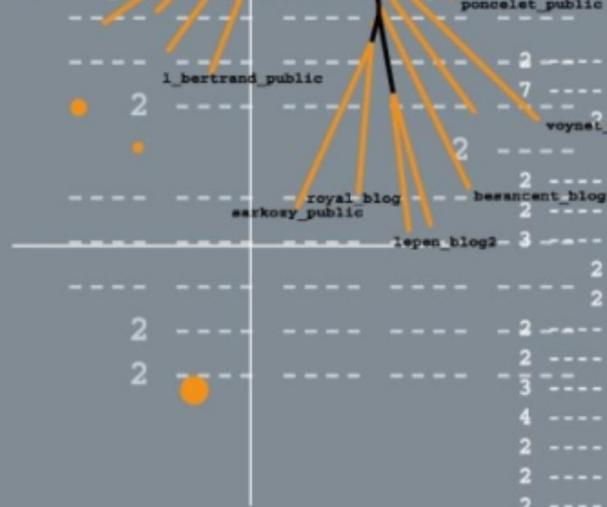


4 juillet 2014

la liberté , la société , et donné à ce
elle a miné la société sous tous ses a
mbreux cas où la société civile ne peut
état m... société civile . cela n
- on dire que la société s' est
au motif que la société spectacul
l'élection... la société se délite , ell
terrorisme , la société civile italienn
niché dans notre société civile ? peut -
isation de ... nous ne sauri
adical de notre société qui , en se tra
eille , et notre société illuminée s' é
al - la première société avancée qui , e

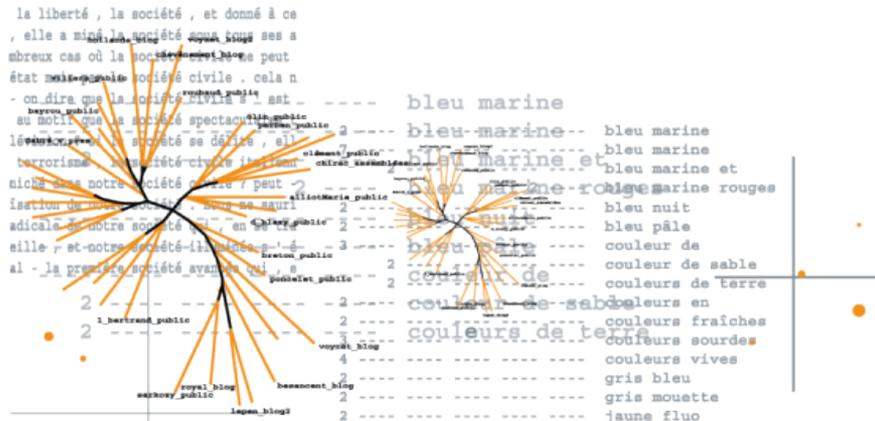


TextObserver/WebObserver: réflexions sur la complémentarité des approches et des méthodes : Lexicométrie, Textométrie, Datasciences, Bigdatas, TAL et... Sciences du langage



JM. Leblanc
UPEC (CEDITEC)

<http://textopol.u-pec.fr/>



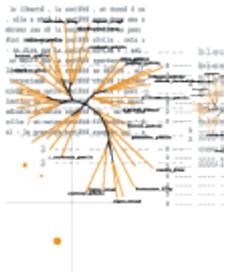
Textométrie/ Lexicométrie/ Visualisations

REPRÉSENTER LES DONNÉES TEXTUELLES

Représenter des données textuelles



- Textométrie/Lexicométrie: une reconstruction/restructuration du texte
 - segmentation/réorganisation
 - Réduction lexicométrique (Tournier)
- Multiplicité des outils logiciels et des représentations
 - Transeo
 - *Vocabulaire de statistique appliquée aux textes et à l'analyse de discours* (Besançon)



- Text
- reco
- s
- R
- Mul
- repl
- I
- v
- à

2	----	----	----	----	----	bleu marine
7	----	----	----	----	----	bleu marine
	2	----	----	----	----	bleu marine et
			2	----	----	bleu marine rouges
2	----	----	----	----	----	bleu nuit
2	----	----	----	----	----	bleu pâle
3	----	----	----	----	----	couleur de
	2	----	----	----	----	couleur de sable
	2	----	----	----	----	couleurs de terre
2	----	----	----	----	----	couleurs en
2	----	----	----	----	----	couleurs fraîches
3	----	----	----	----	----	couleurs sourdes
4	----	----	----	----	----	couleurs vives
2	----	----	----	----	----	gris bleu
2	----	----	----	----	----	gris mouette
2	----	----	----	----	----	jaune fluo
2	----	----	----	----	----	jaune safran
2	----	----	----	----	----	jaune vif
3	----	----	----	----	----	noir d'encre

des
Iles

en , la liberté , la société , et donné à ceu
 ire , elle a miné la société sous tous ses an
 e nombreux cas où la société civile ne peut l
 l ' état mais par la société civile . cela n '
 eut - on dire que la société civile s ' est r
 ue , au motif que la société spectaculaire tr
 a télévision ? si la société se délite , elle
 t du terrorisme , la société civile italienne
 d , niché dans notre société civile ? peut -
 ettoïisation de notre société . nous ne saurio
 on radicale de notre société qui , en se tran
 réveille , et notre société illuminée s ' ét
 torial - la première société avancée qui , en

tes et



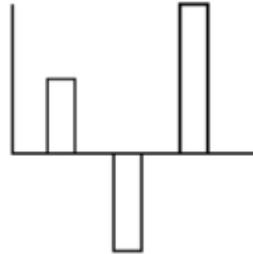
Typologies logicielles

Lexico3, Hyperbase, Txm...

Approche contrastive
Longitudinale



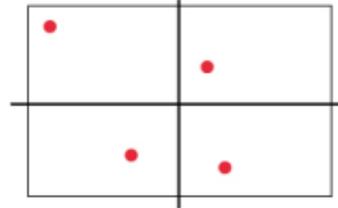
Spécificités



Concordances

rorisme , la **société** civile ital
hé dans notre **société** civile ? pe
tion de notre **société** . nous ne s
cale de notre **société** qui , en se
le , et notre **société** illuminée s
- la première **société** avancée qui

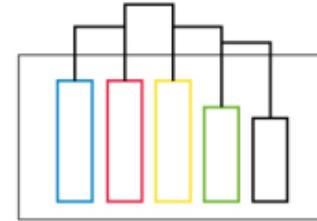
AFC



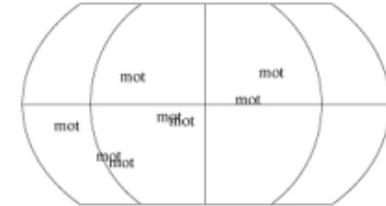
sur la base de partitions
(norme endogène)

Approche structurante

Alceste : cooccurrences entre énoncés

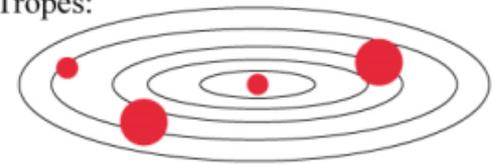


Astartex : cooccurrences généralisées



Ontologies sémantiques

Tropes:



Outils grammaticaux, aide à la rédaction

Cordial:



Alceste, Astartex...

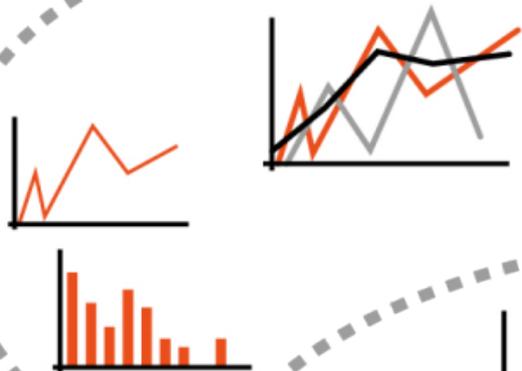
Tropes, Cordial...



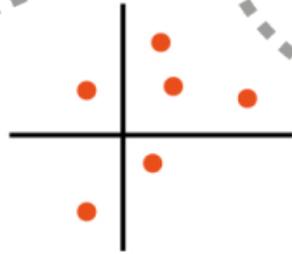
Quelles représentations pour quelles données?

- données chronologiques
 - Courbes
- données quantitatives continues
 - histogramme (surface)
- données quantitatives discrètes
 - diagrammes en bâtons (hauteur)
- données plutôt à caractère qualitatif
 - diagramme en secteur

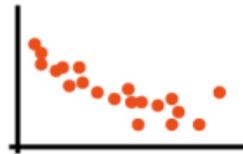
une évolution, une tendance

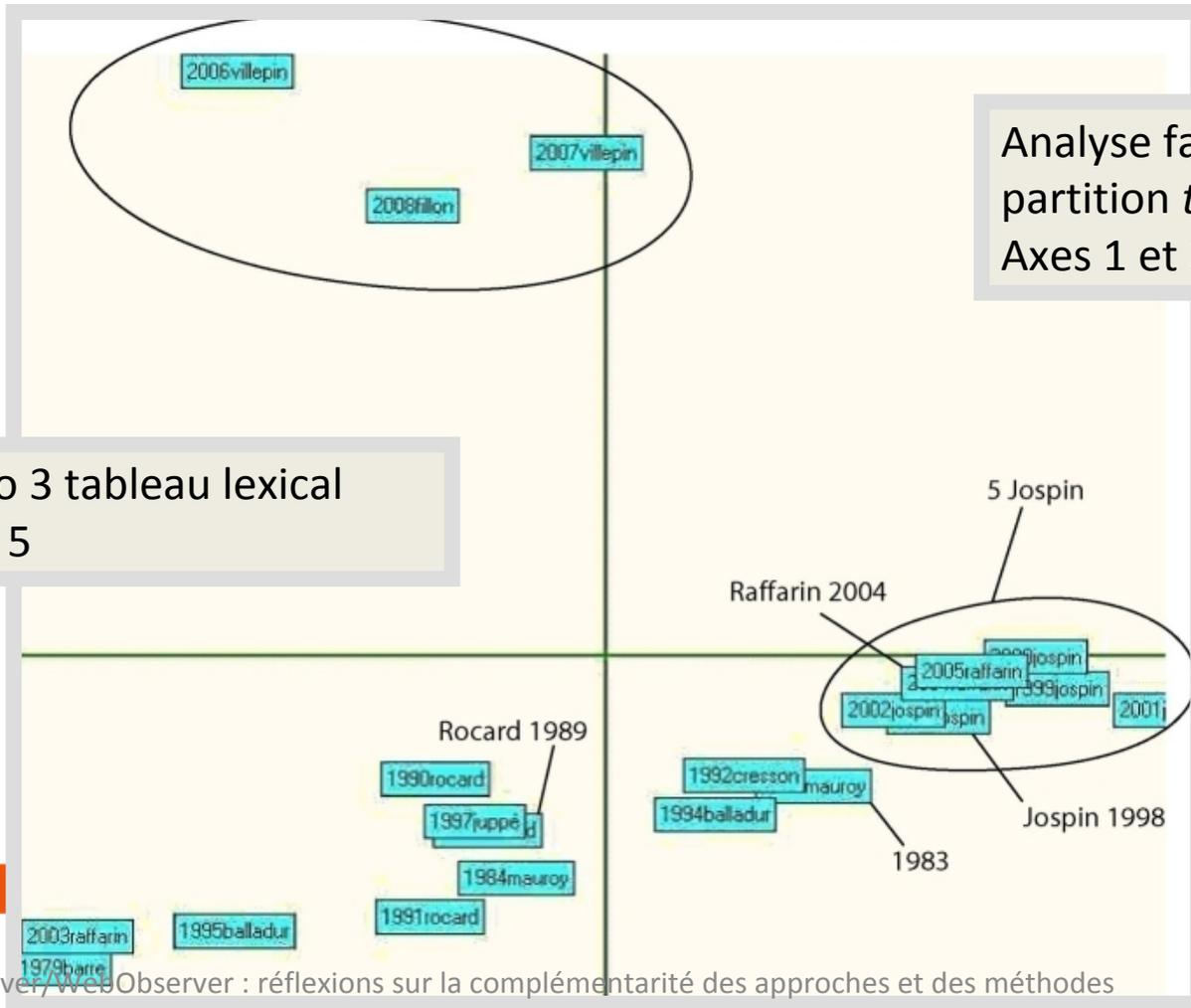


une distribution, une répartition, une composition



une relation

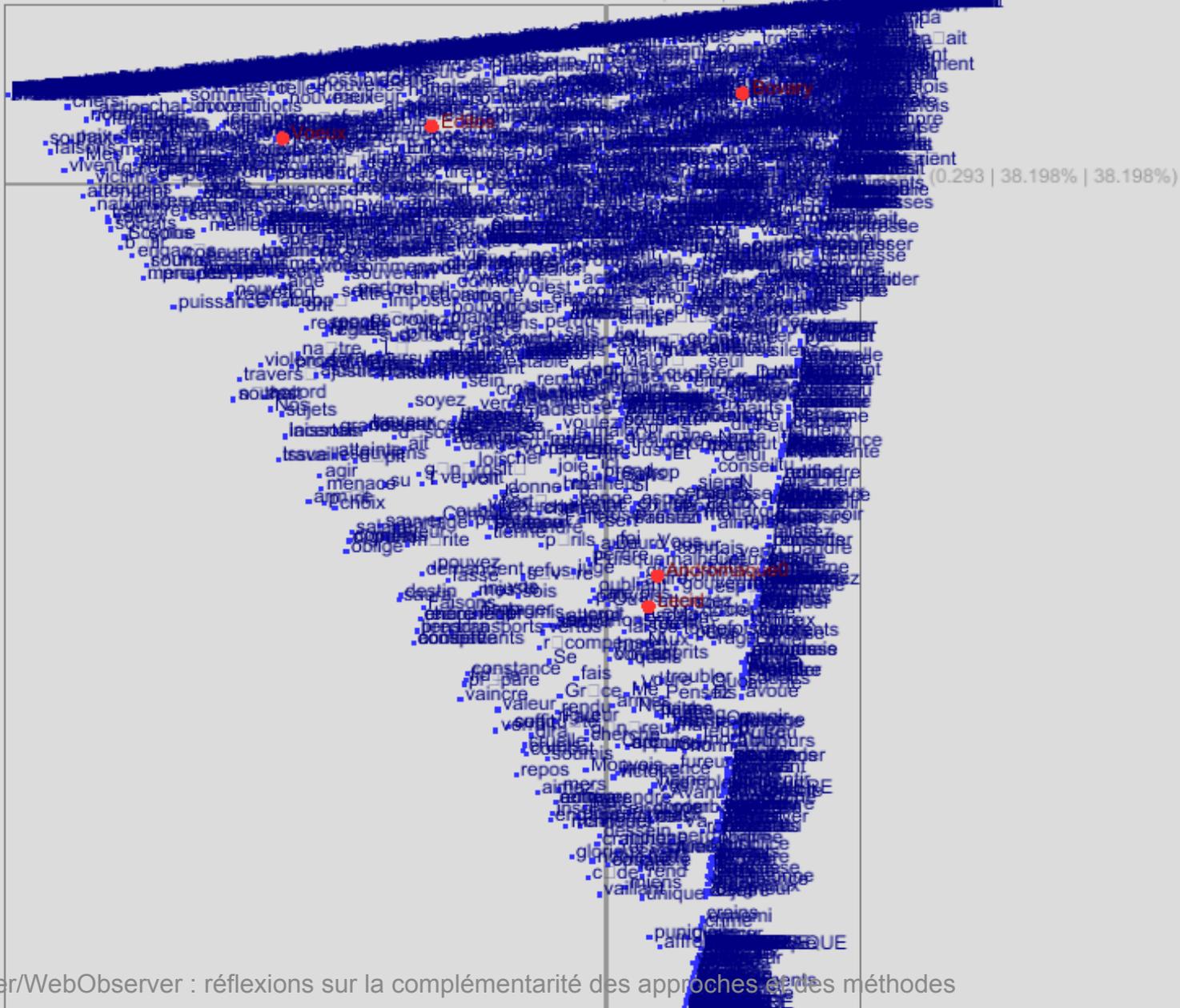


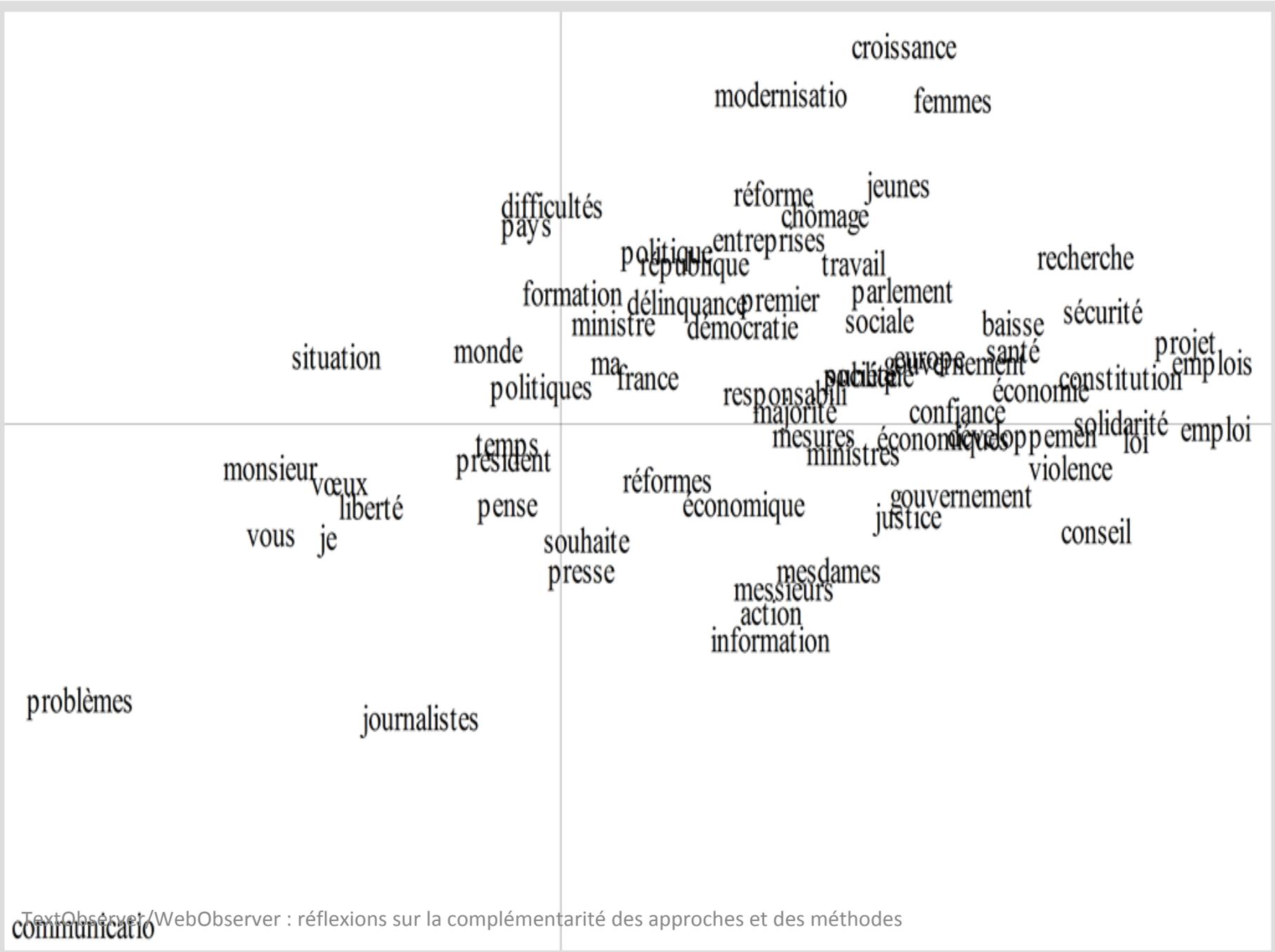


Analyse factorielle de la partition *textes*
 Axes 1 et 2: 12%, 10%

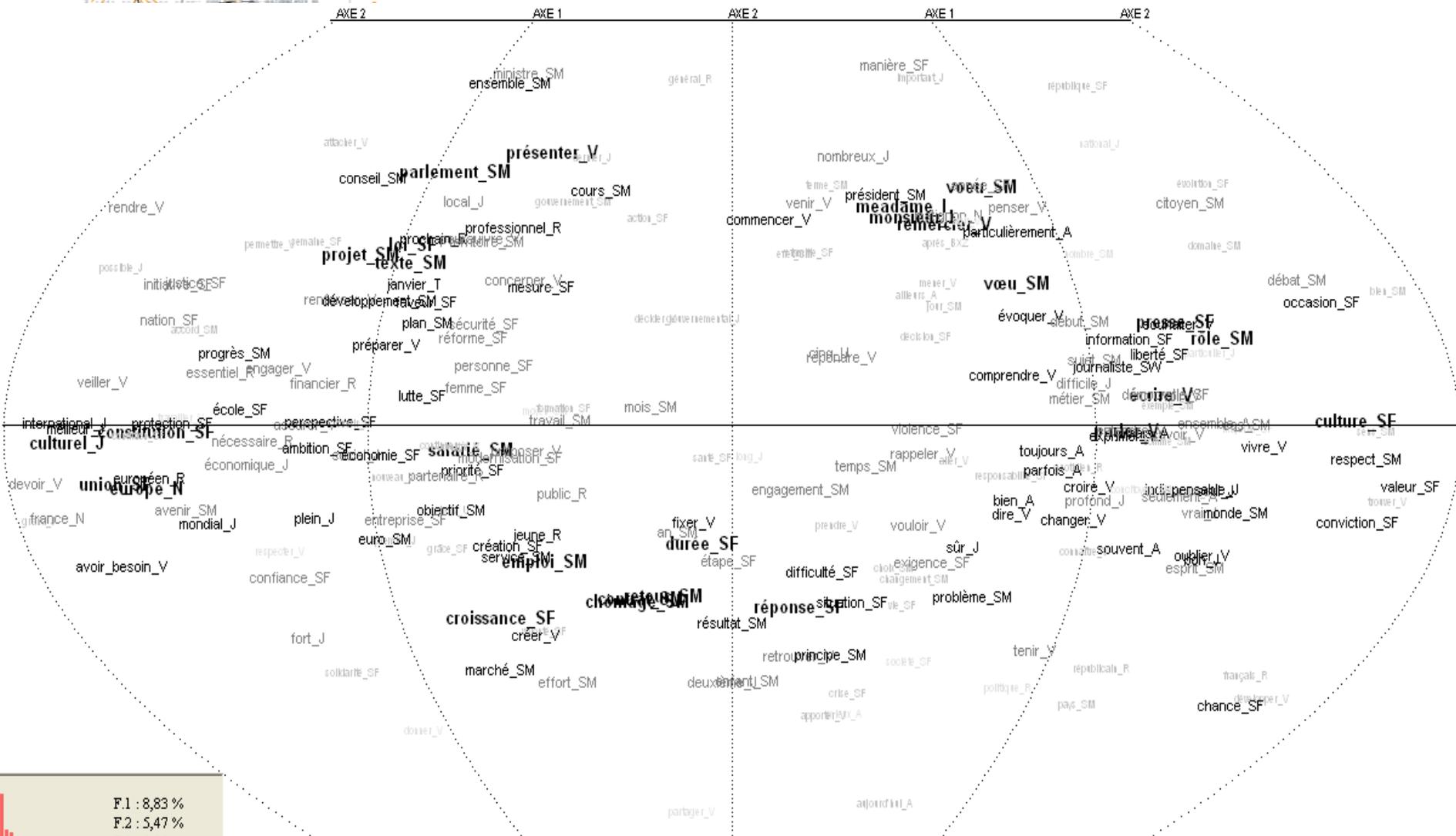
Lexico 3 tableau lexical
 Fréq. 5

axe 2 (0.263 | 34.246% | 72.444%)





