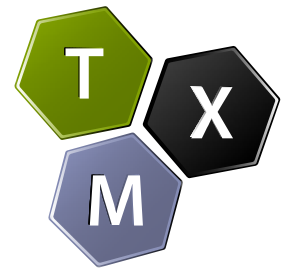


Atelier TXM

Que peut-on faire avec TXM ?

Bénédicte PINCEMIN
ICAR, CNRS & Université de Lyon



Ecole d'été doctorale
« Le doctorat mode(s) d'emploi »



Strasbourg,
7 juillet 2015

Plan

- Contexte de la recherche en SHS
- Illustration de types de questions
 - Q1. La présence des mots
 - Q2. Comment résonne tel mot dans mon corpus ?
 - Q3. Comment caractériser les différents textes (ou genres / périodes / auteurs etc.), et quels pôles les structurent ?
 - Q4. Comment évolue quantitativement l'usage de tel motif ?
- Pendant et après : questions, discussion, démonstration interactive

Textométrie et SHS

- Principes de l'approche
 - Observations basées sur les **contextes** locaux (syntagmatiques) et globaux (textuels, intertextuels)
 - Dictionnaire, grammaire : possible mais facultatif. Importance accordée à l'**expression** (comment c'est dit)
 - Organisation par **fréquence**, forme, catégories
 - Proposer d'**autres** parcours de lecture en pointant des régularités pas toujours consciemment perçues
- Rôle de l'automatisation
 - Mémoire et vitesse de calcul sur un grand corpus
 - Au chercheur revient l'interprétation, de la constitution du corpus à la signification des résultats.
 - Se fonder sur les données ; systématique ; essais/erreur

Logiciel TXM et SHS

- Logiciel libre (gratuit, pas boîte noire, personnalisable, pérennité...)
- Toutes machines : Windows, Mac, Linux, ou en ligne
- Tous types de corpus
 - Texte brut / (très) structuré : nombreux formats d'import
 - Français / autre langue, autre écriture
 - Oral (transcrit) / écrit (roman, poésie, théâtre...)
- Corpus enrichis
 - Possibilité d'étiqueter à la volée (TreeTagger)
 - Utilisation souple et complète de la structuration et de l'annotation du corpus
- Communauté et aide
 - Point d'entrée : <http://textometrie.ens-lyon.fr>
 - Contact : textometrie à ens-lyon *point* fr

Corpus (1)

- FLEURS

- *Fleurs du Mal*, Baudelaire
- Édition numérique préparée par J.-M. Viprey (Besançon)

- VOEUX

- Discours de vœux des présidents français de la Ve République
- Édition numérique préparée par J.-M. Leblanc (Paris 12)

- Ces deux corpus préparés pour TXM sont téléchargeables depuis le site du logiciel TXM sur sourceforge :

- <http://sourceforge.net/projects/txm>

Corpus (2)

- LEMAN2011
 - entretiens semi-directifs avec les pêcheurs français du lac Léman (38/46)
 - Master STADE (géographie), ENS de Lyon, 2011-2012.
 - Oral transcrit (avec Transcriber)
 - Deux versions dans TXM (paramètre d'import)
 - LEMAN2011 : les propos de l'interviewer sont indexés et comptabilisés dans l'analyse
 - LEMAN2011INT : les propos de l'interviewer sont affichés pour la compréhension du contexte, mais ne sont pas en compte dans les recherches et calculs
 - Ce corpus n'est actuellement pas en mesure d'être diffusé.

Q1. La présence des mots

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: [pos="S."...] Requête: [pos="S."] Seuils: Fmin: 1 Fmax: 99999

word	Fréquence T=33233
cœur	110
yeux	92
âme	60
ciel	58
amour	46
œil	46
soleil	44
corps	41
air	36
soir	36
nuit	32
mer	30
beauté	29

FLEURS: [pos="J"]:lemme Requête: [pos="J"] Seuils: Fmin: 1 Fmax: 99999

lemme	Fréquence
grand	72
plein	63
vieux	51
doux	43
long	42
profond	37
autre	29
cher	26
éternel	25
lourd	23
charmant	21
étrange	20
pâle	20

FLEURS: [pos="V"]:le... Requête: [pos="V"] Seuils: Fmin: 1 Fmax: 99999

lemme	Fréquence
être	343
avoir	206
faire	124
voir	73
dire	55
pouvoir	55
vouloir	52
aimer	50
savoir	46
aller	43
prendre	43
venir	39
connaître	31

FLEURS: [pos="A"] Requête: [pos="A"] Seuils: Fmin: 1 Fmax: 99999

word	Fréquence
pas	60
toujours	36
jamais	33
bien	21
donc	19
parfois	15
très	14
encor	13
le plus	12
loin	12
trop	11
souvent	10
à travers	9

Ouverture du tableau d'index

Corpus Fleurs du Mal : noms communs, adjectifs, verbes et adverbes les plus fréquents.

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: [pos="R" ...] >>12

Requête : [pos="R" & lemme="noir"]

Seuils : Fmin: 1 Fmax: 99999

1 -13 / 13

lemme	Fréquence T=33233
noir	61
blanc	17
bleu	14
rose	13
rouge	13
brun	10
jaune	5
vermeil	5
fauve	4
blond	2
violet	2
écarlate	1
roux	1

FLEURS:[lemme="vermeil"]

1 -5 / 5

Afficher paramètres

poeme_titre_ε^	Contexte gauche	Pivot	Contexte droit
BÉNÉDICTION	Retrouve l'ambrosie et le nectar	vermeil	. Il joue avec le vent, ca
ENNEMI	mon jardin bien peu de fruits	vermeils	. Voilà que j'ai touché l'
AUBE	les débauchés l'aube blanche et	vermeille	Entre en société de l'Id
PETITES VIEILLES	tombant Ensanglante le ciel de blessures	vermeilles	, Pensive, s'asseyait à l'
AME DU VIN	prison de verre et mes cires	vermeilles	, Un chant plein de lumi

FLEURS Edition - Page 11

file:///home/bpincemi/TXM/corpora/fleurs6/HTML/FLEURS/default

Ma jeunesse ne fut qu' un ténébreux orage,
 Traversé çà et là par de brillants soleils ;
 Le tonnerre et la pluie ont fait un tel ravage,
 Qu' il reste en mon jardin bien peu de fruits **vermeils** .

Voilà que j' ai touché l' automne des idées,
 Et qu' il faut employer la pelle et les râteaux
 Pour rassembler à neuf les terres inondées,
 Où l' eau creuse des trous grands comme des tombeaux.

11 / 151

Tri terminé.

Corpus Fleurs du Mal : emplois des adjectifs de couleur (recherche sur noir, blanc, rouge, bleu, vert, jaune, rose, violet, marron, orange, brun, blond, roux, fauve, vermeil, écarlate).

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: ".*ténébr.*"%cd... FLEURS: "(pén|s)?ombr.*" FLEURS: ".*clair.*|clarté.*" FLEURS: ".*lumi.*"...

riétés : word Editer Chercher

Résultats par page : 100

t 35 , v 5 , fmin 1 , fmax 19 t 38 , v 9 , fmin 1 , fmax 17 t 48 , v 16 , fmin 1 , fmax 13 t 34 , v 9 , fmin 1 , fmax 13

word	Fréquence
ténèbres	19
Ténèbres	1
ténébreuse	3
ténébreux	11
Ténébreux	1

word	Fréquence
Ombragé	1
ombragea	1
ombre	10
ombres	5
Ombres	1
ombresuses	1
pénombre	1
sombre	17
Sombres	1

word	Fréquence
clair	13
claire	1
clairons	1
clairs	3
Clairs	1
clairvoyance	1
clairvoyant	1
clairvoyants	1
clarté	13
clartés	2
éclair	3
éclaira	1
éclaire	1

word	Fréquence
illuminait	1
illumine	2
illuminé	1
illuminer	3
illuminés	3
lumière	13
lumières	4
lumineuse	2
lumineux	5

Tri terminé.

Corpus Fleurs du Mal : familles de mots -déploiement et quantification des usages.



Corpus

AZ FLEURS: []:pos

AZ FLEURS: ([pos="S."][pos="J"])[pos="J"]([pos="S."):pos

Requête: Propriétés: pos Seuils: Fmin: Fmax: Vmax: Résultats par page:

t 1401 , v 6 , fmin 1 , fmax 556

pos	Fréquence T=33233
SM J	556
J SM	326
SF J	266
J SF	250
J SW	2
SW J	1

AZ FLEURS: [pos="J"]([pos...

