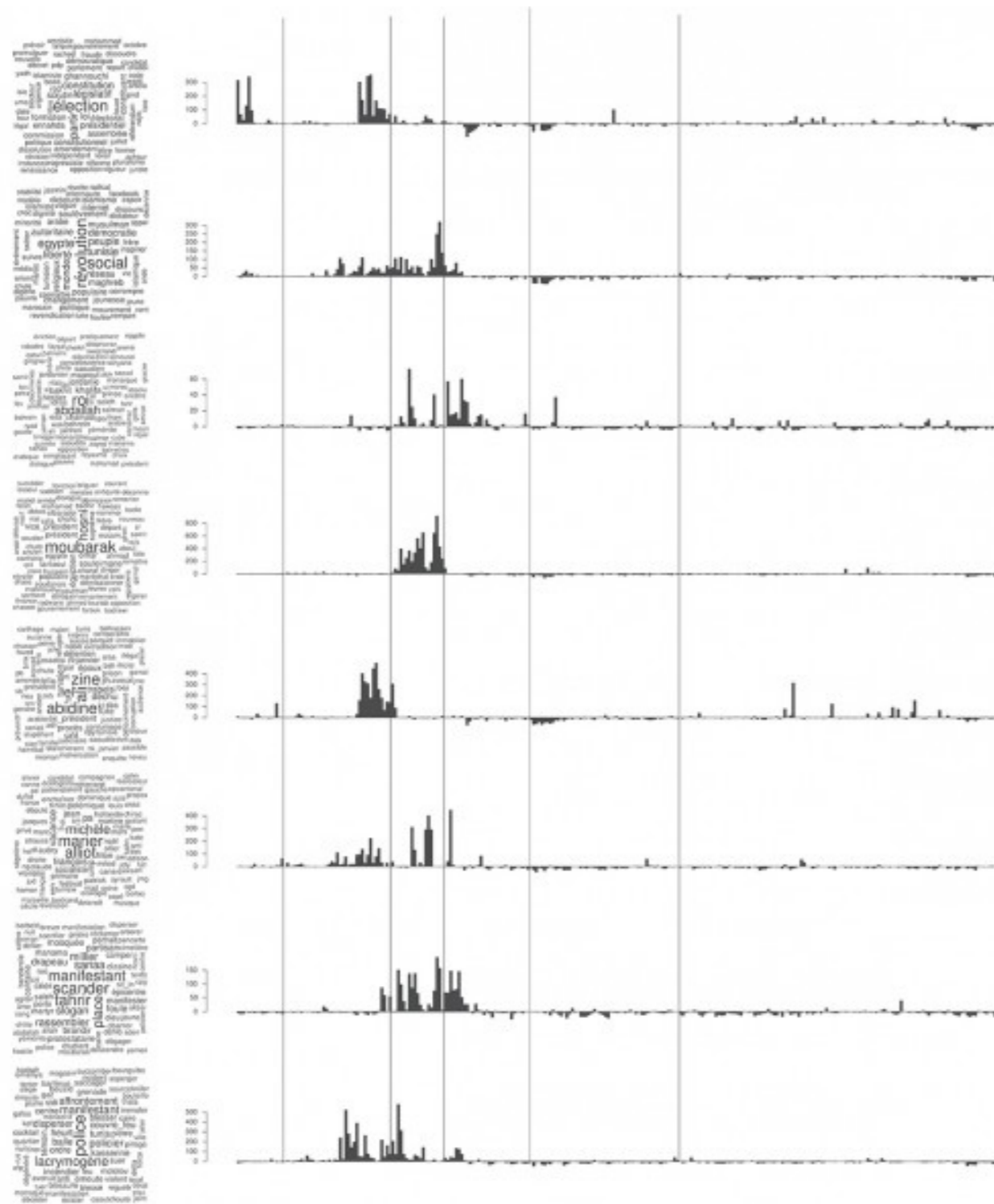
**** *id_10220 *type_PQN *source_la_correspondance_économique *date_2020-02-17 *am_2020-02 *annee_2020
score : 23609.31

la sous direction des **entreprises développe** des **outils** juridiques **économiques** et **sociaux** des méthodes et des **instruments** d **accompagnement** des **acteurs** pour **faciliter** les **démarches** d **innovation** et d **intégration** du **développement durable** des **secteurs** professionnels

