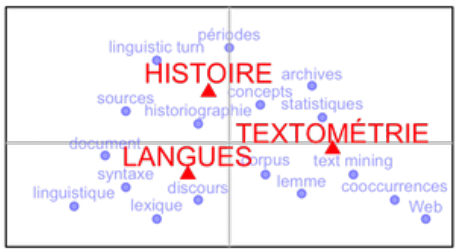
Présentation du logiciel TXM pour la recherche en Histoire

Bénédicte PINCEMIN
(CNRS, IHRIM, Université de Lyon)



CC BY

Colloque Histoire, langues et textométrie

Paris, 16-18 janvier 2019.

Plan

1. Introduction générale et positionnement de TXM
2. Fonctionnalités pour l’étude de corpus diachroniques
3. Des points forts de TXM
   - Textométrie sur corpus annotés et structurés
   - Souplesse et puissance de l’import
   - Du retour au texte au retour au document source
4. Ouverture des traitements : scripts, autres logiciels
5. Ressources en ligne

Historique et positionnement

- Développement commencé en 2009
- Projet ANR Textométrie
  - Fédération des recherches et développements en textométrie autour de la création d'une plateforme logicielle ouverte
- S'inscrit dans la « tradition » textométrique
  - Fonctionnalités centrales : concordances, spécificités, AFC…
  - Importance du retour au texte, articulation quantitatif & qualitatif
- Deux axes d'innovation
  - Textométrie sur des corpus structurés, annotés
  - Développement open-source, mutualisé et ouvert

Deux versions du logiciel

- **Version locale**, pour ordinateur personnel
  - Linux, Windows, Mac OS
- **Version portail**, interrogeable en ligne via un navigateur
  - Mise à disposition du logiciel, à installer sur un serveur pour partager et publier ses propres corpus sur internet (avec contrôle d'accès paramétrable)
  - Un portail de démonstration :
    - http://portal.textometrie.org/demo
  - Pas d'offre d'hébergement de corpus (pas de service en ligne où l'on dépose un corpus puis on l'interroge)
- Chacune des versions a ses particularités de fonctionnalités et d'interface

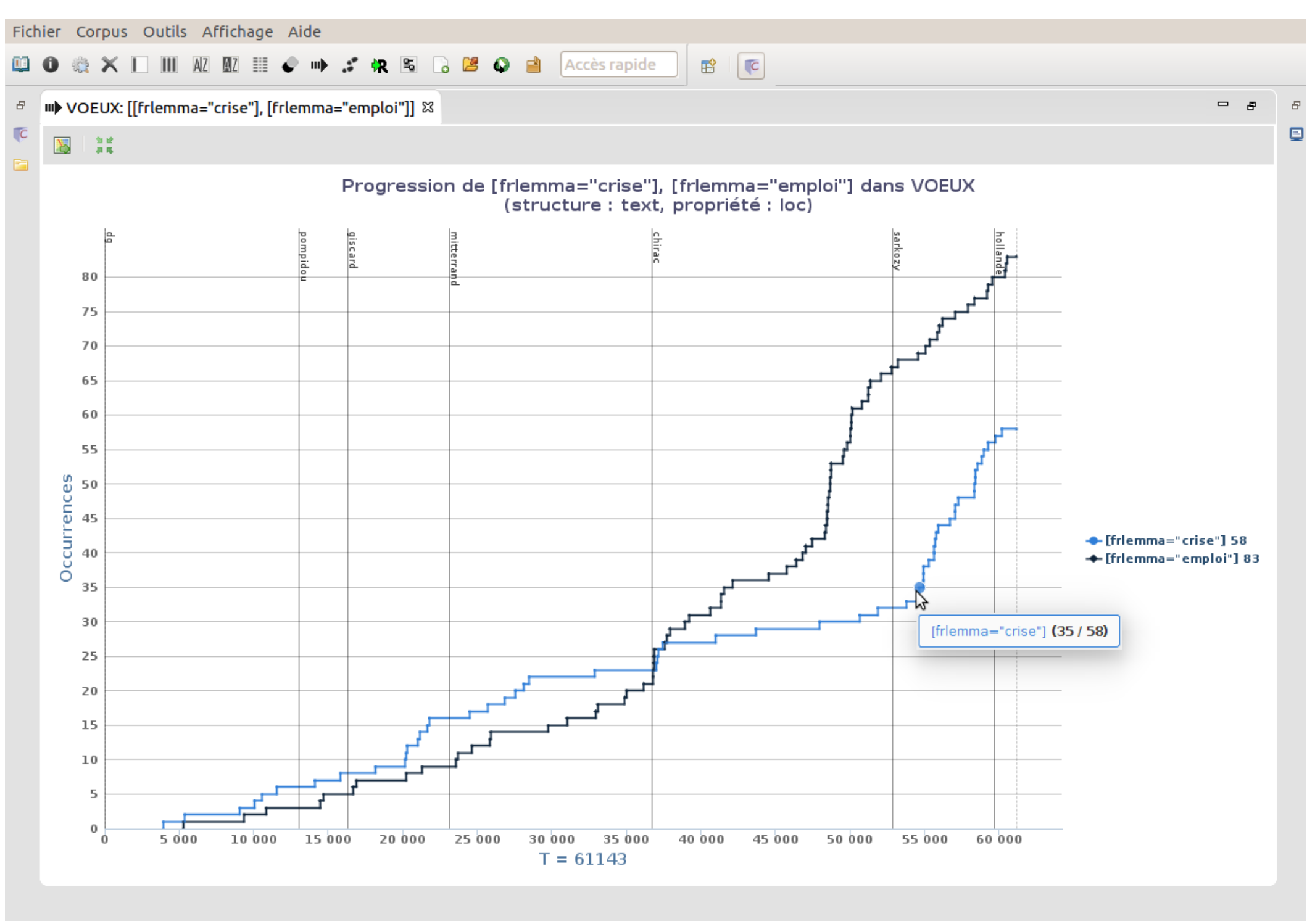
Fonctionnalités pour l'étude de corpus diachroniques