Projection de l'AFC sur trois axes



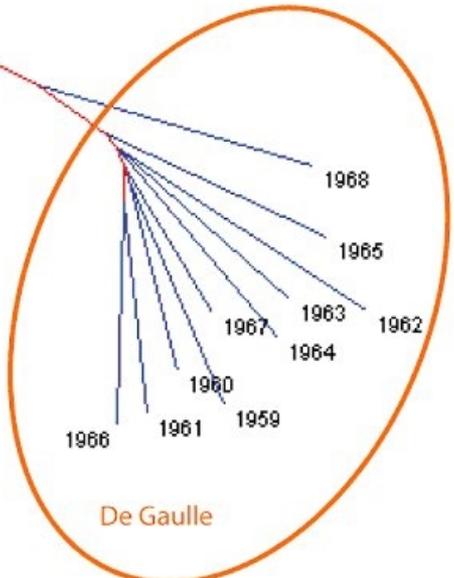
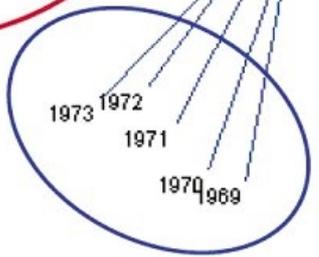
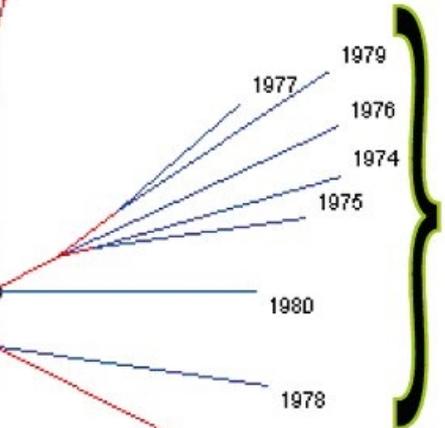
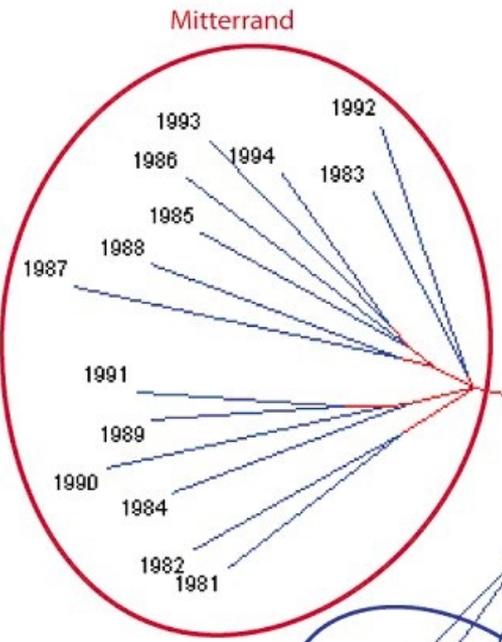
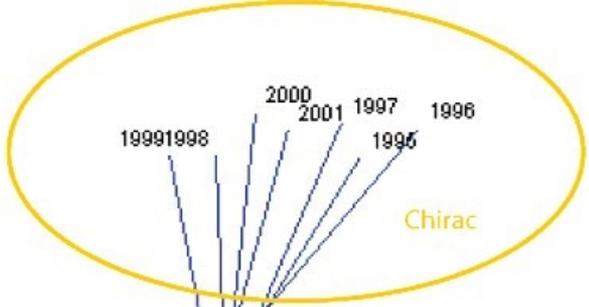
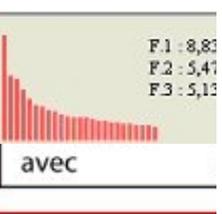
F.1 : 8,83 %
 F.2 : 5,47 %
 F.3 : 5,13 %

TextObserver/WebObserver LIGNES COLONNES RAUTRAICHER GRAPHES Contributions

accord_SM [up arrow] -ROTATION- 1 2 3

Terme
je
vous
souhaite

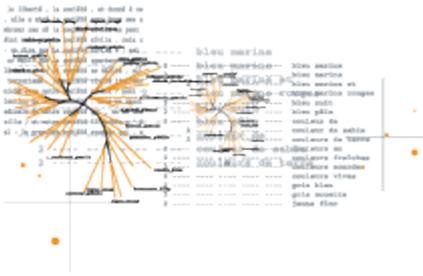
rendre_V
possible_J
infini
natic
veiller_V
intermédiaire_J
culturel_J
devoir_V
france_N
avoir_besoin_V



il_SM
occasion_SF
culture_SF
respect_SM
valeur_SF
conviction_SF

25,29

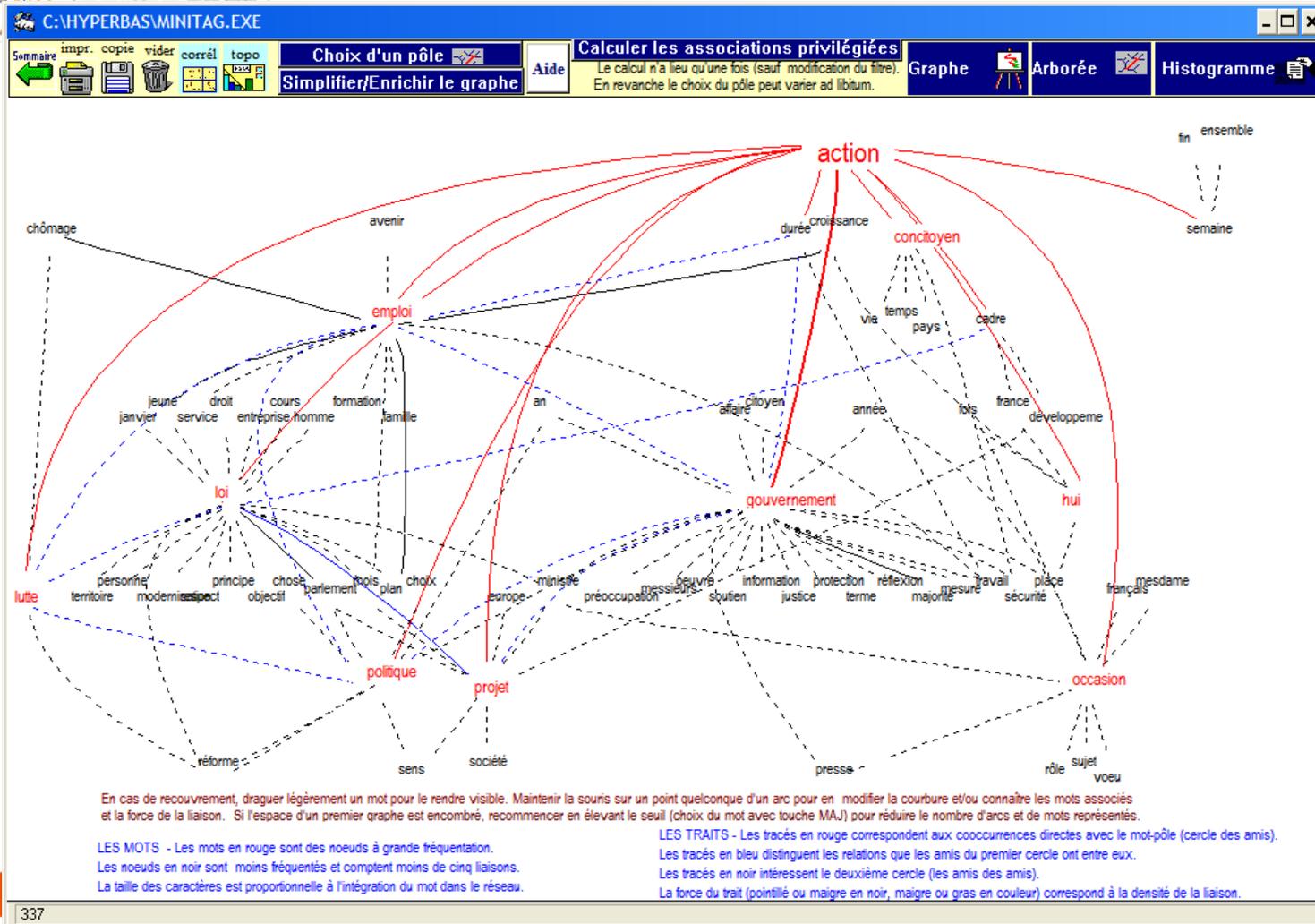
Hyperbase : environnement thématique



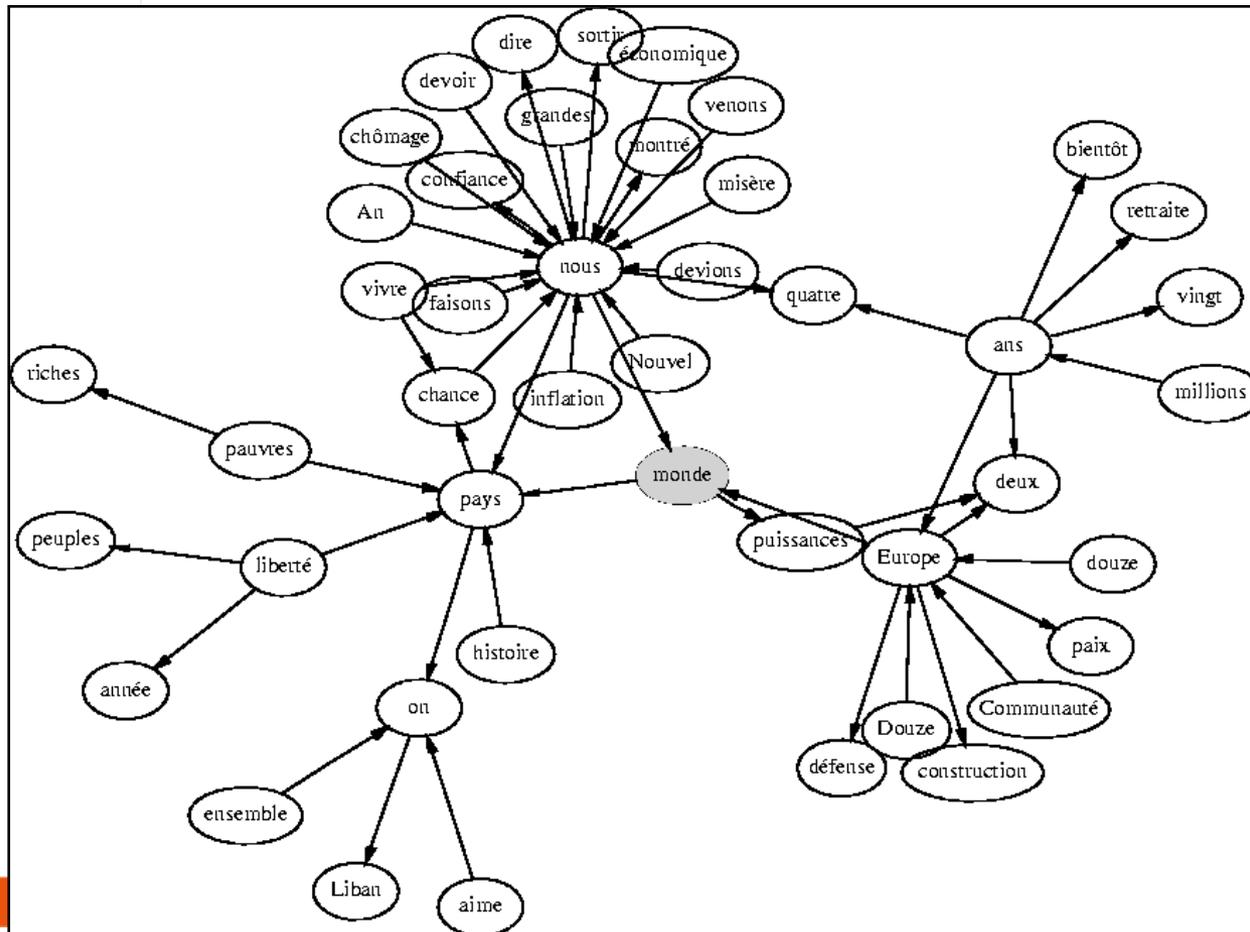
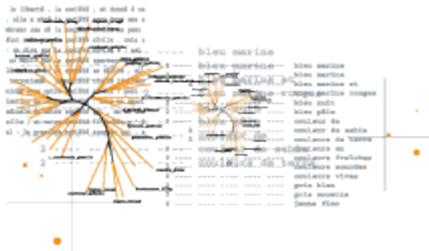
écart	corpus	texte	mot	HIERARCHIQUE	écart	corpus	texte	mot	ALPHABETIQUE
10.11	36	37	jeunes		2.73	163	7	aux	
7.23	5	6	jeune		2.06	24	2	chance	
4.43	55	7	emploi		2.16	21	2	chances	
4.09	15	4	formation		2.41	15	2	conscience	
3.13	5	2	lutter		4.43	55	7	emploi	
2.83	8	2	répondre		2.44	108	5	entre	
2.73	163	7	aux		4.09	15	4	formation	
2.44	239	8	nos		7.23	5	6	jeune	
2.44	108	5	entre		10.11	36	37	jeunes	
2.41	15	2	conscience		3.13	5	2	lutter	
2.32	17	2	système		2.44	239	8	nos	
2.28	46	3	vers		2.06	24	2	notamment	
2.16	21	2	recherche		2.16	21	2	recherche	
2.16	21	2	chances		2.83	8	2	répondre	
2.06	24	2	notamment		2.32	17	2	système	
2.06	24	2	chance		2.28	46	3	vers	

Le lemme « jeune » dans les vœux présidentiels 1959-2010

Associations à partir du pôle *action*

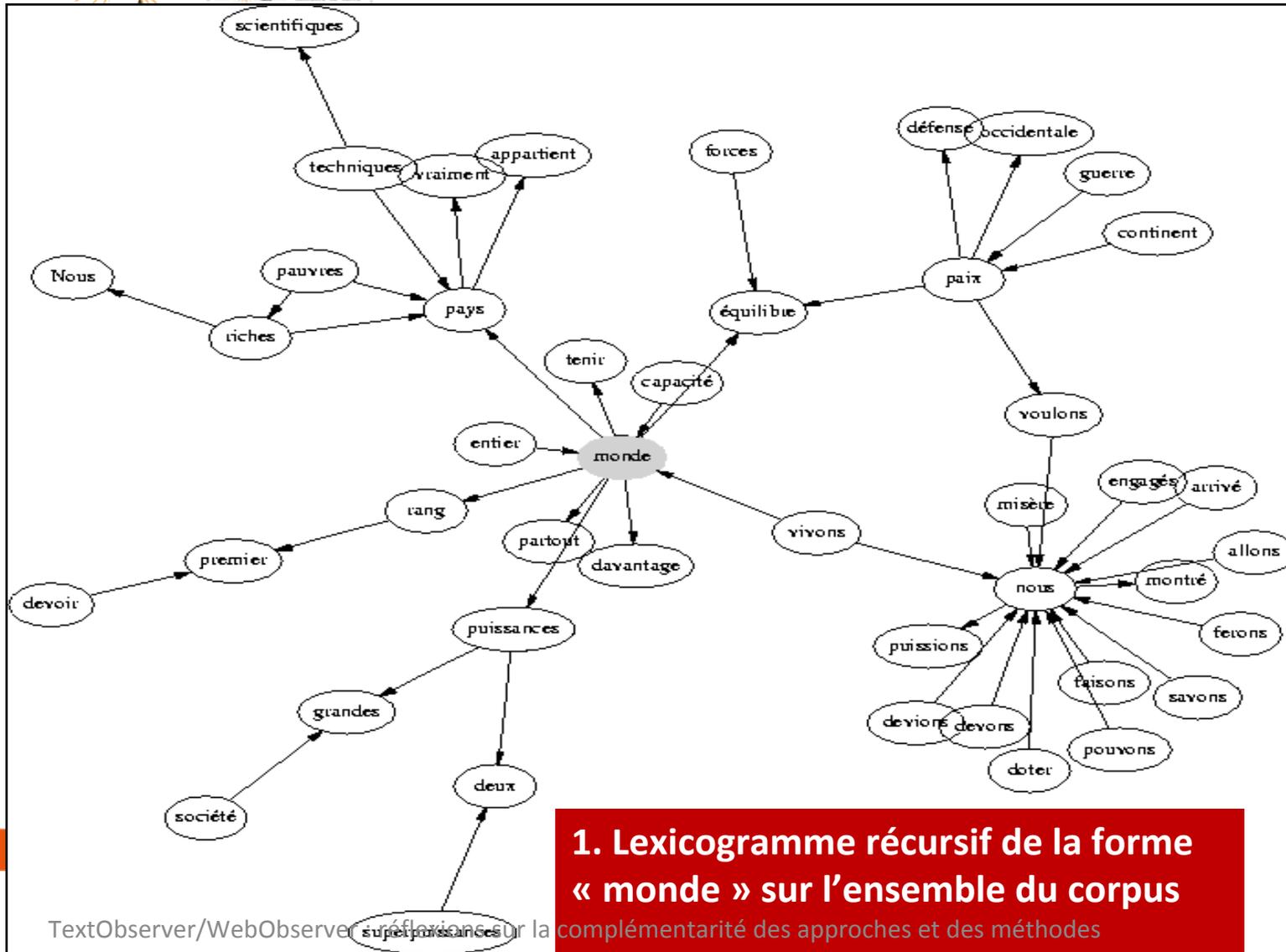


Représenter des données textuelles



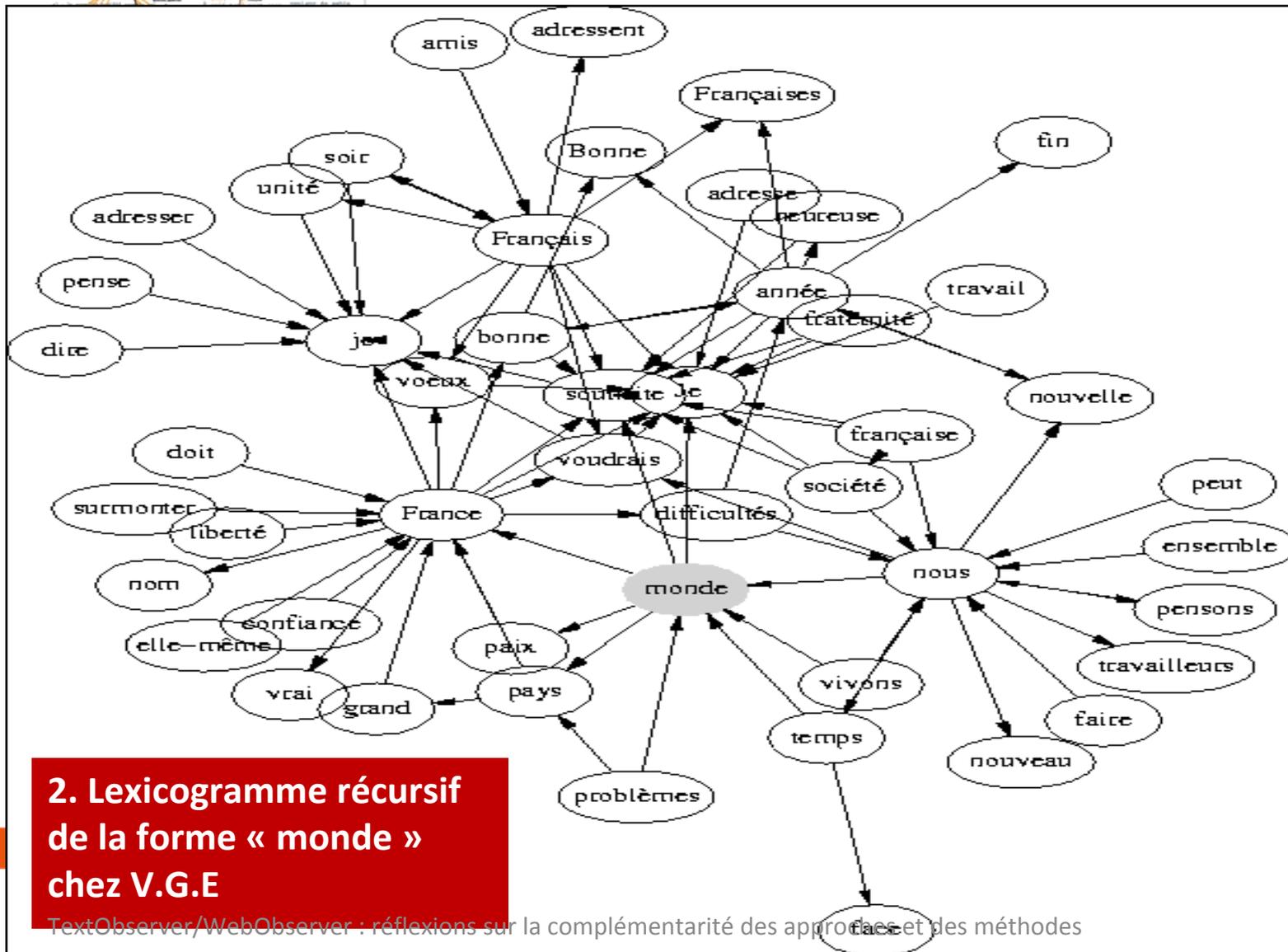
Terme	Frq Tot.	Frq P...	Spécif
pense	29	29	42
pensée	14	14	21
ceux	121	26	13
penser	7	7	11
je	411	41	8
qui	623	52	8
souffrent	12	7	8
ou	130	18	7
maladie	12	6	7
pensées	4	4	7
victimes	15	6	6
solitude	13	6	6
soldats	14	6	6
vont	15	5	5
seuls	16	5	5
ma	25	6	5
pensons	3	3	5
celles	25	6	5
vers	44	9	5
fêter	7	4	5
connaissent	3	3	5
deuil	3	3	5
à	1071	68	5
vos	47	9	5
pensant	2	2	4
aussi	116	12	4
inondations	2	2	4
l'europe	2	2	4
pensais	2	2	4
injustice	6	3	4
soir	48	7	4
abord	35	7	4
soirée	8	3	4
nouvel	20	5	4
du	341	25	4
cruellement	2	2	4

Représenter des données textuelles



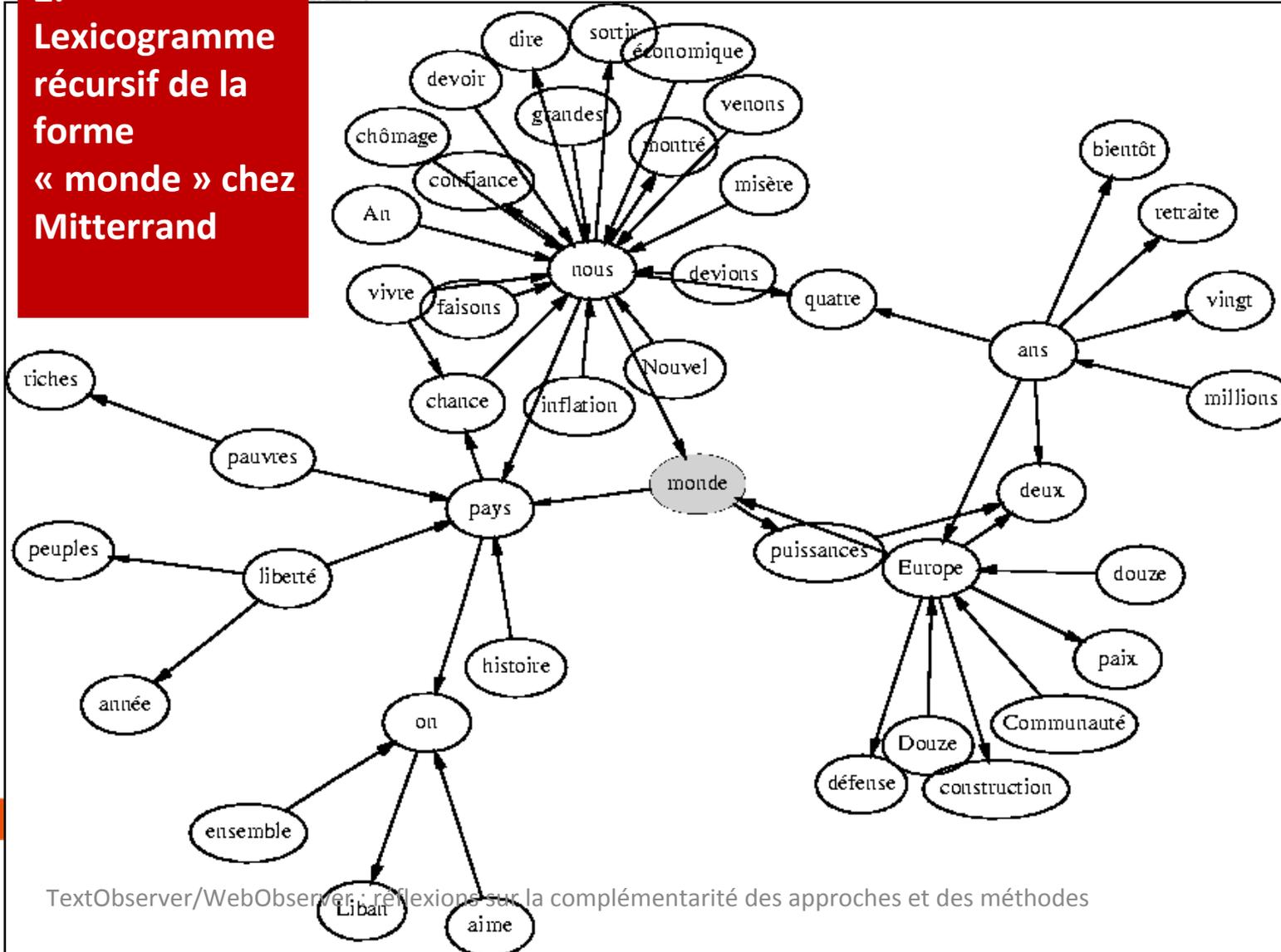
• Élagage des formes outils, numéraires et hapax.
• Seuils : f 3, cf 3, p 5.0E-2, dm 1000.0