Requête: Seuils: Fmin: Fmax:

t 1401 , v 6 , fmin 1 , fmax 556

word	Fréq
longue misère	15
grands yeux	4
larges pleurs	3
Adorable sorcière	2
chaudes soirées	2
diverses beautés	2
éternel martyr	2
étranges grâces	2
grand œil	2
grand reposoir	2
grands flots	2
grand troupeau	2
innocent paradis	2

Terminé : 6 items pour 1401 occurrences.



Corpus Fleurs du Mal : adjectif antéposé ou postposé -quantification et retour au texte.

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: [[]][word="..."] >>2

Requête: [[]][word="\p{P}"]*</vers>

Seuils: Fmin: 1 Fmax: 9999999

word	Fréquence T=33233
misère !	15
...	13
funèbres ,	8
beauté ,	7
fleurs ,	6
seins ,	6
yeux ,	6
âme	5
cœur ,	5
large ,	5
livide ,	5
monde ,	5
mystique	5

FLEURS:[word="fleurs"] [word=","]</vers>

Afficher paramètres

poeme_titre_abr ^	Contexte gauche	Pivot	Contexte droit
J'aime le souvenir	ciel, les oiseaux et les	fleurs,	Ses parfums, ses chansons e
BÉATRICE	grillons, les ruisseaux et les	fleurs,	Et même à nous, auteurs de
CY THERE	la jeune prêtresse, amoureuse des	fleurs,	Allait, le corps brûlé de secr
JET D'EAU	. La gerbe épanouie En mille	fleurs,	Où Phœbé réjouie Met ses c
JET D'EAU	cœur La gerbe épanouie En mille	fleurs,	Où Phœbé réjouie Met ses c
JET D'EAU	. La gerbe épanouie En mille	fleurs,	Où Phœbé réjouie Met ses c

FLEURS Edition - Page 6

file:///home/bpincemi/TXM/corpora/fleurs6/HTML/FLEURS/defa

Et comme qui dirait des beautés de langueur ;
 Mais ces inventions de nos muses tardives
 N'empêcheront jamais les races maldives
 De rendre à la jeunesse un hommage profond,
 - À la sainte jeunesse, à l'air simple, au doux front,
 À l'œil limpide et clair ainsi qu'une eau courante,
 Et qui va répandant sur tout, insouciant
 Comme l'azur du ciel, les oiseaux et les fleurs ,
 Ses parfums, ses chansons et ses douces chaleurs !

6 / 151

Ouverture du tableau d'index

Corpus Fleurs du Mal : les mots à la rime.

TXM

Fichier Outils Affichage Aide

Corp Fichi

- GRAAL
- VOEUX
- BAIP
- CORPUSCONTES
- GRIPARI
- FLEURS
- LEPIGEONXML
- VOEUX
- LEMAN2011
- BFM012DD
- GE
- PERRAULTXML
- TEMOIGNAGES
- LEMAN2011INT
- word
- DISCOURS

Propriétés : word Editer Chercher

Seuils : Fmin : 1 Fmax : 9999999

1 -100 / 11522

word	Fréquence T=367102
,	29312
.	15518
est	8872
de	8110
c'	5922
a	5545
le	5254
il	5247
on	5208
pas	5133
les	5101

Macro

Nouvelle macro

- edition
- misc

Console

TXM

Fichier Outils Affichage Aide

Corp Fichi

- GRAAL
- VOEUX
- BAIP
- CORPUSCONTES
- GRIPARI
- FLEURS
- LEPIGEONXML
- VOEUX
- LEMAN2011
- BFM012DD
- GE
- PERRAULTXML
- TEMOIGNAGES
- LEMAN2011INT
- word
- DISCOURS

Propriétés : word Editer Chercher

Seuils : Fmin : 1 Fmax : 9999999

1 -100 / 11522

word	Fréquence T=367102
les	5101
que	5093
à	4801
en	4746
la	4678
ça	4466
je	4311
euh	4075
y	3907
un	3857
et	3684
des	3579

Macro

Nouvelle macro

- edition
- misc

Console

Corpus Léman : mots les plus fréquents et caractéristiques de l'oral

TXM

Fichier Outils Affichage Aide

GRAAL
VOEUX
BAIP
CORPUSCONTES
GRIPARI
FLEURS
LEPIGEONXML
VOEUX
LEMAN2011
BFM012DD
GE
PERRAULTXML
TEMOIGNAGES
LEMAN2011INT
DISCOURS

Macr
Nouvelle macro

Console
Sortie standard

LEMAN2011INT: [pos="VER:c"]

Requête: [pos="VER:cond"]

Seuils: Fmin: 1 Fmax: 99999

1 -100 / 116

lemme	Fréquence T=3671
pouvoir	92
dire	80
Ca	70
falloir	48
devoir	36
faire	29
Ba	20
aimer	19
vouloir	19
aller	10
ba	10
gagner	6

LEMAN2011INT:[lemme="gagner" & pos="VER:cond"]

Requête: [lemme="gagne"] Pivot: word Chercher

Clés de tri: #1 Contexte #2 Aucun #3 Aucun #4

1 -6 / 6 Cacher paramètres

	Contexte gauche	Pivot	Contexte droit
d ^	les charges gratuites c'est sûr que je	gagnerais	, beaucoup, beaucoup d
	je veux pas rentrer dans ce, je	gagnerais	plus à faire ça, parce qu
	l'eau le soir, parce que je	gagnerais	pratiquement une jour
	ça me rapporterait déjà. Euh, je	gagnerais	des sous hein. Donc je p
	Si on pouvait ne pas cotiser on y	gagnerait	hein. Parce que sur les a
	le grossiste, c'est nous qu'on	gagnerait	les sous. Quelque part

Corpus Léman : les verbes aux conditionnel les plus fréquents → le verbe gagner

TXM

Fichier Outils Affichage Aide

Corpus

VOEUX: [frpos="VER.*"]... Requête: [frpos="VER.*"] Seuils: Fmin: 1 Fmax: 9999999 Vma

frlemma	Fréquence
être	1371
avoir	823
faire	283
devoir	157
pouvoir	151
vouloir	141
aller	101
dire	94
savoir	92
souhaiter	86
falloir	80
vivre	74
prendre	60

VOEUX: [frpos="VER:co..."] Requête: [frpos="VER:cond"] Seuils: Fmin: 1 Fmax: 9999999 Vma

frlemma	Fréquence
vouloir	25
être	22
avoir	10
pouvoir	7
falloir	3
devoir	2
permettre	2
agiter	1
aimer	1
apporter	1
apportera-t-il	1
bouleverser	1
couper	1

VOEUX: [frpos="VER.*"]... Requête: [frpos="VER.*"] Seuils: Fmin: 1 Fmax: 9999999 Vma

frpos	Fréquence
VER:pres	3747
VER:infi	1876
VER:pper	1372
VER:futu	535
VER:subp	210
VER:impf	184
VER:ppre	165
VER:cond	99
VER:simp	34
VER:subi	9
VER:impe	6

VOEUX: [frlemma="p..."] Requête: [frlemma="pouvoir" & frpos="VER.*"] Seuils: Fmin: 1 Fmax: 9999999 Vma

frpos	Fréquence T=61102
VER:pres	83
VER:subp	22
VER:pper	15
VER:futu	11
VER:cond	7
VER:infi	7
VER:impf	6

Ouverture du tableau d'index

Corpus Voeux : verbes et temps verbaux.

TXM

Fichier Outils Affichage Aide

Corpus

Chirac: word		Chirac: word:[]		MitterrandChirac: word:[]		Chirac1vs2: word:[]	
Unités	VOEUX \	Unités	prédeces	Unités	FréquenceT2:	Unités	FréquenceT14791
.	1613	.	1235	une	228	emploi	34
Une	9	La	49	avenir	46	Gouvernement	20
La	70	compatriotes	11	.	1389	enjeu	7
aussi	61	Une	7	Une	28	action	21
emploi	26	valeurs	1	La	84	la	433
Nous	57	emploi	15	chaque	22	:	31
c'	97	avenir	23	valeurs	17	été	22
avenir	38	chers	8	aussi	72	veux	15
une	284	Nous	41	chacun	33	tous	52
solidarité	16	solidarité	9	compatriotes	36	croissance	21
Gouvernement	10	c'	76	progrès	20	2005	5
mondialisation	0	une	224	emploi	41	mobiliser	5
L'	44	aussi	52	action	23	rassemblement	5
ambition	6	croissance	8	chacune	13	véritable	5
croissance	14	L'	33	face	12	avec	50
compatriotes	31	ambition	4	nation	12	pour	204
être	85	faire	40	Nos	12	c'	75
valeurs	10	Gouvernement	8	notre	176	et	409

Corpus Voeux : variation du corpus de référence : Chirac par rapport à (i) tous les présidents Ve République, (ii) ses précécesseurs, (iii) Mitterrand, (iv) 2e mandat par rapport au 1er.