- *Exemples sur le corpus des Voeux présidentiels français (1959-2012) (Jean-Marc Leblanc, univ. Paris-Est Créteil)*

- Progression
- Spécificités
- Analyse factorielle des correspondances

Fonctionnalités pour l'étude de corpus diachroniques

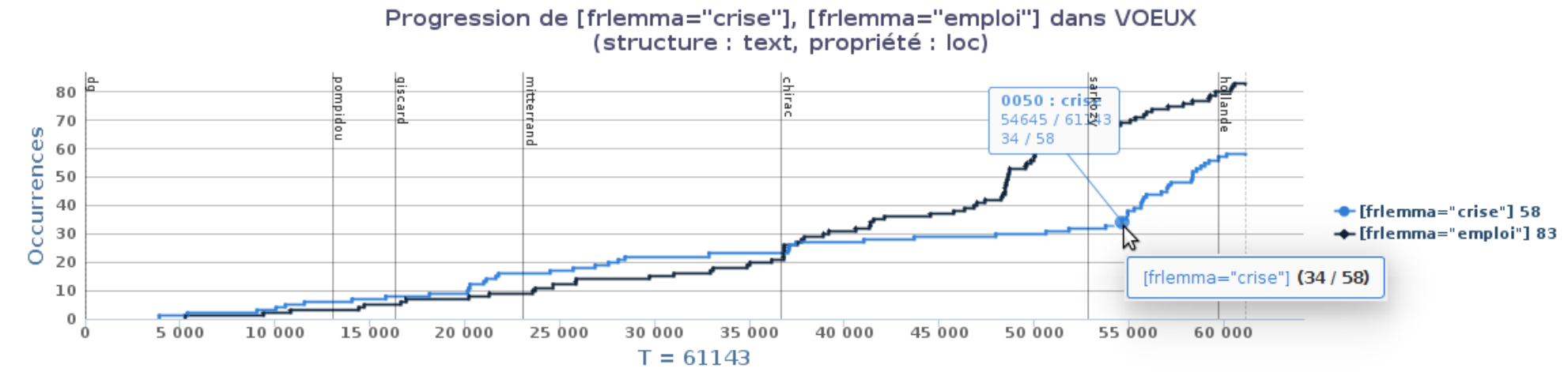
- Progression
  - pour une visualisation globale d'un déroulement (corpus vu comme un continuum) : évolution quantitative de l'usage d'un mot (ou motif linguistique complexe) au fil du temps
  - cf. Carte des sections de Lexico, Topologie dans Hyperbase, Ratinaud JADT 2014 pour Iramuteq
  - dans TXM c'est un graphe *cumulatif* (=> interpréter la *pente* de la courbe)
- Spécificités
- Analyse factorielle des correspondances


Fichier Corpus Outils Affichage Aide
Accès rapide
VOEUX: [[frlemma="crise"], [frlemma="emploi"]]
Progression de [frlemma="crise"], [frlemma="emploi"] dans VOEUX
(structure : text, propriété : loc)
dg
pompidou
giscard
mitterrand
chirac
sarkozy
hollande
Occurrences
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
0
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
50 000
55 000
60 000
T = 61143
[frlemma="crise"] 58
[frlemma="emploi"] 83
[frlemma="crise"] (35 / 58)