Représenter des données textuelles



- Élagage des formes outils, numéraires et hapax.
- Seuils : f 3, cf 3, p 5.0E-2, dm 1000.0

2. Lexicogramme récursif de la forme « monde » chez Mitterrand



- Élagage des formes outils, numéraires et hapax.
- Seuils : f 3, cf 3, p 5.0E-2, dm 1000.0

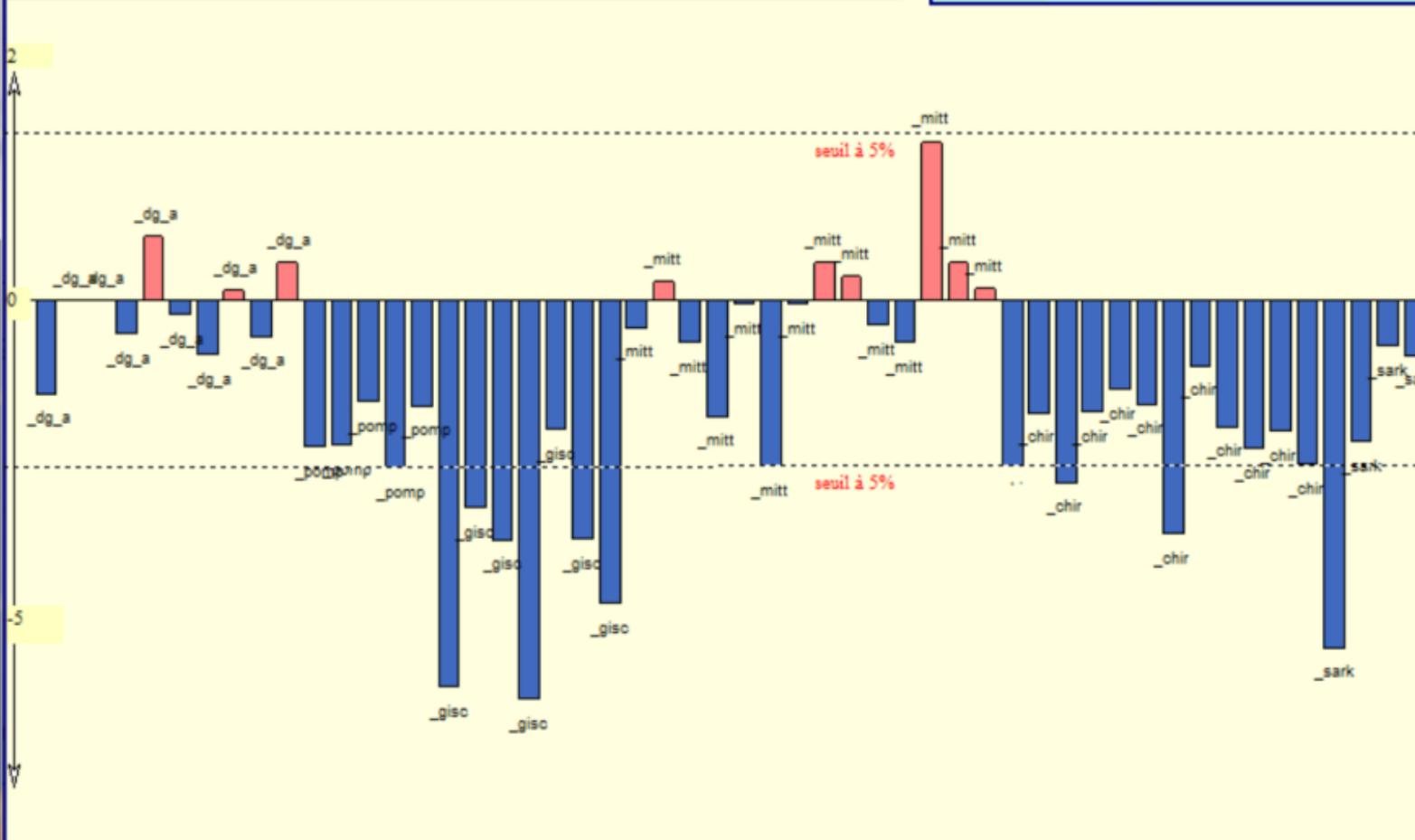
Richesse lexicale

déficits

excédents



Corrélation **-0.136**
Seuil à 5% 0.27

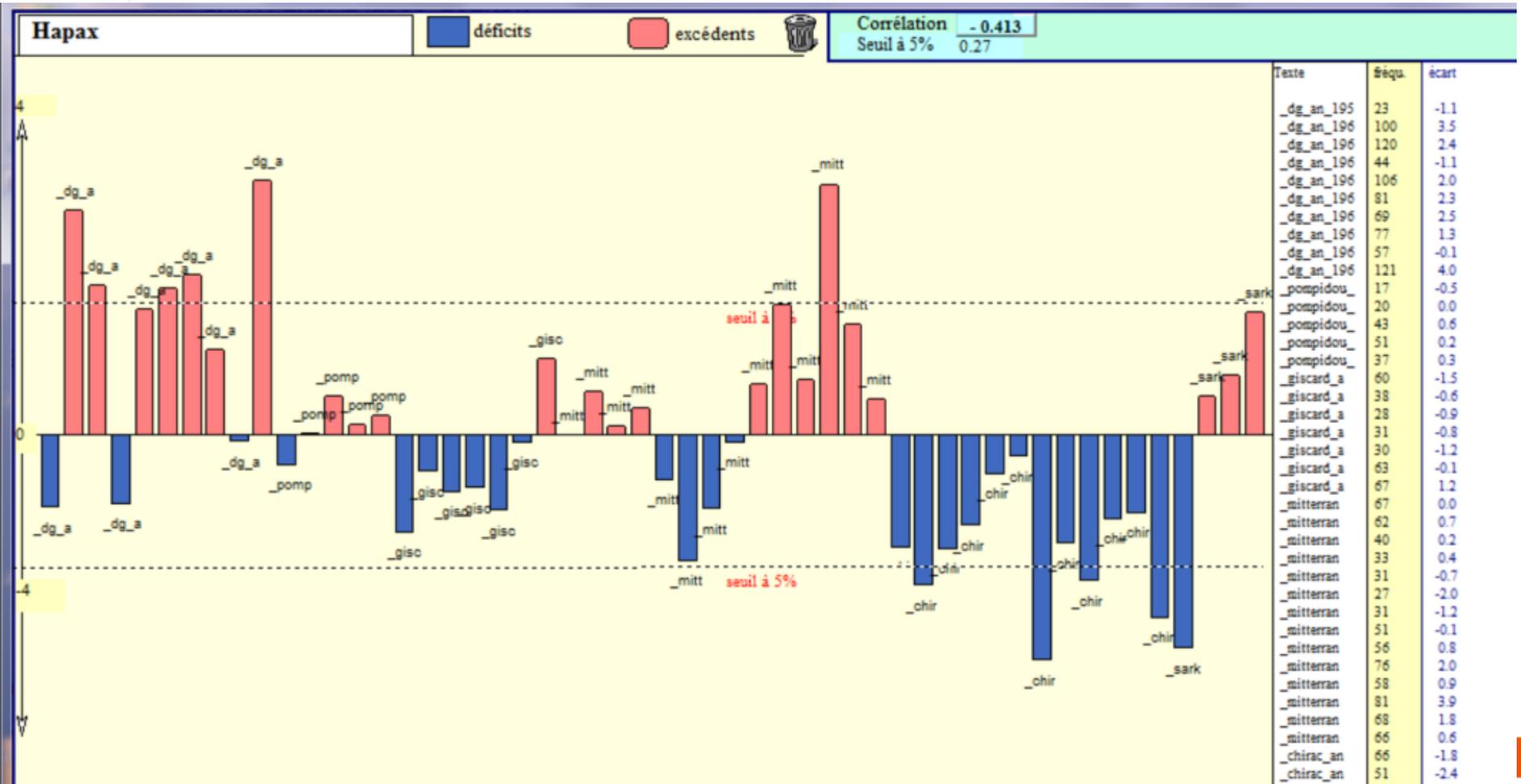
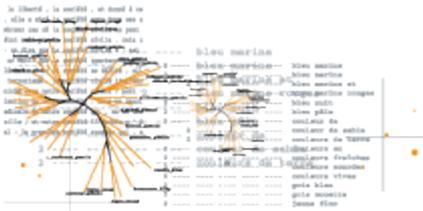


Texte	#éq.	écart
_dg_an_195	265	-1.2
_dg_an_196	539	0.0
_dg_an_196	680	0.0
_dg_an_196	420	-0.4
_dg_an_196	651	0.8
_dg_an_196	490	-0.2
_dg_an_196	411	-0.7
_dg_an_196	517	0.1
_dg_an_196	455	-0.5
_dg_an_196	627	0.5
_pompidou	188	-1.8
_pompidou	194	-1.8
_pompidou	327	-1.3
_pompidou	374	-2.1
_pompidou	301	-1.3
_giscard_a	437	-4.9
_giscard_a	316	-2.6
_giscard_a	258	-3.0
_giscard_a	239	-5.0
_giscard_a	308	-1.6
_giscard_a	433	-3.0
_giscard_a	383	-3.8
_mitterran	510	-0.4
_mitterran	464	0.2
_mitterran	340	-0.5
_mitterran	270	-1.5
_mitterran	323	-0.1
_mitterran	313	-2.1
_mitterran	343	-0.1
_mitterran	439	0.5
_mitterran	427	0.3
_mitterran	472	-0.3
_mitterran	418	-0.5
_mitterran	476	2.0
_mitterran	459	0.5
_mitterran	490	0.1
_chirac_an	549	-2.1
_chirac_an	505	-1.4
_chirac_an	407	-2.3
_chirac_an	459	-1.4
_chirac_an	515	-1.1
_chirac_an	563	-1.3
_chirac_an	514	-2.9

menu retour Noms significatifs

Voir les contextes: CLIC sur un "bâton" du graphique

Préoccupations stylistiques ou stylométriques





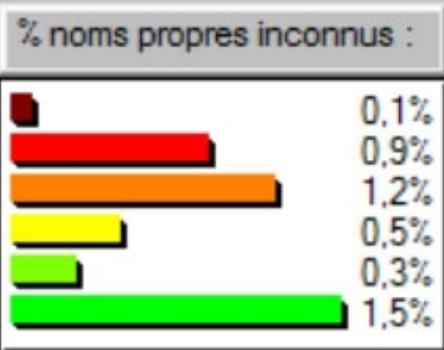
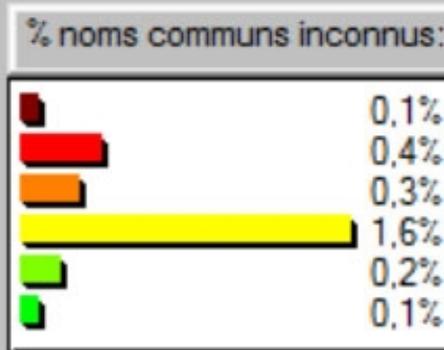
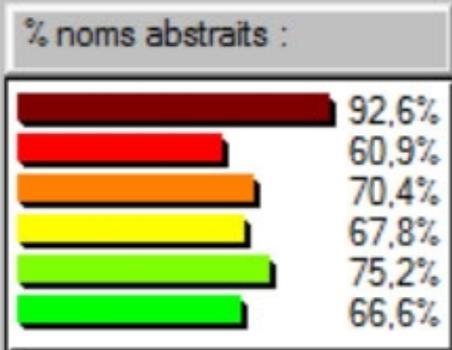
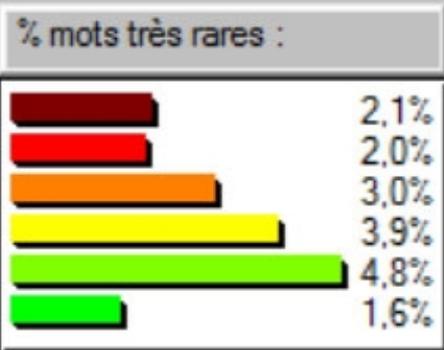
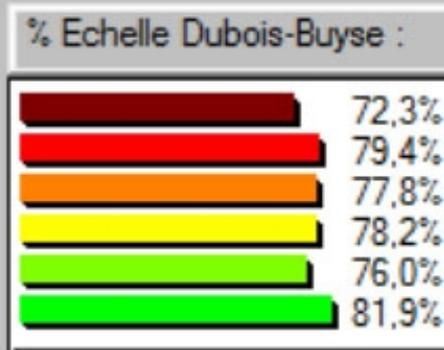
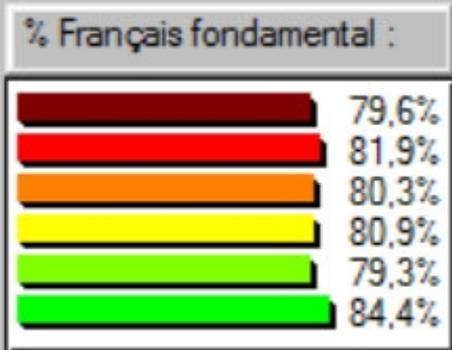
Statistiques | Types grammaticaux | **Types lexicaux** | Thèmes et domaines | Sémantique | Stylistique

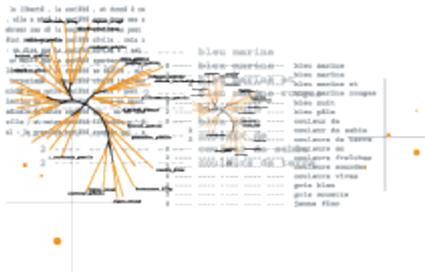
Vocabulaire :

- Vocabulaire de base : 72,8%
- Vocabulaire usuel : 23,9%
- Vocabulaire rare : 3,3%
- Mots inconnus : 0,1%
- % ambigus grammaticalement : 62,3%

Noms :

- % noms communs : 89,2%
- % noms propres : 10,8%
- % noms composés : 0,3%
- % noms ambigus sémantiquement : 52,3%





Tropes: Episodes et rafales

0

Pronom "Je"	troie	Pronom "Je"	destin	seigneur	Pronom "Vous"	cléone	fil	fil	nom	Modalisation	Négation	Pronom "Je"	lieu de culte	Pronom "Nous"			
pylade	pylade	Pronom "Vous"	Verbe Performatif	discours	amour	Modalisation Négation	chagrin	haine	ai	Modalisation	Temp	Pronom "Je"					
Modalisation	Temp	arizona	hermione	Connecteur	Modalisation	Temp	Pronom "Je"	Pronom "Tu"	Pronom "Je"	Pronom "Tu"	madame	Pronom "Nous"	oeur	moment	pylade		
Pronom "Je"	Temp	pyrrhus	ingratitude	oeil	fil	désespoir	souffrance	Pronom "Je"	oreste	Connecteur	Condition	Pronom "Je"	Pronom "Tu"	Pronom "Je"	dieu		
Pronom "Vous"	ector	seigneur	Pronom "Vous"	oreste	Modalisation	Négation	Pronom "Je"	Modalisation	Pronom "Nous"	Modalisation	Négation	Verbe Performatif	amour	Pronom "Je"	de		
Pronom "Je"	amant	fil	Pronom "Je"	Pronom "Vous"	Cliquez moi !	hermione	devoir	sort	époux	croyance	troyenne	chagrin	grèce	Pronom "Vous"			
Pronom "Tu"	grèce	andromaque	Connecteur	Connecteur	Disj	Modalisation	Négation	Pronom "Nous"	modal	Connecteur	Compar	Pronom "Je"	Modalisation	Négation	oeur	autel	
Connecteur	Temp	Pronom "Vous"	cléone	lieu	Pronom "Vous"	Modalisation	Négation	Disjonct	pyrrhus	Pronom "Je"	Pronom "Vous"	ector	Pronom "Je"				
Pronom "Je"	Pronom "Je"	Pronom "Je"	ennui	calme	Pronom "Je"	amitié	Pronom "Nous"	Connecteur	Compar	crime	grèce	Modalisation	Verbe Performatif				
Modalisation	Lieu	hector	Pronom "Vous"	Modalisation	Négation	force	Modalisation	Négation	Pronom "Vous"	céphise	hector	Modalisation	Lieu	promesse	cléone	Pronom "Vous"	
Modalisation	Temp	enfant	Modalisation	Négation	pourvoir	Verbe Performatif	but	Pronom "Nous"	Modalisation	Manière	mort	reste	Pronom "Je"	trahison	oreste	coup	
Pronom "Je"	jour	chagrin	hermione	pyrrhus	ari	Connecteur	Opposit	pyrrhus	Modalisation	Négation	Pronom "Je"	Pronom "Vous"	Modalisation	Doute	Modalisation	Négation	
séduction	Modalisation	Négation	Pronom "Vous"	Modalisation	Lieu	Modalisation	Temp	ingratitude	Pronom "Vous"	Modalisation	Lieu	madame	Pronom "Vous"	Pronom "Je"	Pronom "Vous"	mariage	Pronom "Tu"
Pronom "Je"	sang	Pronom "Tu"	Modalisation	Temp	raisonnement	pylade	esprit	Pronom "Je"	Pronom "Vous"	cléone	cléone	insulte	dentition	Pronom "Je"			

14628

Exemple: Andromaque



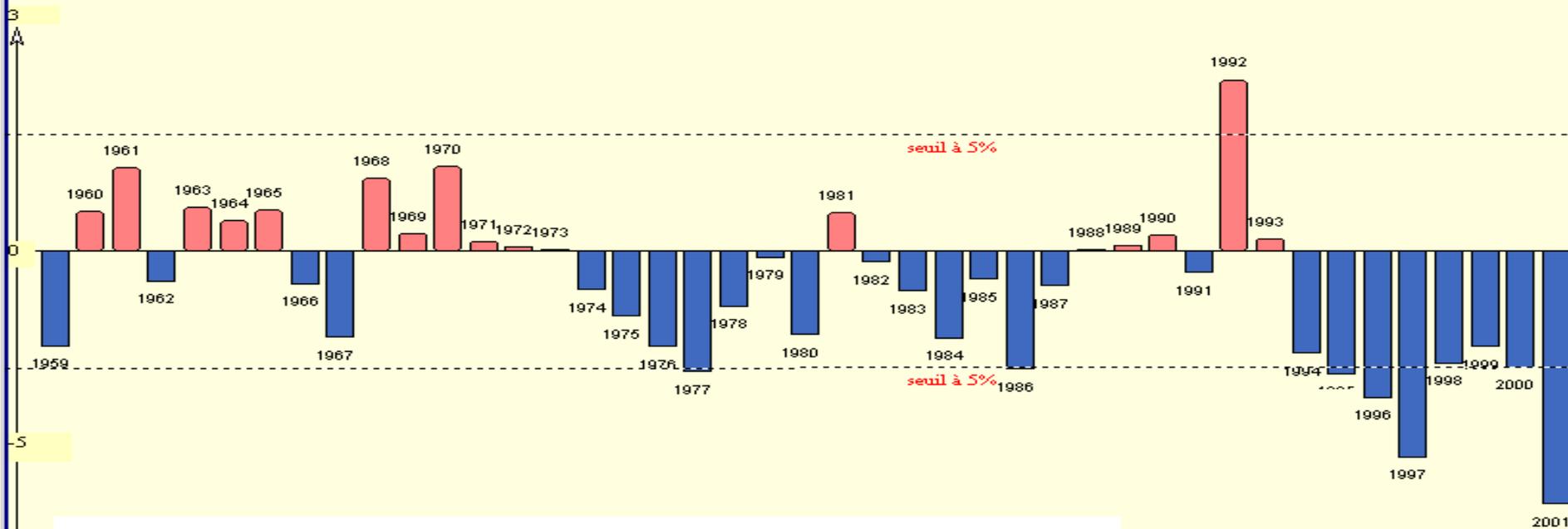
Accroissement lexical

■ déficits

■ excédents

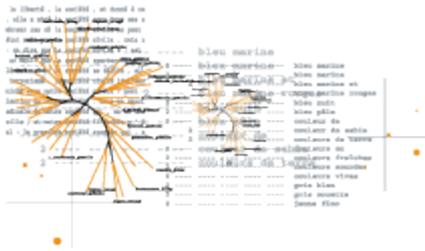


Corrélation **- 0.484**
Seuil à 5% 0.30

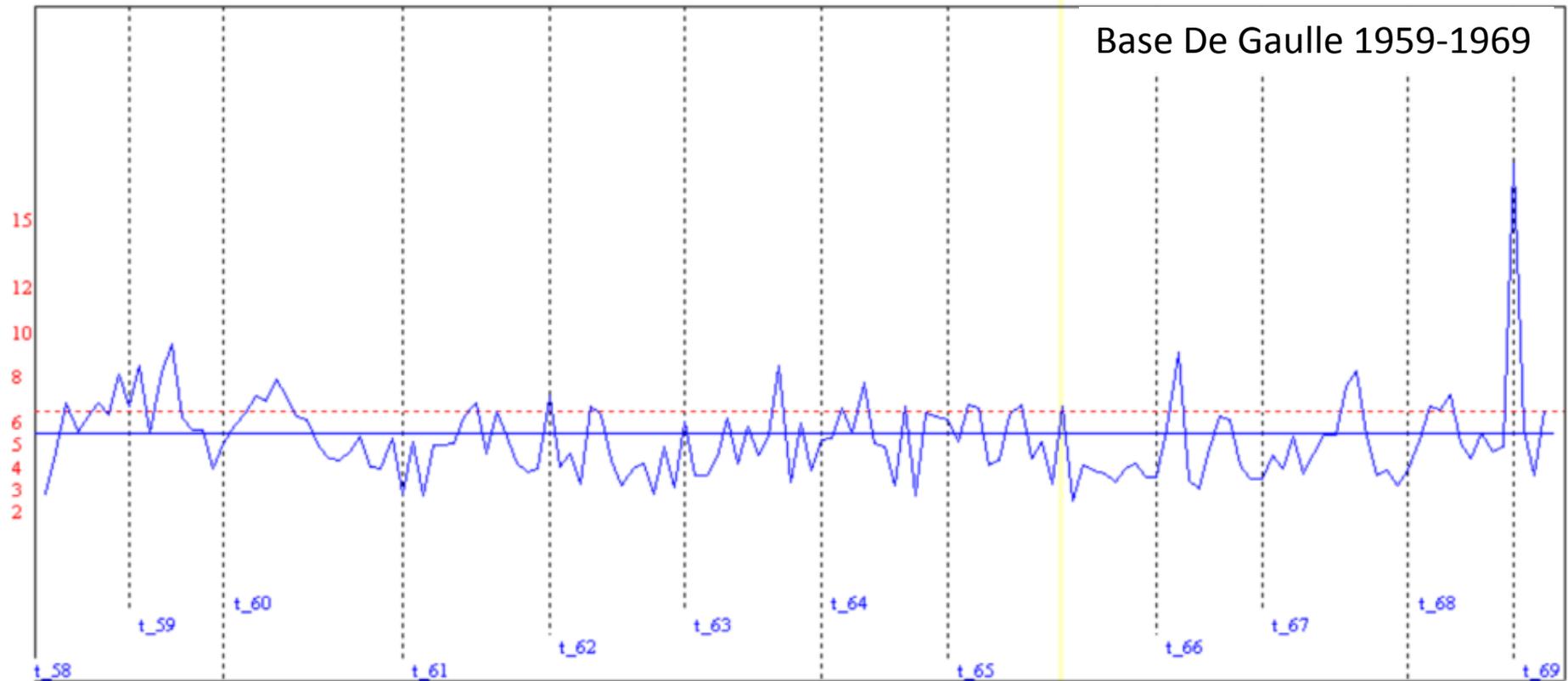


Accroissement lexical partition années- vœux 1959-2001

Accroissement lexical par « tranches »



