

Atelier pratique

Comment explorer des corpus de texte ?

Bénédicte Garnier

Institut National d'Études Démographiques (INED), F-93300 Aubervilliers, France

Lucie Loubère

Lerass – Psycom, Toulouse, France

avec Gaëlle Deletraz

Université de Pau, France

Ecole thématique 2020

EXPLO-SHS

12-16 oct. – La Rochelle

The logo for MATE-SHS features the text 'MATE-SHS' in a bold, sans-serif font. The 'M' and 'A' are orange, while 'TE-SHS' is black. A stylized orange circular graphic element is positioned behind the 'M' and 'A'.

Atelier pratique - Au menu en 2h45

Séquence introductive

- La méthodologie embarquée avec IRaMuTeQ
- Préparer les corpus à analyser
- Les menus et fenêtres d'IRaMuTeQ

Explorer des données textuelles (réponses à des questions ouvertes) avec IRaMuTeQ

- Trouver des thématiques, des spécificités, compter ...
- Interpréter : concordances, classifications, graphes de mots
- Présenter des résultats : visualisations, concordances

Partagez votre expérience par un retour sur cette mise en pratique

A disposition

Documents

- Installer IRaMuTeQ
- Tutos IRaMuTeQ (doc en ligne sur <http://iramuteq.org/documentation>)
- Pas à pas sur le formatage des données avec un tableur :
Corpus_reponses_candidats_Explo-SHS.xlsx

Données – les corpus avec les métadonnées préparés (bruts et avec recodage)

À partir de CorpusComplet2.txt

- th_definition.txt
- th_logiciel_c.txt
- th_materiauRech.txt
- th_matRech.txt
- th_momentExplo.txt
- th_parcForm.txt

A partir de CorpusCompletN.txt

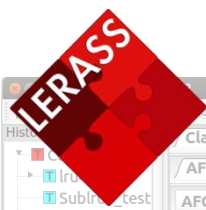
- th_apports.txt
- th_definition.txt
- th_logiciels.txt
- th_materiau.txt
- th_moments.txt
- th_parcour.txt
- th_3discipl.txt
- th_4metho.txt

Séquence introductive

Méthodologie embarquée

Construction des corpus

Présentation d'IraMuTeQ



Lexicométrie : éléments méthodologiques

Lucie Loubère 13/10/2020
Présentation adaptée de celles de Pascal Marchand et Pierre Ratinaud

The screenshot displays the Lexicométrie software interface. On the left is a file explorer with a tree view of folders and files. The main area is divided into several sections: a 3D graph (AFC) showing word clouds for different classes; a table titled 'Profils / Spécificités des TGenS' showing statistical data for five classes; a dendrogram (CHD) showing hierarchical clustering of classes; and a bar chart (Graphique) showing the distribution of specific terms across classes. The word clouds feature terms like 'mouvement', 'étudiant', 'école', 'université', 'excellence', 'janvier', 'décret', 'responsabilité', 'mouvement', 'grand', 'superieur', 'preparatoire', 'manifestation', 'classe', 'syndicat', 'occuper', 'tenter', 'perturber', 'gréviste', 'syndicat', 'jour', 'mouvement', 'étudiant', 'école', 'preparatoire', 'manifestation', 'classe', 'syndicat', 'occuper', 'tenter', 'perturber', 'gréviste', 'syndicat', 'jour'.

1 Classe 1	2 Classe 2	3 Classe 3	4 Classe 4	5 Classe 5
318/1419	249/1419	337/1419	181/1419	334/1419
22.41%	17.55%	23.75%	12.76%	23.54%

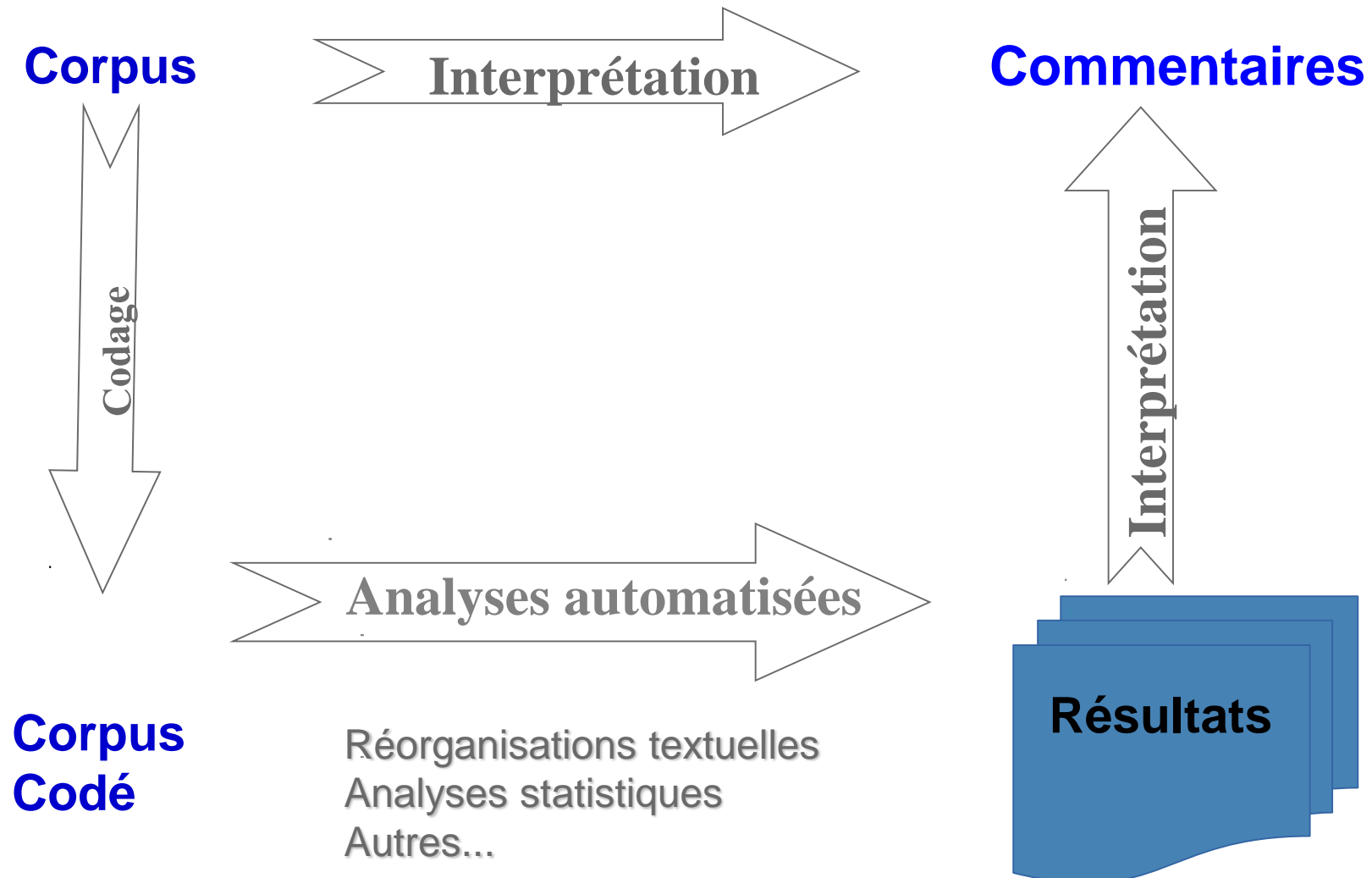
num	eff.	s.t.	eff. tota	pourcenta	chi2	Type	forme	P
0	49	66	74.24	98.88	nom	mouvement	< 0,0001	
1	95	204	46.57	70.22	nom	étudiant	< 0,0001	
2	2	28	85.71	61.36	nom	semaine	< 0,0001	
3	2	30	72.73	54.55	nom	re	< 0,0001	
4	2	44	68.18	0.2	nom	robot	< 0,0001	
5	15	15	100.0	49.25	nom	manifestation	< 0,0001	
6	21	26	80.77	48.2	nom	examen	< 0,0001	
7	18	22	81.82	42.17	nr	unef	< 0,0001	
8	10	24	79.17	41.67	nom	renne	< 0,0001	
9	14	22	78.57	36.12	nom	rentrer	< 0,0001	
10	18	22	81.82	42.17	nr	unef	< 0,0001	
11	10	24	79.17	41.67	nom	renne	< 0,0001	
12	14	22	78.57	36.12	nom	rentrer	< 0,0001	
13	18	22	81.82	42.17	nr	unef	< 0,0001	
14	10	24	79.17	41.67	nom	renne	< 0,0001	
15	14	22	78.57	36.12	nom	rentrer	< 0,0001	
16	18	22	81.82	42.17	nr	unef	< 0,0001	
17	10	24	79.17	41.67	nom	renne	< 0,0001	
18	14	22	78.57	36.12	nom	rentrer	< 0,0001	
19	18	22	81.82	42.17	nr	unef	< 0,0001	
20	10	24	79.17	41.67	nom	renne	< 0,0001	
21	14	22	78.57	36.12	nom	rentrer	< 0,0001	
22	18	22	81.82	42.17	nr	unef	< 0,0001	
23	10	24	79.17	41.67	nom	renne	< 0,0001	
24	14	22	78.57	36.12	nom	rentrer	< 0,0001	
25	18	22	81.82	42.17	nr	unef	< 0,0001	
26	10	24	79.17	41.67	nom	renne	< 0,0001	
27	14	22	78.57	36.12	nom	rentrer	< 0,0001	
28	18	22	81.82	42.17	nr	unef	< 0,0001	
29	10	24	79.17	41.67	nom	renne	< 0,0001	
30	14	22	78.57	36.12	nom	rentrer	< 0,0001	
31	18	22	81.82	42.17	nr	unef	< 0,0001	
32	10	24	79.17	41.67	nom	renne	< 0,0001	
33	14	22	78.57	36.12	nom	rentrer	< 0,0001	
34	18	22	81.82	42.17	nr	unef	< 0,0001	
35	10	24	79.17	41.67	nom	renne	< 0,0001	
36	14	22	78.57	36.12	nom	rentrer	< 0,0001	
37	18	22	81.82	42.17	nr	unef	< 0,0001	
38	10	24	79.17	41.67	nom	renne	< 0,0001	
39	14	22	78.57	36.12	nom	rentrer	< 0,0001	
40	18	22	81.82	42.17	nr	unef	< 0,0001	
41	10	24	79.17	41.67	nom	renne	< 0,0001	
42	14	22	78.57	36.12	nom	rentrer	< 0,0001	
43	18	22	81.82	42.17	nr	unef	< 0,0001	
44	10	24	79.17	41.67	nom	renne	< 0,0001	
45	14	22	78.57	36.12	nom	rentrer	< 0,0001	
46	18	22	81.82	42.17	nr	unef	< 0,0001	

Text snippets from the interface:
"chécée comme l'un des socles de la législature selon le premier ministre la loi pompeusement ptisée libertés et **responsabilités** des universités connaitrait elle à petit bruit le même sort et le trop fameux bouclier fiscal adopté dans le même temps
** *s_liberation *date_2009-04-09
comités d'experts nommés par les présidents d'universités tels que les ont envisagés les
** *s_lemonde *date_2010-06-22
** *s_lemonde *date_2010-12-10
imé des iut assure l'unipuit qui estime que depuis la **sponsabilités** des universités tru il y a deux ans et
tation qui gouvernent le capitalisme doit être libertés et aux **responsabilités** des universités tru
es libertés et **responsabilités** des universités qui ion cinquante et une universités sur

Quelques précautions d'usage...