Fichier Outils Affichage Aide

Accès rapide

VOEUX: [[frlemma="crise"], [frlemma="emploi"]]



VOEUX:[frlemma="crise"]

Requête : [frlemma="crise"] Pivot: word Editer **Chercher**

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

|< < 1 - 58 / 58 > >| Cacher paramètres

text_id	Contexte gauche	Pivot	Contexte droit
0040	monde, c'est vrai, où les	crises	, notamment financières, se propagent très vite. Mais aussi et
0042	la marée noire de l'Erika ; la	crise	de la vache folle, dont l'Europe, à l'initiative
0045	par l'instabilité, les tensions et les	crises	: la guerre en Irak, le terrorisme, les attentats et
0047	société : le non au référendum et la	crise	des banlieues en portent le témoignage. Avec en arrière-plan, une
0048	terrorisme, la guerre en Irak, la	crise	au Proche-Orient, la tentation absurde et irresponsable du choc des civilisations
0049	, malgré une conjoncture internationale freinée par la	crise	financière, les premiers résultats de l'action entreprise devraient se faire
0050	, cette année a été difficile. La	crise	économique et financière mondiale est venue ajouter son lot de peines et

Fonctionnalités pour l’étude de corpus diachroniques

- Progression
- Spécificités
  - Pour la caractérisation de périodes (corpus divisé en parties individualisées, nommées)
  - Mots des périodes, périodes des mots
- Analyse factorielle des correspondances