Nombre moyen de mots nouveaux par page = 5.99



Fonction $y=a(x \text{ exposant } b)$: $a=0.00003$ $b=2.00147$
 $r^2=0.98781$ $r=0.99389$

en rouge: courbe brute
en bleu: courbe pondérée

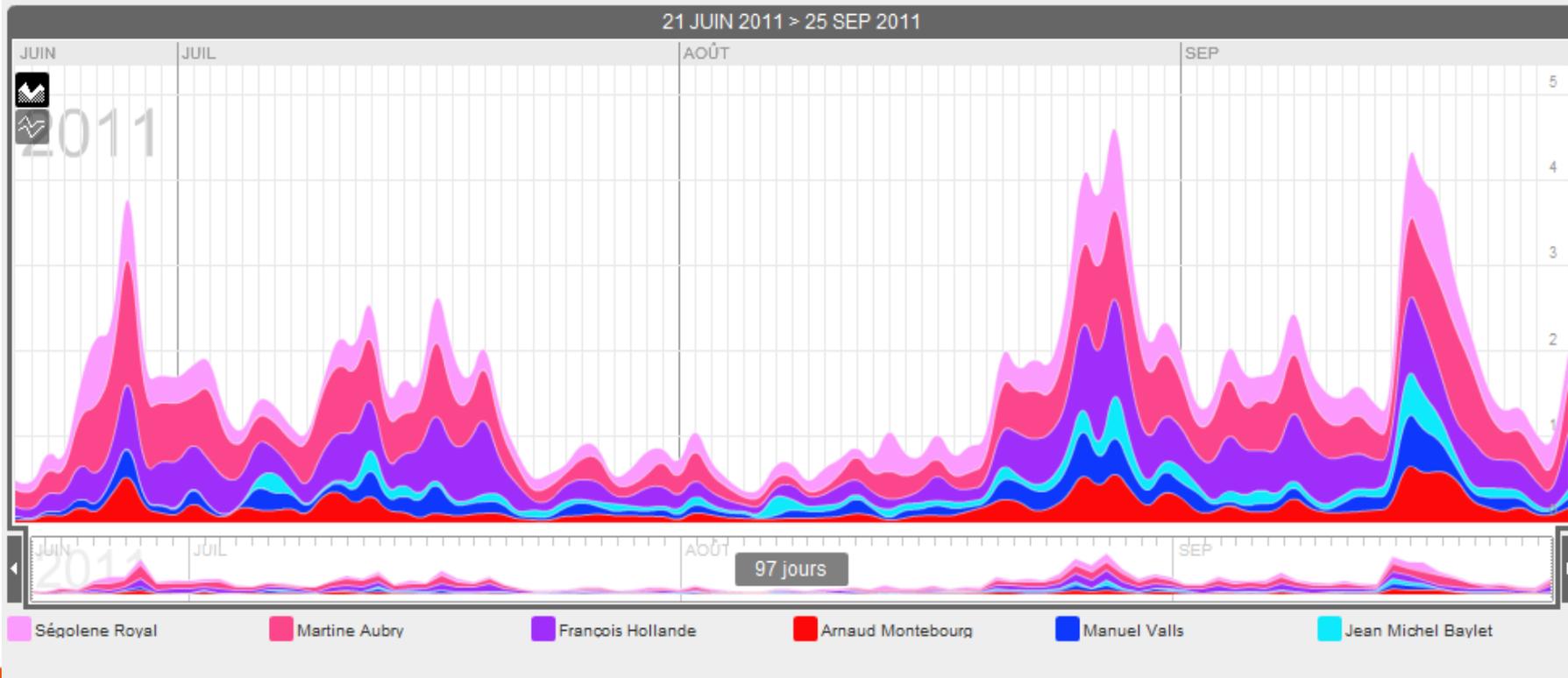
Buzz à la une

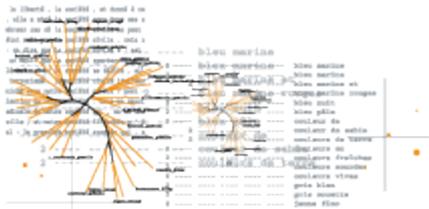
La Primaire du PS depuis le 21 juin

<http://fr.linkfluence.net/>



La présence des candidats dans les conversations sur le Net

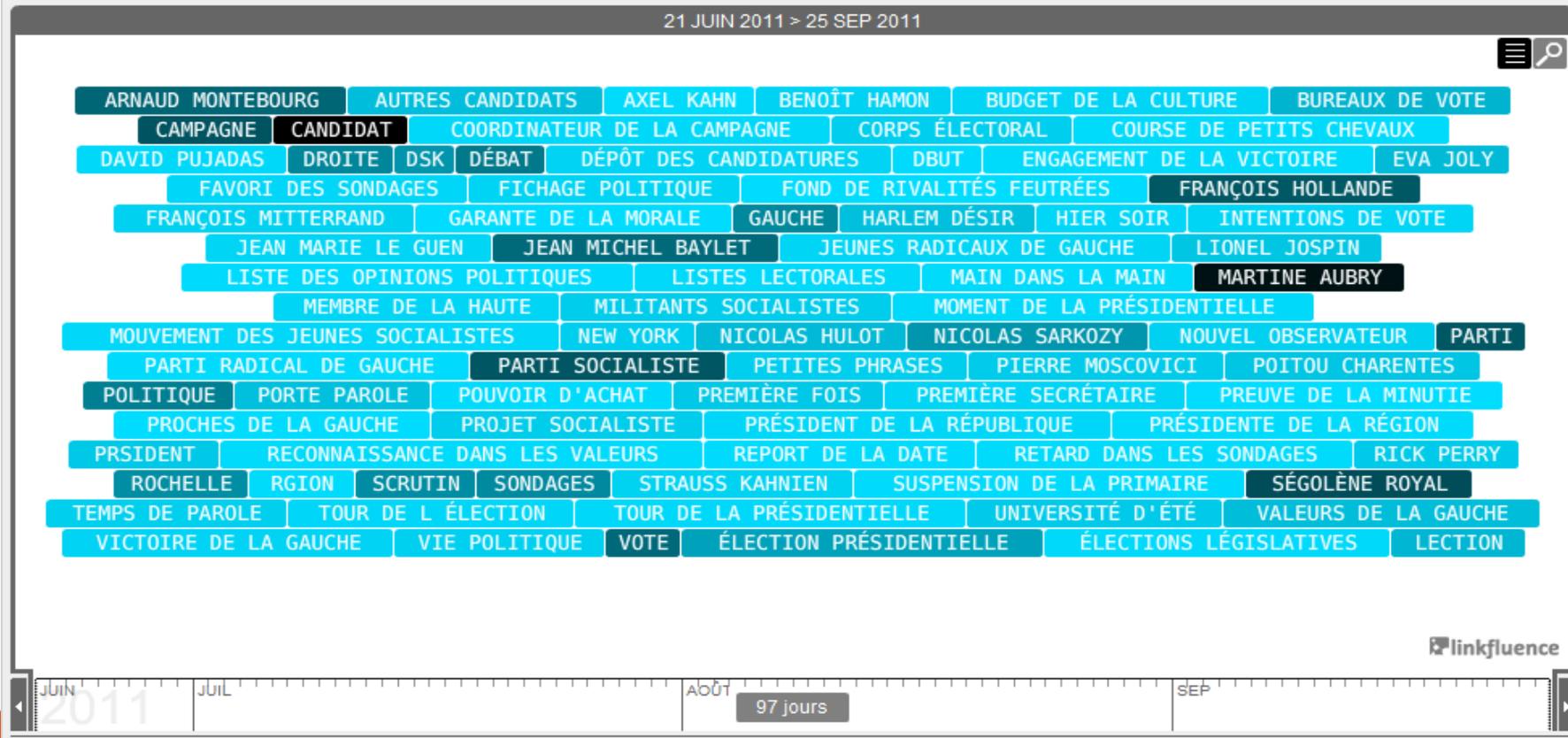




La Primaire du PS depuis le 21 juin

<http://fr.linkfluence.net/>

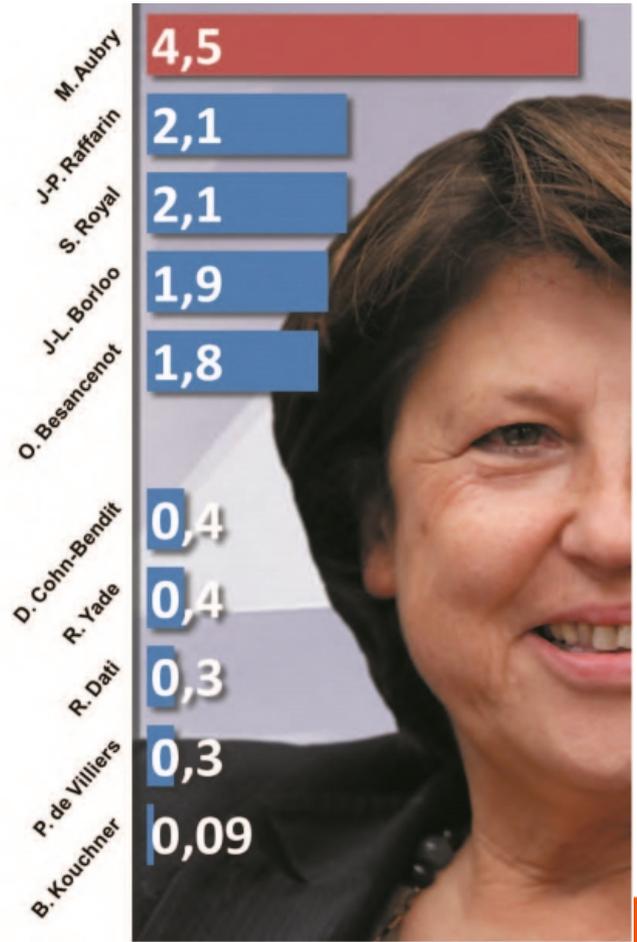
Les mots associés à la primaire





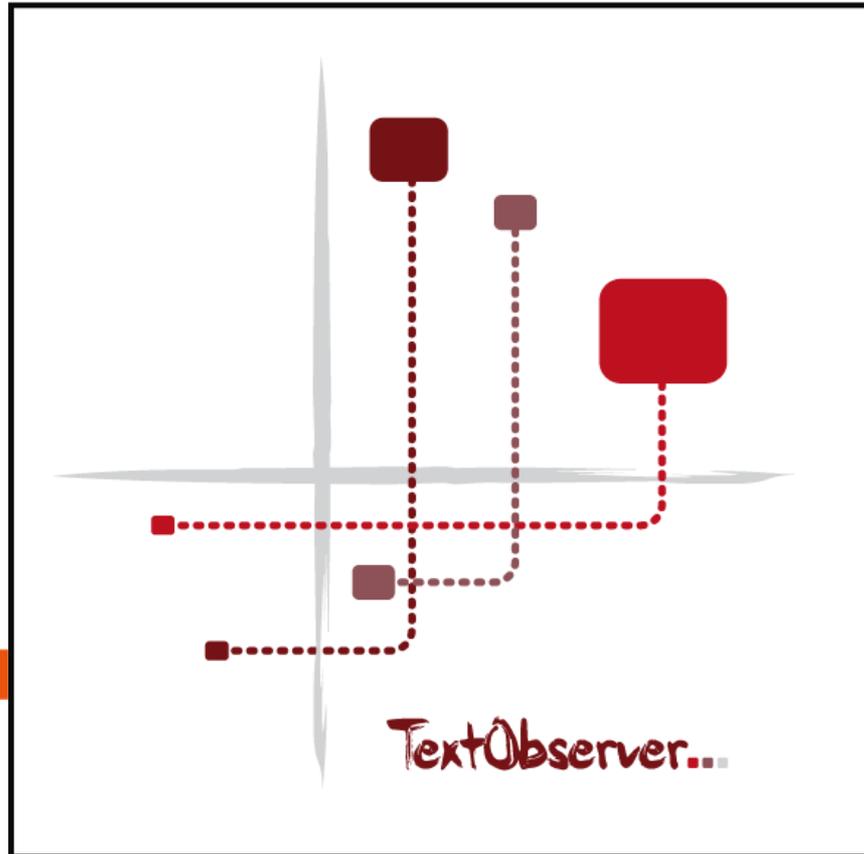
JE

SOCIAL





TEXTOBSERVER, UN OUTIL D'OBSERVATION ET D'EXPLORATION DES DONNÉES TEXTUELLES ET MULTIMODALES



TextObserver/WebObserver : réflexions sur la complémentarité des approches et des méthodes

Représenter des données textuelles

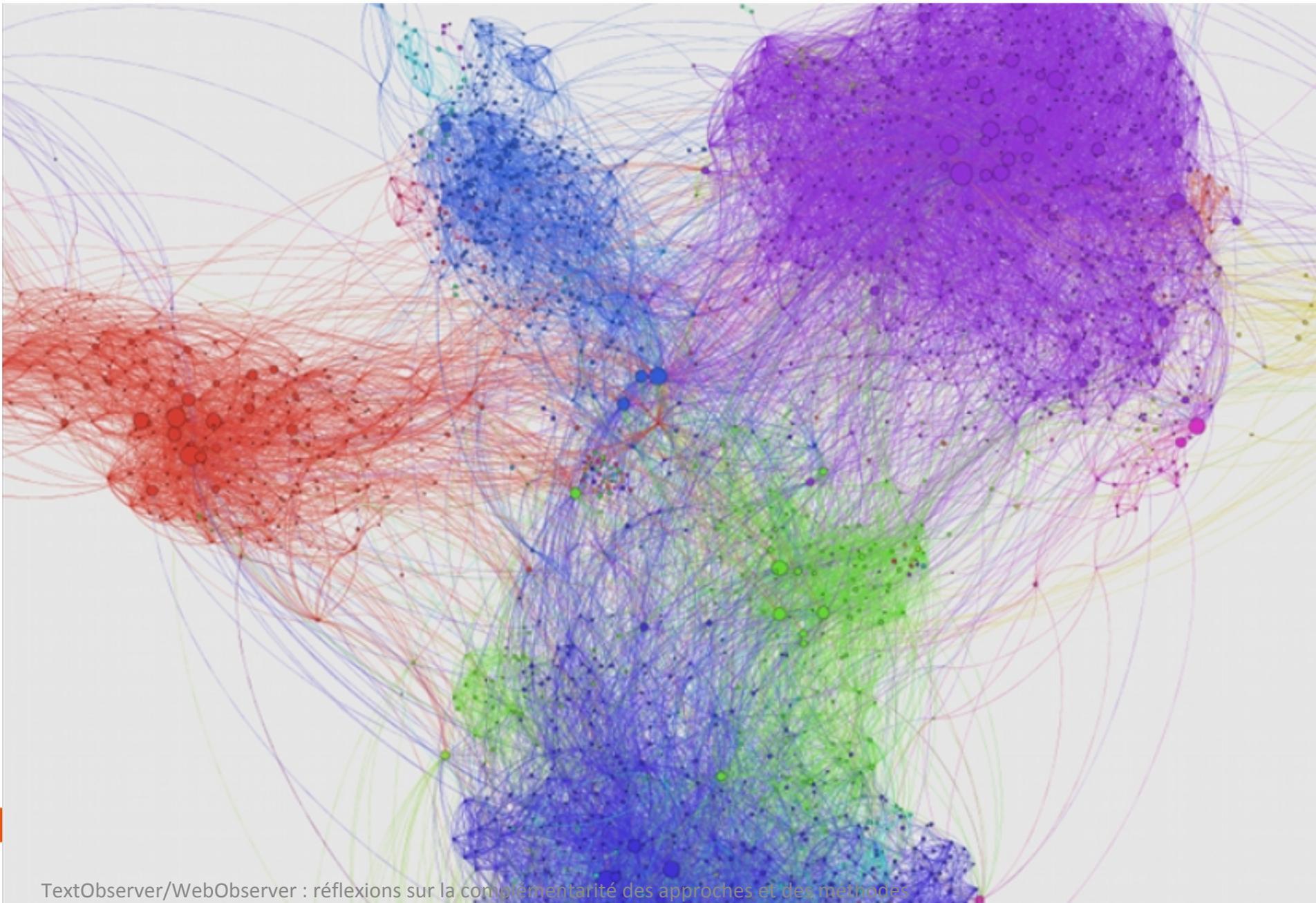


- Visualisation/interprétation
- Données, recueil, mesures...

Représenter des données textuelles



- Nouveaux outils, nouvelles visualisations, apprentissage, interprétation...
- « Belles images »



TextObserver/WebObserver : réflexions sur la complémentarité des approches et des méthodes



TEXTOSERVER :

POUR UNE APPROCHE ERGONOMIQUE DES DISPOSITIFS LEXICOMÉTRIQUES



Démarche

- Equipe Textopol du Céditec
- Pôle d'expertise lexicométrique/textométrique
- Recherche et formation à la recherche
- Démarche expérimentale, comparaison des outils



Le dispositif textométrique

*Lexico 3, Hyperbase, Astartex, TXM, Iramuteq
Weblex, Alceste*

Sphinx

Tropes, Cordial, treetagger

TextObserver/webObserver

XL stat, coocs, Edxml...



- Pourquoi développer un nouvel outil logiciel?
 - Fonctionnalités originales
 - Répondre à des questions de recherche
 - Textométrie/Multimédia/Infographie



Une certaine conception de la lexicométrie

- Privilégiant le travail sur la forme graphique sans recours systématique à la lemmatisation ou catégorisation
- Retour au texte,
- Des tendances au « micro-textuel »
- Validation de la démarche sur des petits corpus : Big Datas / small corpus
- Inscription dans un genre
- Comparer ce qui est comparable
- Une conception « large » de la lexicométrie
- Expérimentation, utiliser les résultats d'un outils pour les soumettre aux autres...



Une certaine conception de la lexicométrie

- Lexicométrie/Textométrie: Visualisations, cartographies, topographies textuelles, résonnance...
- Logométrie (Mayaffre)



Une certaine conception de la lexicométrie

- TAL/Lexico-Textométrie:
- Fonctionnalités approchantes mais objectifs différents: attribution d'auteur, analyse ou repérage des genres, résumés automatiques ou paragraphes spécifiques



Lexicométrie/textométrie

- Evolution des pratiques et des interprétations
 - Tournier, projection sur l'axe 1
 - TXM, TextObserver: prise en considération des points contributifs des axes
 - Généralisation de la visualisation Tridimensionnelle (TextObserver, Iramuteq, Astartex)
 - Méthodes de validation : ellipses de confiance...

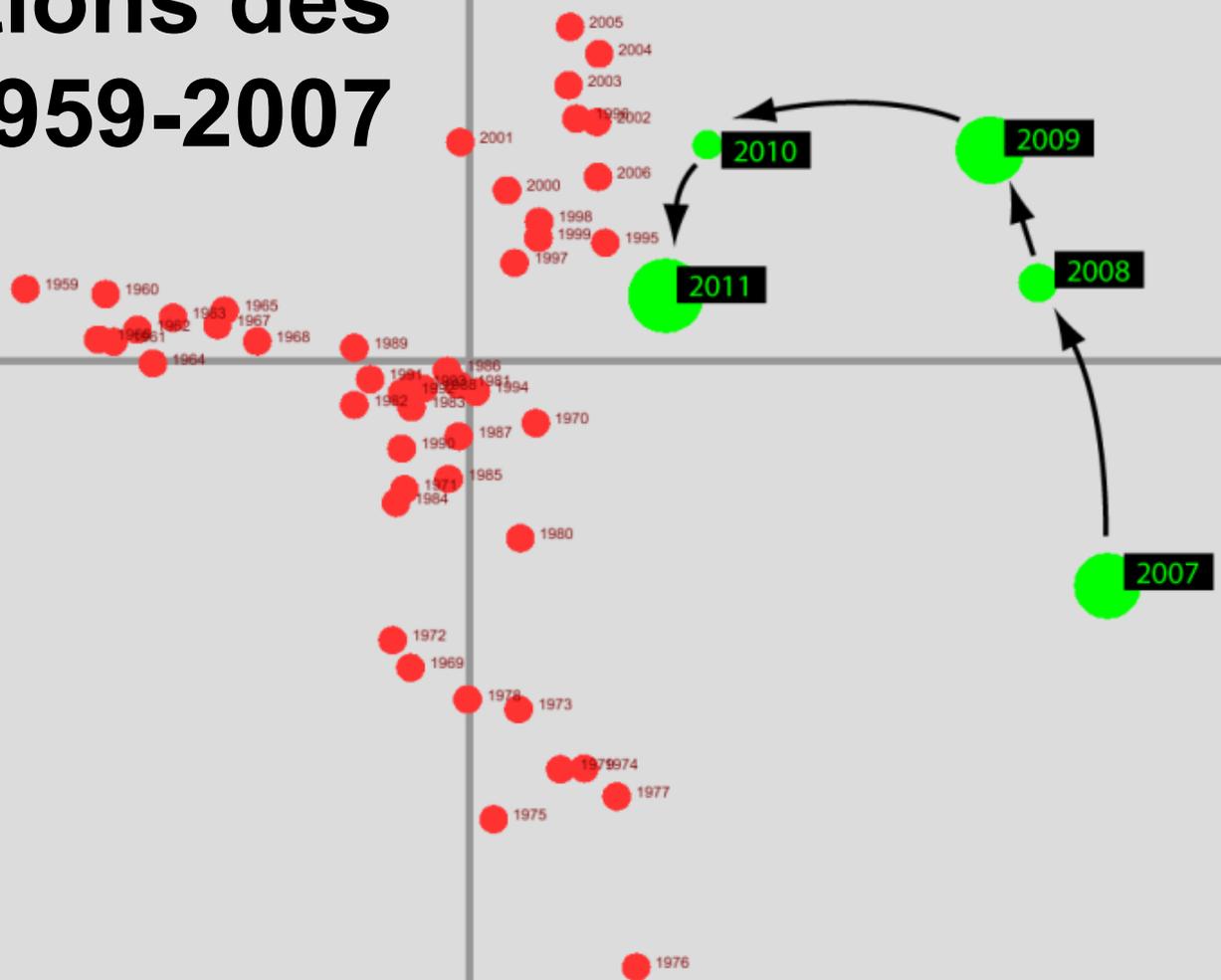


TextObserver

- Questions de recherche
 - Vœux
 - Énonciation
 - Évolutions du corpus vœux
 - Phénomènes liés à la variation
- Projet interdisciplinaire
 - Ergonomie
 - Multimedia
 - IHM



Evolutions des vœux: 1959-2007





TextObserver

- JADT 2010: Pour une approche ergonomiques des dispositifs lexicométriques
 - Interfaces, ergonomie
 - Trois propositions
- L'interactivité comme vecteur cognitif majeur
 - Mouvement ou 3D
 - Implémenter le multimédia dans les outils textométriques
 - Navigation, Zoom



Propositions JADT 2010

- Représenter l'analyse factorielle des correspondances sur trois axes (lorsque cela faisait sens)
- Implémenter le mouvement dans les représentations des résultats => mieux saisir les phénomènes de variations.
- Proposer une navigation entre les différentes partitions d'un corpus.