« Nullax » des présidents dans Voeux

- De Gaulle : *votre/vos, solidarité, chômage, croissance, pense, sais...*
- Pompidou : *sans, on, République, Mes, Vive...*
- Giscard : *Europe, jamais, solidarité, Vive...*
- Mitterrand : *nation, chacun, valeurs, voudrais, pouvons...*
- Chirac : *bonheur, doute, travailleurs...*
- Sarkozy : *bonne, progrès, aujourd'hui, peu...*

Q1. La présence des mots

- Qu'est-ce qui est (étonnamment) fréquent ?
 - Liste générale [LEXIQUE] ou focalisée [INDEX]
 - On peut lister et compter des mots, des catégories grammaticales, des constructions, des temps verbaux... [CQL]
 - Sélection sur la fréquence [LEXIQUE] [INDEX], ou par contraste dans le corpus [SOUS-CORPUS] [SPÉCIFICITÉS]
- Qu'est-ce qui n'est pas dit ?
 - Ce qui n'est pas attesté malgré l'intuition [INDEX]
 - Ce qui semble évité (« nullax ») [SPÉCIFICITÉS]

Q2. Comment résonne tel mot
dans mon corpus ?

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS:[lemme="soleil"]

Requête : [lemme="soleil"] Pivot: word Chercher

Clés de tri : #1 Contexte d #2 Aucun #3 Aucun #4 Aucun Tri

1 - 59 / 59 Cacher paramètres

poeme_titre_abr	Contexte gauche	Pivot	Contexte droit
AUBE	yeux agrandis voitige incessamment. Le	soleil	a noirci la riamme des dougtes Ainsi, toujours vainqueur, ton
SONNET	marguerite ! Comme moi n'es -tu pas un	soleil	automnal, Ô ma si blanche, ô ma si froide Marguerite
COUVERCLE	re, Sous un climat de flamme ou sous un	soleil	blanc, Serviteur de Jésus, courtisan de Cythère, Mendiant ténébre
GRAVURE	sans horizon, Oû gisent, aux lueurs d'un	soleil	blanc et terne, Les peuples de l'histoire ancienne et moderne
CRÉOLE	ervoir de larmes. Au pays parfumé que le	soleil	caresse, J'ai connu, sous un dais d'arbres tout
VOYAGE	la mer violette, La gloire des cités dans le	soleil	couchant, Allumaient dans nos cœurs une ardeur inquiète De plon
CHANT	phémère D'un glorieux automne ou d'un	soleil	couchant. Courte tâche ! La tombe attend ; elle est avide
POISON	Dans l'or de sa vapeur rouge, Comme un	soleil	couchant dans un ciel nébuleux. L'opium agrandit ce qui n'a
INVITATION	ir Qu'ils viennent du bout du monde. -Les	soleils	couchants Revêtent les champs, Les canaux, la ville entière,
SOLEIL	annes, abri des secrètes luxures, Quand le	soleil	cruel frappe à traits redoublés Sur la ville et les champs,
AME DU VIN	en flamme, De peine, de sueur et de	soleil	cuisant Pour engendrer ma vie et pour me donner l'âme ;
CHATIMENT	sa raison s'en alla. L'éclat de ce	soleil	d'un crêpe se voila ; Tout le chaos roula dans cette intelligence
CHANT	, labeur dur et forcé, Et, comme le	soleil	dans son enfer polaire, Mon cœur ne sera plus qu'un
PROFUNDIS	nde qui surpasse La froide cruauté de ce	soleil	de glace Et cette immense nuit semblable au vieux Chaos ; Je

Tri terminé.

Corpus Fleurs du Mal : concordance du mot « soleil »

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS:[lemme="soleil"]

Requête : [lemme="sol"] Pivot: word Chercher

Clés de tri : #1 Contexte d #2 Aucun #3 Aucun

1 - 59 / 59 Cacher paramètres

poeme_titre	Contexte gauche	Pivot	Contexte droit
AUBE	riche incessamment. Le	soleil	a noirci la tiamme
SONNET	me moi n'es -tu pas un	soleil	automnal, Ô ma si
COUVERCLE	t de flamme ou sous un	soleil	blanc, Serviteur de
GRAVURE	gisent, aux lueurs d'un	soleil	blanc et terne, Les
CRÉOLE	Au pays parfumé que le	soleil	caresse, J'ai connu
VOYAGE	gloire des cités dans le	soleil	couchant, Allumaie
CHANT	rieux automne ou d'un	soleil	couchant. Courte t
POISON	peur rouge, Comme un	soleil	couchant dans un c
INVITATION	du bout du monde. -Les	soleils	couchants Revêter
SOLEIL	rètes luxures, Quand le	soleil	cruel frappe à trait
AME DU VIN	de peine, de sueur et de	soleil	cuisant Pour enger
CHATIMENT	n s'en alla. L'éclat de ce	soleil	d'un crêpe se voila
CHANT	-et forcé, Et, comme le	soleil	dans son enfer pol
PROFUNDIS	La froide cruauté de ce	soleil	de glace Et cette ir

FLEURS Edition - Page 129

file:///home/bpincemi/TXM/corpora/fle

Faites, pour égayer l' ennui de nos prisons,
Passer sur nos esprits, tendus comme une toile,
Vos souvenirs avec leurs cadres d' horizons.

Dites, qu' avez-vous vu ? [Nous avons vu des astres
Et des flots ; nous avons vu des sables aussi ;
Et, malgré bien des chocs et d' imprévus désastres,
Nous nous sommes souvent ennuyés, comme ici.

La gloire du soleil sur la mer violette,
La gloire des cités dans le soleil couchant,
Allumaient dans nos cœurs une ardeur inquiète
De plonger dans un ciel au reflet alléchant.

Les plus riches cités, les plus grands paysages,
Jamais ne contenaient l' attrait mystérieux
De ceux que le hasard fait avec les nuages.
Et toujours le désir nous rendait soucieux !

-La jouissance ajoute au désir de la force.
Désir, vieil arbre à qui le plaisir sert d' engrais,
Cependant que grossit et durcit ton écorce,
Tes branches veulent voir le soleil de plus près !

129 / 15

Corpus Fleurs du Mal : double accès aux contextes du mot « soleil » (concordance et édition)

TXM

Fichier Outils Affichage Aide

Corpus

Requête : [lemme="soleil"][pos="J|R"]

Seuils : Fmin: 1 Fmax: 9999999

1 -14 / 14

word	Fréquence T=33233
soleil couchant	3
soleil blanc	2
soleil automnal	1
soleil cruel	1
soleil cuisant	1
soleil monotone	1
Soleil moribond	1
soleil nouveau	1
soleil oblique	1
soleil radieux	1
soleil rare	1
soleils couchants	1
soleils malçais	1

Requête : ([pos="S.")[lemme="noir"]]

Seuils : Fmin: 1 Fmax: 9999999

1 -37 / 37

word	Fréquence T=33233
noir océan	2
noirs bataillons	2
noir tableau	2
couleur noire	1
couvercle noir	1
drapeau noir	1
flot noir	1
gouffre noir	1
horizon noir	1
ifs noirs	1
jour noir	1
marbre noir	1
noire colère	1

Requête : [pos="D1"]?[pos="S."]

Seuils : Fmin: 1 Fmax: 9999999

1 -100 / 174

word	Fréquence
mon cœur	32
mon âme	23
ma longue misère	15
mes yeux	11
Mon cœur	6
ma tête	5
Mon âme	5
mon esprit	5
ma sœur	4
mon amour	4
Mon esprit	4
mon sein	4
ma main	3

Ouverture du tableau d'index

*Corpus Fleurs du Mal : recherche de paradigmes : quels adjectifs suivent « soleil » ?
qu'est-ce qui est « noir » ? quels noms détermine le possessif de première personne du
singulier ?*

TXM

Fichier Outils Affichage Aide

LEMN2011INT: [lemme="mon"][pos="ADV"?[pos="ADJ"?[pos="NOM"]]:word

Requête: [lemme="mon"][pos="ADV"?[pos="ADJ"?[pos="NOM"]] Propriétés: word

Seuils: Fmin: 1 Fmax: 9999999 Vmax: 9999999 Résultats par page: 100

|< < 1 -100 / 340 > >| t 977 , v 340 , fmin 1 , fmax 68

word	Fréquence T=367102
mon père	68
mes filets	50
mon avis	50
mon bateau	35
mon poisson	25
mes parents	22
ma vie	18
mes collègues	16
mon frère	16
mon permis	15
mon collègue	13
ma famille	12
mes clients	11
mon activité	11
mon métier	10

Macr

Corpus Léman : quels noms détermine le possessif de première personne du singulier ?