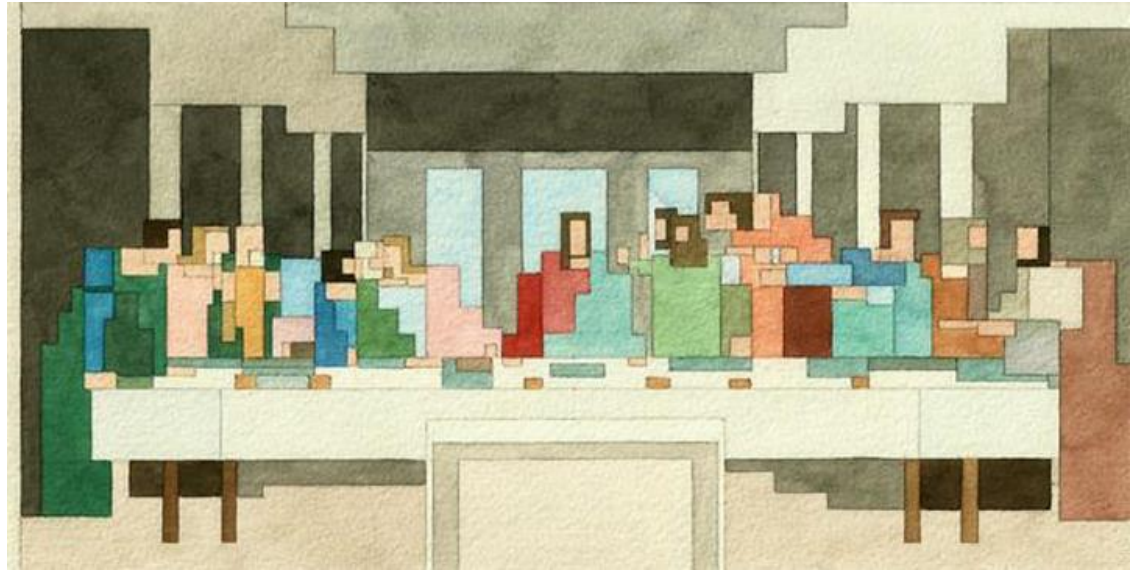
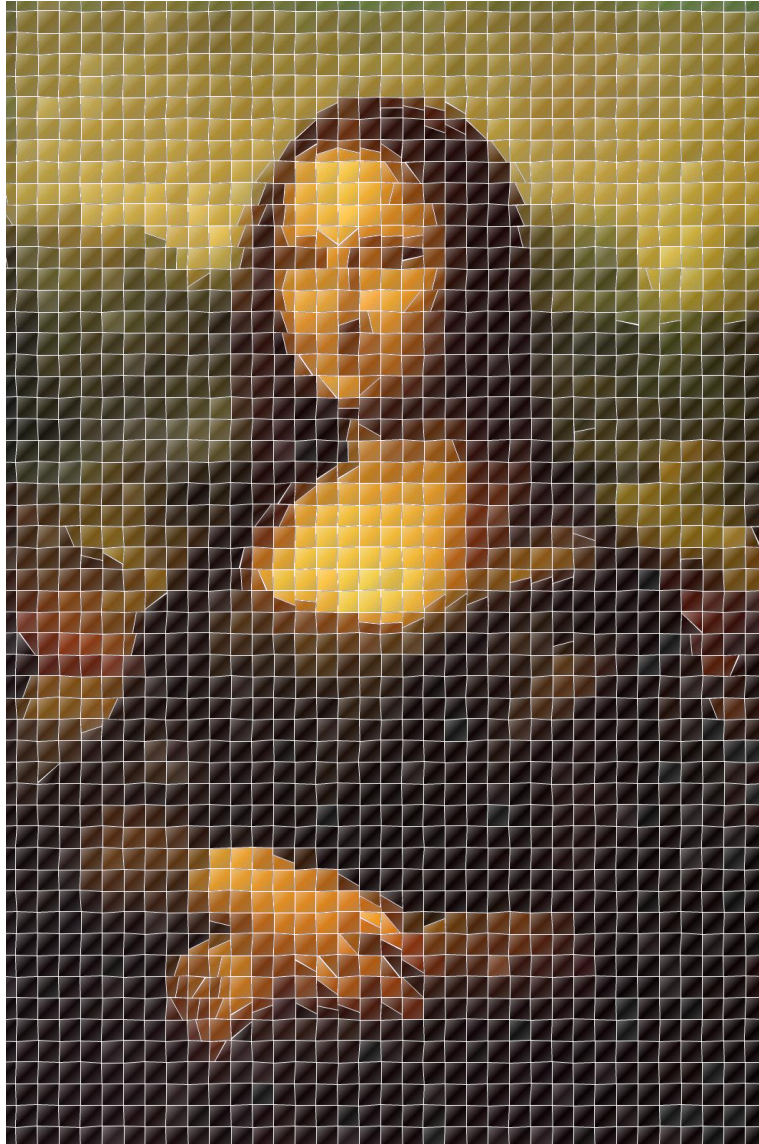
- ✓ La statistique n'est pas la seule approche possible ...
- ✓ *Aux chiffres, on leur fait dire n'importe quoi !*
- ✓ La statistique ne peut pas tout faire: il faut prévoir ce qu'on lui demandera :
HYPOTHÈSES
- ✓ Ce n'est pas la statistique qui garantit la qualité d'une recherche, mais le protocole.

L'interprétation en ADT



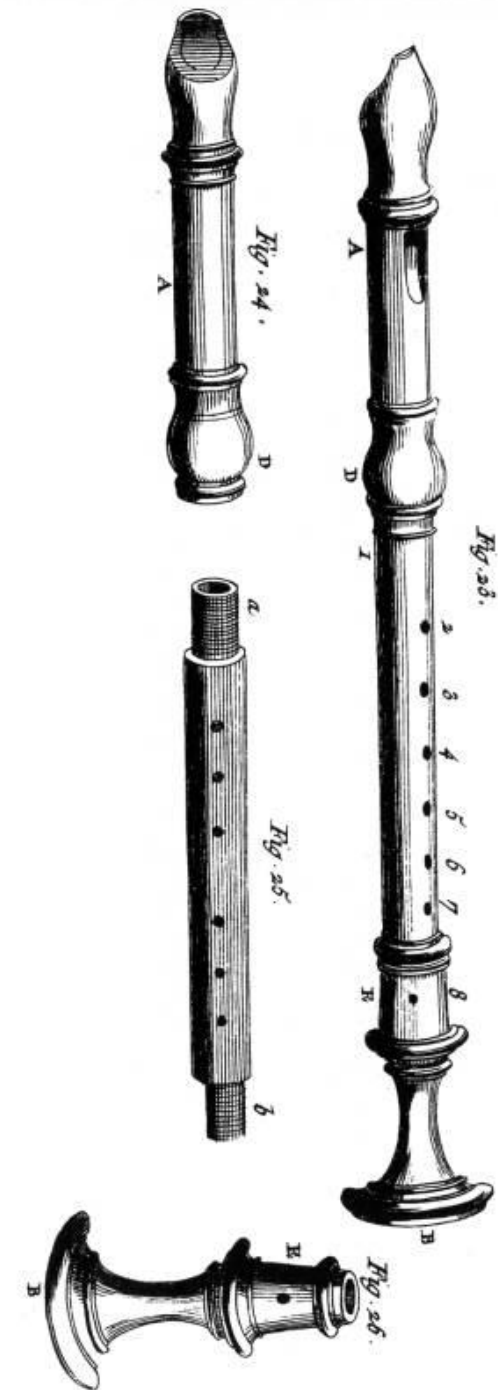
Merci à André Salem

Les résultats... une portion de l'objet de départ



Lebart & Salem (1994)

« Supposons (...) que l'on étudie les histogrammes des longueurs d'ondes correspondant aux couleurs d'un tableau de Rembrandt (pour chacun des pixels d'une reproduction). Il va de soi que l'on utilise une fraction dérisoire de l'information contenue dans l'image d'origine. Il est cependant possible que la forme de l'histogramme (ou d'une fonction plus élaborée des mêmes mesures et données de base) permette de distinguer un Rembrandt d'un Rubens ou d'un Van Dyck » (p.21).



Quelques logiciels de lexicométrie

- **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/>)

- **IRAMuTeQ** ➤ P. Ratinaud (Win, Mac, Linux)
(<http://www.iramuteq.org>)

Quelques 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).

- **Tableau lexical** (*formes * textes*)
- 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)

Analyse lexicale : 1 - *Tokenization*

Analyse lexicale : 1 - *Tokenization*

- Une suite de caractères bornée par deux caractères délimiteurs est une **occurrence** (*word-tokens*) : *Taille*.
 - espace, retour à la ligne, [(« ,.;?;!'/- _ »)]
- Deux suites identiques constituent deux occurrences d'une même **forme graphique** (*word-type*) : *Index*
- Normes de saisie (Labbé, 1990)
 - <http://halshs.archives-ouvertes.fr/docs/00/43/71/50/PDF/LabbeNormes.pdf>

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 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 a	533 tous	342 sa	235 président	195 croissance

Formes initiales / réduites

Lemmatisation

- On cherche à réduire les déclinaisons d'un mot en les ramenant à leur racine :
 - ✓ les verbes → infinitif
 - ✓ Les noms et adjectifs → masculin singulier
- Exemple :

Le petit chat est mort, c' est dommage il était sympa le chat	13 occurrences 10 formes
Le petit chat être mort, c être dommage il être sympa le chat	13 occurrences 9 formes

Formes initiales / réduites

Lemmatisation

Lemmatisation = réduction information...

Réponses des conducteurs RATP suisses sur leur métier :

Conduire Boire → Lemme du pp
Passager bus et
Nom du véhicule bus

Réponses des conducteurs sur leur rapport à leur voiture :

prendr rout
e e
vole → Lemme du pp volant
Nom : le volant de la voiture...

Formes initiales / réduites

Lemmatisation

- **Comment juger de la catégorie grammaticale d'un mot ?**
- Dictionnaire à étiquettes.

pariğots	pariğot	adj	m	p	0.4	1.22	0.14	0.14	
parions	parier	ver			81.96	15.61	0.38	0.2	imp:pre:lp;ind:imp:lp;ind:pre:lp;
paris	pari	nom	m	p	26.61	12.57	12.68	7.97	
paris_brest	paris_brest	nom	m			0.19	0.2	0.19	0.2
parisien	parisien	adj	m	s	2.73	30.61	0.94	12.09	
parisienne	parisien	adj	f	s	2.73	30.61	1.11	10.07	
parisiennes	parisien	adj	f	p	2.73	30.61	0.16	2.3	
parisiens	parisien	nom	m	s	2.73	30.61	1.26	6.40	

➤ **Autres solutions :**

TreeTagger - a language independent part-of-speech tagger

<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

Cordial Analyseur

<http://www.synapse-fr.com/>

Lexique 3 (Paris 5)

<http://www.lexique.org/>

Analyse lexicale : 2 – Partition / segmentation

Analyse lexicale : 2 – Partition / segmentation

- Le corpus = l'ensemble des textes (dans un fichier)
- Le texte (uci dans alceste) = un entretien, un chapitre, un livre, discours...
- Le segment de texte (uce dans alceste) = une portion du texte la taille est choisie par le chercheur
- Chacune de ces délimitations est choisie en fonction des éléments à étudier et des hypothèses du chercheur.

Les textes...

Le corpus

**** *source_bulletin_quotidien *date_2020-03-09 *am_2020-03
*annee_2020

La France fourbit ses armes sanitaires face à la propagation du coronavirus. La France fourbit ses armes sanitaires face à la propagation du coronavirus La France fourbit ses armes sanitaires face à la propagation du coronavirus Le président de la République Emmanuel MACRON a réuni hier en fin de journée à l'Élysée un conseil de Défense alors que le pays est le deuxième le plus affecté par le coronavirus en Europe, après l'Italie dont le nord a été placé en quarantaine. En France, cinquième pays le plus touché dans le monde, le dernier bilan officiel fait état de 1126 personnes contaminées au Covid-19 depuis le début de l'épidémie et 19 décès.

**** *source_lest_republicain *date_2020-03-07 *am_2020-03
*annee_2020

Confinement total à l'Ehpad de Bellevaux et au centre Weinman d'Avanne Afin de « prévenir le risque épidémique » lié au coronavirus -COVID-19, le centre de Long Séjour de Bellevaux, situé au centre-ville de Besançon, et le centre de soins Jacques-Weinman, situé à Avanne-Aveney, ont été placés en « confinement total » ce vendredi 6 mars, à partir de 12 h. Les visites des familles sont désormais proscrites. Les fournisseurs ou livreurs sont également invités à ne plus franchir le seuil de ces établissements. Ce sont les employés, habilités, qui viendront à leur rencontre.

2 textes

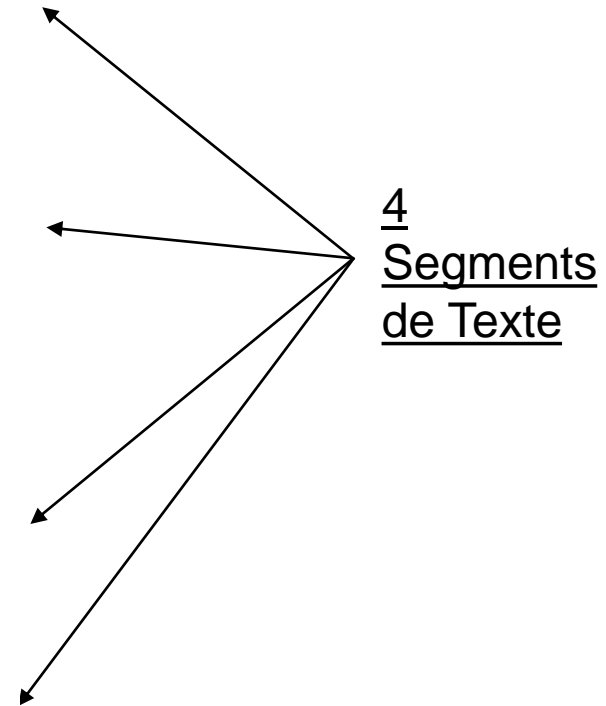
Les segments de textes ...

**** *source_bulletin_quotidien *date_2020-03-09 *am_2020-03
*annee_2020

La France fourbit ses armes sanitaires face à la propagation du coronavirus. La France fourbit ses armes sanitaires face à la propagation du coronavirus La France fourbit ses armes sanitaires face à la propagation du coronavirus Le président de la République Emmanuel MACRON a réuni hier en fin de journée à l'Elysée un conseil de Défense alors que le pays est le deuxième le plus affecté par le coronavirus en Europe, après l'Italie dont le nord a été placé en quarantaine. En France, cinquième pays le plus touché dans le monde, le dernier bilan officiel fait état de 1126 personnes contaminées au Covid-19 depuis le début de l'épidémie et 19 décès.