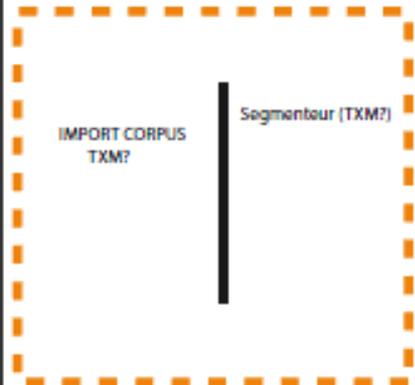
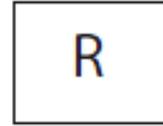


Tableau lexical

Modifiable, tris, seuils, suppression ou ajout de lignes et de colonnes

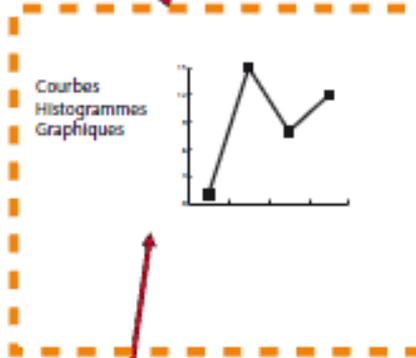
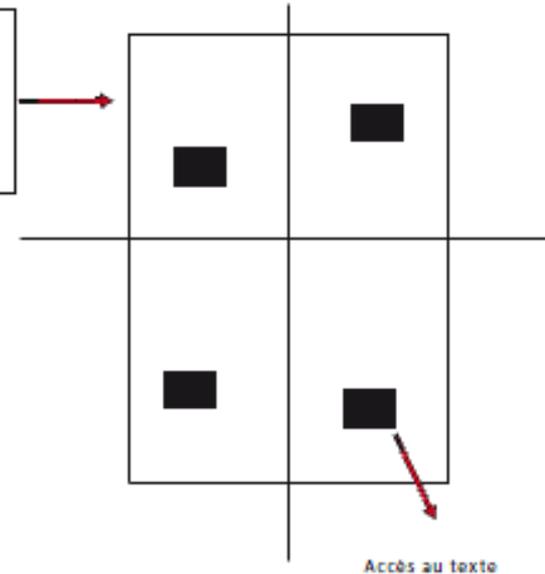
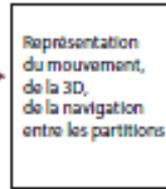
Tableau bloqué sur vœux pour l'instant

Effet immédiat sur les AFC et plus tard sur d'autres représentations



Calcul AFC

Moteur 3D



UTILISATEUR

Action sur le tableau de départ

TextObserver/WebObserver

Action sur les AFC qu'il peut multiplier, superposer, comparer, modifier.

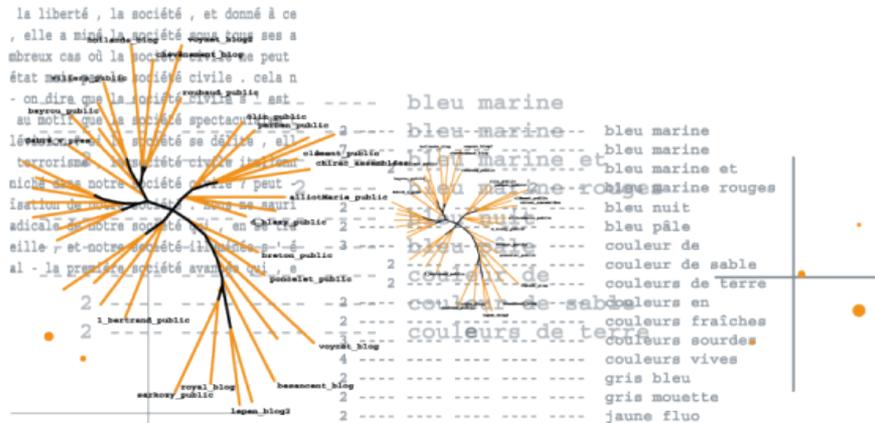
A partir de l'AFC il peut aussi afficher le texte qui correspond à la partie sélectionnée

TextObserver/WebObserver : complémentarité des approches et des méthodes



TextObserver

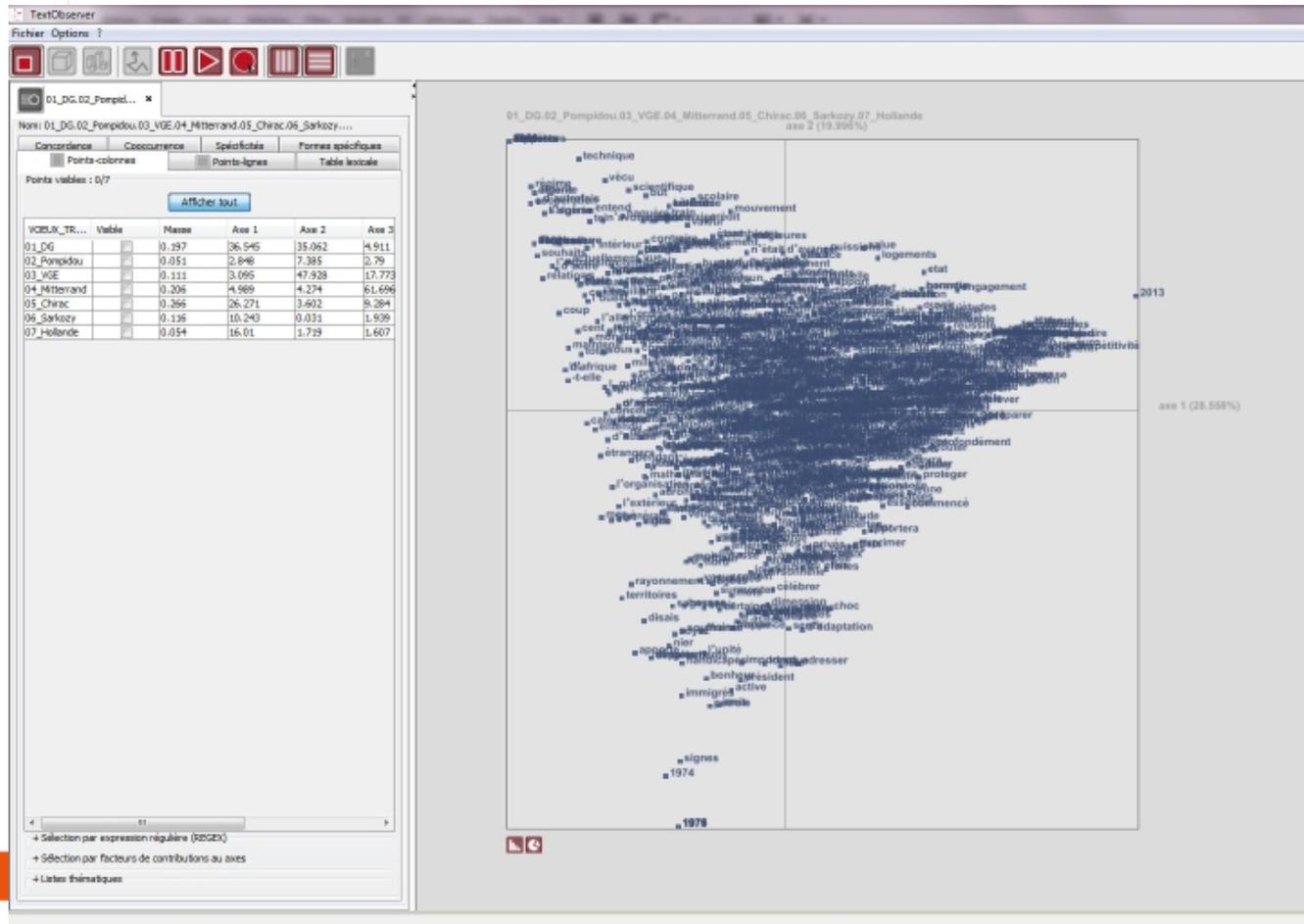
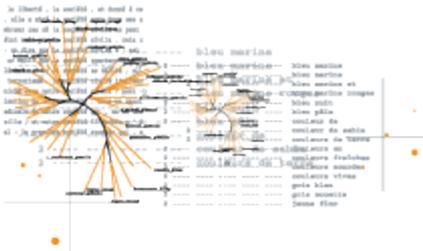
- Expérimenter les phénomènes de variation
- L'interactivité comme vecteur cognitif majeur
 - Mouvement ou 3D, interactivité
- Fonctionnalités de visualisation (pt ligne, pt colonnes, pt contributifs, marquages)
- Dupliquer, déplacer, comparer les AFC



Une aide à l'interprétation

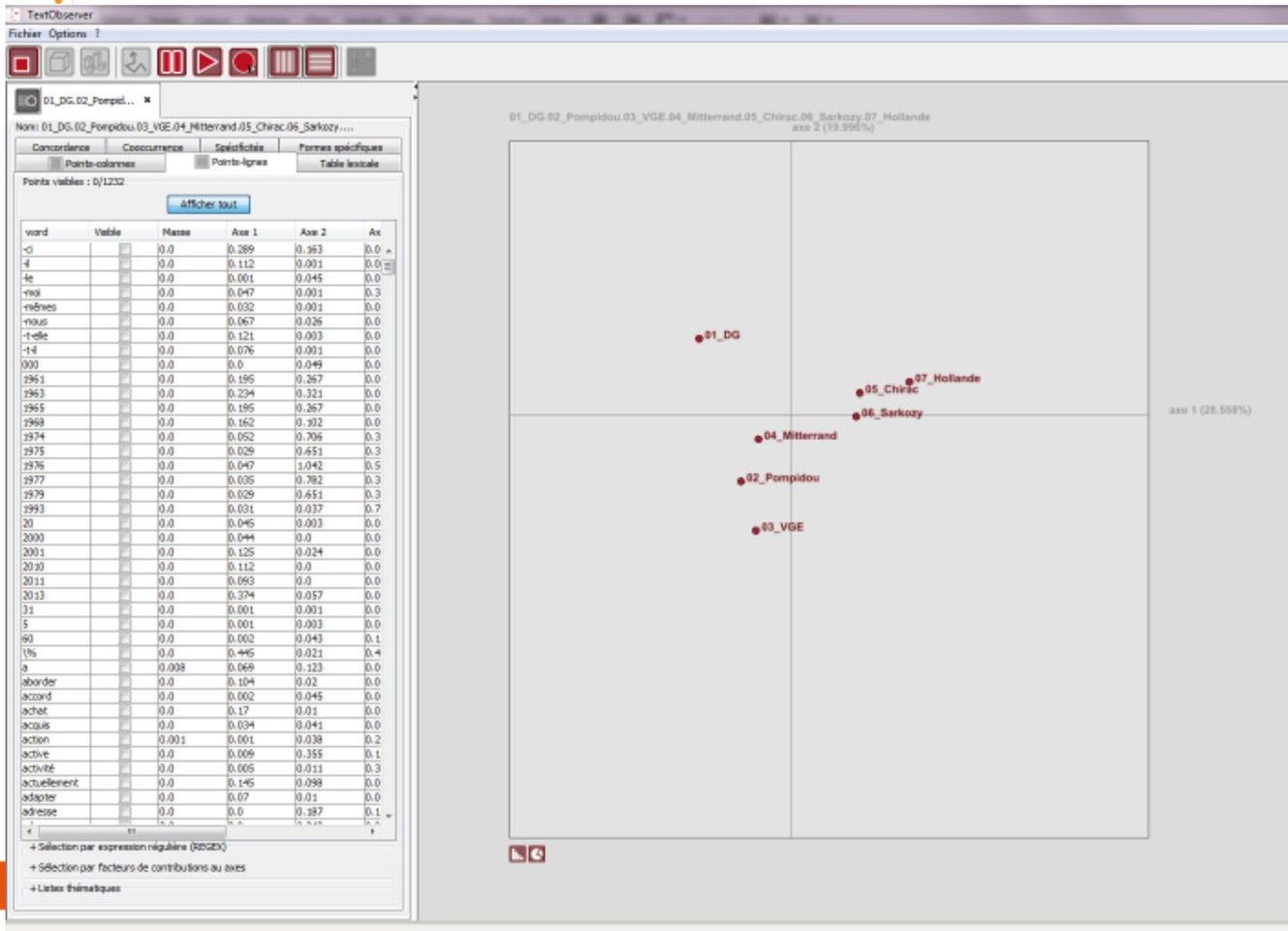
FONCTIONNALITÉS DE LECTURE/AFFICHAGE

Points lignes



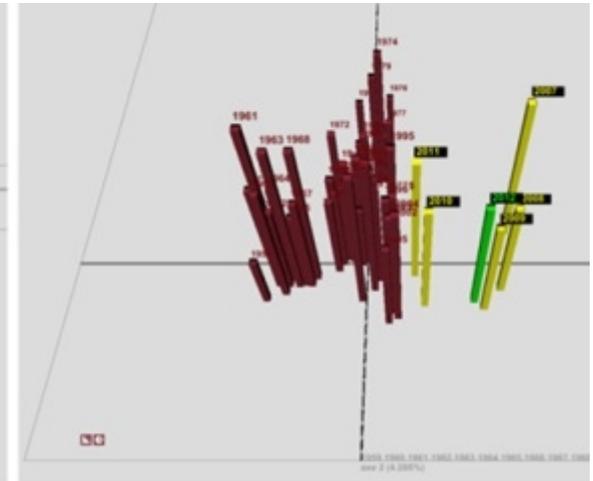
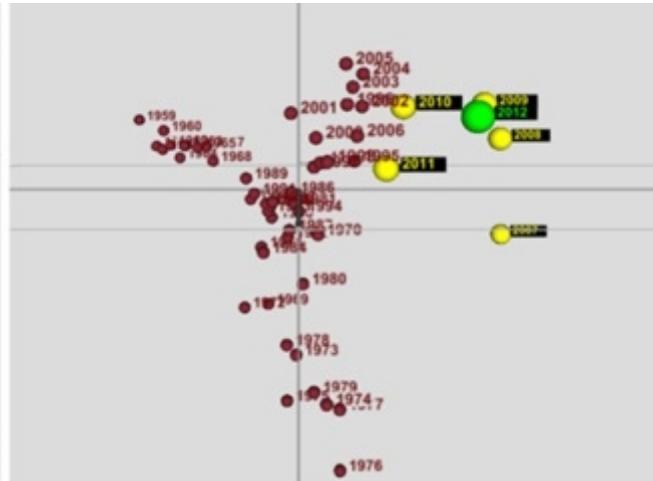
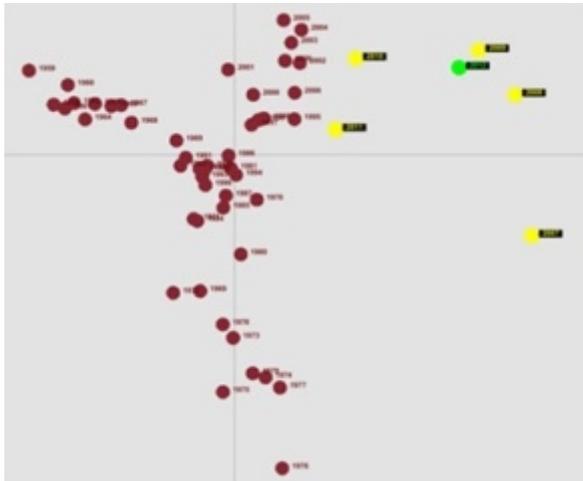


Points colonnes





Affichage des points colonnes, en 2D; 3D et version extrudée





Andromaque0.Bovary...
Nom: Andromaque0.Bovary.Editos.Leoel.Voeux

Points-colonnes | Points-lignes | Table lexicale

Points visibles : 30/30

Masquer tout

Forme	Visible	Masse	Axe 1	Axe 2	Ax
elle	<input checked="" type="checkbox"/>	0.136	12.259	5.551	0.9
elles	<input checked="" type="checkbox"/>	0.0060	0.03	0.617	0.1
il	<input checked="" type="checkbox"/>	0.192	9.397	0.549	0.3
ils	<input checked="" type="checkbox"/>	0.024	0.935	0.052	0.1
je	<input checked="" type="checkbox"/>	0.09	6.313	7.942	8.3
leur	<input checked="" type="checkbox"/>	0.022	0.545	0.16	0.0
leurs	<input checked="" type="checkbox"/>	0.013	0.015	0.183	0.8
men	<input checked="" type="checkbox"/>	0.0	0.013	0.22	0.0
mienne	<input checked="" type="checkbox"/>	0.0	0.028	0.024	0.1
miens	<input checked="" type="checkbox"/>	0.0010	0.06	0.623	0.4
moi	<input checked="" type="checkbox"/>	0.02	0.017	4.919	0.4
mon	<input checked="" type="checkbox"/>	0.027	1.019	23.513	5.4
nos	<input checked="" type="checkbox"/>	0.018	8.51	2.475	0.1
notre	<input checked="" type="checkbox"/>	0.027	17.939	7.91	5.3
nous	<input checked="" type="checkbox"/>	0.059	27.679	30.729	3.9
sa	<input checked="" type="checkbox"/>	0.066	2.482	0.056	0.0
se	<input checked="" type="checkbox"/>	0.096	5.117	3.504	0.8
ses	<input checked="" type="checkbox"/>	0.047	1.285	0.777	0.9
son	<input checked="" type="checkbox"/>	0.0	0.0030	0.0040	0.0
sienna	<input checked="" type="checkbox"/>	0.0010	0.073	0.0020	0.0
siens	<input checked="" type="checkbox"/>	0.0010	0.0020	0.058	2.5
ta	<input checked="" type="checkbox"/>	0.0060	0.04	6.72	13.
te	<input checked="" type="checkbox"/>	0.0060	0.0050	3.917	8.4
tes	<input checked="" type="checkbox"/>	0.0020	0.067	1.543	0.0
toi	<input checked="" type="checkbox"/>	0.0050	0.026	1.605	0.2
ton	<input checked="" type="checkbox"/>	0.0060	0.0010	5.271	16.
tu	<input checked="" type="checkbox"/>	0.021	0.236	5.409	1.9

- Sélection par expression régulière (REGEX)

Je
je
..

Afficher Masquer

Marquer : [palette de couleurs]

Supprimer le marquage Supprimer tous les marquages

+ Sélection par facteurs de contributions aux axes

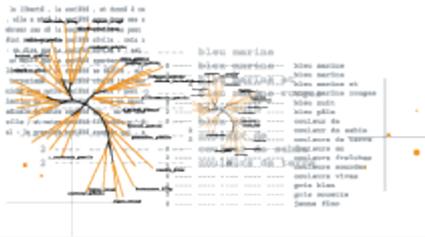
- Listes thématiques (pronomperso_ansi.it)