TXM

Fichier Outils Affichage Aide

Corpus

LEMAN2011INT age: lemme:[] LEMAN2011INT: ([L... »s

Requête: Propriétés: word Editer

Seuils: Fmin: 1 Fmax: 9999999 Vmax: 9999999 Résultats par page: < < 1 -38 / 38 > > t 66 , v 38 , fmin 1 ,

word	Fréquence T=367102
beau métier	7
métier dur	6
autre métier	4
même métier	4
joli métier	3
métier physique	3
ancien métier	2
métier artisanal	2
métier complet	2
métier difficile	2
métier traditionnel	2
métier très individualiste	2
petit métier	2
autres métiers	1
beaux métiers	1
bon métier	1
dernier métier	1
mêmes métiers	1
métier ancestral	1
métier assez différent	1
métier bucolique	1
métier dangereux	1
métier extrêmement dur	1
métier idyllique	1
métier local	1

Terminé : 38 items pour 66 occurrences.

LEMAN2011INT:[word="beau"] [word="métier"]

Requête: Pivot: word Editer Chercher

Clés de tri: #1 Aucun #2 Aucun #3 Aucun #4 Aucun

< < 1 -7 / 7 > > Cacher paramètres

text_id	Contexte gauche	Pivot	Contexte droit
int06	moi, c'est le plus	beau métier	du monde pêcheur. Alors en fait, quand mo
int10	est vrai que c'est un	beau métier	mais c'est vrai que c'est une denrée qui est
int12	joli métier, on a un	beau métier	. Bon, un métier de travail, faut pas rêver.
int12	on a quand même un	beau métier	, on a quand même des bonnes choses. Mai
int13	and même comme un	beau métier	quoi. Ouais il y a une bonne image, et ça
int13	non bah voilà c'est un	beau métier	, que quand même quelques inquiétudes pa
int15	métier, qui est le plus	beau métier	du monde, pour rien au monde je ne ferais

LEMAN2011INT:[word="métier"] [word="dur"]

Requête: Pivot: word Editer Chercher

Clés de tri: #1 Aucun #2 Aucun #3 Aucun #4 Aucun

< < 1 -6 / 6 > > Cacher paramètres

text_id	Contexte gauche	Pivot	Contexte droit
int07	une image que c'est un	métier dur	parce qu'effectivement l'hiver bor
int10	. Non que c'est un	métier dur	, pas toujours évident. C'est bien c
int13	pêcheur professionnel c'est un	métier dur	, difficile, parce que malgré tout o
int13	donc ça reste malgré tout un	métier dur	, un métier physique. Euh, il y a, il
int13	, c'est malgré tout un	métier dur	. C'est un métier de passion si on a
int43	. Parce que c'est un	métier dur	quand même. Et puis quand on voi

Corpus Léman : comment les pêcheurs qualifient leur métier

TXM

Fichier Outils Affichage Aide

Corpus

LEMAN2011INT: "seul.*|solit.*|isol.*|individualis.*":lemme

Requête : seul.*|solit.*|isol.*|individualis.* Propriétés : lem

Seuils : Fmin : 1 Fmax : 9999999 Vmax : 9999999 Résu

t 220

lemme	Fréquence T=367102
seul	185
individualiste	19
seulement	14
individualisme	1
solitaire	1

LEMAN2011INT: [pos="VER.*"] [lemme="tout"] ? [lemme="seul"] ...

Requête : [pos="VER.*"] [lemme="tout"] ? [le

Seuils : Fmin : 1 Fmax : 9999999 Vmax : 9999999 Résu

t 78

lemme	Fréquence T=367102
être tout seul	19
suivre être tout seul	12
travailler seul	7
aller tout seul	3
apprendre tout seul	3
être seul	3
faire tout seul	3
travailler tout seul	3
aller seul	2
lancer tout seul	2
rester seul	2

Tri terminé.

LEMAN2011INT: [lemme="seul"]

Requête : [lemme="seul"] Pivot: word Editer Chercher

Clés de tri : #1 Contexte g #2 Aucun #3 Aucun #4 Aucun Tri

t 101 / 185

Cacher paramètres

text_id	Contexte gauche	Pivot	Contexte droit
int04	ui est une pisciculture domaniale, qui son	seul	but est de réempoissonner le lac Léman. D
int05	d'ici et puis je suis tranquille : je suis	seul	, je vois pas pourquoi j'irais à Lugrin ou à T
int20	au ventre d'aller au milieu du lac, tout	seul	. Euh, je vous parle en tant que femme. Il
int22	, qui va, euh, se purger, tout	seul	. C'est que le lac, à un moment donné,
int20	pêcheur, seul, qui va faire tout, tout	seul	, euh, alors à ce moment là, oui, il
int19	... Alors d'un côté je fais tout, tout	seul	, ou je suis tout seul et puis euh, [il
int08	ux sur le même bateau. Après j'allais tout	seul	avec mon bateau. Voilà oui. On va parler d
int19	voulez progresser, [il] faut y aller tout	seul	, se faire son exprience euh, se jeter à l'eau
int33	ri qui me suis démener pendant 7 ans tout	seul	pour avoir tous ces clients là. Donc j'ai dû é
int19	encore euh, je continue d'apprendre tout	seul	, quoi, en fait euh, j'ai encore euh,
int17	quelque chose qu'on peut apprendre tout	seul	automatiquement quand on arrive sur zon
int33	ienda tout seul. Si, vous apprendrez tout	seul	. Si, ben je me suis chopé la grêle, euh
int33	demi, un fois qu'il aura fait ça tout	seul	, euh, il peu se lancer y'a pas de soucis
int29	uis voilà. Et puis après j'ai commencé tout	seul	. Alors on va prendre une journée d'été, c'e
int39	s bien passé, puis après j'ai continué tout	seul	. Non je m'étais jamais intéressé à, je n'y
int02	. Mais là oui, on s'est débrouillé tout	seul	. Si, même, vous prenez la maison Pertuise
int19	ris plein de choses, cette année déjà tout	seul	, mais euh, et j'ai encore plein, plein,
int22	e vieille camionnette, je me démerde tout	seul	, voilà hein, perso. C'est pas mon truc,
int25	ignée et maintenant le filet se dévide tout	seul	dans le bateau, je n'ai plus d'effort à faire
int22	matique, c'est mathématique ! Si t'es tout	seul	, et que tu pêches, tu pêches mal un matin,
int33	pas, c'est ton droit, tu es tout	seul	. C'est ton boulot. T'es, t'es ton
int38	ossait. Ouais. Que maintenant on est tout	seul	. Ouais. Les filets. Puisque c'est plus fin déj
int33	orce que bon on couvre, quand on est tout	seul	on est tout seul. On est tout seul on est tou
int42	s employés, mais sinon, quand on est tout	seul	on peut pas. Et vous pensez que c'est un m
int18	est ça le problème. Puis quand on est tout	seul	, quand il faut tout faire tout seul ça occup

Corpus Léman : approche du thème de la solitude : mots pour la dire, constructions et contextes

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: [lemme="soleil"] (10, 10)

Requête : [lemme="soleil"]

Propriétés des cooccurents : word Editer Seuils : Fréq ≥ 2

Contexte : forme structure div

de - 10 à - 0

Cooccurrent	Fréq	Cofréq	Indi	Distance moyenne
flamme	13	5	4	5,0
couchant	5	3	3	,0
s'	116	13	3	4,1
Par-delà	2	2	2	3,0
crêpe	2	2	2	3,5
noyé	2	2	2	2,0
fige	2	2	2	8,0
soirées	3	2	2	5,0
climat	3	2	2	4,0
six	3	2	2	6,5

Tri terminé.

FLEURS: ([lemme="soleil"] []* [word="flamme"]) | ([word="fla...]