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



6. Focus lexico-chronologique



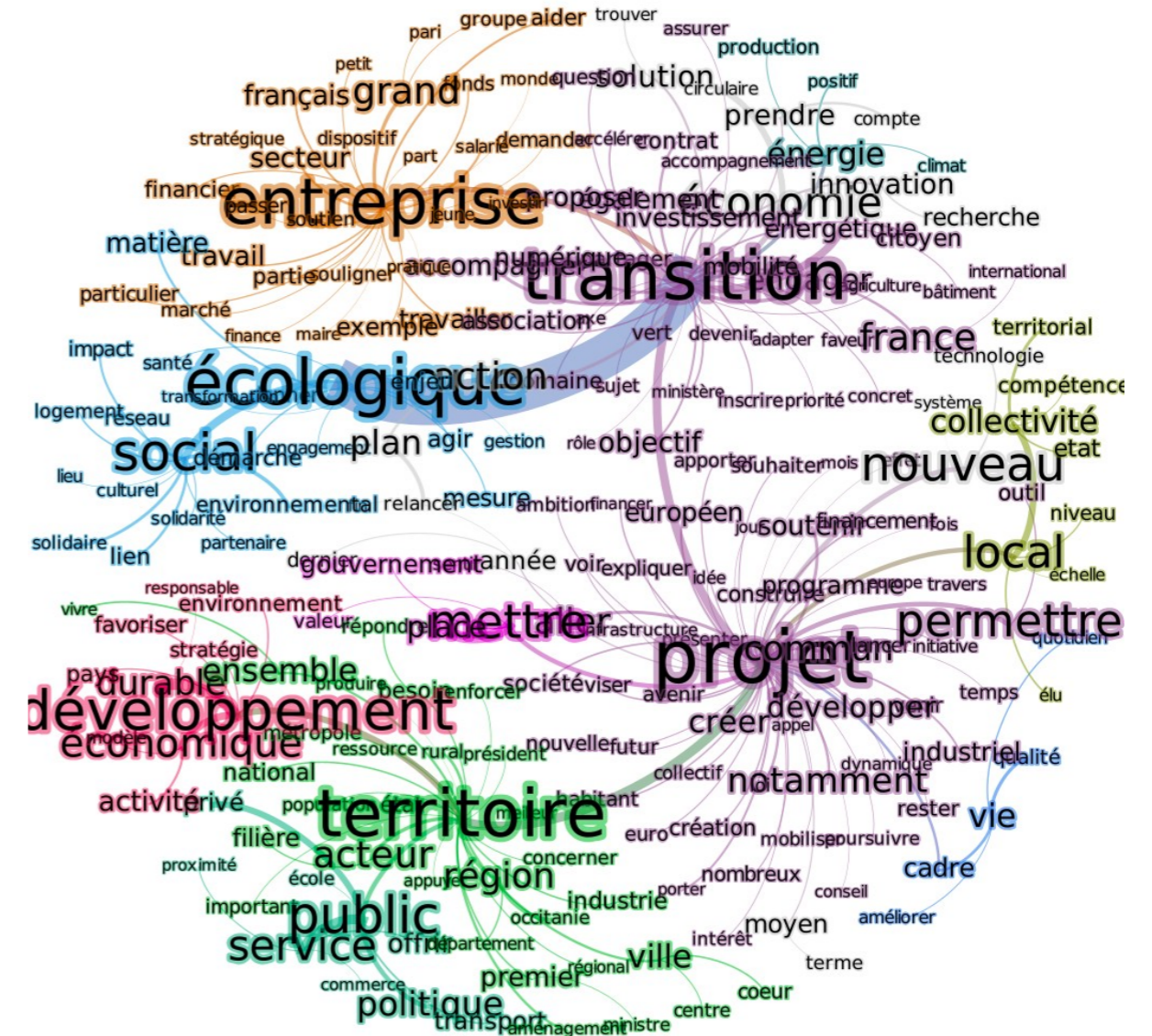
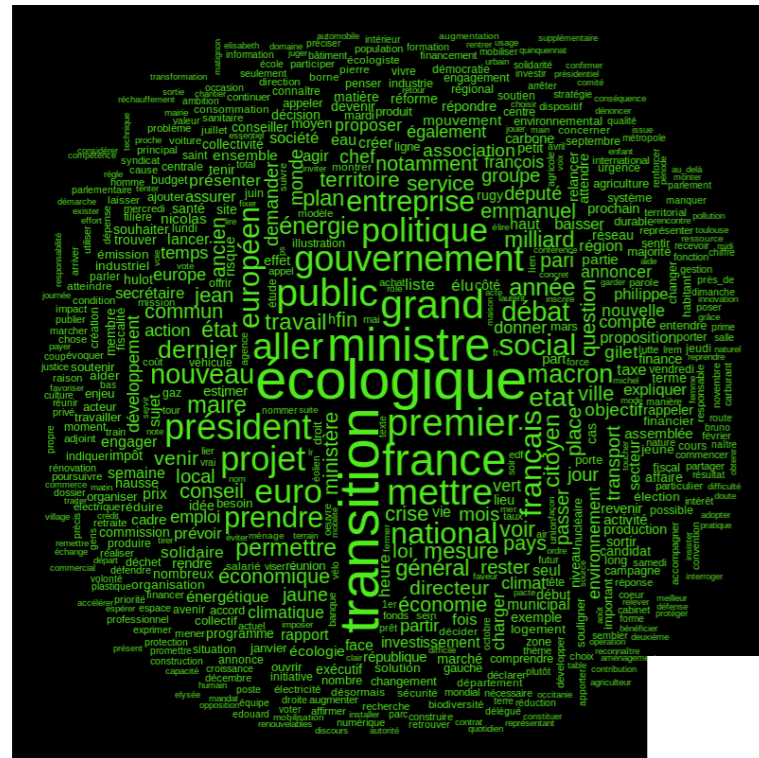
Les Révolutions Arabes d'après les Dépêches de l'AFP