Leurs
leurs

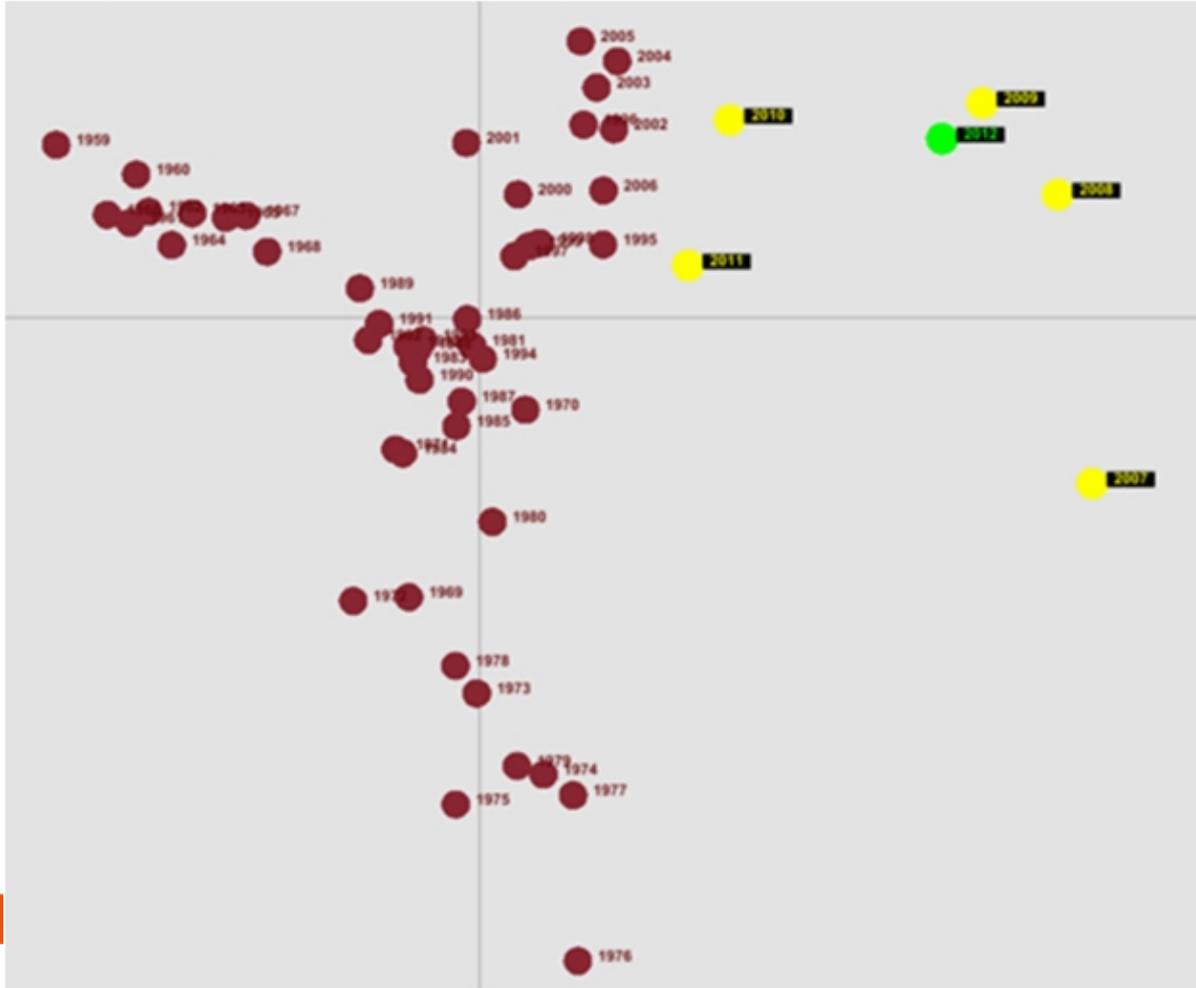
Importer Exporter

Copier vers REGEX Supprimer

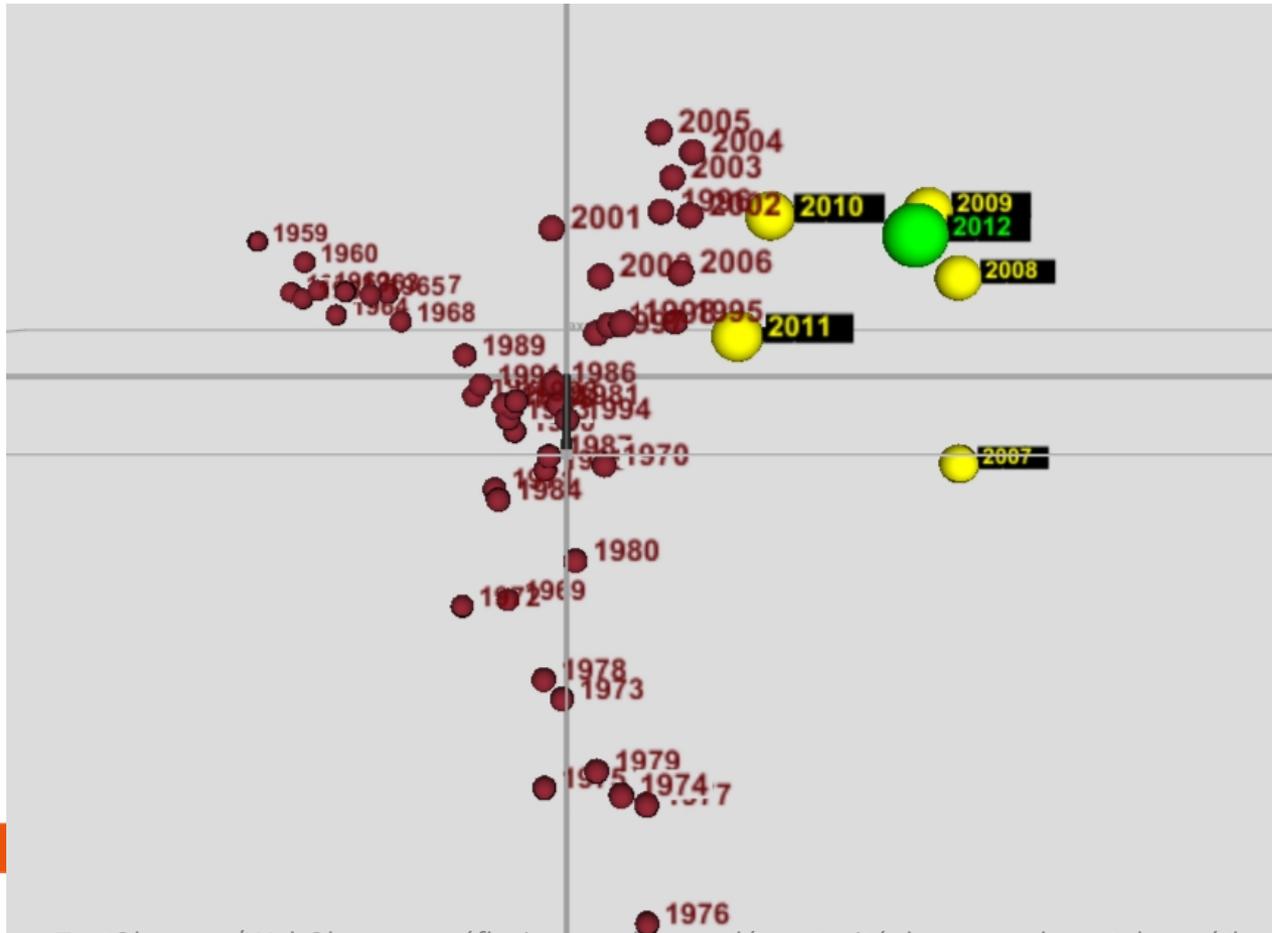




Mise en évidence par la coloration

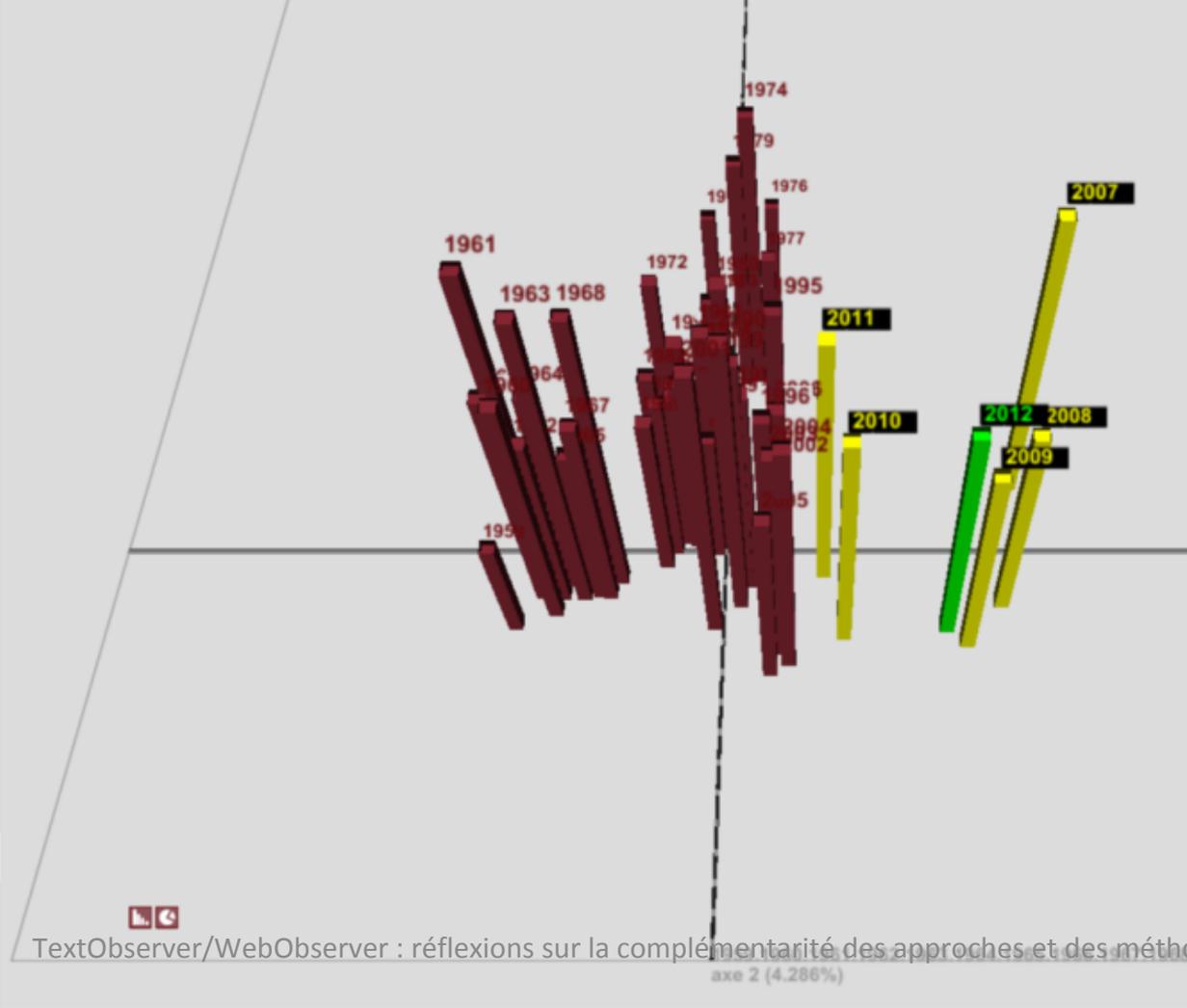
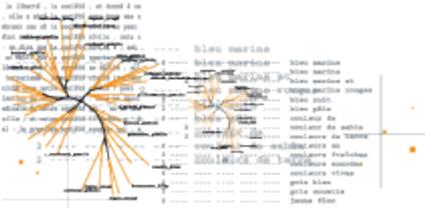


Fonctionnalités d'affichage



TextObserver/WebObserver : réflexions sur la complémentarité des approches et des méthodes

Fonctionnalités d'affichage



TextObserver/WebObserver : réflexions sur la complémentarité des approches et des méthodes

TextObserver

Fichier Options ?

01_DG.02_Pompidou... *

Nom: 01_DG.02_Pompidou.03_VGE.04_Mitterrand.05_Chirac.06_Sarkozy....

Concordance Cooccurrence Spécificités Formes spécifiques

Points-colonnes Points-lignes Table lexicale

Points visibles : 0/1232

Afficher tout

word	Visible	Masse	Axe 1	Axe 2	Axe 3
-d	<input type="checkbox"/>	0.0	0.289	0.163	0.0
-il	<input type="checkbox"/>	0.0	0.112	0.001	0.0
-le	<input type="checkbox"/>	0.0	0.001	0.045	0.0
-moi	<input type="checkbox"/>	0.0	0.047	0.001	0.3
-mêmes	<input type="checkbox"/>	0.0	0.032	0.001	0.0
-nous	<input type="checkbox"/>	0.0	0.067	0.026	0.0
-t-elle	<input type="checkbox"/>	0.0	0.121	0.003	0.0
-t-l	<input type="checkbox"/>	0.0	0.076	0.001	0.0
000	<input type="checkbox"/>	0.0	0.0	0.049	0.0
1961	<input type="checkbox"/>	0.0	0.195	0.267	0.0
1963	<input type="checkbox"/>	0.0	0.234	0.321	0.0
1965	<input type="checkbox"/>	0.0	0.195	0.267	0.0
1968	<input type="checkbox"/>	0.0	0.162	0.102	0.0
1974	<input type="checkbox"/>	0.0	0.052	0.706	0.3
1975	<input type="checkbox"/>	0.0	0.000	0.000	0.0

- Sélection par expression régulière (REGEX)

`.*isme|`

Afficher Masquer

Aide-mémoire REGEX

- `.` N'importe quel caractère
- `[]` Classe de caractères, ex : `[a-z]` - tous les caractères de "a" à "z"
`[abz]` - les caractères "a", "b" et "z"
- `[^]` Négation de classe de caractères, ex : `[^a-z]` - tous les caractères exceptés ceux de "a" à "z"
- `{}` Quantificateur, ex : `[a-z]{2,4}` - tous les caractères de "a" à "z" répétés de 2 à 4 fois
- `*` 0, 1 ou plusieurs occurrences du caractère ou de la classe précédente
- `+` 1 ou plusieurs occurrences du caractère ou de la classe précédente
- `?` 0 ou 1 occurrence du caractère ou de la classe précédente
- `|` OU logique, ex : `il|elle` - recherche les mots "il" ou "elle"

Exemples

- `.*` Recherche n'importe quel mot
- `a.*` Recherche n'importe quel mot commençant par "a"
- `.*ant` Recherche n'importe quel mot terminant par "ant"
- `[0-9]{4}` Recherche n'importe quel nombre de 4 chiffres

La recherche est insensible à la casse.