Requête : ([lemme="soleil"] []] Pivot: word Editer Chercher

Clés de tri : #1 Aucun #2 Aucun #3 Aucun #4

1 - 5 / 5 Cacher paramètres

eme_titre	Contexte gauche	Pivot	C
AMBEAU	t en plein jour ; le	soleil Rougit, mais n'éteint pas leur flamme	f
AMBEAU	e, Astres dont nul	soleil ne peut flétrir la flamme	!
BE	ncessamment. Le	soleil a noirci la flamme	d
ME DU VIN	t, sur la colline en	flamme, De peine, de sueur et de soleil	c
UVERCLE	Sous un climat de	flamme ou sous un soleil	b

Corpus Fleurs du Mal : cooccurrences du mot « soleil » comme synthèse des contextes, et retour au texte.

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: [word="œil"] (10, 10) ✖

Requête : [word="œil"]

Propriétés des cooccurents : word Seuils : Fréq ≥ 2

Contexte : forme structure div

de - 10 à - 0 et c

Cooccurrent	Fréquence	Cofréquence	Indice	Distance moy
terreur	2	2	3	
Platon	2	2	3	
froncer	2	2	3	
cherchait	2	2	3	
l'	525	29	3	
Ciel	8	3	2	
souliers	3	2	2	
Laisse	3	2	2	
austère	3	2	2	
tires	3	2	2	

FLEURS: [word="yeux"%c] (10, 10) ✖

Requête : [word="yeux"%c]

Propriétés des cooccurents : word Seuils : Fréq ≥ 2

Contexte : forme structure div

de - 10 à - 0 et c

Cooccurrent	Fréquence	Cofréquence	Indice	Distance moy
mes	57	17	7	
Tes	18	8	5	
sont	45	12	5	
faits	5	4	4	
visions	3	3	3	
tes	87	14	3	
cette	36	8	3	
larges	5	3	2	
pleins	11	4	2	
vos	34	7	2	

Ouverture du tableau d'index

Corpus Fleurs du Mal : cooccurrences d' « oeil » et « yeux », pour comparer leur usage.

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: "brill.*|. *lumi.*|. *flamm.*":word

Requête : Propriétés

Seuils : Fmin: Fmax: Vmax: Ré

1 -20 / 20

word	Fréquence T=33233
flamme	13
lumière	13
lumineux	5
brillants	4
flammes	4
lumières	4
illuminer	3
illuminés	3
brille	2
illumine	2
lumineuse	2
brillaient	1
brillant	1

Ouverture du tableau d'index

FLEURS: "brill.*|. *lumi.*|. *flamm.*|feux" (10, 10)

Requête :

Propriétés des cooccurents : word Seuils : Fréq ≥

Contexte : forme structure

de - à - et c

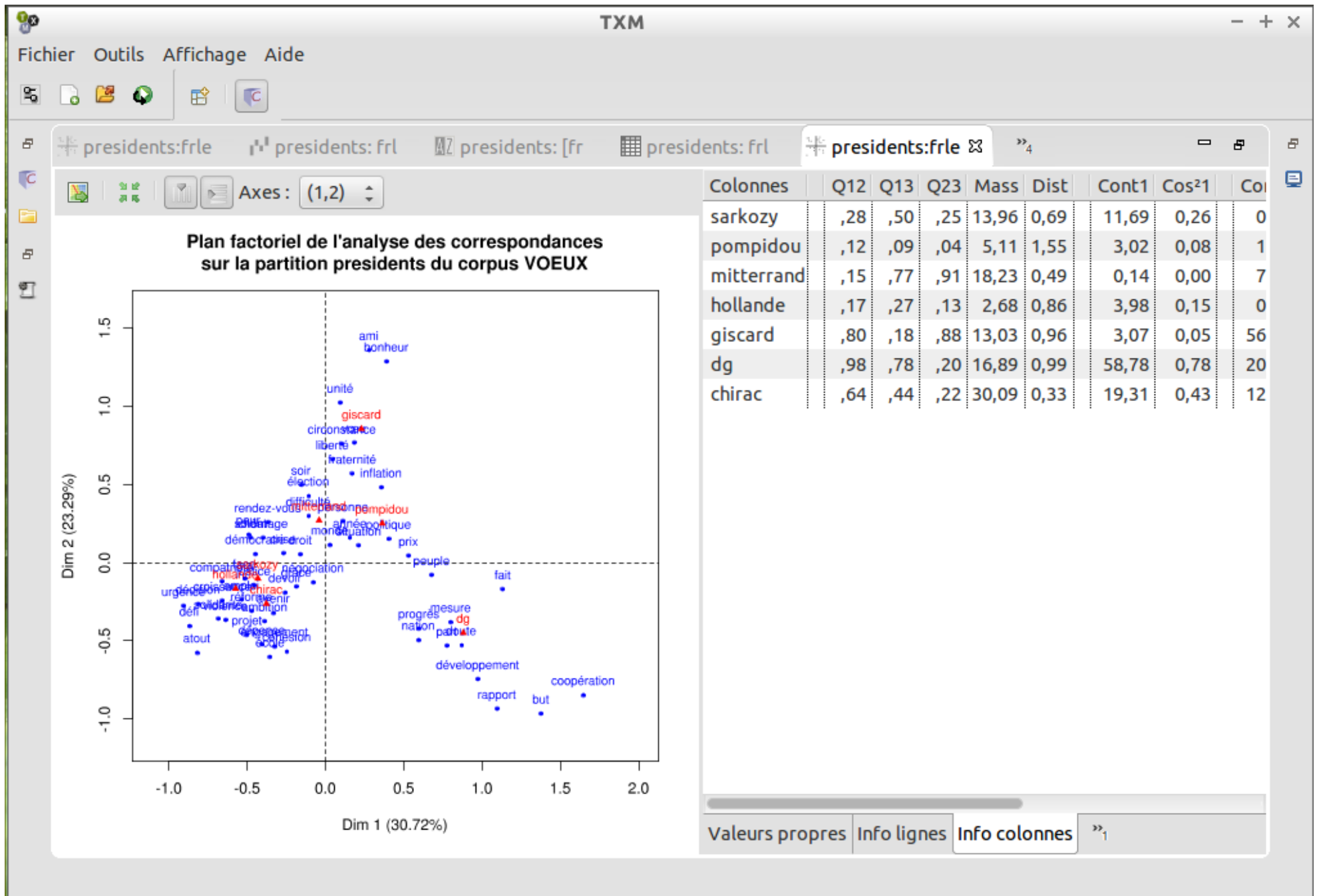
Cooccurrent	Fréquence	Cofréquence	Indice	Distance moy
soleil	44	9	3	
Peut	4	3	3	
soirs	5	3	3	
Un	49	8	2	
-on	6	3	2	
Allume	2	2	2	
Yeux	2	2	2	
bougies	2	2	2	
taudis	2	2	2	
blafarde	2	2	2	

Corpus Fleurs du Mal : recherche des mots liés à un thème (lumière) par définition d'une première liste dans un index puis recherche des cooccurents de cette liste.

Q2. Comment résonne tel mot dans mon corpus ?

- **[CONCORDANCE] et [ÉDITION]**
 - Contexte local et global
 - Technique herméneutique des passages parallèles
 - Phraséologie
- **[INDEX]**
 - Approche distributionnelle et observation de paradigmes
- **[COOCCURRENCES]**
 - Synthèse statistique des attirances

Q3. Comment caractériser les différent(e)s
textes / genres / périodes / auteurs etc.
et quels pôles les structurent ?



Corpus Voeux : Analyse Factorielle des Correspondances sur les présidents (caractérisés par les lemmes nominaux les plus spécifiques ($S+>2$, 10 max / président).