- 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>

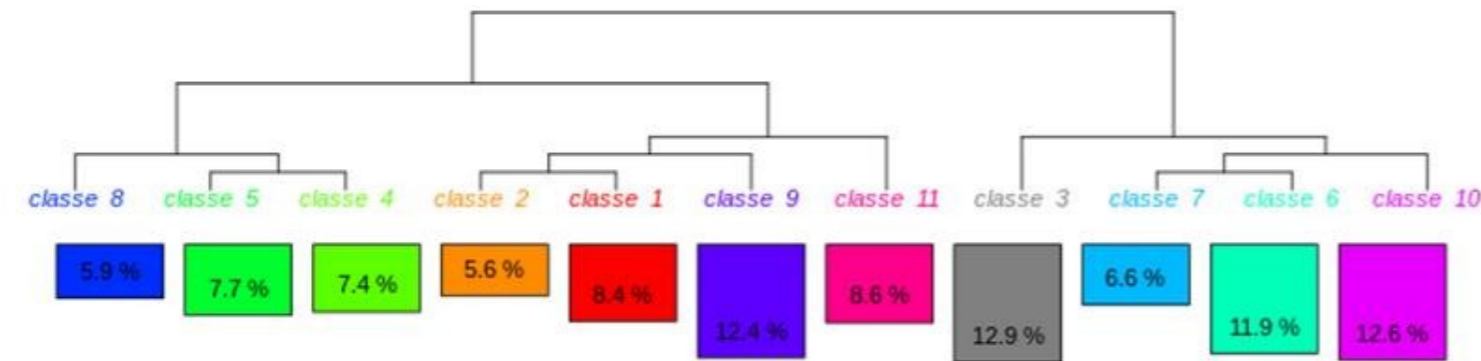
le 17 décembre 2010, immolation par le feu de Mohammed Bouazizi ; **le 25 janvier 2011**, début des mouvements de contestation au Caire ; **le 15 février 2011**, début des mouvements de contestation en Libye ; **le 17 mars 2011**, résolution 1973 de l'ONU création d'une zone d'exclusion aérienne au-dessus 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.