**** *source_lest_republicain *date_2020-03-07 *am_2020-03
*annee_2020

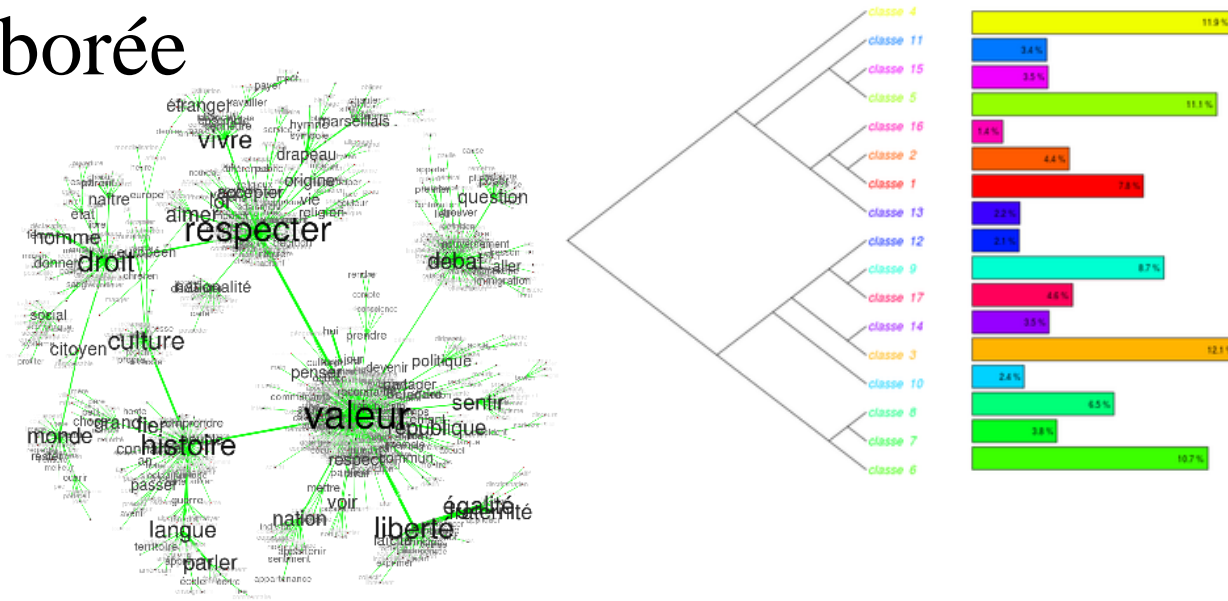
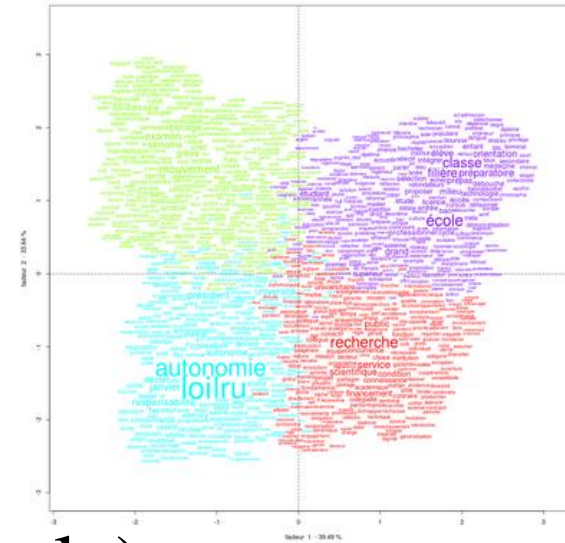
Confinement total à l'Ehpad de Bellevaux et au centre Weinman d'Avanne Afin de « prévenir le risque épidémique » lié au coronavirus -COVID-19, le centre de Long Séjour de Bellevaux, situé au centre-ville de Besançon, et le centre de soins Jacques-Weinman, situé à Avanne-Aveney, ont été placés en « confinement total » ce vendredi 6 mars, à partir de 12 h. Les visites des familles sont désormais proscrites. Les fournisseurs ou livreurs sont également invités à ne plus franchir le seuil de ces établissements. Ce sont les employés, habilités, qui viendront à leur rencontre.



Analyse lexicale : 3 - Statistiques

Analyse lexicale : 3 - Statistiques

- *Formes* et réponses caractéristiques, **spécificités** (profil)
- Méthodes factorielles
- Classification automatique
- Calculs de distance (connexion lexicale) et représentation arborée
- Cooccurrences et similitudes



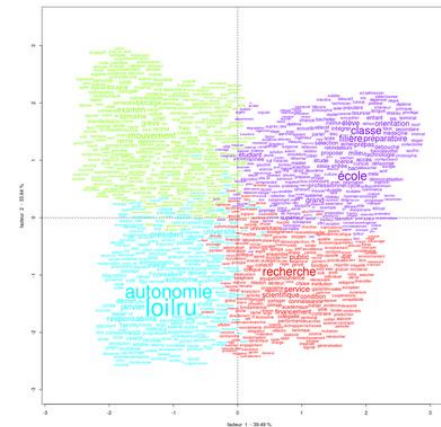
AFC et spécificités

Tableau de contingence

	*source la croix	*source le figaro	*source le monde	*source l'humanité	*source libération
sécurité	15	46	28	43	15
devenir	33	27	35	17	40
établissement	1	7	12	4	3
tirer	13	4	14	6	12
pencher	0	4	3	5	2
brexit	5	7	4	0	3
reporter	3	5	4	0	0
évolution	4	13	7	3	4
chine	1	5	2	0	2
connaissance	1	2	7	5	1
naturel	0	2	3	1	4
attente	4	5	8	3	3
unir	11	9	15	5	11

Calcul probabiliste des répartitions → Lexique spécifique des sous-ensembles

Analyse factorielle de correspondances →



CHD : classification hiérarchique descendante

- Méthode Reinert
 - Initialement dans le logiciel Alceste,
 - Classification hiérarchique descendante
 - Construction tableau lexical (croise les lemmes avec les segments)

Formes actives (adjectifs,
nom, verbes actions)

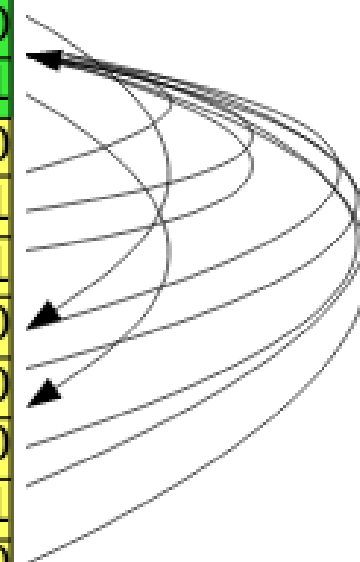
le petit chat être mort →
c être dommage il être sympa le chat →

	petit	chat	mort	dommage	sympa
le petit chat être mort	1	1	1	0	0
c être dommage il être sympa le chat	0	1	0	1	1

CHD : classification hiérarchique descendante

- Méthode Reinert
 - Sur tableau lexical, comparaison des segments sur présence et absence des formes.
 - Découpe le corpus en deux classes pour extraire le maximum d'inertie

id	a	b	c	d	e	f	g	h	i
1	1	1	1	1	0	0	0	0	0
6	1	1	1	1	0	0	0	0	1
8	1	0	1	0	1	0	0	0	0
4	1	0	1	0	1	0	0	0	1
9	0	0	1	0	1	0	1	0	1
3	0	0	1	0	1	0	1	0	0
5	0	0	1	0	1	0	1	0	0
10	0	0	1	0	1	0	1	0	0
2	0	0	0	0	1	1	1	1	1
7	0	0	0	0	1	1	1	1	0



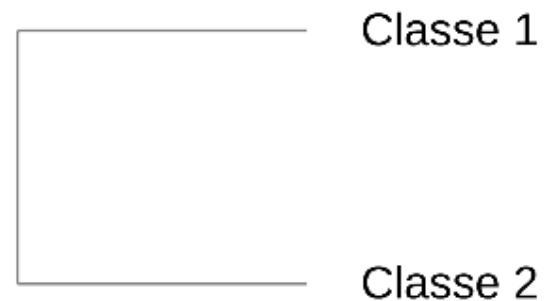
CHD : classification hiérarchique descendante

- Méthode Reinert

id	a	b	c	d	e	f	g	h	i
1	1	1	1	1	0	0	0	0	0
6	1	1	1	1	0	0	0	0	1
8	1	0	1	0	1	0	0	0	0
4	1	0	1	0	1	0	0	0	1
9	0	0	1	0	1	0	1	0	1
3	0	0	1	0	1	0	1	0	0
5	0	0	1	0	1	0	1	0	0
10	0	0	1	0	1	0	1	0	0
2	0	0	0	0	1	1	1	1	1
7	0	0	0	0	1	1	1	1	0