TXM

Fichier Outils Affichage Aide

presidents: frl

Unités	Fréc	dg t=	score
développement	28	18	7,6
coopération	14	12	7,5
peuple	91	36	6,9
but	12	10	6,1
progrès	58	25	5,7
part	36	18	5,4
nation	47	21	5,2
fait	22	13	5,1
rapport	14	10	5,0
doute	21	12	4,5
mesure	37	17	4,5
politique	73	20	1,8
prix	20	5	0,6
année	281	51	0,5
engagement	10	2	0,3
négociation	10	2	0,3
cohésion	17	3	0,2
dépense	11	2	0,2
inflation	12	2	-0,2
école	19	3	-0,2
monde	139	23	-0,3

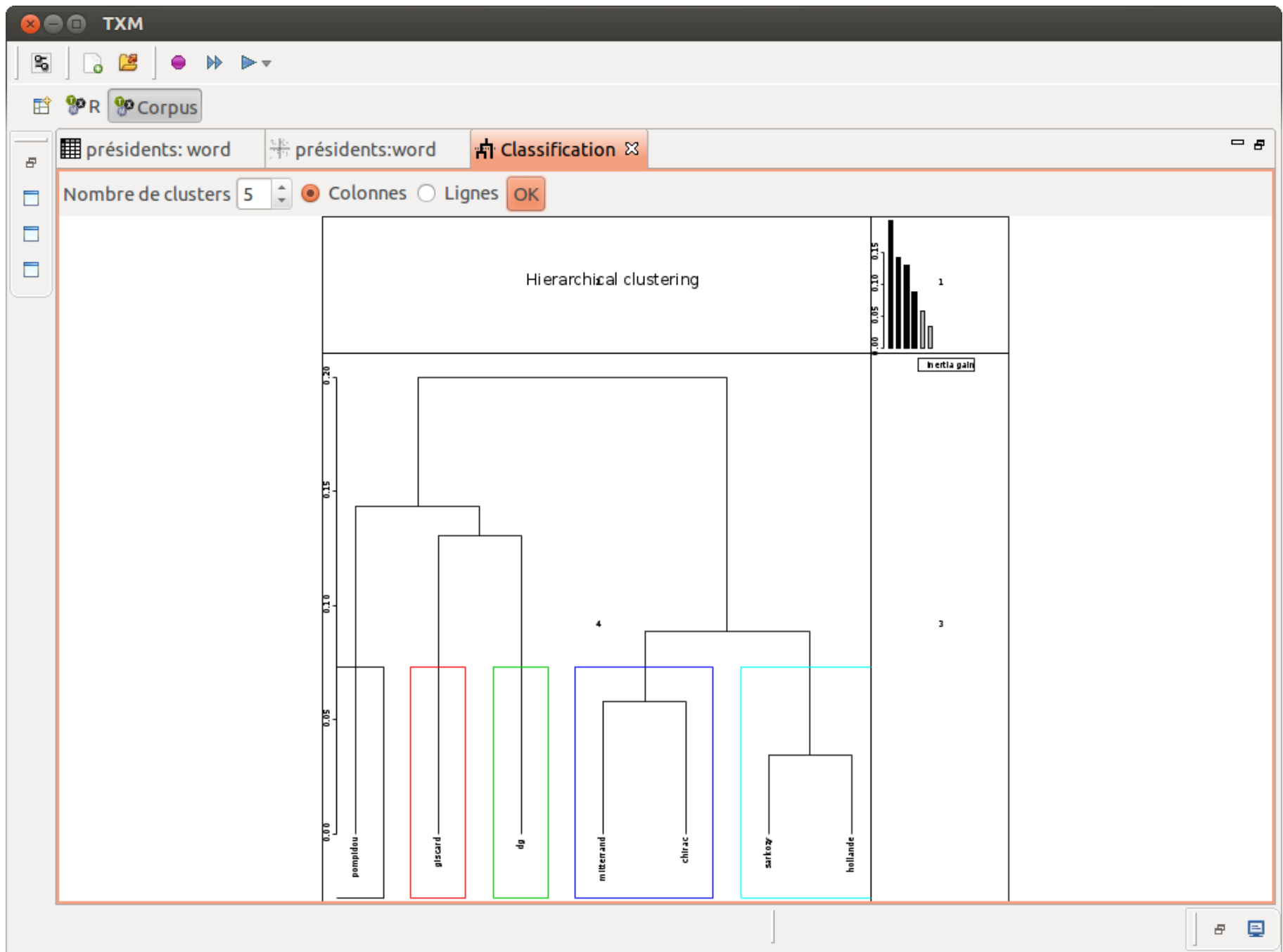
presidents: frlemma

Unités	Fréq	giscarc	score
vœu	100	37	9,4
bonheur	27	17	8,8
unité	30	15	6,0
liberté	62	21	4,8
ami	13	8	4,3
fraternité	23	9	2,8
difficulté	43	13	2,6
soir	53	14	2,2
année	281	50	2,0
monde	139	28	2,0
circonstance	10	3	0,9
fait	22	5	0,8
politique	73	11	0,5
crise	58	8	0,3
face	29	4	0,3
élection	15	2	0,2
inflation	12	1	-0,3
progrès	58	7	-0,3
prix	20	2	-0,3
compatriote	65	7	-0,4
personne	43	4	-0,5

presidents: frlemma

Unités	Fréq	chirac t=	score
solidarité	50	34	7,6
emploi	83	46	6,0
projet	19	15	4,8
avenir	79	41	4,5
ambition	27	18	4,1
atout	12	10	3,7
croissance	35	21	3,7
école	19	13	3,2
cohésion	17	12	3,2
violence	17	12	3,2
compatriote	65	31	2,7
devoir	186	68	1,5
face	29	13	1,2
réforme	46	19	1,2
progrès	58	22	0,9
démocratie	19	8	0,7
nation	47	17	0,7
chômage	36	13	0,6
défi	10	4	0,5
justice	33	11	0,4
monde	139	43	0,4

Corpus Voeux : utilisation des spécificités sur le même tableau de données pour interpréter l'analyse factorielle.



Corpus Voeux : Classification Ascendante Hiérarchique sur les mêmes données (sur les 5 premières dimensions de l'AFC)

TXM

Fichier Outils Affichage Aide

Corpus

taille: lemme				taille: lemme					taille: lemme			
Unités	Fréquence	gros t=14468	score	Unités	Fréquence	moyen t=31468	score	peti	Unités	Fréquence	petit t=11389	score
équiper	36	26	8,4	pêcher	718	463	7,0		apprendre	138	58	8,6
étudier	10	9	4,5	respecter	51	43	5,1		sauver	24	16	6,1
prévoir	23	15	4,2	concerner	19	19	5,0		espérer	29	17	5,3
surgeler	24	15	3,9	associer	22	21	4,5		lever	239	75	4,8
suivre	68	31	3,7	vendre	527	332	4,0		subir	11	9	4,7
placer	12	9	3,4	lier	31	27	3,8		travailler	416	118	4,7
HAPAX	580	180	3,0	savoir	887	539	3,7		former	53	24	4,6
confondre	5	5	3,0	parler	188	126	3,3		dépêcher	20	12	4,0
rééquilibrer	5	5	3,0	transformer	68	50	2,9		permettre	112	38	3,5
créer	33	17	3,0	payer	208	136	2,9		appréhender	7	6	3,4
diminuer	28	15	2,9	connaître	235	151	2,7		progresser	7	6	3,4
lancer	39	19	2,9	recupérer	46	35	2,6		enlever	48	20	3,3
adapter	34	17	2,8	marier	10	10	2,6		jeter	30	14	3,1
marquer	29	15	2,7	racheter	24	20	2,5		préférer	48	19	2,9
attaquer	14	9	2,6	attacher	13	12	2,3		entourer	6	5	2,8
pousser	14	9	2,6	frayer	23	19	2,3		retraire	6	5	2,8
développer	53	23	2,5	démarrer	29	23	2,2		imprégner	4	4	2,8
embêter	22	12	2,5	perdre	62	44	2,2		essayer	153	46	2,8
donner	197	67	2,4	évaluer	12	11	2,1		préserver	11	7	2,7
									gagner	134	41	2,7

Corpus Léman : verbes caractéristiques des pêcheurs selon la taille de leur exploitation

TXM

Fichier Corpus Outils Affichage Aide

thematiques: wo "1

Unités	Fréq	acteurs	score
ils	2354	765	41,0
nous	985	350	26,7
suisses	77	60	26,1
[1242	417	25,8
]	1243	417	25,8
Suisses	51	43	21,5
pêcheurs	864	299	21,0
eux	360	154	20,7
amateurs	63	46	18,2
ont	971	302	14,0
commune	45	33	13,3
législation	29	25	13,2
Suisse	107	56	12,4
concordat	20	19	11,8
règlement	26	22	11,3
pas	5133	1245	9,9
n'	994	289	9,8
sont	1058	304	9,6
Lausanne	14	14	9,6
français	51	31	9,3
règlementa	51	31	9,3

thematiques: word:[]

Unités	Fréq	images	score	n
gens	382	123	40,4	
image	58	46	38,5	
touristes	60	31	17,7	
"	426	93	17,5	
métier	304	75	17,4	
lac	841	144	16,4	
savent	53	23	11,3	
positive	8	8	8,7	
du	2138	253	8,4	
est	8872	867	7,3	
connaissent	27	12	6,3	
questions	37	14	6,2	
oui	775	102	5,8	
disent	53	16	5,5	
posent	22	10	5,5	
voient	27	11	5,4	
ah	73	19	5,3	
plage	23	10	5,2	
pensent	19	9	5,1	
viennent	128	26	4,9	
appelé	12	7	4,9	

thematiques: word:[]