- Sélection par facteurs de contributions au axes

- Axe 1

0.0 3.54 7.081

- Axe 2

0.0 2.244 4.487

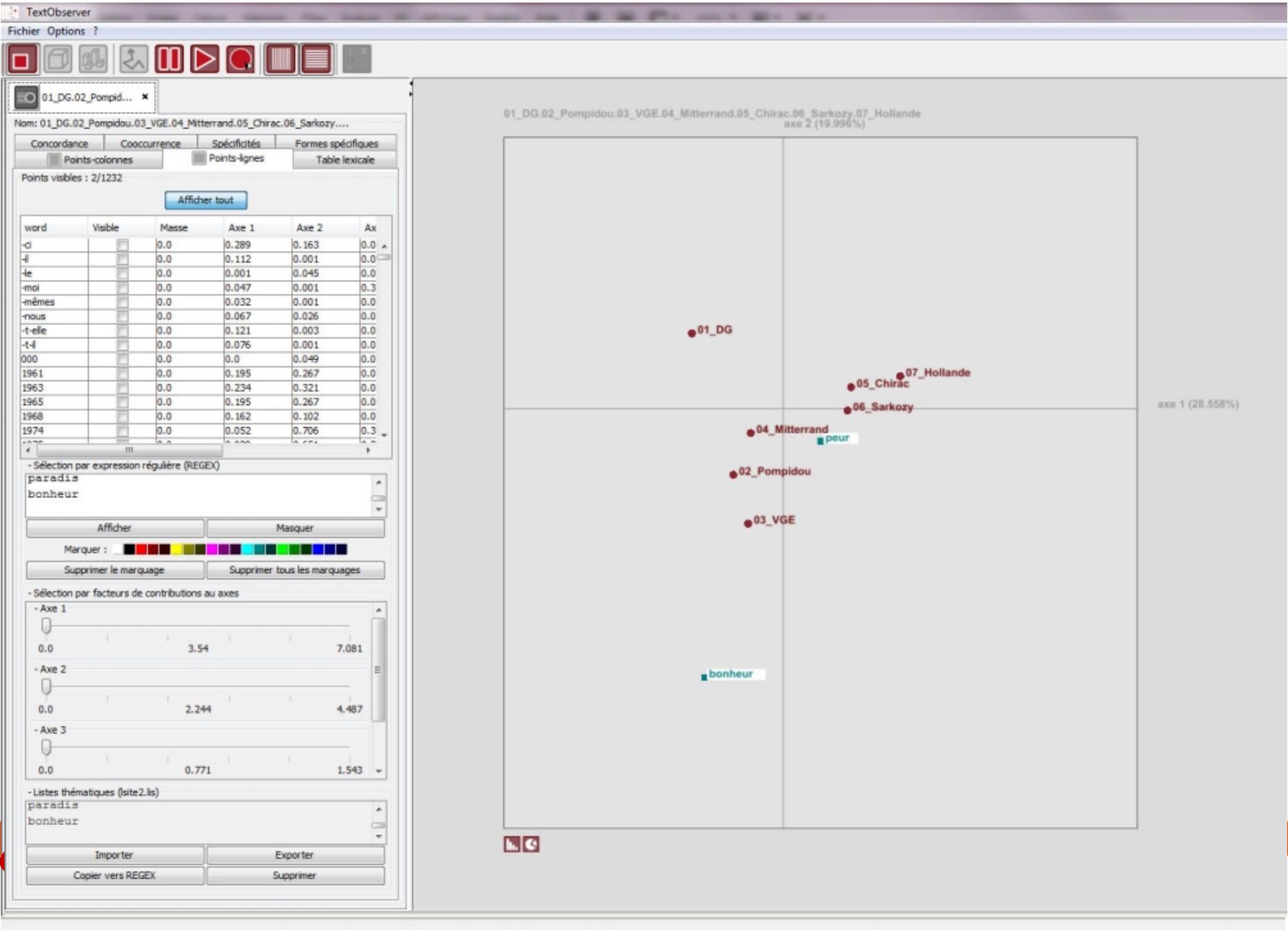
- Axe 3

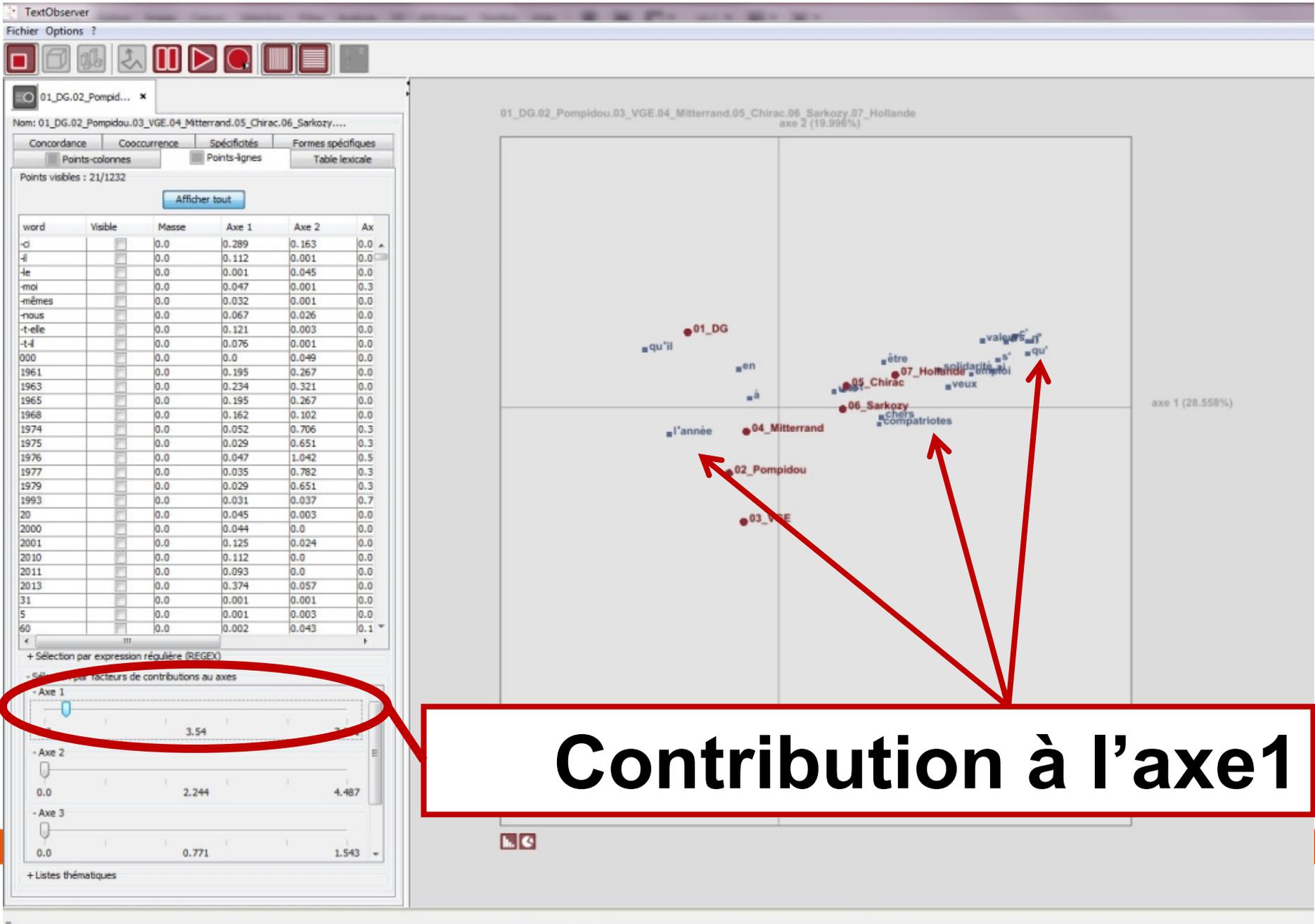
0.0 0.771 1.543

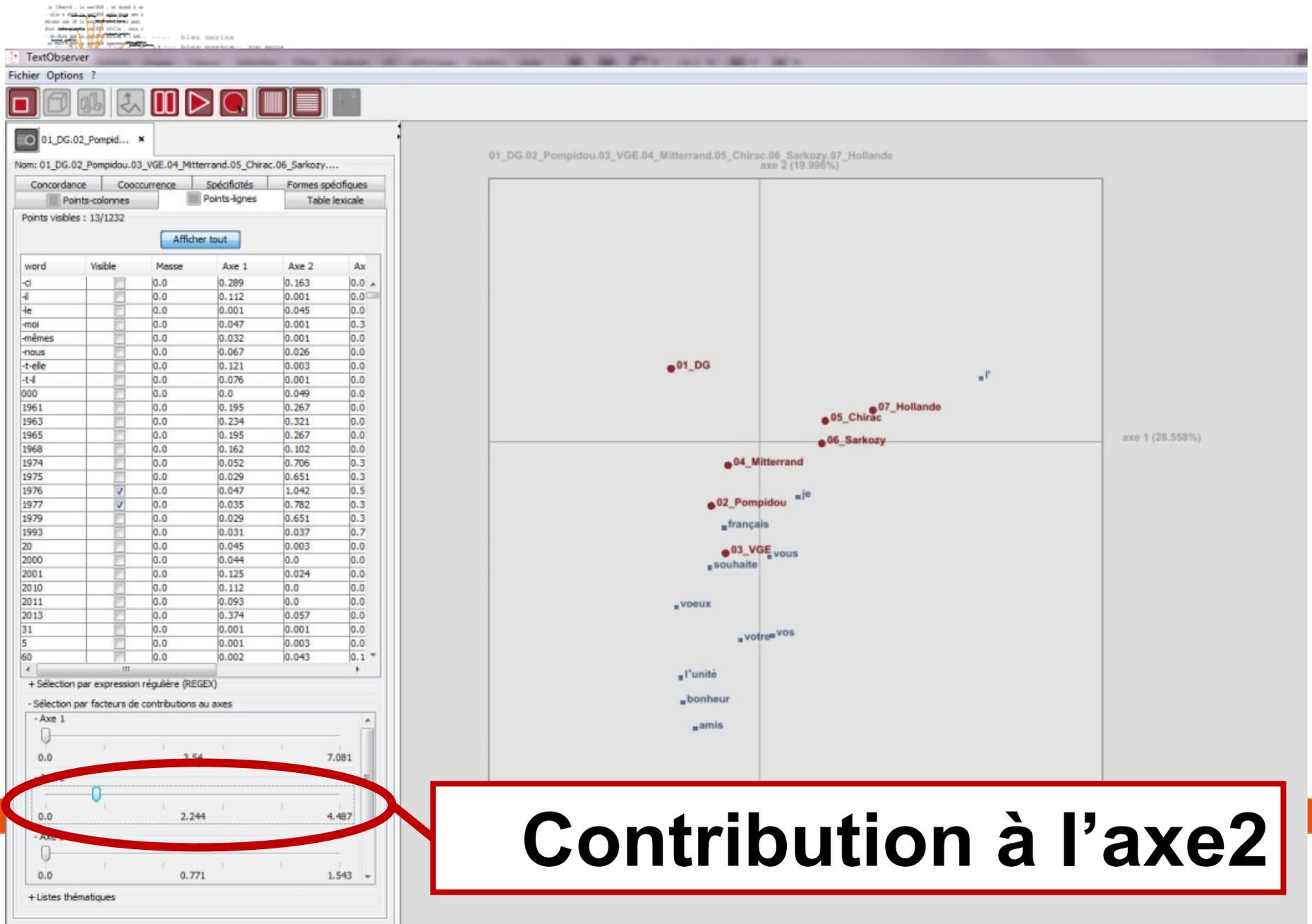
- Listes thématiques

Importer Exporter

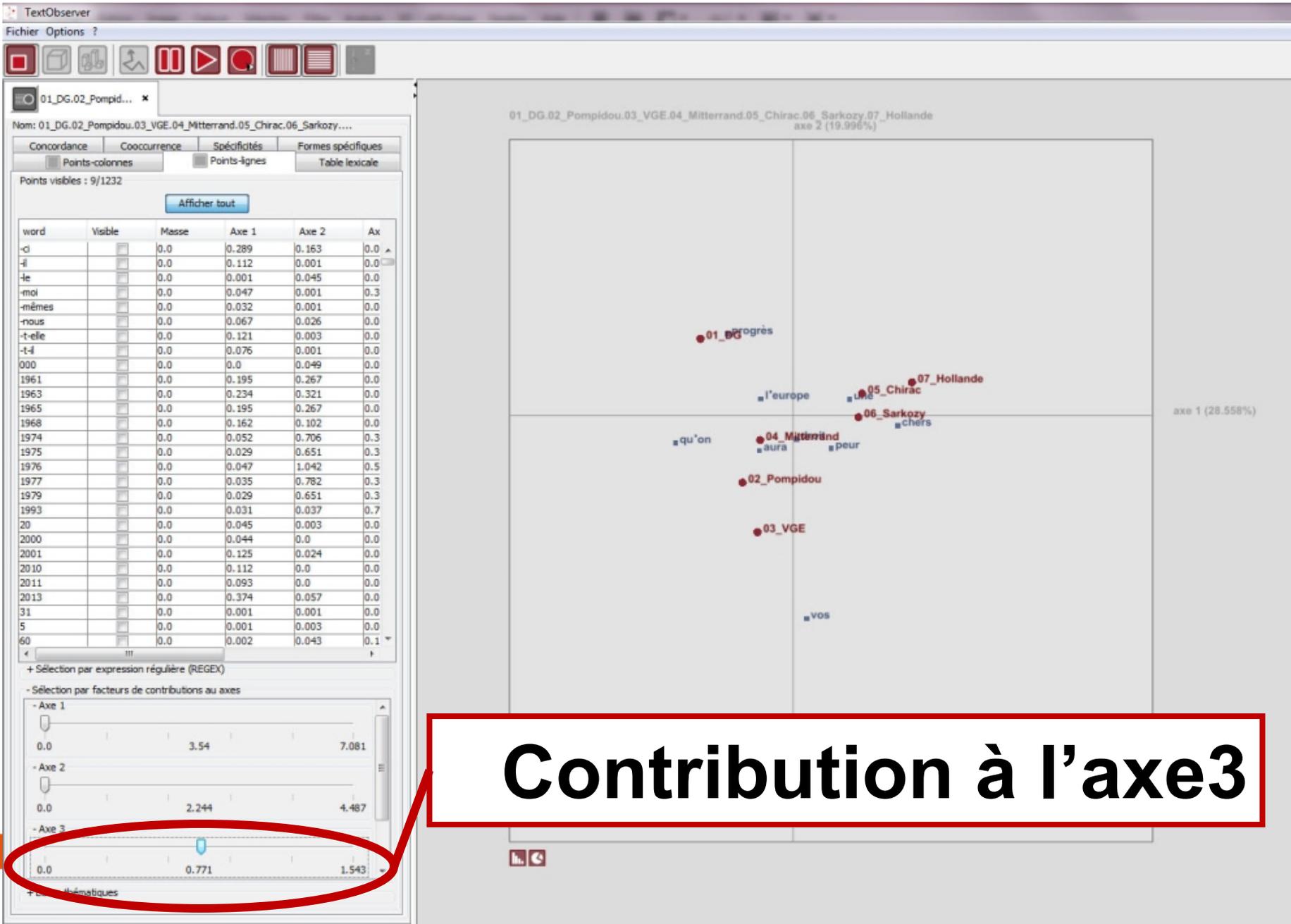
Copier vers REGEX Supprimer



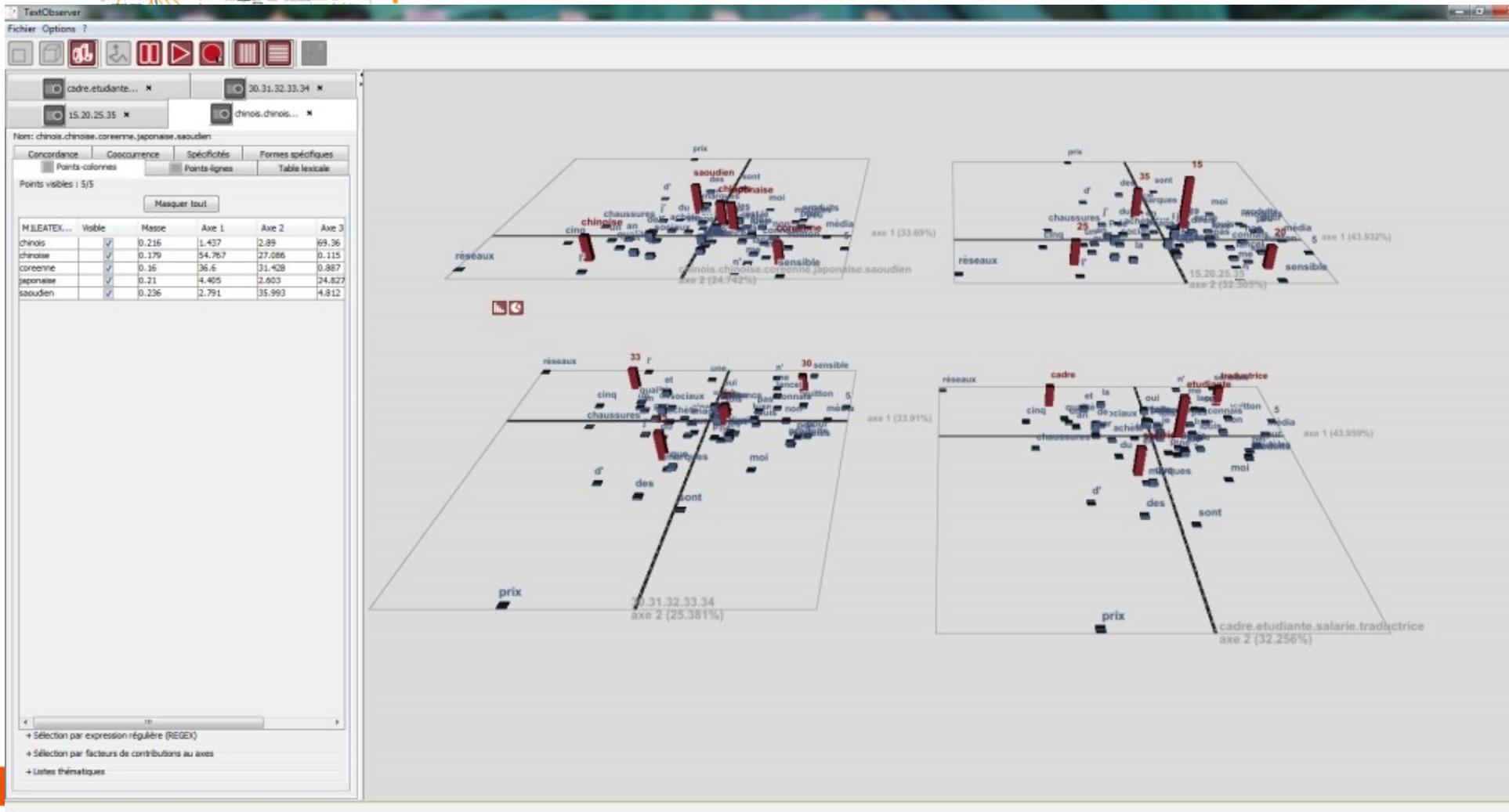




Contribution à l'axe2



Affichage simultané de 4 partitions d'un même corpus xml



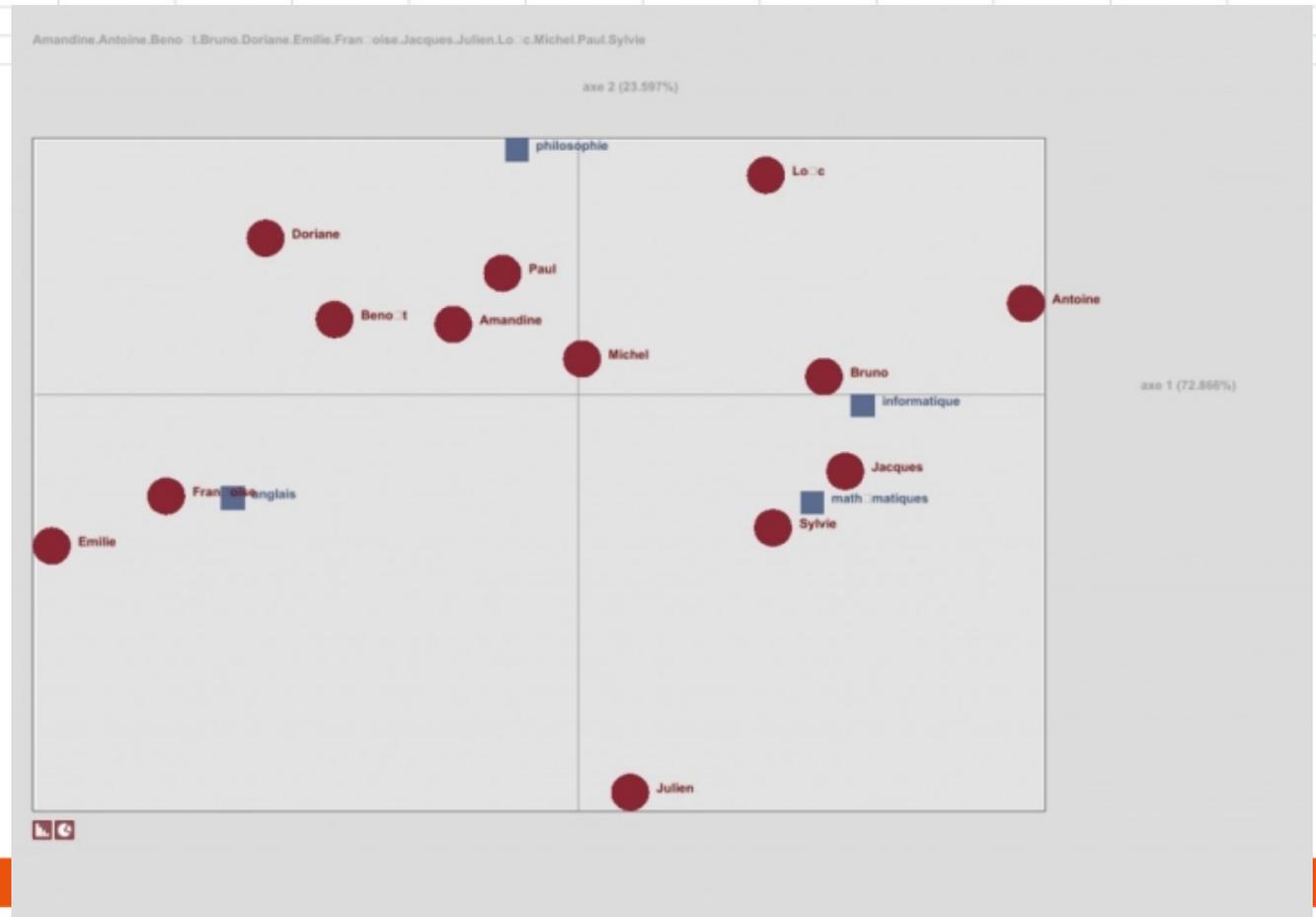
TextObserver/WebObserver : réflexions sur la complémentarité des approches et des méthodes

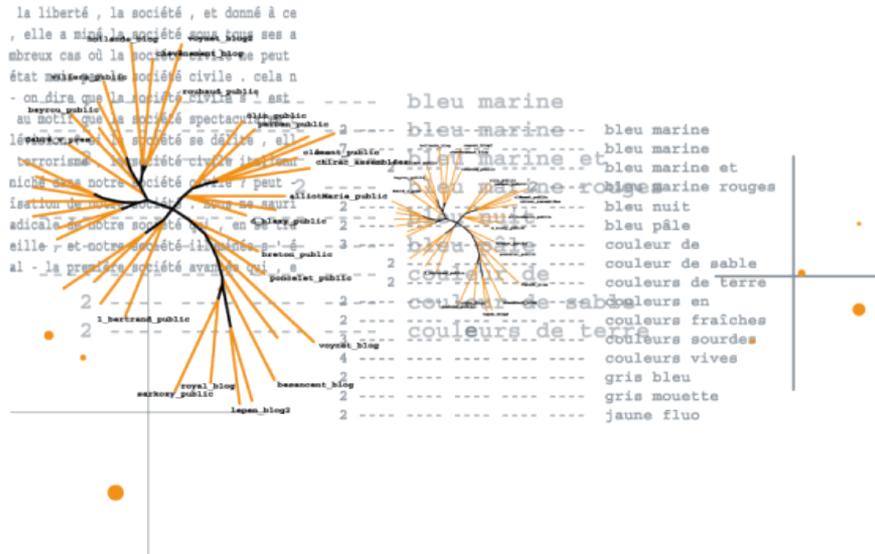


soumettre à *TextObserver* des tableaux de toutes sortes

- mieux comprendre le fonctionnement de l'analyse factorielle
- croiser les résultats issus d'un corpus textuel à des données autres

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Benoît	Antoine	Françoise	Antoine	Michel	Doriane	Emilie	Amandine	Jacques	Sylvie	Loïc	Julien	Paul	Bruno
2	Philosophie	14	12	8	6	17	16	7	14	8	9	17	3	18	12
3	Anglais	17	10	17	2	18	18	18	15	9	12	8	13	17	9
4	Mathématiques	8	15	5	7	14	6	4	10	11	15	12	13	11	14
5	Informatique	7	18	5	8	15	7	4	8	15	14	13	8	12	14
6															
7															
8															
9															
10															





FONCTIONNALITÉS DE CALCUL

01_DG.02_Pompid... x

Nom: 01_DG.02_Pompidou.03_VGE.04_Mitterrand.05_Chirac.06_Sarkozy....

Concordance Cooccurrence Spécificités Formes spécifiques

Points-colonnes Points-lignes Table lexicale

14 lignes pour 5 colonnes (VOEUX_TREETAGG_XML_2013_discours_loc_word.tab)

Modifier l'AFC courante Transformer en AFC interpolée

word	Actif	Total	01_DG	02_Pompidou
[Toutes les li...	<input checked="" type="checkbox"/>	52401	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
pour	<input checked="" type="checkbox"/>	731	109	38
la	<input checked="" type="checkbox"/>	1960	383	83
métropole	<input type="checkbox"/>	24	6	0
française	<input type="checkbox"/>	40	15	0
l'algérie	<input type="checkbox"/>	10	8	0
communauté	<input type="checkbox"/>	30	9	1
je	<input checked="" type="checkbox"/>	497	34	32
forme	<input type="checkbox"/>	18	3	0
des	<input checked="" type="checkbox"/>	714	142	27
voeux	<input type="checkbox"/>	82	10	3
ardents	<input type="checkbox"/>	3	3	0
et	<input checked="" type="checkbox"/>	1627	359	93
confiants	<input type="checkbox"/>	4	4	0
au	<input type="checkbox"/>	288	74	8
premier	<input type="checkbox"/>	38	3	1
jour	<input type="checkbox"/>	41	14	6
de	<input checked="" type="checkbox"/>	2536	531	118
1960.	<input type="checkbox"/>	2	2	0
suis	<input type="checkbox"/>	39	3	1
rempli	<input type="checkbox"/>	4	3	0
l'espoir	<input type="checkbox"/>	8	3	0

- Fréquences (minimum = 459; maximum = 2536)

1 254 507 760 1013 1266 1519 1772 2025 2278 2531

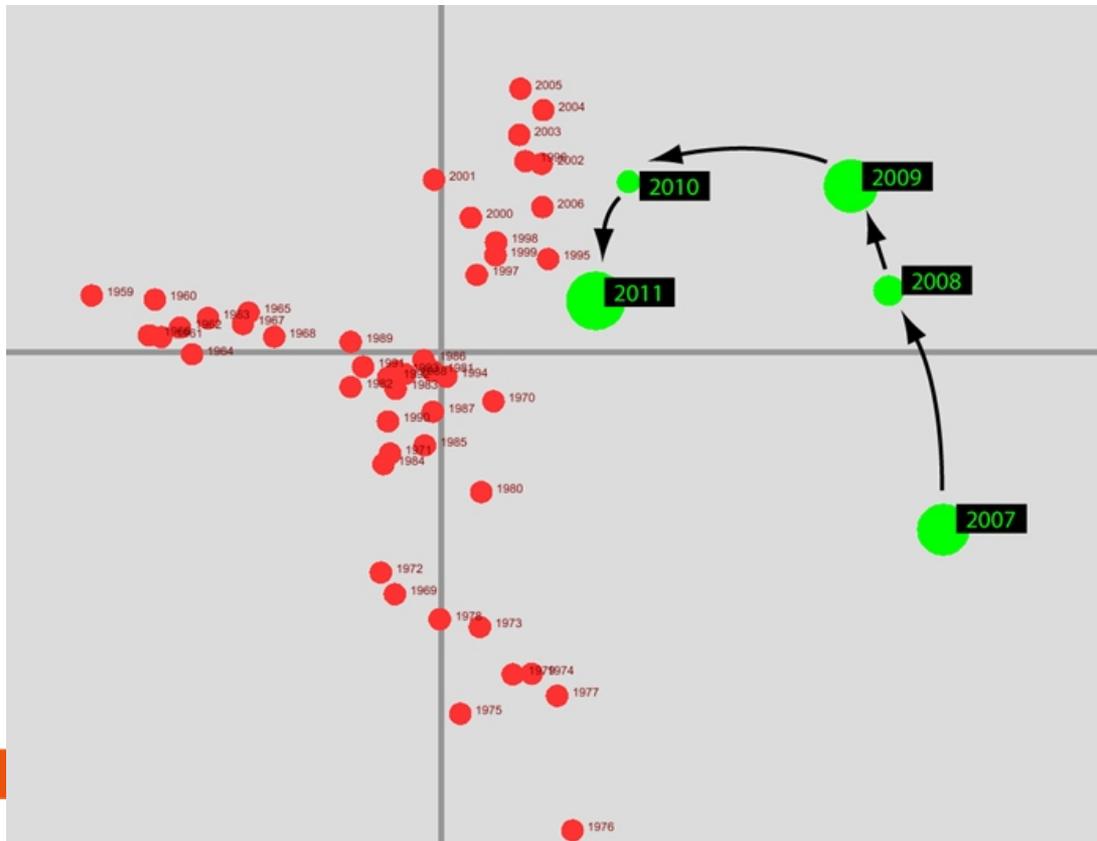
+ Sélection par expression régulière (REGEX)

+ Listes thématiques

Modulation à partir d'un tableau lexical entier



interpolation de mouvement : mesurer en temps réel l'évolution de la configuration factorielle



Fréquence absolue

TextObserver

Fichier Options ?

01_DG.02_Pompidou... x

Nom: 01_DG.02_Pompidou.03_VGE.04_Mitterrand.05_Chirac.06_Sarkozy....

Concordance Cooccurrence Spécificités Formes spécifiques

Points-colonnes Points-lignes Table lexicale

Points visibles : 17/1232

Masquer tout

word	Visible	Masse	Axe 1	Axe 2	
-ci	<input type="checkbox"/>	0.0	0.289	0.163	0.0
-il	<input type="checkbox"/>	0.0	0.112	0.001	0.0
-le	<input type="checkbox"/>	0.0	0.001	0.045	0.0
-moi	<input type="checkbox"/>	0.0	0.047	0.001	0.3
-mêmes	<input type="checkbox"/>	0.0	0.032	0.001	0.0
-nous	<input type="checkbox"/>	0.0	0.067	0.026	0.0
-t-elle	<input type="checkbox"/>	0.0	0.121	0.003	0.0
-t-il	<input type="checkbox"/>	0.0	0.076	0.001	0.0
000	<input type="checkbox"/>	0.0	0.0	0.049	0.0
1961	<input type="checkbox"/>	0.0	0.195	0.267	0.0
1963	<input type="checkbox"/>	0.0	0.234	0.321	0.0
1965	<input type="checkbox"/>	0.0	0.195	0.267	0.0

+ Sélection par expression régulière (REGEX)

- Sélection par facteurs de contributions au axes

- Axe 1
0.0 | 3.54 | 7.081

- Axe 2
0.0 | 2.244 | 4.487

- Axe 3
0.0 | 0.771 | 1.543

+ Listes thématiques

Point-ligne : solidarité

- Fréquences
- Absolues
- Calcul Spécificité
- Concordance
- Cooccurrence

Point-ligne : solidarité

Répartition :

- Total = 50
- 01_DG = 0
- 02_Pompidou = 1
- 03_VGE = 0
- 04_Mitterrand = 8
- 05_Chirac = 31
- 06_Sarkozy = 5
- 07_Hollande = 5

Fréquence relative

TextObserver

Fichier Options ?

01_DG.02_Pompidou... x

Nom: 01_DG.02_Pompidou.03_VGE.04_Mitterrand.05_Chirac.06_Sarkozy....

Concordance Cooccurrence Spécificités Formes spécifiques

Points-colonnes Points-lignes Table lexicale

Points visibles : 17/1232

Masquer tout

word	Visible	Masse	Axe 1	Axe 2	
-ci	<input type="checkbox"/>	0.0	0.289	0.163	0.0
-il	<input type="checkbox"/>	0.0	0.112	0.001	0.0
-le	<input type="checkbox"/>	0.0	0.001	0.045	0.0
-moi	<input type="checkbox"/>	0.0	0.047	0.001	0.3
-mêmes	<input type="checkbox"/>	0.0	0.032	0.001	0.0
-nous	<input type="checkbox"/>	0.0	0.067	0.026	0.0
-t-elle	<input type="checkbox"/>	0.0	0.121	0.003	0.0
-t-il	<input type="checkbox"/>	0.0	0.076	0.001	0.0
000	<input type="checkbox"/>	0.0	0.0	0.049	0.0
1961	<input type="checkbox"/>	0.0	0.195	0.267	0.0
1963	<input type="checkbox"/>	0.0	0.234	0.321	0.0
1965	<input type="checkbox"/>	0.0	0.195	0.267	0.0

+ Sélection par expression régulière (REGEX)

- Sélection par facteurs de contributions au axes

- Axe 1

0.0 | 3.54 | 7.081

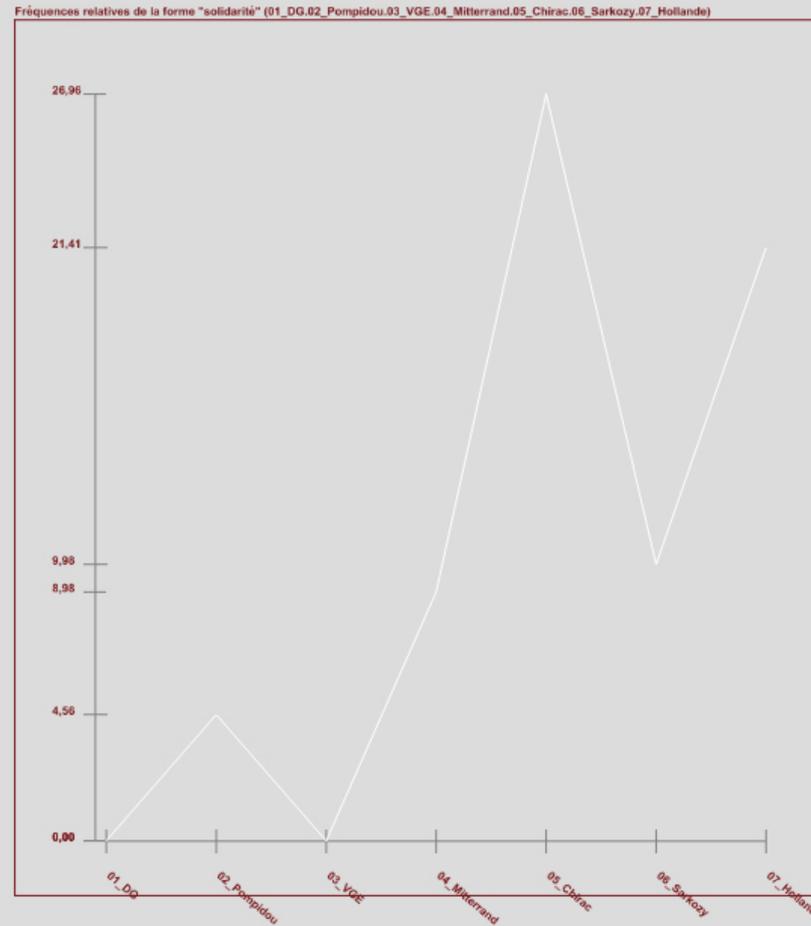
- Axe 2

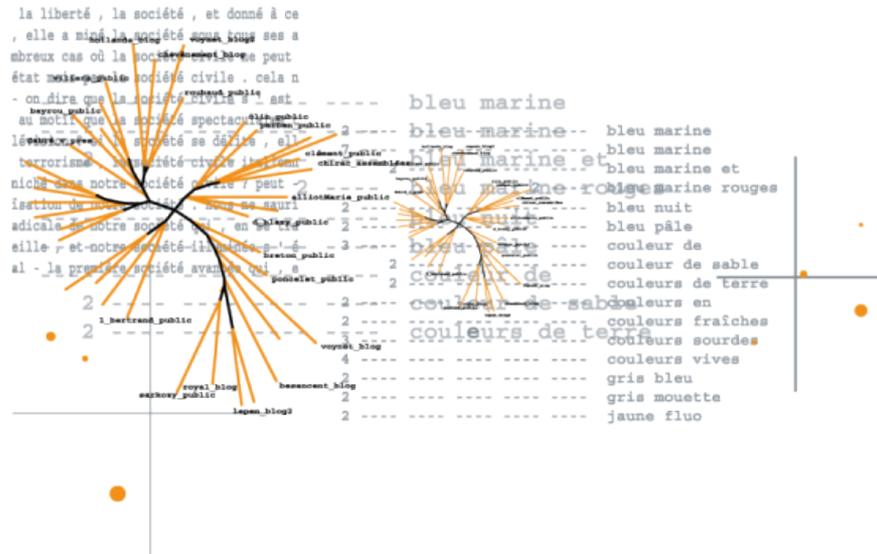
0.0 | 2.244 | 4.487

- Axe 3

0.0 | 0.771 | 1.543

+ Listes thématiques





RETOUR AU TEXTE

Concordance

01_DG.02_Pompidou... x

Nom: 01_DG.02_Pompidou.03_VGE.04_Mitterrand.05_Chirac.06_Sarkozy,...