6. Analyses de similitude \neq nuage de mots

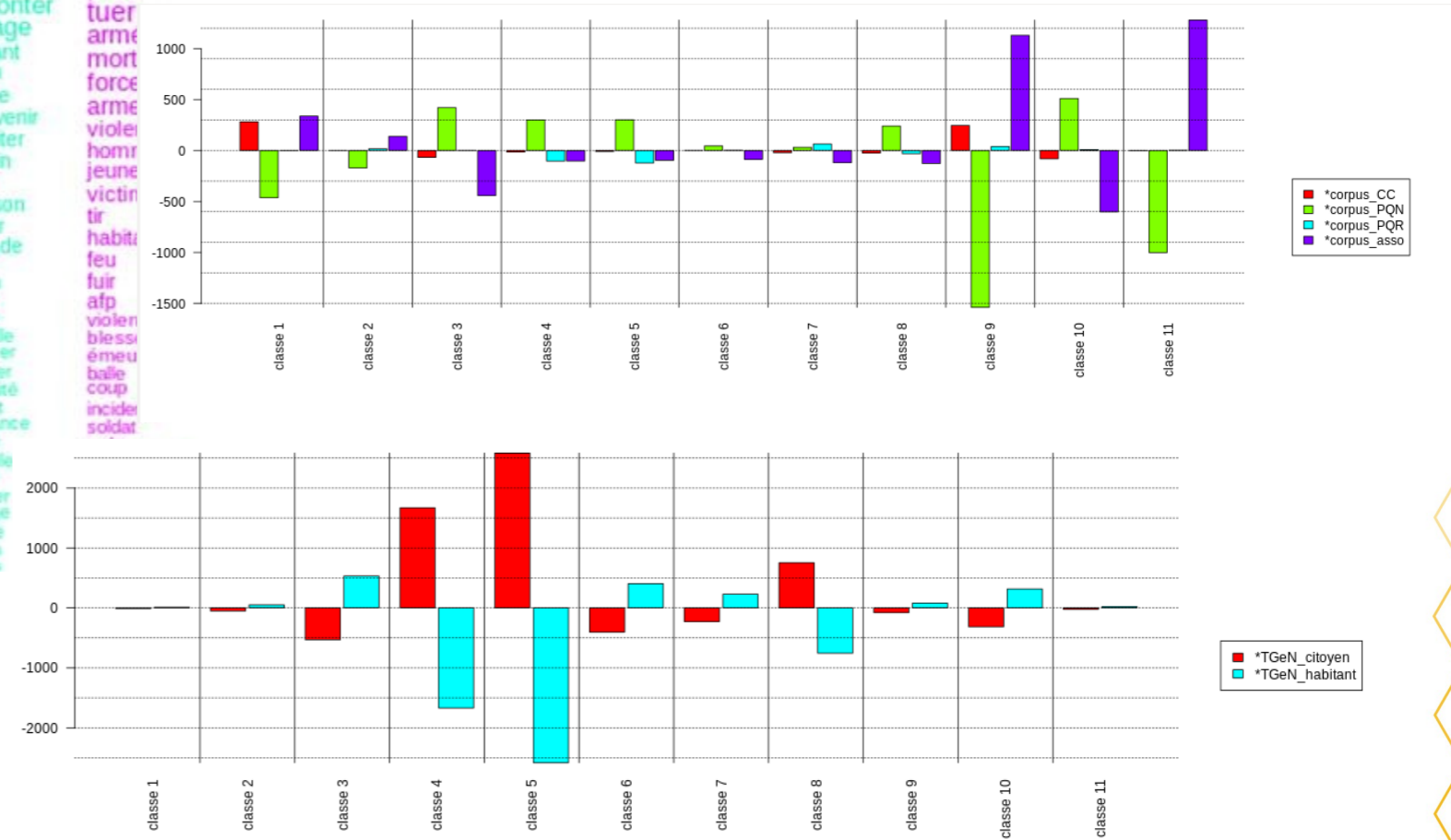


Graph retravaillé avec le logiciel Gephi