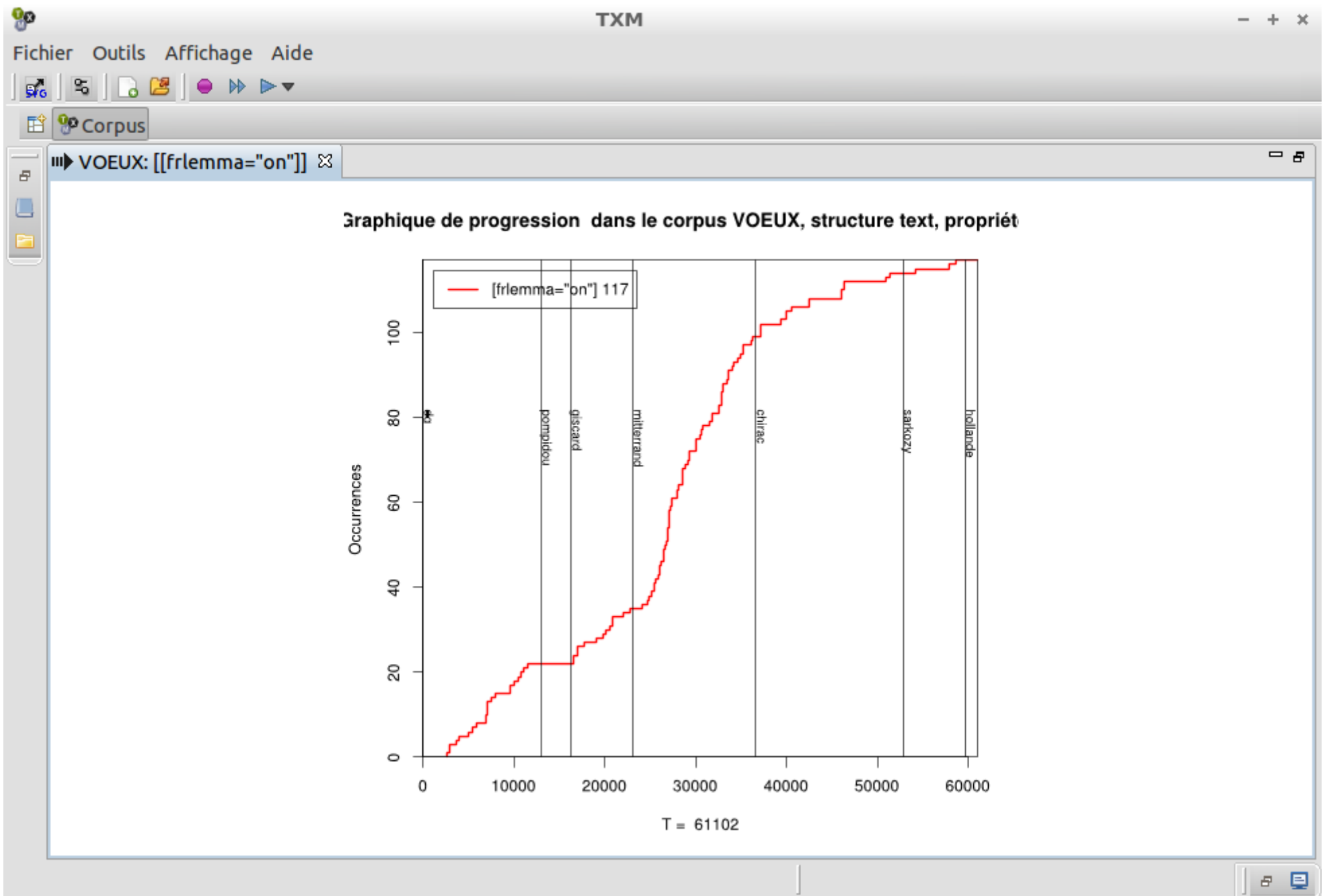
Unités	Fréq	parcour	score
j'	2243	523	97,3
ai	1824	406	69,6
père	157	84	46,4
mon	580	165	42,9
pêcheur	507	152	42,6
avais	207	85	35,6
étais	137	66	33,0
était	819	179	30,2
suis	681	153	27,3
formation	60	39	26,4
m'	398	104	24,2
appris	43	31	23,3
école	52	31	19,5
ans	492	102	16,0
J'	282	69	14,8
je	4311	523	14,5
famille	63	29	14,3
aquaculture	13	13	13,8
études	31	20	13,8
travaillé	49	25	13,7
apprendre	42	23	13,5

Corpus Léman : vocabulaire caractéristique de différentes sections de l'entretien

Q3. Comment caractériser les différent(e)s textes / genres / périodes / auteurs etc. et quels pôles les structurent ?

- [PARTITION] pour définir les parties à contraster
- [SPÉCIFICITÉS] pour repérer les sur- ou sous-emplois dans une partie, *par rapport à l'ensemble du corpus*
- [AFC] pour une cartographie positionnant les différentes parties les unes par rapport aux autres
- [CAH] pour observer comment les parties peuvent se regrouper

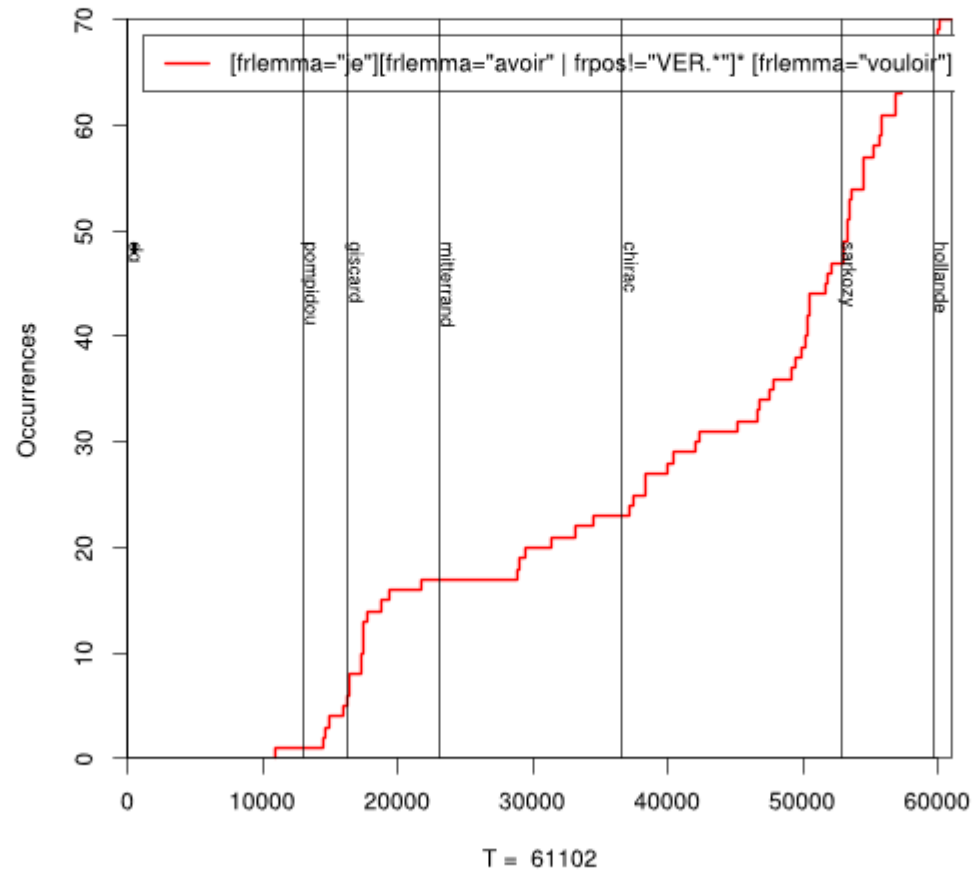
Q4. Comment évolue quantitativement l'usage de tel motif ?