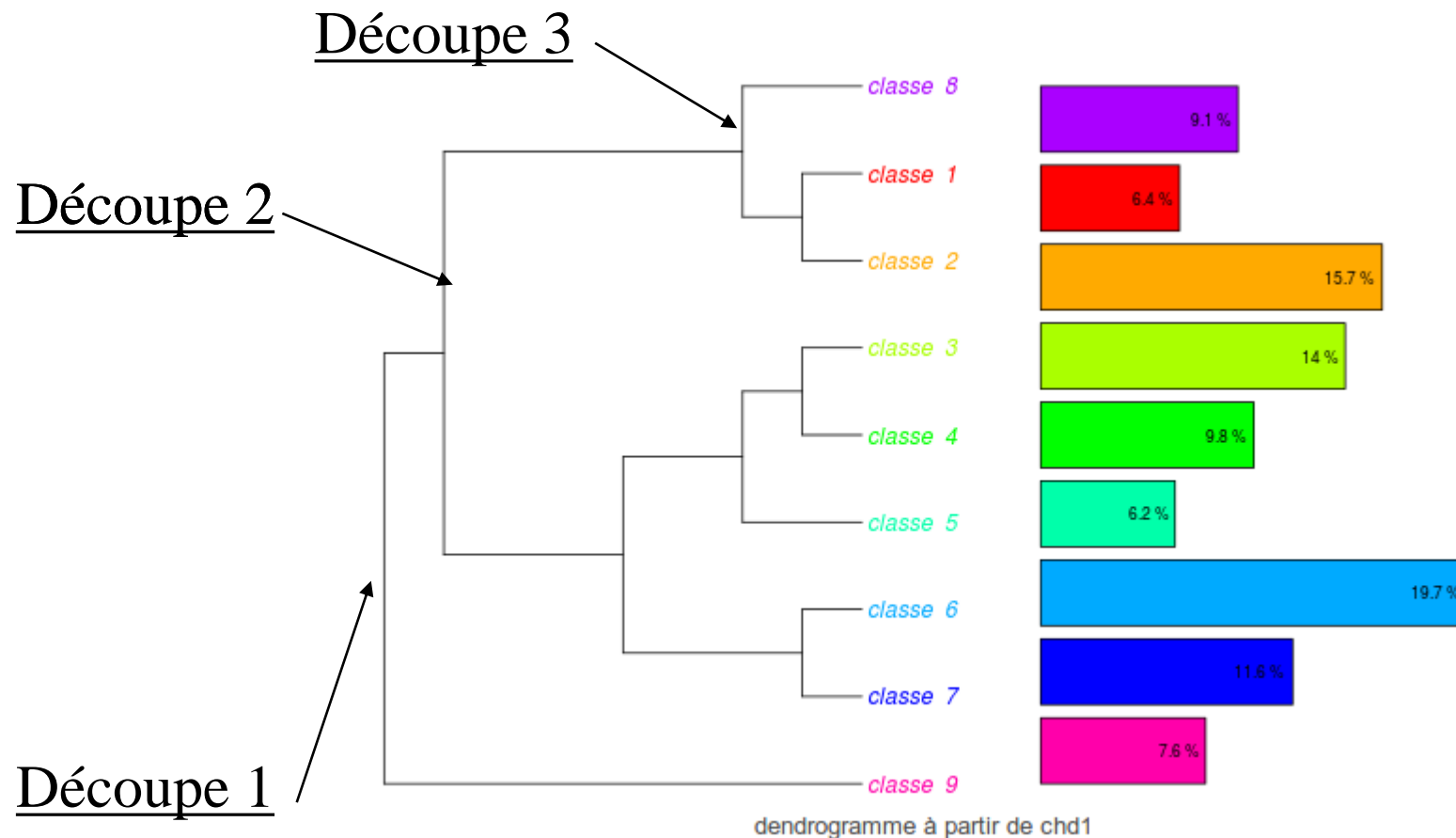
Classe 1

Classe 2



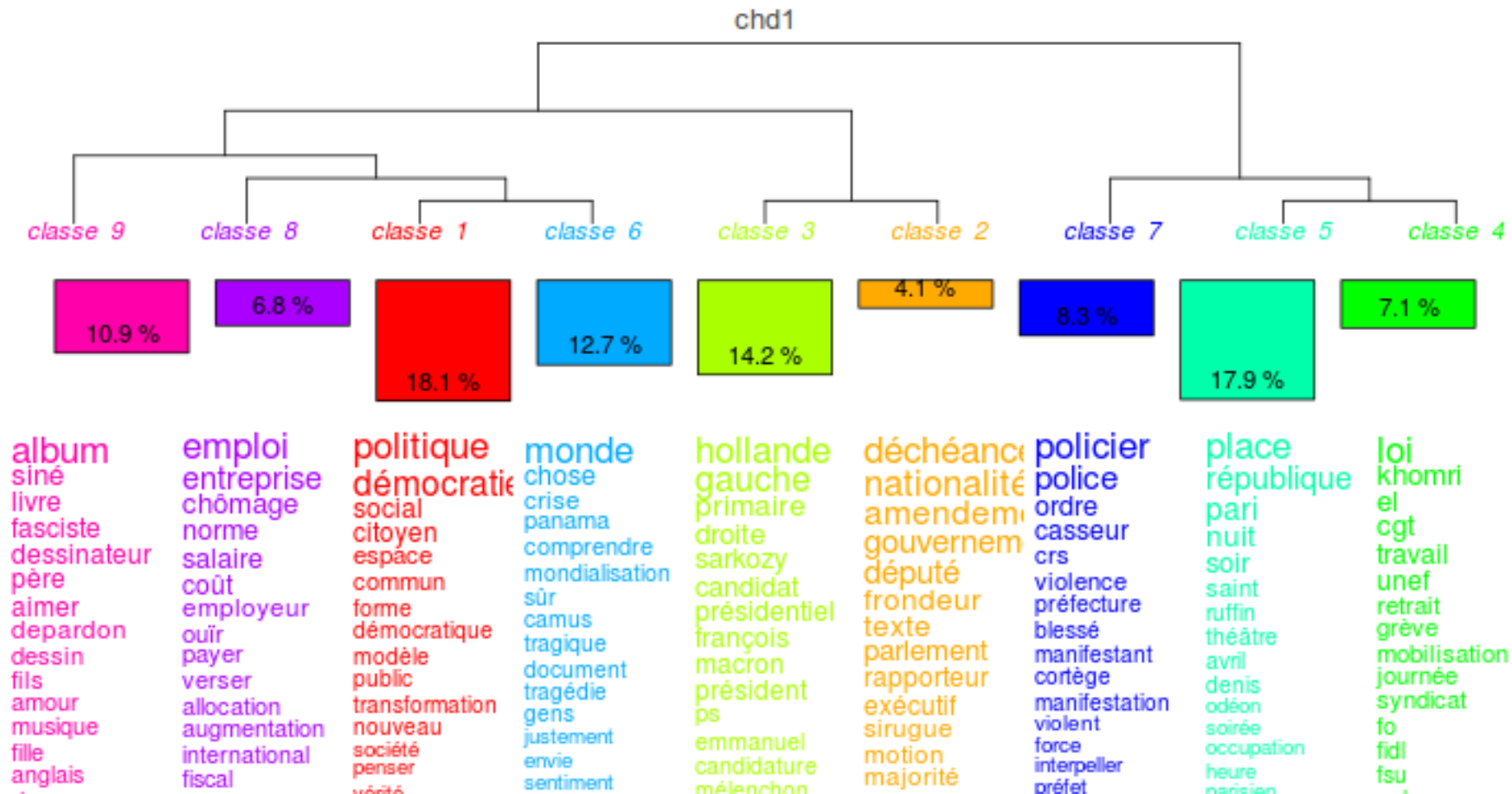
CHD : classification hiérarchique descendante

- Prend la classe la plus grosse, et refait le découpage
- Recommence jusqu'à avoir le nombre de classe défini en phase 1



CHD : classification hiérarchique descendante

Édition des profils : formes significativement sur représentées dans la classe.



CHD : classification hiérarchique descendante

- Méthode Reinert

- Édition des profils :

Classe 1						
	eff.	eff. Tot.	pourc.	chi2	forme	p
	2	2	100	10.00	b	0.001565402
	4	2	2	100	d	0.001565402
	1	2	4	50	a	0.052807511
	3	2	8	25	c	0.429195300
	5	1	4	25	i	0.746885633

	Classe 1	Classe 6	Classe 5	Classe 3	
a	2	2	0	0	4
b	2	0	0	0	2
c	2	2	4	0	8
d	2	0	0	0	2
e	0	2	4	2	8
f	0	0	0	2	2
g	0	0	4	2	6
h	0	0	0	2	2
i	1	1	1	1	4
	9	7	13	9	38

	dans la Classe	dans l'ensemble Des autres classes	
nb d'unité Avec la forme	2	0	2
nb d'unité Sans la forme	0	8	8
	2	8	10

Nbr total d'unités classées

Effectifs observés

Nbr d'unités dans la classe

	dans la Classe	dans l'ensemble Des autres classes	
nb d'unité Avec la forme	0,4	1,6	2
nb d'unité Sans la forme	1,6	6,4	8
	2	8	10

Si Eff. Obs. > Eff. Théo. → profil
Sinon → antiprofil

Chi2 = 10

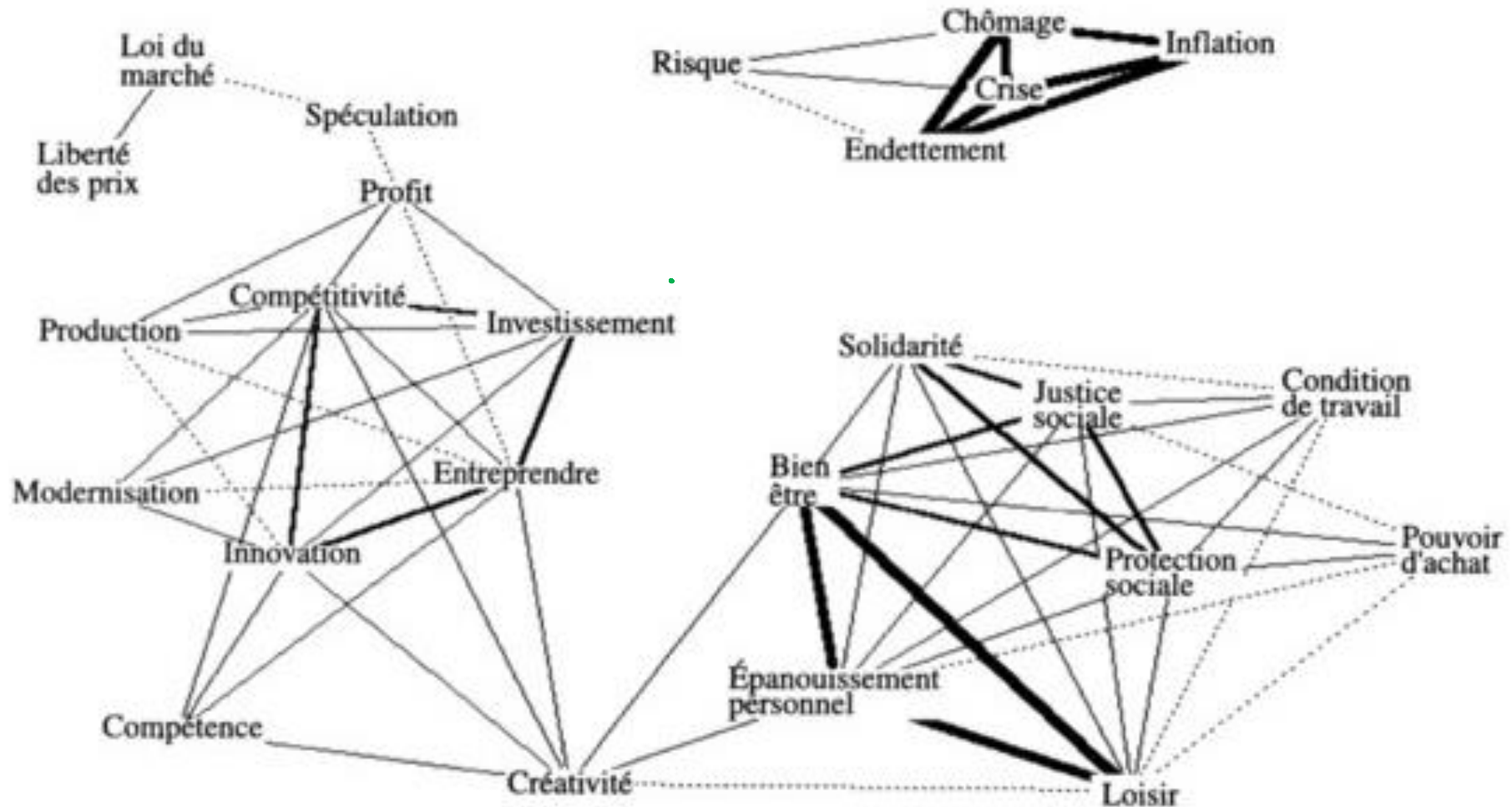
Effectifs théoriques

L'analyse de similitude (ADS)

- Méthode(s) graphique(s) pour l'étude des relations entre les parties d'un ensemble.
- Les matrices :
 - Distances (euclidiennes, maximum, Jaccard, Manhattan...)
 - Similitude (cooccurrences, Jaccard, simple matching, phi...)
- L'ADS est classiquement utilisée pour décrire des représentations sociales, sur la base de questionnaires d'enquête.
 - Flament, 1962 ; Flament, 1981 ; Vergès & Bouriche, 2001.

P. Vergès, 2001, « L'analyse des représentations sociales par questionnaires », *Revue française de sociologie*, 42 (3), 537-561

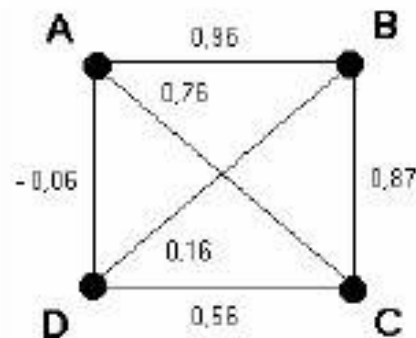
FIGURE XII. – France, graphe des relations données
(entre 32 termes décrivant l'univers socio-économique) par au moins 30 % des étudiants



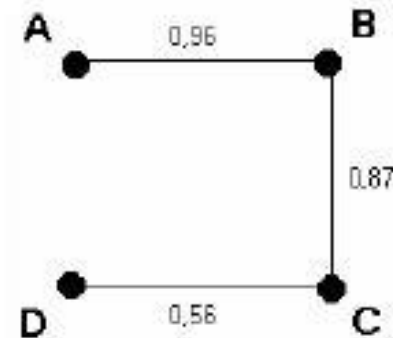
L'analyse de similitude (ADS)

- étudier la proximité et les relations entre les éléments d'un ensemble, généralement sous forme d'*arbres maximum* :
 - le nombre de liens entre deux items évoluant « comme le carré du nombre de sommets » (Flament & Rouquette, 2003 : 88), l'ADS cherche à réduire le nombre de ces liens pour aboutir à « un graphe connexe et sans cycle » (Degenne & Vergès, 1973 : 473).
 - L'« arbre maximum ») est créé par les arêtes les plus fortes du graphique. C'est l'arbre le plus simple que l'on peut obtenir, mais c'est aussi le plus lourd (en termes d'information).
 - On considère toutes les « cliques » possibles (ex. ABCA, BCDB) et on élimine les liens les plus faibles (ex. entre A et C et entre B et D).