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

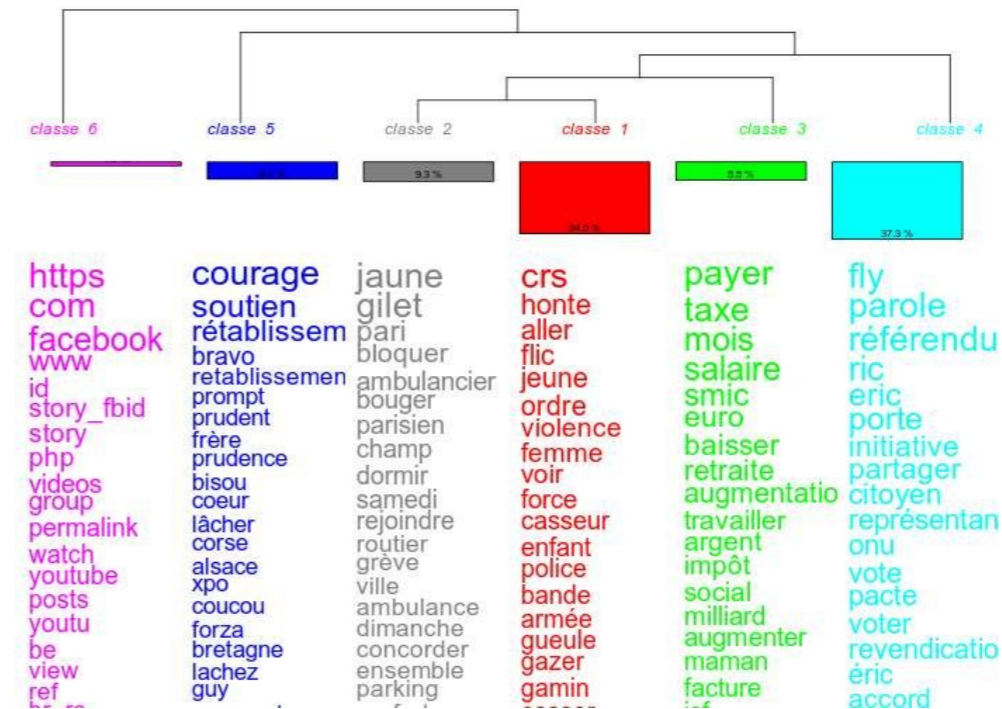
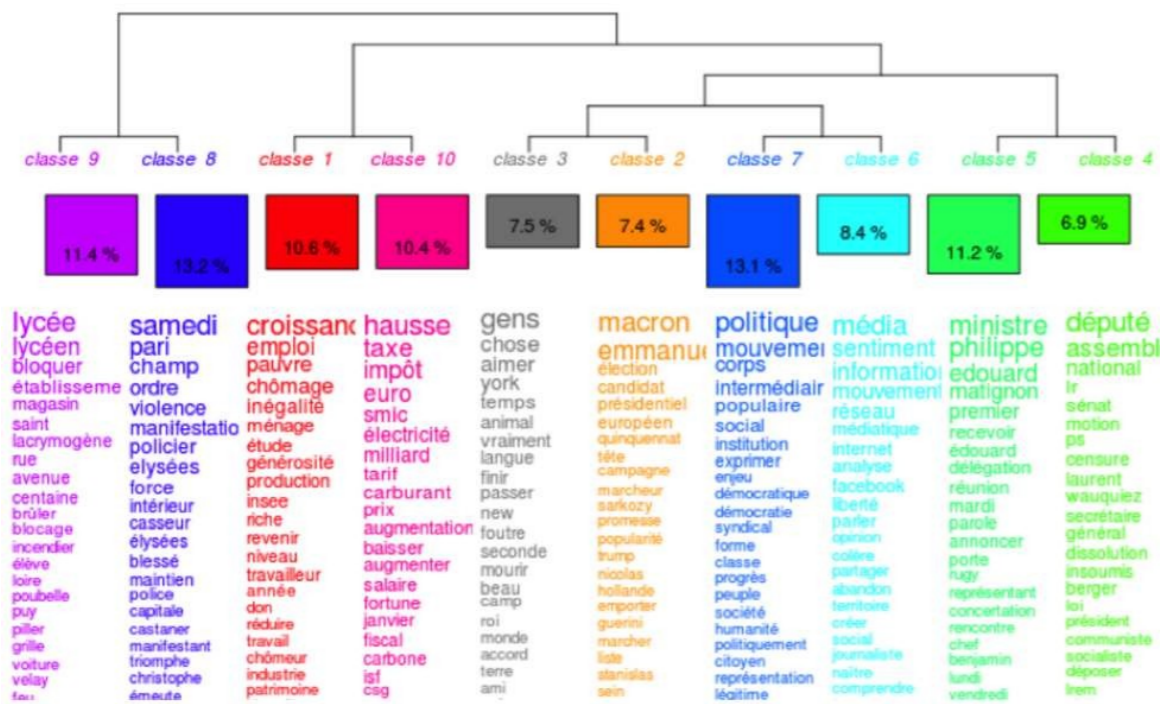