Points-colonnes Points-lignes Table lexicale

Concordance Cooccurrence Spécificités Formes spécifiques

Choisissez la partition **VOEUX_TREETAGG_XML_2013**

Gauche 10 Requête solidarité Droite 10 Appliquer

Fréquence **44** Spécificité

LeftContext	Keyword	RightContext
la culture, n'ont cessé d...	solidarité	. La France a apporté u
dans cette contradiction...	solidarité	que celle qui nous unit à
pour nous un devoir nati...	solidarité	. Quiconque est seul da
. A condition, évidemment...	solidarité	et de dialogue inspire e
fruits du travail doivent ...	solidarité	nationale doit inspirer n
les bases dès le printem...	solidarité	accrue à l'égard des pa
comprendre que l'intérêt...	solidarité	sur les corporatismes, l
: que les préférences pol...	solidarité	nationale lorsque l'intér
ne songe à s'attaquer à ...	solidarité	des Français entre eux,
jour, un esprit de respon...	solidarité	exemplaires. Des million
confiance, parce qu'elle ...	solidarité	. Mes Chers Compatriot
libertés, l'indépendance ...	solidarité	, de tolérance et de res
voudrait intimider et qui ...	solidarité	. Je vous demande de r
République, au premier r...	solidarité	. Responsable de l'aver
notre démocratie. C'est ...	solidarité	, une solidarité respons
C'est ainsi que nous inve...	solidarité	responsable pour rame
d'être en famille, entour...	solidarité	des leurs, comme à ceu
mon émotion mais aussi ...	solidarité	qui anime tant de bénévo
essentielles : le service p...	solidarité	. Un Etat auquel il appart
allante, plus mobile, plus...	solidarité	. Une solidarité plus respons
mobile, plus optimiste. L'...	solidarité	plus responsable où chacu
au temps. Conscience, e...	solidarité	. Solidarité entre le Nord
voisin et tout ce qui don...	solidarité	. Le progrès n'est rien sans
je veux dire, en ce 31 d...	solidarité	et l'amitié de la nation. Et
, et qui provoque chez n...	solidarité	. On le voit aussi dans la

Exporter

Choisissez la partition V2012ZMAJ_MIN

Gauche 5 Requête "an"[]*"disais"%c Droite 5 Appliquer

LeftContext	Keyword	RightContext
plus qu'un million par	an. une amélioration aussi considérable tient sans doute, d	à une large mesure, à l'expansion qui est en cours e...
ayons confiance, que l'	an nouveau soit celui de l'espérance. bonne année l une année s'achève sous la neige, qui a été à plusieurs reprises	« nous ne sommes
. Il y a un	an, je vous disais	encore : « nous ne
. Il y a un	an, en vous offrant mes vœux pour l'année 1973, je vous disais	que ce serait une année
il y a moins d'un	an. je voudrais que cette certitude vous détourne du désouragement ou de la résignation, et vous ouvre les voie...	que nous avions de la

Entrer

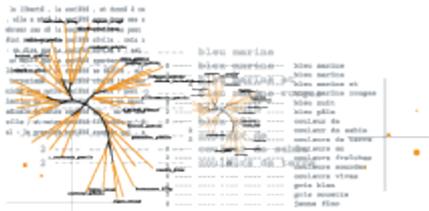
"an"[]*"disais"%c

des approches et des méthodes

The screenshot shows the TextObserver application window. At the top, there is a menu bar with 'Fichier' and 'Options ?'. Below it is a toolbar with various icons. The main area is divided into several sections. A search configuration section is highlighted with a red border, containing a dropdown menu for 'Choisissez la partition' set to 'V2012XMAJ_MIN', a search query input field containing '"an"[*]"disais"%c', and an 'Appliquer' button. Below this is a table with three columns: 'LeftContext', 'Keyword', and 'RightContext'. The table contains several rows of search results. A red arrow points from the search query input field to a larger red-bordered box below the table containing the same query: '"an"[*]"disais"%c'. At the bottom of the application window, there is an 'Exporter' button.

LeftContext	Keyword	RightContext
plus qu'un million par	an. une amélioration aussi considérable tient sans doute, dans une large mesure, à l'expansion qui est en cours e...	: dans le calme.
ayons confiance. que l'	an nouveau soit celui de l'espérance. bonne année ! une année s'achève sous la neige, qui a été à plusieurs repris...	: « nous ne sommes
, il y a un	an, je vous disais	encore : « nous ne
, il y a un	an, en vous offrant mes vœux pour l'année 1973, je vous disais	que ce serait une année
y a moins d'un	an. je voudrais que cette certitude vous détourne du désespoir ou de la résignation, et vous ouvre les voie...	que nous avons de la

"an"[*]"disais"%c



TextObserver

Fichier Options ?

a.de.gaulle.b.p...

Nom: a.de.gaulle.b.pompidou.c.giscard.d.mitterrand.e.chirac.f.sar...

Choisissez la partition V2012XMAJ_MIN

Gauche 5 Requête **"an"[](2,5)"disais"%c** Droite 5 Appliquer

LeftContext	Keyword	RightContext
. il y a un	an, jour pour je vous disais	: « nous ne sommes
. il y a un	an, je vous disais	encore : « nous ne

Exporter

b.pompidou

c.giscard

"an"[](2,5)"disais"%c

Retour au texte

TextObserver
Fichier Options ?

Points-colonnes Points-lignes Table lexicale Concordance

Choisissez la partition V2012XMAJ_MIN

Gauche 5 Requête :confiance |raison |raisons Droite 5 Appliquer

LeftContext	Keyword	RightContext
un bond en avant,	nous avons toutes raisons	de compter que dans les
nous sommes persuadés - et	nous avons de bonnes raisons	pour l'être - que
et lorsque nous constatons que	nous avons été capables de surmonter ces d...	en elle - même pour
en elle - même.	nous avons toutes les raisons	d'être fiers de notre
ensemble, oui ensemble,	nous avons voulu au contraire qu'elle aille d...	en elle, que son

Exporter

FR 20:44 21/05/2013

"nous | Nous" "avons" []{1,20} "espérer | espoir | confiance | raison | raisons"



Aide à l'interprétation et formation à la recherche

- Aide à l'interprétation
 - exemple des points contributifs
- Mouvement
 - mieux saisir les phénomènes liés à la variation
 - 2D 3D, navigation zoom : lisibilité
 - Interface: ergonomie simplifiée et retour au texte (AfC dynamique et interactive)
 - Formation à la recherche (exemple des tableaux de données de type carnet de notes)



[pos="PRP"][lemme="le"]

détergent quand	sur les	maines t'
sang Diplômé	de l'	école buissonnière
Foolek imprévisible	depuis l'	époque de
plus fort	depuis l'	époque de
toujours joué	contre les	bleus,
sales faces	pour les	crânes balafrés
fréquenter Aventurier	de la	jungle,
leurs tiroirs	dans le	noir pendant
amour Spécial	pour les	quartiers chauds
parle pue	de la	gueule,
revoir Allahouahlam	pour la	suite de
la suite	de l'	histoire On

Cooccurrences

01_DG.02_Pompid... x

Nom: 01_DG.02_Pompidou.03_VGE.04_Mitterrand.05_Chirac.06_Sarkozy....

Points-colonnes Points-lignes Table lexicale

Concordance Cooccurrence Spécificités Formes spécifiques

Choisissez la partition **VOEUX_TREETAGG_XML_2013**

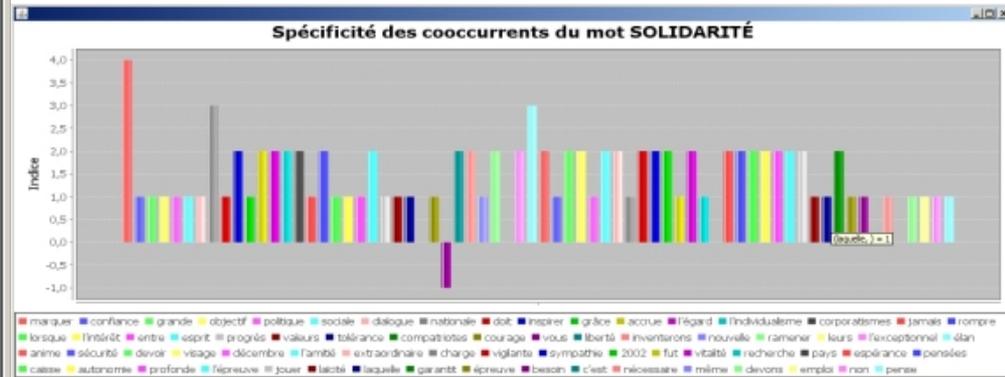
POLE Freq Gauche Droite récursivité

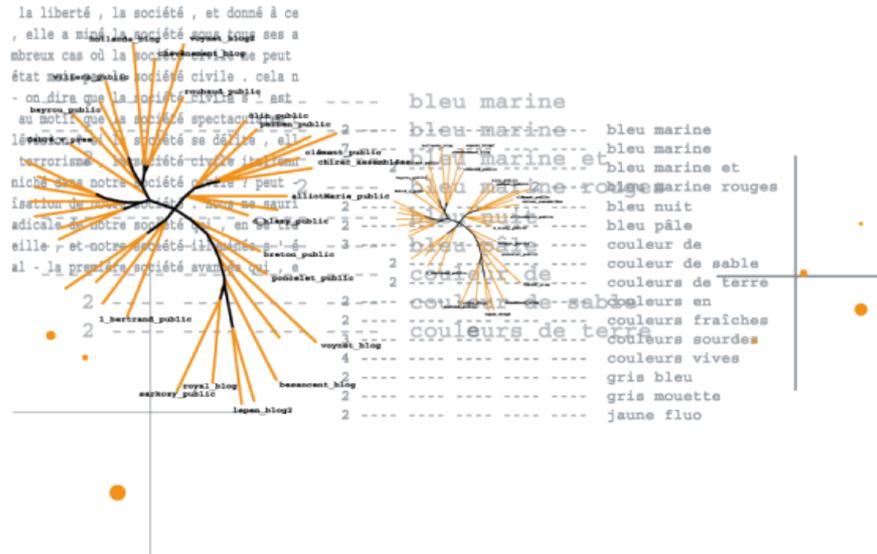
Seuil Appliquer Exporter

Mots cooccurents	Fréquence	Co-Fréquence	Spécificité
marquer	4	2	4.0
confiance	61	1	1.0
grande	29	1	1.0
objectif	12	1	1.0
politique	65	1	1.0
sociale	35	1	1.0
dialogue	23	1	1.0

+ Listes de filtres

```
graph TD;
  solidarite((solidarité)) --> elan((élán));
  solidarite --> anime((anime));
  solidarite --> securite((sécurité));
```





QUELQUES SORTIES...

H:\portable\2013-14\JML\plaquette-TO-mai-2014\images\AFC.eps - Inksc...

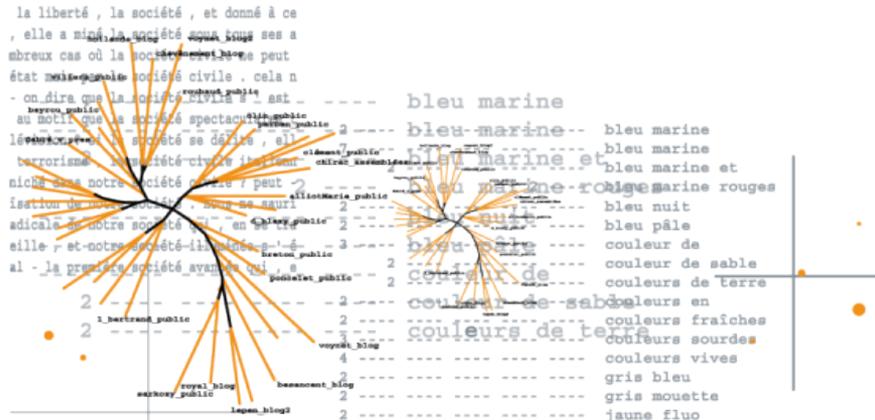
Fichier Édition Affichage Calque Objet Chemin Texte Filtres Extensions Aide

	A	B	C	D	E	F	G	H
1	SOLIDARITÉ:XT= 10 RIGHTCONTEXT= 10							
2	LeftContext	Keyword	RightContext					
3	la culture, n'ont cessé de nous marquer confiance et	solidarité	. La France a apporté un appui considérable pour la					
4	dans cette contradiction. Il n'est pas de plus grande	solidarité	que celle qui nous unit au peuple de Pologne.					
5	pour nous un devoir national. Troisième objectif, la	solidarité	. Quiconque est seul dans la vie, quiconque est					
6	. A condition, évidemment, qu'une politique sociale de	solidarité	et de dialogue inspire et accompagne la politique économique.					
7	fruits du travail doivent être justement partagés, que la	solidarité	nationale doit inspirer nos lois, que la protection sociale					
8	les bases dès le printemps prochain, grâce à la	solidarité	accrue à l'égard des pays de l'Est. Tout cela					
9	comprendre que l'intérêt général doit l'emporter sur l'individualisme, la	solidarité	sur les corporatismes, le sens national sur l'esprit partisan					
10	: que les préférences politiques s'affirment sans jamais rompre la	solidarité	nationale lorsque l'intérêt commun est en jeu. Je veillerai					
11	ne songe à s'attaquer à nos acquis sociaux. La	solidarité	des Français entre eux, entre bien-portants et malades,					
12	jour, un esprit de responsabilité, un esprit de	solidarité	exemplaires. Des millions d'entre vous se sont levés très					
13	confiance, parce qu'elle a choisi le progrès et la	solidarité	. Mes Chers Compatriotes, en cette soirée de voeux					
14	libertés, l'indépendance de la justice, les valeurs de	solidarité	, de tolérance et de respect de l'autre, qui					
15	voudrait intimider et qui suscite chez nos compatriotes courage et	solidarité	. Je vous demande de rester vigilants. La délinquance					
16	République, au premier rang desquelles la liberté et la	solidarité	. Responsable de l'avenir de la Nation, j'interviendrai chaque					
17	notre démocratie. C'est ainsi que nous inventerons une nouvelle	solidarité	, une solidarité responsable pour ramener vers l'emploi ceux qui					
18	C'est ainsi que nous inventerons une nouvelle solidarité, une	solidarité	responsable pour ramener vers l'emploi ceux qui en sont exclus					
19	d'être en famille, entourés de l'affection et de la	solidarité	des leurs, comme à ceux qui sont seuls ce					
20	mon émotion mais aussi ma fierté devant l'exceptionnel élan de	solidarité	qui anime tant de bénévoles et d'associations mobilisés aux côtés					
21	essentielles : le service public, la sécurité, la	solidarité	. Un Etat auquel il appartient de prévoir, de					
22	allante, plus mobile, plus optimiste. L'exigence de	solidarité	. Une solidarité plus responsable où chacun s'efforcerait de prendre					
23	mobile, plus optimiste. L'exigence de solidarité. Une	solidarité	plus responsable où chacun s'efforcerait de prendre sa part du					
24	au temps. Conscience, enfin, du devoir de	solidarité	. Solidarité entre le Nord et le Sud, indispensable					
25	voisin et tout ce qui donne un visage à la	solidarité	. Le progrès n'est rien sans la fraternité. Face					
26	je veux dire, en ce 31 décembre, la	solidarité	et l'amitié de la nation. Et à vous tous					
27	, et qui provoque chez nous un extraordinaire élan de	solidarité	. On le voit aussi dans la joie, chaque					

Concordance

TextObserver/WebC

Remplissage : Indéfini O : 100 (racine) X : 601,71 Y : 825,40 Z : 45%



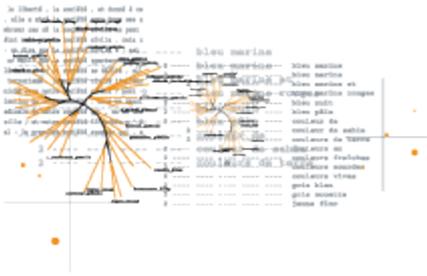
>> corpus multimodaux

WEBOBSERVER

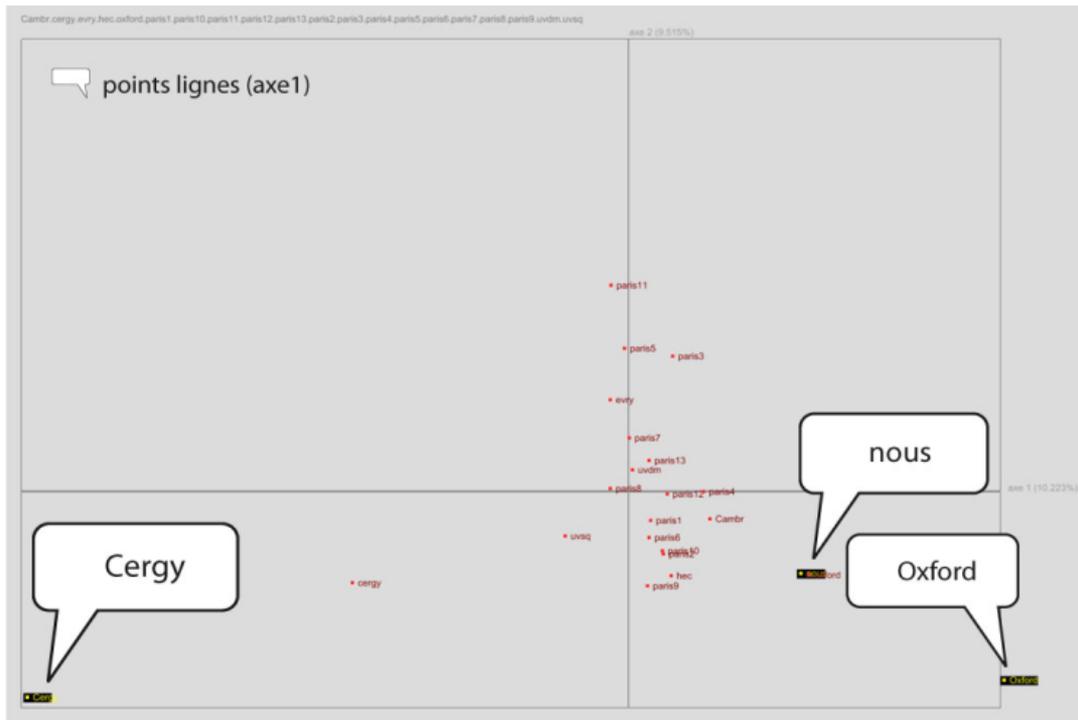


Corpus texte

- **SOURCE** : site officiel de 12 universités de la région parisienne ; racine ou premier niveau : page d'accueil ou de présentation ou de bienvenue ; texte avec données chiffrées
- **DATE** : 10 octobre 2011
- **RECUEIL** : manuel
- **FILTRAGE** : Critères non strictement formels ; le choix tient compte de l'architecture des sites et vise une comparabilité ; Si la présentation n'est pas synthétique, il ne dépasse pas 2 niveaux de liens dans le recueil.
- **MISE EN FORME** : suppression des liens ; sans illustration ni mise en page ; 19 partitions ; suppression de la signature du président.
- **CORPUS** : 20 sites : 13 uni parisiennes ; une grande école : HEC ; 4 université de la RP non numérotées ; 2 université anglaises



Corpus texte





Roborealm™
vision for machines

Screenshots Download Purchase Documentation Tutorials Resources Contact Forum Search

AdChoices ▶



**Composants
Electroniques**

Plus de 480 000
produits en stock de
1500 fabricants
leaders.
fr.farnell.com

**Visicon
Technologies**

High Performance
Vision Solutions For
Metrology and Defect
Detection
www.VisiconTech.com

**Robotics and
mechatronics**

Consulting and R&D
service. Accelerate
your innovation.
www.icarustechnology.com

**Face Detection in
Image**

Detect Human Faces
and Recognize Facial
Features in Your
Software
www.Luxand.com/FaceSDK

Vision Industrielle

Vente de Matériel
Optoélectronique
Caméra, Eclairage,
Câble, Optique..
www.framos.eu

Geometric Statistics

The Geometric Statistics module adds a couple of image statistic variables centered around image geometry. These variables can then be used in RoboRealm [conditional statement](#), [VBScript module](#) or exported to external programs using [RoboRealm Plugins](#).

Interface

Name	Value
AREA	10386
PERIMETER	611
EQUIV_DIAMETER	114.9951
AP_RATIO	16.9984
COMPACTNESS	35.9447
COMPACTNESS_2	0.7173
CIRCULARITY	0.3497
CIRCULARITY_2	0.7529
CIRCULARITY_3	0.4375
MAX_RADIUS	96
MIN_RADIUS	42
FERET	160.3278
FERET_AREA	20361.6290
FERET_ORIENTATION	13.3429
FERET_FIRST_X	56
FERET_FIRST_Y	128
FERET_SECOND_X	212
FERET_SECOND_Y	165
BREATH	127.0001
ROUNDFNESS	0.5145

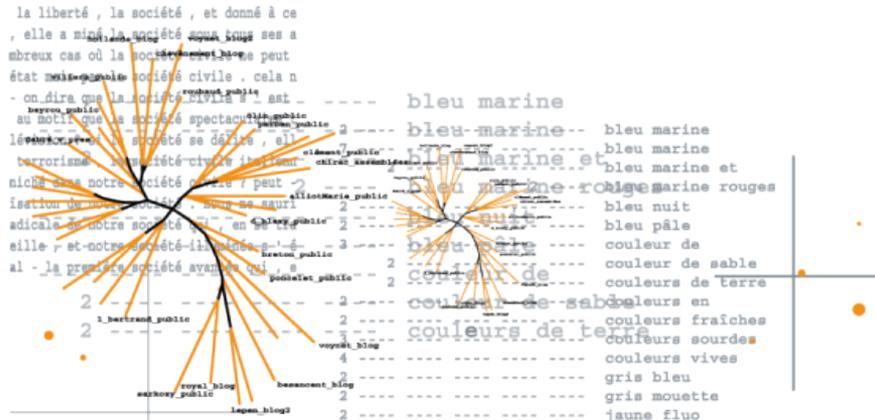
1. Scope - Click on the appropriate radio button to select if you want the statistics to be generated from the entire image or from an individual blob. If you select individual blobs non-black pixels in the current image will be used to form blobs. To see the statistics for an individual blob click on the blob within the main RoboRealm GUI



Perspectives webObserver

- Développer interface d'aide à l'encodage
- Développer la cartographie
- Tester l'outil statistique (AFC, ACP?)
- Module de recueil de sites et automatisations de la reconnaissance automatique d'images

-
- Partenariat industriel



TextObserver

PERSPECTIVES TEXTOBROWSER

TextObserver

Fichier Options ?

2081.2082.2083.... x

Nbre: 2081.2082.2083.2084.2085.2086.2087.2088.2089.2090.2091.2092....

Points-colonnes Points-lignes Table lexicale Concordance Cooccurrence

Choisissez la partition MITT_JML

POLE europe Freq 3 Gauche 15 Droite 15 Appliquer

calcul hyper
calcul brut

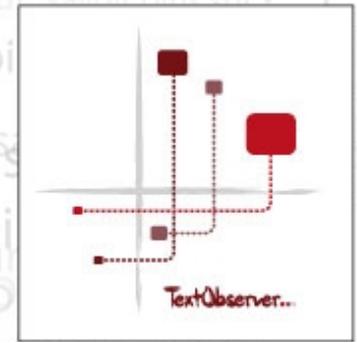
POLE

Exporter

Mots cooccurents	Co-Fréquence
une	3
France	7
nous	15
monde	5
paix	5
sera	7
monde	3
bars	4
armes	3
nudistes	3
tendez	3
vous	7
bonne	3
ans	3
encore	3
nous	3
construction	3
pays	3
perce	3
europe	3
car	4
propose	3
enjeu	3
celui	3
douze	5

DG 1 POMP 2 VGE 3 MITT 4 CHI 5 SARK 6

TextObserver



- **Conception**

- JM Leblanc Pierre Fiala, UPEC, Céditec

- **Développement**

- Sébastien Jacquot , Université de Franche Comté

- Amani Daknou, Post-doc Céditec – 2012-2013

- Julien Bonneau, Post-doc Céditec – 2013-2014

- Marie Pérès, UPEC, Céditec