Arbre de similitude



Arbre maximum



En amont : Préparer les données

= Etape de l'analyse

- Diffère selon les types de corpus (questions ouvertes, entretiens, romans, articles, pages Web etc..) ;
- Pré-traiter les textes et métadonnées = nettoyer, normaliser

La transformation des données textuelles en tableaux lexicaux (objets numériques) est faite par les logiciels

Changer la « structure » du fichier : pour la mise en forme IRaMuTeQ

Format particulier = les lignes étoilées

Corpus_complet2.txt

- Marqueurs entre les individus
- Métadonnées (variables qualitatives) : Sexe, Statut, Discipline 1 à 3, ...
- Préfixe ->variable Suffixe --> modalité

Plusieurs questions ouvertes dans le même fichier → plusieurs variables textuelles

- th_def = Pour vous, en quoi consiste l'exploration et quelle(s) difficulté(s) soulève-t-elle ?-> **Definition_Explorer**
- th_momentExplo= A quel(s) moment(s) et de quelle(s) façon(s) faites-vous de l'exploration dans votre pratique professionnelle ?-> **Moments_explo**
- th_parcForm = Décrivez en quelques lignes votre parcours de formation (initiale et - éventuellement - complémentaire)-> **Parcours_formation**
- th_matRech = Avec quels matériaux et/ou jeux de données travaillez-vous régulièrement ? _Précisez la forme, le contenu et la masse des jeux de données →**Materiaux_recherche**
- th_Logiciel= Avec quels logiciels traitez-vous et analysez-vous vos jeux de données, avec quel niveau de maîtrise (débutant, confirmé, expert) ? ->**Logiciels**

Mise en forme « Alceste »

Marqueurs entre les individus

Métadonnées

Thématiques

```
*corpus_complet2.txt - Bloc-notes
Fichier Edition Format Affichage Aide
**** *id_1 *sex_Femme *statut_Doctorante *disc1_Sociologie *disc2_Anthropologie *disc3_Santé_Publique
-*th_def
L'exploration relève à la fois d'une étape et d'une démarche du travail de recherche, ainsi que d'un
elle, je ne veille pas toujours à enregistrer et traiter sérieusement ces données. De nombreuses fois
-*th_momentExplo
En tant que chercheuse du Sud menant une enquête sur un terrain du Nord, je me vois désormais exposée
-*th_parcForm
2010 - 2015 : Double licence internationale en sciences sociales (Parcours Sociologie, Anthropologie
-*th_matRech
1. Entretiens semi-directifs 2. Observation non-participante et participante 3. Archives (rapports of
-*th_logiciel
1. Entretiens semi-directifs 2. Observation non-participante et participante 3. Archives (rapports of
**** *id_2 *sex_Femme *statut_Post-doctorante *disc1_Sociologie *disc2_Géographie *disc3_Anthropologi
-*th_def
L'ethnographie donne une large place à l'exploration et le théorise. Dans cette approche, le régime d
endant le tournage d'une vidéo ou en participant à une situation ou un échange sur les réseaux sociaux
-*th_momentExplo
En tant qu'ethnographe, je me considère comme un acteur au milieu d'autres acteurs, dont je suis les
-*th_parcForm
J'ai soutenu une thèse en sociologie à l'EHESS en 2016 sur des mobilisations protestataires à Moscou.
-*th_matRech
Dans mon postdoctorat j'étudie des mobilisations d'habitants contre un projet de décharge dans le Gra
-*th_logiciel
Excel (expert) ; ignorante sur les autres. J'aimerais comprendre ce que je pourrais utiliser.
**** *id_3 *sex_Femme *statut_Post-doctorante *disc1_Géographie *disc2_Démographie *disc3_ *methode1_
...

```

Corpus_complet2.txt

Préparer les corpus à analyser

Un seul fichier à disposition au format tableur (.csv, .xlsx)

- 105 Réponses au questionnaire et descriptif (2 onglets)
- Caractéristiques des répondants : variables qualitatives
- Réponses aux questions ouvertes : textes courts

→ Formater les données avec un tableur puis les coller dans un éditeur de texte et générer un fichier texte (.txt)

Créer autant de fichiers que de textes courts à explorer

→ Utiliser l'outil extraire à partir des thématiques d'IRaMuTeQ

Les données sont dans un tableau

Réponses aux questions ouvertes (surlignées en jaune) et métadonnées

Attention au nombre de modalités et à l'écriture des métadonnées → Recoder les variables Statut et Discipline1

	A	B	C	D	E	F	G	H	I	J	M	N	O	P	Q	R	S	T	U	
1	N°Obs	Sexe	Statut	Statut_Autre	Discipline1	Discipline2	Discipline3	Parcours_for	Materiaux_r	Logiciels	Methode2	Frequence_r	Methode3	Frequence_r	Methode4	Frequence_r	Freque	Definition_E	Moments	
2		Identité de g	Statut*	Précisez votr	Discipline 1	Discipline 2	Discipline 3	Décrivez en r	Avec quels r	Avec quels lc	Méthode	A quelle fré	Méthode	A quelle fré	Méthode	A quelle fré	A quell	Pour vous, e	A quel(s) r	
3	1	Femme	Doctorant.e		Sociologie	Anthropolog	Santé Public	2010 – 2015 : 1. Entretien	1. Entretien	ois par semaine)									L'explorati	En tant qu
4	2	Femme	Post-doctorant.e		Sociologie	Géographie	Anthropolog	J'ai soutenu	Dans mon pc	Excel (exper	analyse spat	Régulièreme	statistiques	Régulièrement	(1 fois par mois)				L'ethnograph	En tant qu'
5	3	Femme	Post-doctorant.e		Géographie	Démographie		Domaine d'e	Méthodes m	R niveau déb	stat. inféren	De temps en	analyse spat	Rarement	(1 ou 2 fois par an)				Stricto sensu	J'accorde u
6	4	Femme	Assistant.e	ii	Topographe	Géomatique	Topographie	Formation												
7	5	Homme	Ingénieur.e	de recherche	Archéologie	Epistémolog	Histoire des	Géographe-c	J'utilise d'un	QGIS, ArcGIS	analyse spat	Régulièrement	(1 fois par mois)						Le premier p	Aussi bien
8	6	Femme	Professeur.e	des universi	Etudes Cultu	Humanités n	Théâtrologie	J'ai fait des é	Base de don	Logiciels Exo	analyse com	Rarement (1	analyse quar	Régulièrement	(1 fois par mois)				- en quoi cor	- Dans la cc
9	7	Homme	Doctorant.e		Sociologie	Anthropologie		Je suis agrég	Les archives:	Sonal (exper	analyse text	Souvent (1 f	analyse de d	Régulièrement	(1 fois par mois)				L'explorati	Lors de séj
10	8	Femme	Chargé.e	de recherche	Anthropolog	Science Polit	Economie	J'ai en docto	Entretien	fo Excel	ois par semaine)								L'explorati	Je fais de l
11	9	Homme	Ingénieur.e	de recherche	sociologie			Double Deug	Je travaille s	SAS, Nvivo, S	analyse des	De temps en	analyse d'en	De temps en	temps (5 ou 6 fois par an)				je la vois sur	en fait je c
12	10	Homme	Ingénieur.e	d'études	Informatique	Linguistique		Je suis titulai	Je travaille p	Python (conf	Analyse de r	Rarement (1	Analyse qual	Régulièrement	(1 fois par mois)				Je dirais que	Comme in
13	11	Femme	Doctorant.e		Géographie	Tourisme	Big Data	J'ai obtenu u	Ma thèse cor	Excel (exper	Analyses Mu	De temps en	Analyse de r	Rarement (1	Statistiques	Rarement (Rareme		Selon moi, l'	Ma thèse c
14	12	Femme	Doctorant.e		Etudes des r	Science du je	Sciences de	Formation in	Données issi	Encore à déf	Analyse text	Régulièreme	Analyse mul	De temps en	temps (5 ou 6 fois par an)				L'explorati	Doctorante
15	13	Femme	Doctorant.e		Urbanisme	Géographie	ethnologie													
16	14	Femme	Ingénieur.e	d'études	Géomatique	Cartographie	Géographie	Avec une for	Des fichiers	QGIS (expert	Analyse qual	De temps en	Analyse qual	Souvent (1 f	Webscraping		De tem		L'explorati	Je fais de l
17	15	Homme	Ingénieur.e	d'études	Sociologie			Master profe	Résultats d'e	Python (exp	statistiques	Régulièreme	statistiques	De temps en	analyse de réseau		Rareme		Il s'agit d'ext	Dans le cac
18	16	Femme	Ingénieur.e	d'études	Sciences-humaines			Issue d'un pa	Je document	SAS	ois par semaine)								C'est une ph	Pour veille

Corpus_complet2.txt

Créer la lignes contenant les métadonnées

Utiliser la fonction Concaténer pour générer les lignes étoilées

Ajout de colonnes (ici azerty) que l'on remplacera par des sauts de ligne dans le fichier .txt à importer dans IRaMuTeq

Préparation du fichier
CorpusCompletN.txt
à partir du fichier
Corpus_reponses_candidats_Explo-
SHS.xlsx

**** *n_10 *Sex_Hom *Stat_Ingénieur *Disc1_Autre *Meth1_An_textuelle *Freq_meth1_Régult	azerty	-*th_parcour	azerty	Je suis titulaire	azerty	-*th_materiau	azerty	Je travaille pri	azerty	-*th_l
**** *n_11 *Sex_Fem *Stat_Doctorant_e *Disc1_Geo *Meth1_An_Spatiale *Freq_meth1_Svt	azerty	-*th_parcour	azerty	J'ai obtenu un	azerty	-*th_materiau	azerty	Ma thèse con	azerty	-*th_l
**** *n_12 *Sex_Fem *Stat_Doctorant_e *Disc1_Autre *Meth1_An_quali *Freq_meth1_Régult	azerty	-*th_parcour	azerty	Formation init	azerty	-*th_materiau	azerty	Données issue	azerty	-*th_l
**** *n_13 *Sex_Fem *Stat_Doctorant_e *Disc1_Geo *Meth1_*Freq_meth1_	azerty	-*th_parcour	azerty		azerty	-*th_materiau	azerty		azerty	-*th_l
**** *n_14 *Sex_Fem *Stat_Ingénieur *Disc1_Geo *Meth1_An_Spatiale *Freq_meth1_Svt	azerty	-*th_parcour	azerty	Avec une form	azerty	-*th_materiau	azerty	Des fichiers g	azerty	-*th_l
**** *n_15 *Sex_Hom *Stat_Ingénieur *Disc1_Sociologie *Meth1_An_textuelle *Freq_meth1_Tps_en_tps	azerty	-*th_parcour	azerty	Master profes	azerty	-*th_materiau	azerty	Résultats d'er	azerty	-*th_l
**** *n_16 *Sex_Fem *Stat_Ingénieur *Disc1_SHS *Meth1_Stat_desc *Freq_meth1_Svt	azerty	-*th_parcour	azerty	Issue d'un par	azerty	-*th_materiau	azerty	Je documente	azerty	-*th_l
**** *n_17 *Sex_Fem *Stat_Ingénieur *Disc1_Geo *Meth1_An_Spatiale *Freq_meth1_Svt	azerty	-*th_parcour	azerty	J'ai suivi un p	azerty	-*th_materiau	azerty	Je travaille av	azerty	-*th_l
**** *n_18 *Sex_Hom *Stat_Ingénieur *Disc1_Geo *Meth1_An_Spatiale *Freq_meth1_Svt	azerty	-*th_parcour	azerty	Formation de	azerty	-*th_materiau	azerty	Des bases de	azerty	-*th_l
**** *n_19 *Sex_Fem *Stat_Ingénieur *Disc1_Linguis *Meth1_An_textuelle *Freq_meth1_Svt	azerty	-*th_parcour	azerty	- DESS en lexi	azerty	-*th_materiau	azerty	- des bases de	azerty	-*th_l
**** *n_20 *Sex_*Stat_Doctorant_e *Disc1_Anthropo *Meth1_An_quali *Freq_meth1_Svt	azerty	-*th_parcour	azerty	Après avoir ob	azerty	-*th_materiau	azerty	Mes méthode	azerty	-*th_l
**** *n_21 *Sex_Fem *Stat_Doctorant_e *Disc1_SciencesPo *Meth1_Stat_desc *Freq_meth1_Svt	azerty	-*th_parcour	azerty	J'ai commenc	azerty	-*th_materiau	azerty	J'ai été recrut	azerty	-*th_l
**** *n_22 *Sex_Fem *Stat_Ingénieur *Disc1_Linguis *Meth1_An_textuelle *Freq_meth1_Svt	azerty	-*th_parcour	azerty	formation uni	azerty	-*th_materiau	azerty	j'étudie le lexi	azerty	-*th_l
**** *n_23 *Sex_Fem *Stat_Doctorant_e *Disc1_Educ *Meth1_*Freq_meth1_	azerty	-*th_parcour	azerty		azerty	-*th_materiau	azerty		azerty	-*th_l
**** *n_24 *Sex_Fem *Stat_Ingénieur *Disc1_Sociologie *Meth1_Stat_desc *Freq_meth1_Svt	azerty	-*th_parcour	azerty	Master 2 # Ex	azerty	-*th_materiau	azerty	Majoritairem	azerty	-*th_l

Lecture des onglet de la droite vers la gauche pour suivre les étapes de codage

Au l'éditeur de texte : Insérer les sauts de ligne

**** *id_ 1 *sex Femme *statut Doctorante *disc1_Sociologie *disc2_Anthropologie *disc3_Santé Publique *methode1_Analyse qualitative manuelle *freqM1 Souvent 1 fois par semaine *methode2_ *FreqM2_ *methode3_ *freqM3_ *methode4_ *freqM4_ azerty -*th def azerty "L'exploration relève à la fois d'une étape et d'une démarche du travail de recherche, ainsi que d'un état d'esprit du chercheur-explorateur. Une étape de la recherche, dans le sens où, pour moi, elle fait référence à un moment (ou des moments) de « tâtonnement » d'un inconnu (sujet, base de données, etc.) qui nous interpelle. Une démarche, dans le sens où elle implique des actions précises, en même temps que, d'une certaine manière, expérimentales, intuitives. Enfin, elle implique, voire exige, un état d'esprit particulier du chercheur où il se livre aux « petits riens » du social qui s'insinuent. Ceci dit, paradoxalement, la logique de l'exploration ne s'arrête pas à une étape, à une démarche ou à un état d'esprit : elle traverse, avec plus ou moins d'intensité selon les moments, le travail du