gauche élection liste citoyen front candidat vote voter hollande français électoral socialiste électeur jean campagne municipal présidentiel communiste sarkozy abstention législatif ps président droite bacqué pcf être politique scrutin militant conseiller alternative	citoyen citoyenn égalité droit français républicain société laïcité religion liberté valeur discrimination origine immigration respect républicain loi fraternité égal ethnique respecter nation nationalité racisme religieux enter question musulman principe fonder laïque politique	citoyen politique démocratie européen public réforme confiance considérer agir participation uvre éducation parlement droit mesure entreprise lutter justice responsabilité contribuer contrôle mégalié loi identité gouvernement critique autonomie fonction transparence légitimité individu capacité système pleinement rapport	espace logement équipeme urbain habitat renovatio social vert qualité crèche priorité architecture service proximité mixité transport environnement public projet bailleur condition matière vie durable développement aménagement construction accessible démolition intégrer	lien action associati projet lien_social initiative social améliorer prévention vie insertion mission bénévole acteur essayer conseils_cit participation médiation tissu développer accompagne fiph caf créer chose démarche favoriser quartier	réunion concerta associati mairie demander commissio projet comité information élu conseil_citoy informer représentant urbanisme conseil avis quartier municipalité question répondre dossier stationnement proposition lettre mètre enquête auprès_des questionnaire besoin adhérent	fête repas organiser atelier animation artiste associati spectacle art rencontre journée rendez_vou occasion exposition concert juin club quartier mjc activité soirée mercredi convivial musique manifestation inviter partenaire samedi salle	ville populati euro habitant commun aggloméra taux chômage kilomètre moyenne pari insee france région département moitié élevé peupler chiffre seul démographique près_de croissance rue banlieue compte année situer pauvreté parisien revenir	immeub étage eau escalier appartement cage évacuer explosion locataire bâtiment barre nuit toit azl dégât ascenseur habitant garden bloquer hlm cave fenêtre pompiers mètre lumière nettoyer rez_de_chauss catastrophe rez_de_chauss usine électrique chalet	habitant rue raconter village enfant nom mère souvenir habiter voisin jour maison venir façade aller beau août terre famille passer arriver retraité client vacance hiver installé père fermer visage rueille main arthur	policier police tuer armé mort force armé viole homr jeune victin tir habiti feu fuir afp violen blessi émeu balle coup incider soldat
--	--	--	---	--	--	---	---	---	--	--