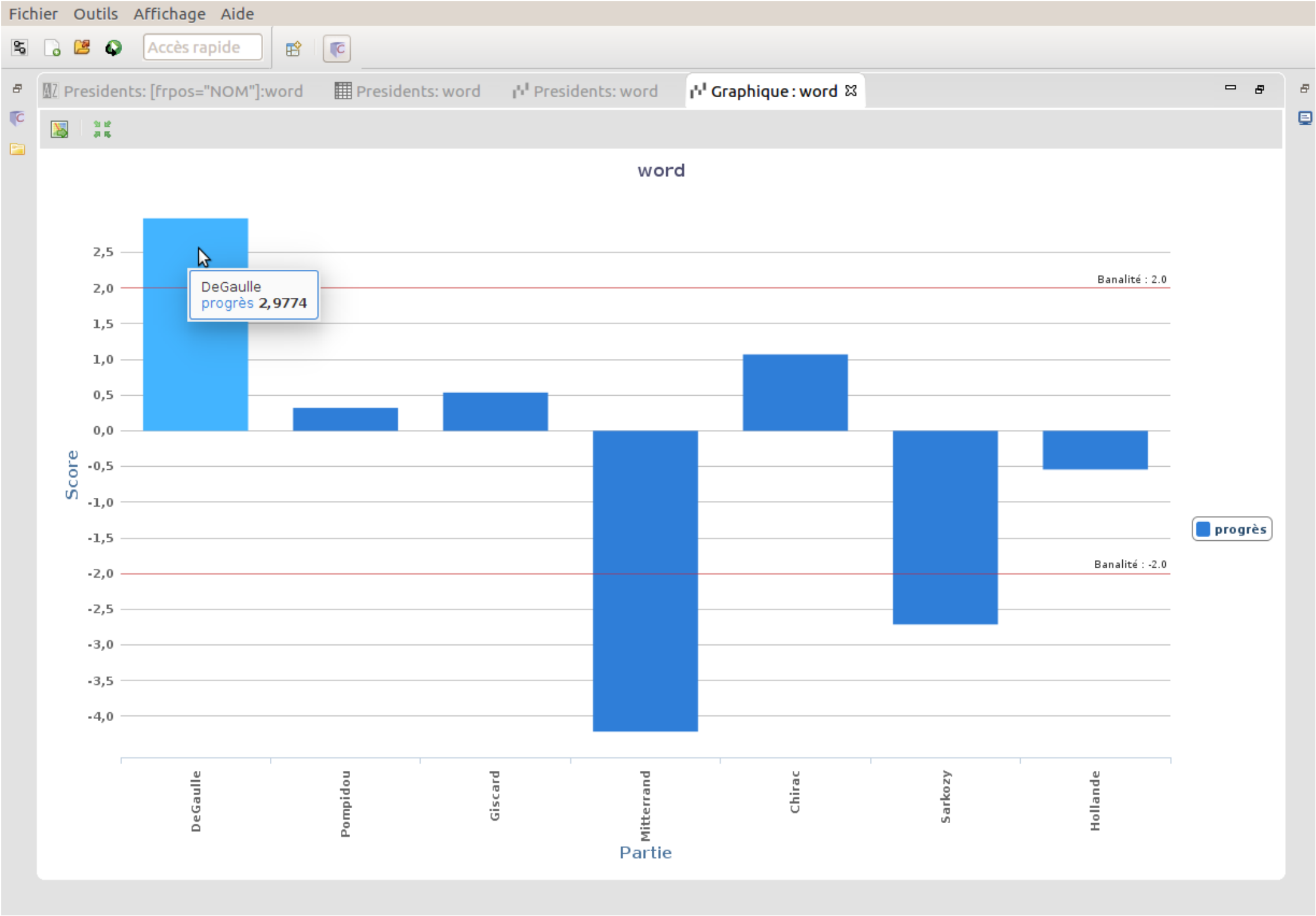
Fichier Outils Affichage Aide

Accès rapide

Presidents: [frpos="NOM"]:word | Presidents: word | **Presidents: word**

Unités	Fréqu	DeGaulle t=2222	sc	Pompidou t=524	sco	Giscard t=1100	scor	Mitterrand t=2245	sco	Chirac t=2874	scor	Sarkozy t=1197	score	Ho
coopération	14	12	6,3	0	0,3	0	-0,7	1	-0,8	0	-2,0	0	-0,7	
développement	28	18	6,0	0	-0,6	1	-0,7	1	-2,0	5	-0,8	3	-0,2	
rapports	8	8	5,4	0	-0,2	0	-0,4	0	-0,8	0	-1,1	0	-0,4	
œuvre	15	11	4,6	0	-0,3	0	-0,7	0	-1,6	0	-2,1	4	1,1	
nation	33	18	4,5	1	-0,3	1	-0,9	0	-3,5	12	0,8	1	-1,0	
fait	20	13	4,5	0	-0,4	4	0,8	1	-1,3	2	-1,2	0	-1,1	
mesure	18	12	4,4	1	0,2	1	-0,4	1	-1,1	2	-1,0	1	-0,4	
part	35	18	4,1	3	0,6	1	-1,0	4	-1,0	8	-0,5	1	-1,1	
régime	8	7	3,9	0	-0,2	0	-0,4	1	-0,3	0	-1,1	0	-0,4	
doute	16	10	3,4	0	-0,4	1	-0,3	3	-0,3	0	-2,2	2	0,2	
ardeur	5	5	3,4	0	-0,1	0	-0,2	0	-0,5	0	-0,7	0	-0,3	
essor	5	5	3,4	0	-0,1	0	-0,2	0	-0,5	0	-0,7	0	-0,3	
détente	7	6	3,3	0	-0,2	0	-0,3	1	-0,3	0	-1,0	0	-0,4	
univers	7	6	3,3	0	-0,2	1	0,3	0	-0,7	0	-1,0	0	-0,4	
peuples	46	20	3,2	3	0,4	1	-1,4	17	1,9	2	-4,2	1	-1,6	
progrès	51	21	3,0	3	0,3	7	0,5	1	-4,2	19	1,1	0	-2,7	
intérieur	10	7	2,9	1	0,4	0	-0,5	1	-0,5	1	-0,7	0	-0,5	
but	8	6	2,8	0	-0,2	0	-0,4	0	-0,8	0	-1,1	1	0,2	
contraire	13	8	2,7	0	-0,3	0	-0,6	2	-0,4	3	-0,3	0	-0,7	
raison	13	8	2,7	0	-0,3	1	-0,2	0	-1,4	2	-0,6	1	-0,3	
au-dehors	4	4	2,7	0	-0,1	0	-0,2	0	-0,4	0	-0,6	0	-0,2	
buts	4	4	2,7	0	-0,1	0	-0,2	0	-0,4	0	-0,6	0	-0,2	
souhaits	6	5	2,7	0	-0,1	1	0,3	0	-0,6	0	-0,8	0	-0,3	
prospérité	16	9	2,6	2	0,7	0	-0,8	1	-1,0	3	-0,5	1	-0,4	
puissance	22	11	2,6	0	-0,5	1	-0,5	3	-0,6	7	0,4	0	-1,2	
agriculture	11	7	2,5	1	0,4	0	-0,5	2	-0,2	1	-0,8	0	-0,6	
coup	11	7	2,5	3	1,8	0	-0,5	1	-0,6	0	-1,5	0	-0,6	
expansion	11	7	2,5	2	1,0	0	-0,5	1	-0,6	1	-0,8	0	-0,6	
évolution	9	6	2,4	0	-0,2	2	0,6	0	-0,9	1	-0,6	0	-0,5	