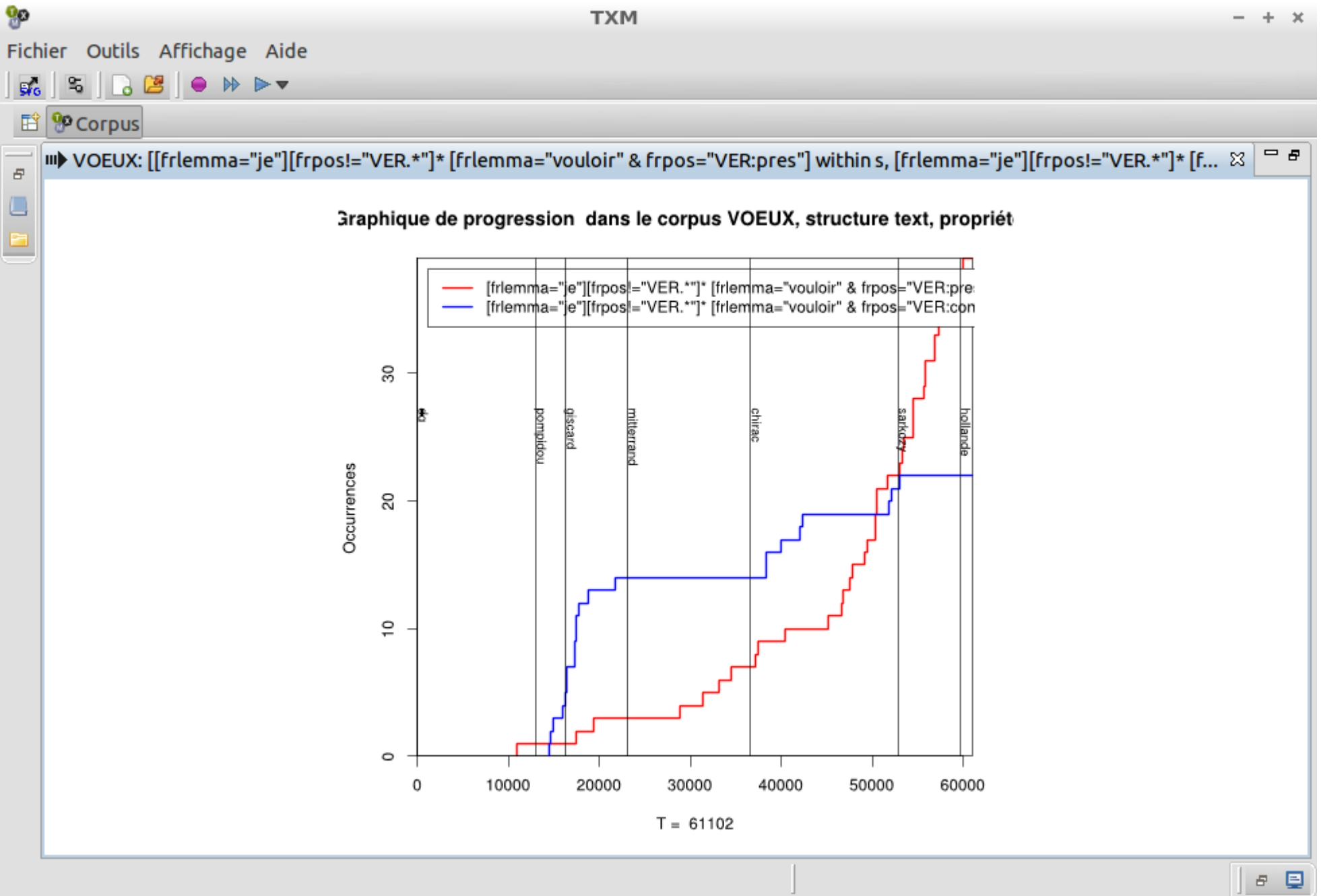
Corpus Voeux : évolution dans le temps de l'usage du pronom « on ».

VOEUX: [[frlemma="je"]][frlemma="avoir" | frpos!="VER.*"]* [frlemma="vouloir"] within s

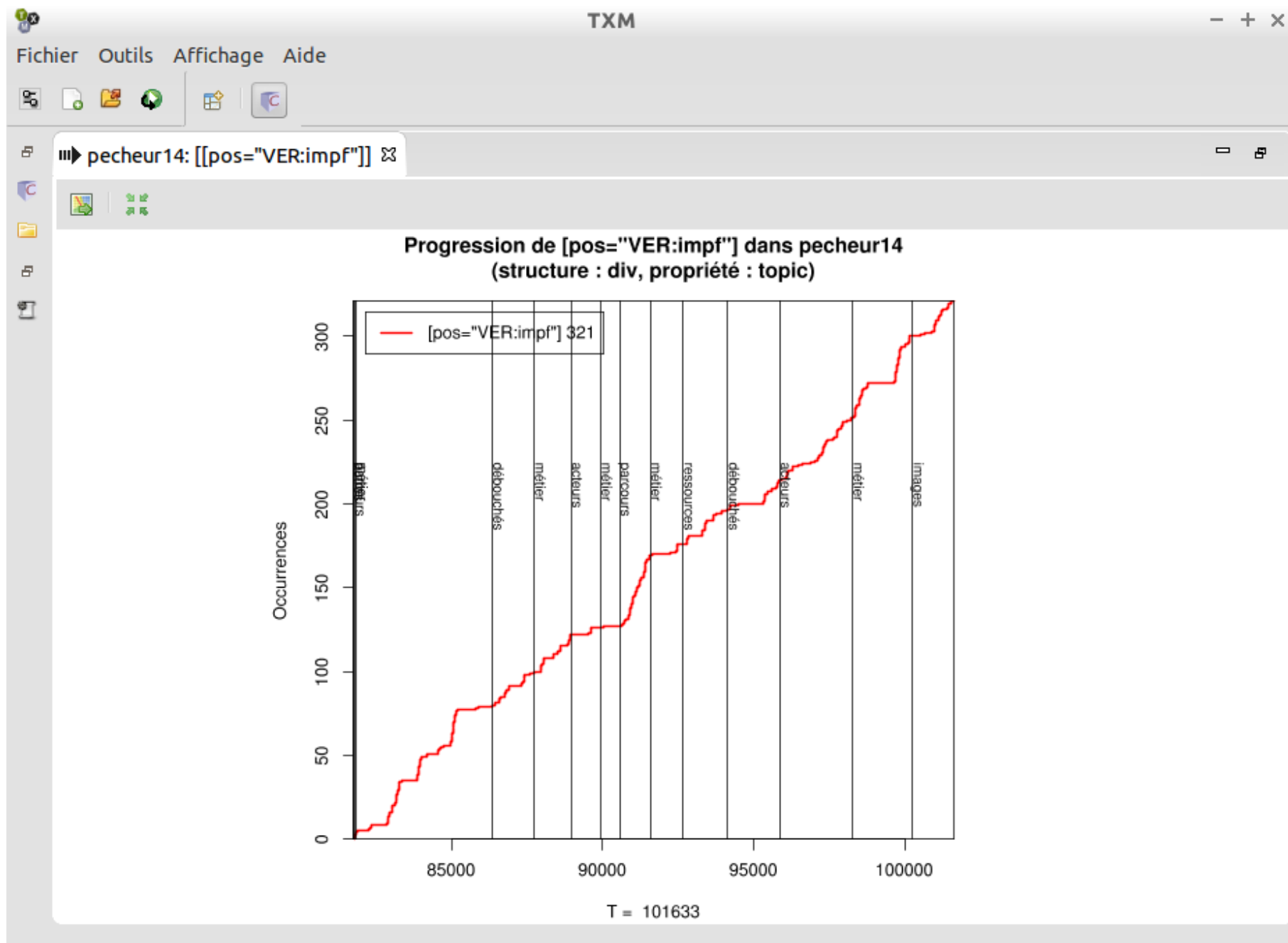
Graphique de progression dans le corpus VOEUX, structure text, propriét



Corpus Voeux : évolution dans le temps de l'usage du motif « je ... vouloir »



Corpus Voeux : progressions comparées de « je ... veux » et « je ... voudrais ».



Corpus Léman : pour le pêcheur 14, usage de l'imparfait au fil des différentes sections de l'entretien

Q4. Comment évolue quantitativement l'usage de tel motif ?

- La fonctionnalité [**PROGRESSION**] en donne une visualisation graphique
- Au fil du temps (corpus chronologique) ou pour un (long) texte

Bilan

- Les questionnements présentés
 - Q1. La présence des mots
 - Q2. Comment résonne tel mot dans mon corpus ?
 - Q3. Comment caractériser les différents textes (ou genres / périodes / auteurs etc.), et quels pôles les structurent ?
 - Q4. Comment évolue quantitativement l'usage de tel motif ?
- En (sa)voir plus ?
 - Démonstration interactive
 - Pour TXM, point d'entrée : <http://textometrie.ens-lyon.fr>