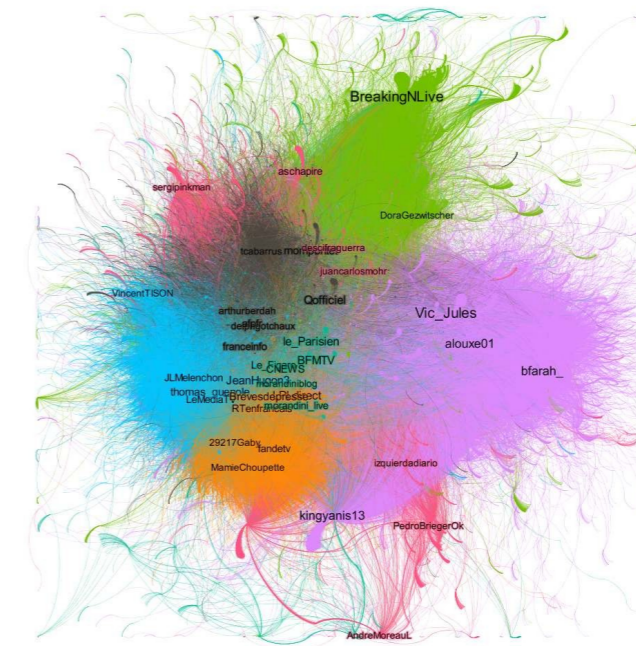
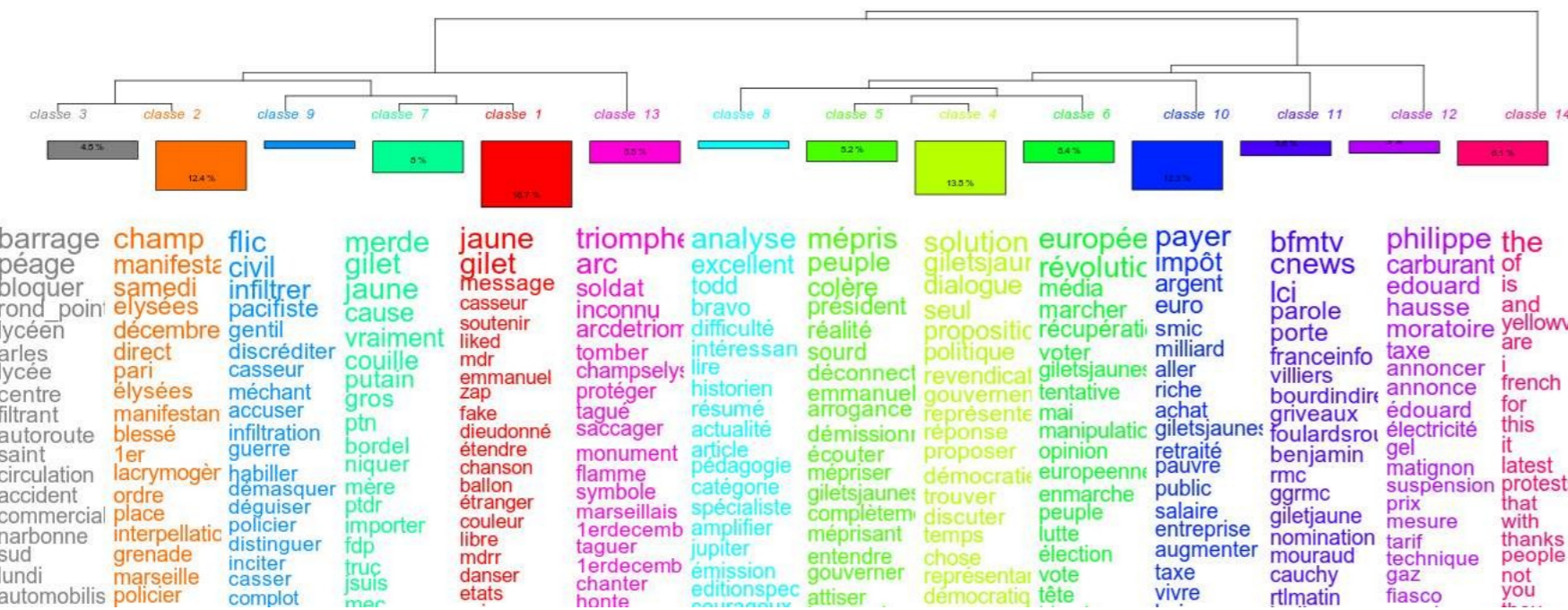
6. Corpus réseaux socio-numériques : comparer des « arènes » médiatiques d'un mouvement social (Gilets jaunes)

Natacha Souillard, Brigitte Sebbah, Lucie Loubère, Laurent Thiong-Kay et Nikos Smyrnaio



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

Dendrogramme résultat de la CHD, Corpus Facebook (Iramuteq)



Dendrogramme résultat de la CHD, Corpus Twitter (Iramuteq)

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

Une revue en ligne

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