Les plus grades difficultés qu'elle soulève pour moi en ce moment font référence à l'archivage, pu Comme j'explore de manière très intuitive et informelle, je ne veille pas toujours à enregistrer et t suis, par exemple, contrainte de relancer une recherche pour retrouver des éléments que j'avais « fois je ne retrouve même plus ce que je souhaitais..." azerty -*th momentExplo azerty En dans ce cadre de devoir explorer encore et encore des facettes cachées qui peuvent me révéler dans un cadre d'intimité et dépaysement qui m'invite à garder l'état d'esprit de l'exploration et la les questions ou inconnus qui m'interrogent j'utilise notamment des moteurs de recherches et *th parcForm azerty 2010 – 2015 Double licence internationale en sciences sociales à l'Univ Sociologie XXX avec un séjour à l'Université de XXX. 2017-2018 Master 2 Recherche en sociologie sociologie à l'université de azerty -*th matRech azerty "1. Entretiens semi-direct

**** *id_ 2 *sex Femme *statut Post-doctorante *disc1_Sociologie *d *methode1_analyse qualitative manuelle *freqM1 Souvent 1 fois par semaine *FreqM2 Régulièrement 1 fois par mois *methode3_statistiques_descriptives_ *fr *freqM4_ azerty -*th def azerty L'ethnographie donne une large pl le régime de l'exploration provoque tout le long du terrain des surprises et des découvert des hypothèses par une exploration exhaustive et une analyse contextuelle des situations en puis s'agit alors de tester ces conjectures de manière processuelle dans le temps et l'écologie de l'enqu Dewey disaient que les théories sont des « instruments » temporaires et faillibles qui permettent jusqu'à plus ample information ou jusqu'à preuve du contraire une explication ou une int démarche exploratoire sans prédéfinir le contexte pertinent des situations a priori. L'enjeu de ce t situations (pendant le tournage d'une vidéo ou en participant à une situation ou un échange sur le

Rechercher et remplacer

Rechercher Remplacer Atteindre

Rechercher : azerty

Remplacer par : ^\n

<< Moins

Options de recherche

Rechercher : Tous

Respecter la casse

Mot entier

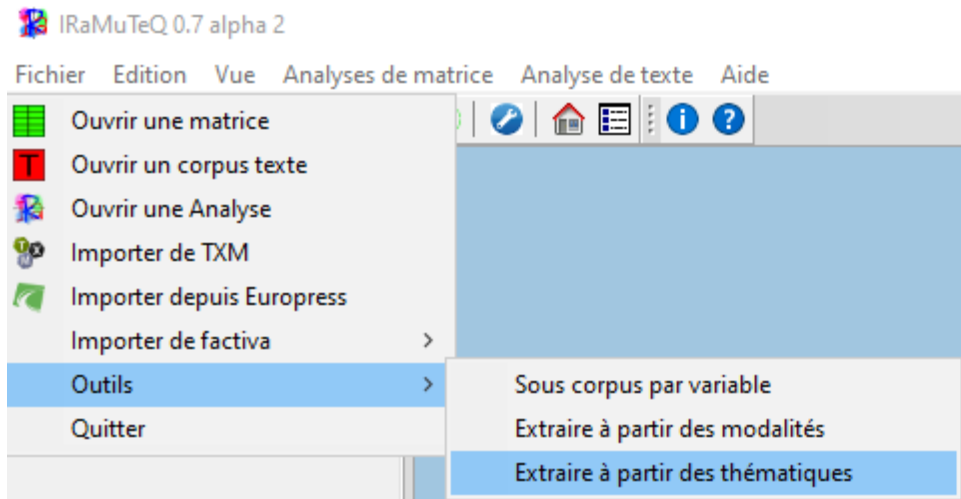
Utiliser les caractères génériques

Remplacer

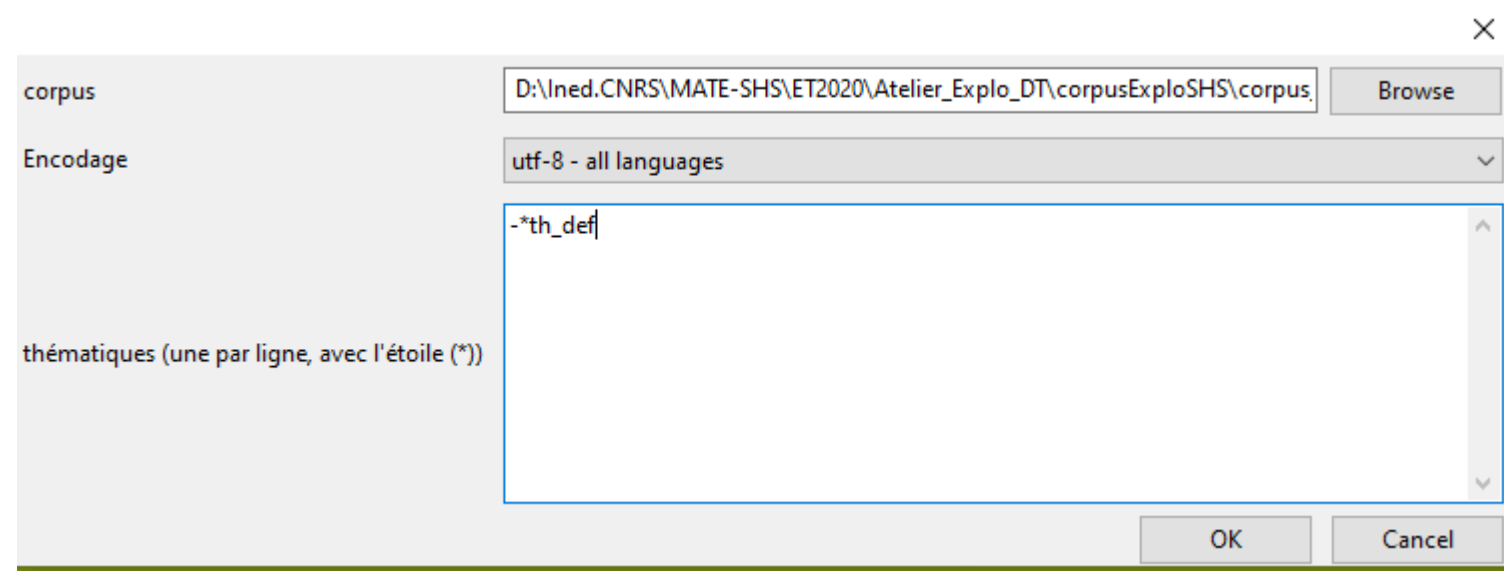
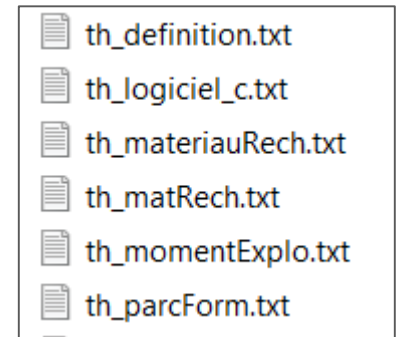
Format Spécial Sans attributs

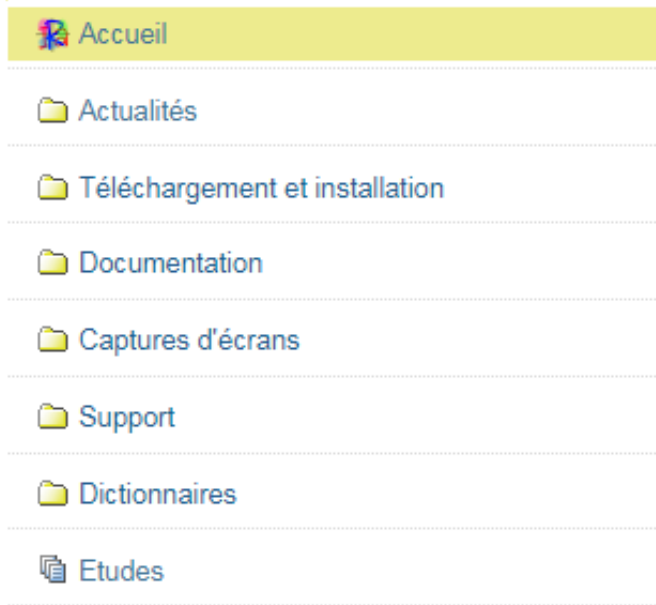
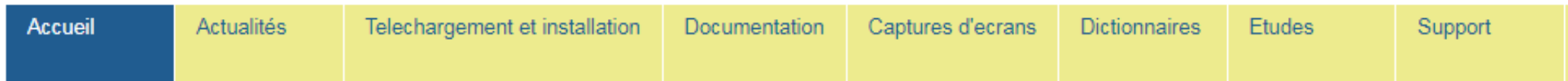
Corpus_complet2.txt

Avec IRaMuTeQ : Créer des sous corpus



à partir de
Corpus_complet2.txt





Vous êtes ici : Accueil

Iramuteq

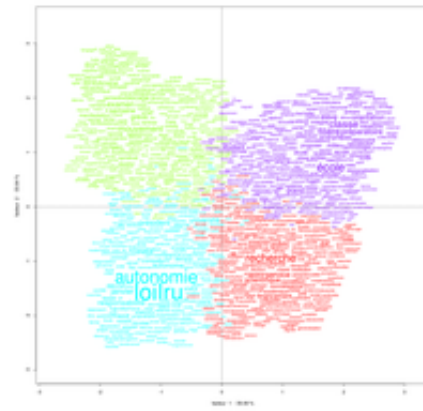
Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Un logiciel libre construit avec des logiciels libres.

[Aussi disponible en mode diaporama...](#)

Une nouvelle version d'iramuteq est en ligne - IRaMuTeQ 0.7 alpha 2 - 22/12/2014

De nombreuses ressources

logiciel libre construit avec des logiciels libres : R, python



<http://www.iramuteq.org/>

Explorer des données textuelles

Compter les mots (occurrences)

trouver des thématiques à l'aide de classifications, graphes de mots (cooccurrences)

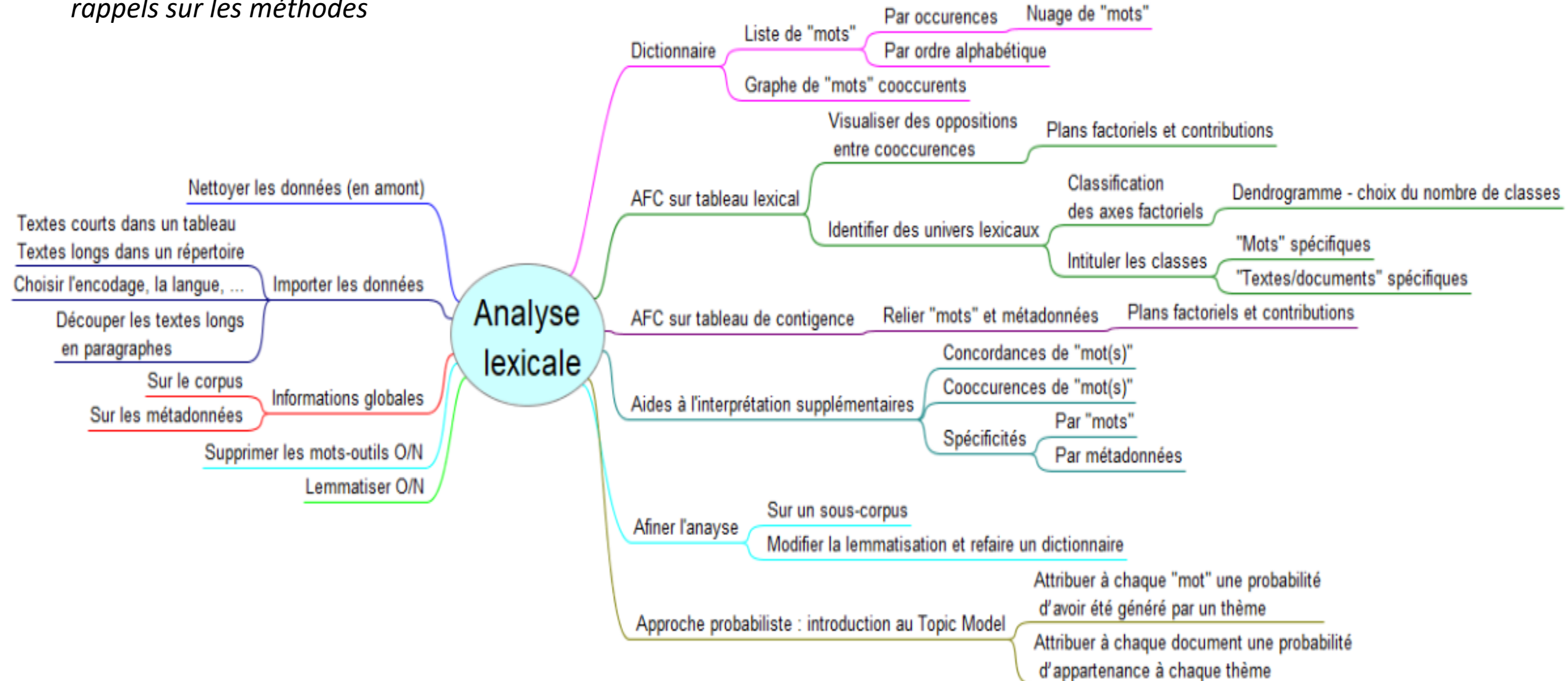
Calcul de spécificités

Interpréter : retour aux concordances

Extraire les résultats, les visualisations pour les présenter

La démarche de l'exploration

rappels sur les méthodes



Corpus Distribués

- th_definition/ th_def = Pour vous, en quoi consiste l'exploration et quelle(s) difficulté(s) soulève-t-elle ?-> **Definition_Explorer**
- th_moment= A quel(s) moment(s) et de quelle(s) façon(s) faites-vous de l'exploration dans votre pratique professionnelle ?-> **Moments_explo**
- th_parcours/th_parcForm = Décrivez en quelques lignes votre parcours de formation (initiale et - éventuellement - complémentaire)-> **Parcours_formation**
- th_materiau/th_matRech = Avec quels matériaux et/ou jeux de données travaillez-vous régulièrement ? Précisez la forme, le contenu et la masse des jeux de données →**Materiaux_recherche**
- th_apports= De quelle manière et sur quels aspects l'école thématique Explo-SHS peut vous **aider dans votre pratique** professionnelle ?
- **th_3disc**=concaténation des 3 disciplines renseignées et **th_4métho**=concaténation des 4 méthodes pratiquées pour en faire deux nouvelles variables « textuelles »

Les menus et fenêtre d'IRaMuTeQ

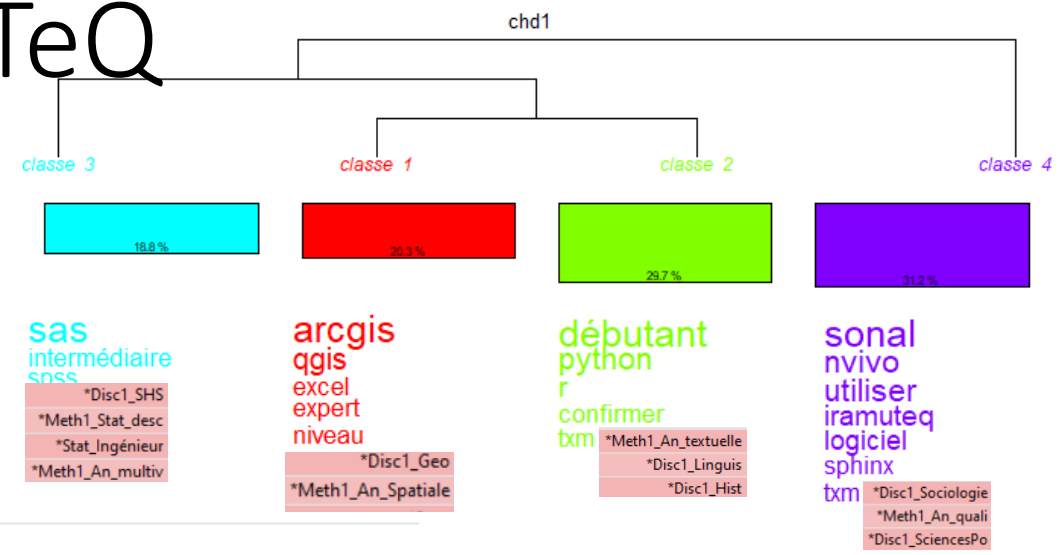
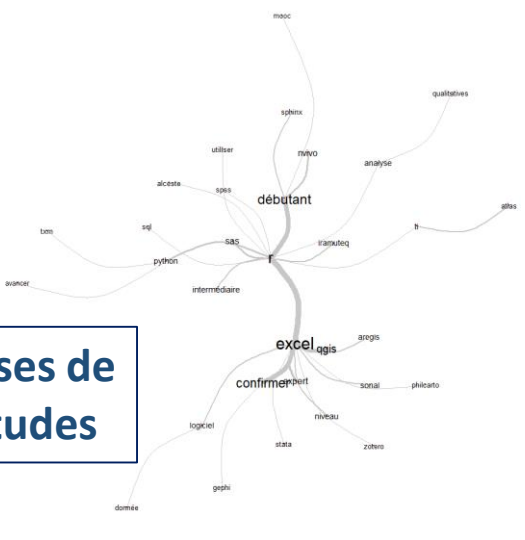
Description th_logiciels_corpus_1 / th_logiciels_stat_1 / Classification - th_logiciels_corpus_1 / Classification - th_logiciels_corpus_1

Résumé Formes actives x Formes supplémentaires Total Hapax

Forme	Freq.	Types
confirmé	80	ver
débutant	73	nom
excel	64	nr
r	54	nr
qgis	30	nr
expert	25	nom
nvivo	16	nr
niveau	15	nom
sas	15	nom
python	13	nom
intermédiaire	11	nom
arcgis	9	nr
iramuteq	9	nr
sonal	9	nr
ti	7	nr
avancé	6	ver
logiciels	6	nom
atlas	5	nom
sphinx	5	nom
spss	5	nr
stata	5	nr
analyses	4	nom
débutante	4	nom
qualitatives	4	nr
txm	4	nr
alceste	3	nr
analyse	3	nom
données	3	nom
gephi	3	nr
mooc	3	nr
philcarto	3	nr
sql	3	nr
zotero	3	nr
adobe	2	nom

Statistiques

Analyses de similitudes



Classification

Demo avec l'analyse du corpus « Logiciels »

Concordances - ti / th_logiciels_corpus_1 / Classification - th_logiciels_corpus_1 / Spécificités - th_logiciels

**** *n_43 *Sex_Fem *Stat_MCF_Prof *Disc1_Sc_language *Meth1_Stat_infer *Freq_meth1_Régult
 désormais r avant atlast **ti** maintenant rqda en r pour les analyses qualitatives

**** *n_44 *Sex_Fem *Stat_MCF_Prof *Disc1_Sc_language *Meth1_Stat_infer *Freq_meth1_Régult
 désormais r pour les analyses quantitatives et qualitatives package rqda avant 2004 2014 c était atlast **ti** et spss

**** *n_25 *Sex_Fem *Stat_Ingénieur *Disc1_Sociologie *Meth1_An_quali *Freq_meth1_Svt
 nvivo formée sur atlas **ti**

**** *n_61 *Sex_Fem *Stat_Ingénieur *Disc1_Sociologie *Meth1_ *Freq_meth1_Svt
 caqdas nvivo atlas **ti** depuis un certain temps je me questionne sur l'intérêt du recours aux logiciels de traitement
 qualitatives initiation en cours

**** *n_100 *Sex_Fem *Stat_CR *Disc1_Geo *Meth1_An_multiv *Freq_meth1_Tps_en_tps
 sas faux débutant r confirmé nvivo débutant atlas **ti** débutant excel confirmé expert qgis arcgis confirmé python

**** *n_101 *Sex_ *Stat_Doctorant_e *Disc1_Anthropo *Meth1_An_quali *Freq_meth1_Svt
 atlas **ti** niveau début excel niveau confirmé

Concordances

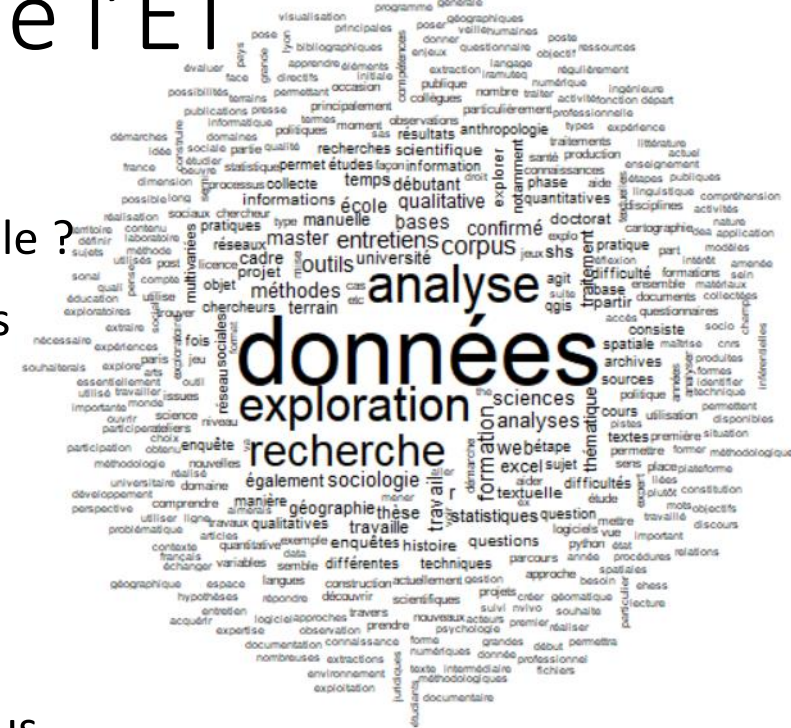
relatives des formes	Fréquences relatives des types	AFC
Geo *Disc1_Hist	*Disc1_Linguis	*Disc1_SHS *Disc1_Sc_language
-0.0402	2.2939	-0.062 -0.051
0.5832	2.2466	-0.2028 -0.167
-0.359	2.2009	0.8684 -1.0856
-0.0301	1.0527	-0.0464 -0.0382
-0.1889	0.6552	0.2382 0.7011
-0.2551	0.264	-0.3937 -0.3242
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
1.1743	-0.0403	0.994 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0301	-0.0403	-0.0464 -0.0382
-0.0402	-0.0537	-0.062 -0.051
1.054	-0.0537	-0.062 -0.051
-0.0402	-0.0537	-0.062 -0.051

Spécificités

Explorer le corpus de l'ensemble des réponses au questionnaire posé aux participants de l'ET

Textes courts / réponses libres

- Pour vous, **en quoi consiste l'exploration** et quelle(s) difficulté(s) soulève-t-elle ?
- A quel(s) moment(s) et de quelle(s) **façon(s) faites-vous de l'exploration** dans votre pratique professionnelle ?
- Décrivez en quelques lignes votre **parcours de formation** (initiale et - éventuellement - complémentaire)
- Avec **quels matériaux** et/ou jeux de données travaillez-vous régulièrement ?
_Précisez la forme, le contenu et la masse des jeux de données
- De quelle manière et sur quels aspects l'école thématique Explo-SHS peut vous **aider dans votre pratique professionnelle** ?



Données (753), Exploration (344), connaissances (33)

Métadonnées

V3. Sexe V8. Statut V13. Discipline1 V14. Discipline2 V15. Discipline3 V16.
Parcours_formation V17. Matériaux_recherche V19. Methode1 V20.
Frequence_meth1 V21. Methode2 V22. Frequence_meth2 V23. Methode3 V24.
Frequence_meth3 V25. Methode4 V26. Frequence_meth4 V28. Frequence_meth5

Extrait du lexique associé au corpus des réponses aux questions ouvertes (mots cités plus de 10 fois) – mots différents 5679

Recherche de thématiques avec la classification d'IRaMuTeQ

Description CorpusCompleat_N_corpus_1 Classification - CorpusC

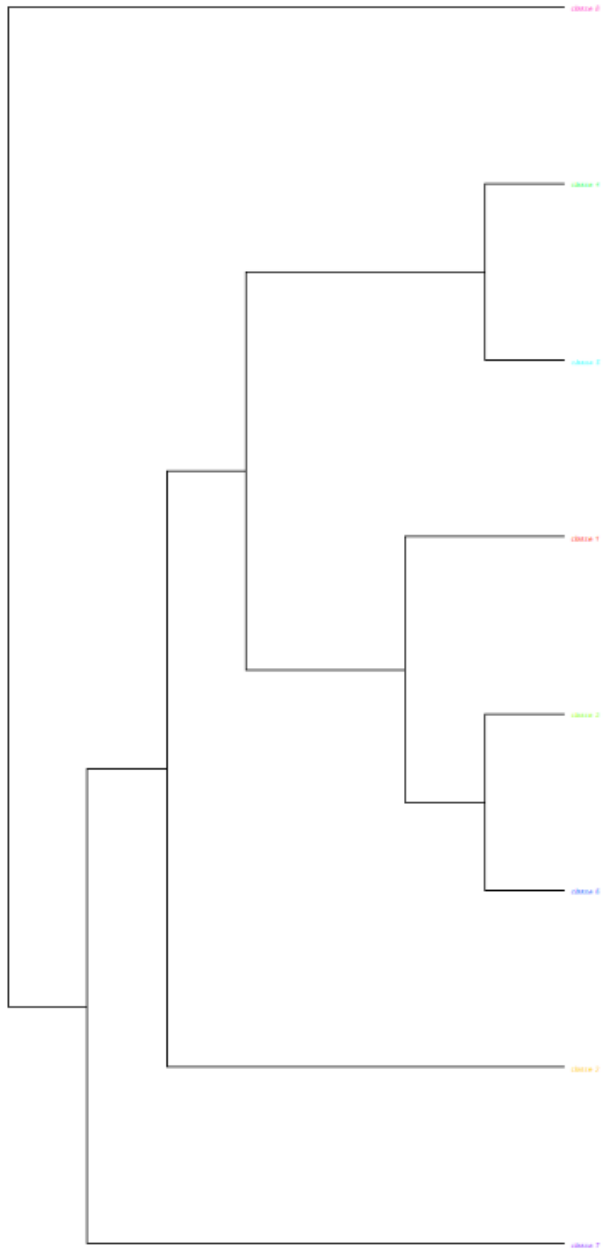
CHD x Profils AFC

- Nombre de textes: 105
- Nombre de segments de texte: 1608
- Nombre de formes: 5679
- Nombre d'occurrences: 46184
- Nombre de lemmes: 4157
- Nombre de formes actives: 3784
- Nombre de formes supplémentaires: 373
- Nombre de formes actives avec une fréquence >= 3: 1360
- Moyenne de formes par segment: 28.721393
- Nombre de classes: 8
- 1450 segments classés sur 1608 (90.17%)

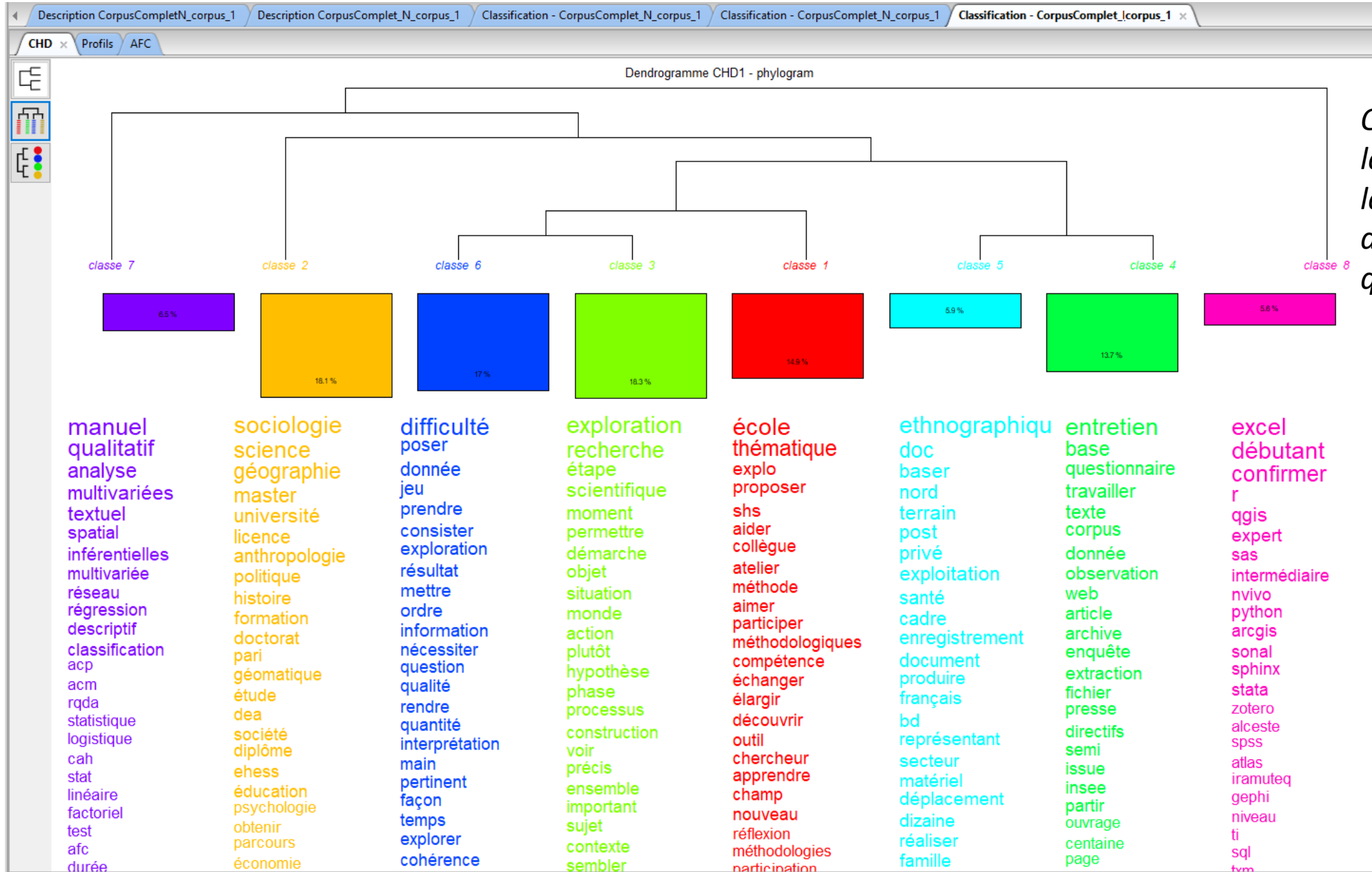
#####

temps : 0h 1m 1s

#####



Identifier les thématiques à l'aide du vocabulaire spécifique



On retrouve dans les thématiques les réponses aux différentes questions posées

Croiser avec les métadonnées et retour au contexte (concordancier)

Formes	Formes banales	Types	Fréquences des formes	Fréquences des types	Fréquences relatives des formes	Fréquences relatives des types	AFC					
formes			*Disc1_Anthropo	*Disc1_Educ	*Disc1_Geo	*Disc1_Hist	*Disc1_Linguis	*Disc1_SHS	*Disc1_Sc_language	*Disc1_SciencesPo	*Disc1_Sociologie	*Disc1_Autre
sociologie			-0.3672	-0.9568	-3.6449	-0.5818	-1.1741	-0.4307	-0.2888	0.5374	12.0667	-3.13
enquête			0.351	-0.967	-2.4556	-1.0059	-1.1866	0.3865	-1.1621	-0.3206	7.7335	-1.0476
matériau			-1.0668	-0.2441	-0.8892	-0.4307	-0.2995	-0.3635	-0.2933	-0.3066	7.2254	-1.3117
ligne			-0.8889	-0.2034	-2.4256	-0.3589	-0.2496	-0.3029	-0.2444	-0.2554	5.4202	0.2876
quali			-0.5333	-0.122	-0.768	-0.2153	0.5352	-0.1817	-0.1466	-0.1533	3.7901	-1.0592
me			-2.308	-1.2781	0.6709	-0.7531	-0.7468	0.8222	-1.0864	1.4929	3.6137	-2.2627
observation			2.5187	-0.1745	-2.1182	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861	-3.4068
site			-0.3316	-0.183	-0.8273	-0.323	-0.2246	0.3315	0.4008	-0.2299	2.9553	-0.8849
idée			-0.2758	1.2858	-1.1511	-0.2871	-0.1996	-0.2423	0.4407	-0.2043	2.9087	-1.4123
questionnaire			-1.1246	0.2034	-0.2591	-0.7538	-0.5242	-0.6362	0.9571	0.4626	2.8621	-1.439
donc			-0.8167	-0.226	0.2979	-0.2517	-0.7491	0.4703	-0.7336	0.2823	2.8599	-1.6041
je			-0.893	-0.3251	-0.2976	-0.2517	-0.7491	0.4703	-0.7336	0.2823	2.8599	-1.6041
exploratoire			-0.5913	0.4803	-0.4072	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861	-3.4068
d			0.4034	0.6377	-0.6791	-0.3411	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
bibliographique			0.4403	-0.2237	-1.1621	-0.3911	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
sembler			-0.9468	0.2369	-0.3616	-0.6611	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
objet			0.2565	-0.4069	-1.2839	0.3141	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
ne			-2.0607	-0.9644	0.3781	0.6751	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
archive			-0.4119	0.1974	-5.217	9.1111	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
hypothèse			-0.7111	-0.1627	-0.3682	0.3151	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
approcher			-0.7899	-0.4679	-0.5439	0.2511	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
démarche			-1.1969	-0.4476	0.4187	0.2701	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
quanti			-0.4444	-0.1017	-0.5872	-0.1711	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
tout			-0.6831	-0.6817	-0.8637	1.2701	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
recueillir			1.1186	-0.1729	-2.0617	-0.3011	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
social			2.102	-0.9466	-3.1077	0.4891	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
traiter			-1.2002	-0.2746	-0.2994	0.5241	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
situation			0.7959	-0.2237	-0.746	-0.3911	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
ils			-0.6667	-0.1525	-1.8191	0.9151	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
sans			0.3959	0.269	-0.2312	-0.5911	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
mal			-0.1974	-0.1322	-0.4532	0.3811	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
intéressant			-0.1974	0.581	-0.8615	-0.2311	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
nombre			-0.4808	-0.2339	-1.2494	0.6231	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861
ouvrir			-0.5756	-0.2644	-0.2629	0.5471	-0.9155	-0.6367	-0.7726	1.2781	-0.2496	3.4861

Concordancier - matériau

**** *n_30 *Sex_Fem *Stat_Doctorant_e *Disc1_Sociologie *Meth1_An_quali *Freq_meth1_Svt

pour moi l exploration de données consiste dans un premier temps à identifier des tendances sur un sujet à partir des **matériaux** disponibles c est aussi l occasion de découvrir des signaux faibles mais important auxquels je n aurai pas pensé dans un premier temps

**** *n_69 *Sex_Hom *Stat_Doctorant_e *Disc1_Sociologie *Meth1_An_quali *Freq_meth1_Svt

en troisième année de doctorat je dispose aujourd'hui d une quantité assez importante de **matériaux** empiriques autant qualitatifs que quantitatifs

**** *n_78 *Sex_Hom *Stat_Doctorant_e *Disc1_Autre *Meth1_Stat_infer

j ai déjà pu mener ce genre d analyse lors de mon mémoire qui reposait également sur différents **matériaux** des entretiens ainsi que des bases de données

**** *n_69 *Sex_Hom *Stat_Doctorant_e *Disc1_Sociologie *Meth1_An_quali *Freq_meth1_Svt

ce **matériau** est donc à l état brut et n est pas encore l objet d un traitement spécifique ni d une analyse précise

**** *n_69 *Sex_Hom *Stat_Doctorant_e *Disc1_Sociologie *Meth1_An_quali *Freq_meth1_Svt

l école thématique explo shs me permettrait donc d approfondir mes premières analyses tout autant que d ouvrir de nouvelles perspectives quant à ce **matériau** pour l instant en somme il

**** *n_61 *Sex_Fem *Stat_Ingénieur *Disc1_Sociologie *Meth1_*Freq_meth1_Svt

autre analyse qualitative de corpus d entretiens découpage thématique du **matériau** attribution de thèmes ou codage attribution de catégories annotations rédaction exploration de textes en projet

Relancer une analyse par sous corpus

The screenshot displays the IRaMuTeQ 0.7 alpha 2 software interface. The main window shows a menu with 'Sous-corpus' selected, which has opened a sub-menu with two options: 'Sous-corpus par méta-données' and 'Sous-corpus par thématique'. The 'Sous-corpus par thématique' option is highlighted. In the foreground, a dialog box titled 'Description corpus_complet2_corpus_2' is open, showing a list of parameters for a sub-corpus. The 'Nom' field is set to 'Sous-corpus'. The 'Définition' field contains a list of meta-data patterns: '-*th_def', '-*th_logiciel', '-*th_matRech', '-*th_momentExplo', and '-*th_parcreateForm'. The dialog box has 'OK' and 'Cancel' buttons at the bottom.

IRaMuTeQ 0.7 alpha 2

Fichier Edition Vue Analyses de matrice Analyse de texte Aide

Statistiques
Spécificités et AFC
Classification
Analyses de similitudes
Nuage de mots
Sous-corpus
Exporter la table des méta-données

Sous-corpus par méta-données
Sous-corpus par thématique

Historique

Corpus textuel

- corpus_complet2_co
- Subcorpus_complet2
- th_logiciel_corpus_3
- corpus_complet2_co
- corpusComplet_corp

Description corpus_complet2_corpus_2

Description du corpus

Nom
Langue
Encodage
originalpath
pathout
date
time
Paramètres
ucemethod
ucesize
keep_caract
expressions

Sous-corpus

Nom

Définition

- *th_def
- *th_logiciel
- *th_matRech
- *th_momentExplo
- *th_parcreateForm

Sélectionnez une ou plusieurs méta-données

OK Cancel

Partagez votre expérience

Les + et les – d'IRaMuteQ

- ✓ La lemmatisation automatique (paramétrable par le « rôle » de certains mots) mais non modifiable
- ✓ Les découpage de textes longs
- ✓ Les classifications pour faire émerger des thématiques
- ✓ Les graphes de mots associés
 - Nœuds=*mots*
 - Arêtes= *relation entre les mots (cooccurrence par défaut) ici*
- ✓ Les regroupements de « mots » → Tgen : *Types généralisés = Sous ensemble d'occurrences d'un corpus (Salem)*
- Installer R, IRaMuTeQ (et multiples versions de R et packages)
- Analyser ensemble les mots dans les textes et les métadonnées sur les textes : ex. savoir *Qui dit quoi ?*
- La reproductibilité