DeGaulle t=2222


Fichier Outils Affichage Aide
Accès rapide
Presidents: [frpos="NOM"]:word
Presidents: word
Presidents: word
Graphique : word
word
DeGaulle
progrès 2,9774
Banalité : 2.0
Banalité : -2.0
progrès
Score
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-2,5
-3,0
-3,5
-4,0
DeGaulle
Pompidou
Giscard
Mitterrand
Chirac
Sarkozy
Hollande
Partie

Fonctionnalités pour l’étude de corpus diachroniques

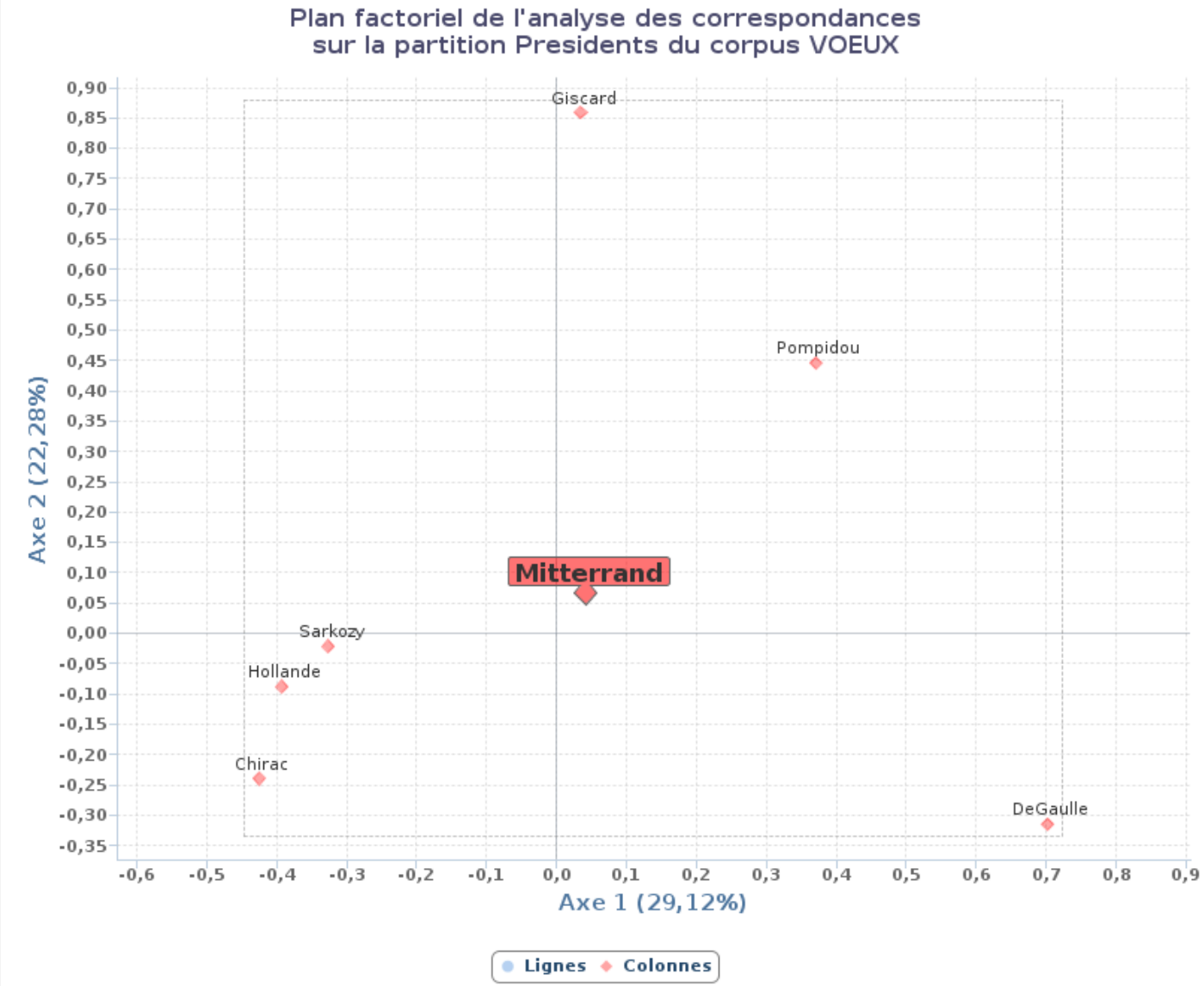
- Progression
- Spécificités
- Analyse factorielle des correspondances
  - Pour une représentation cartographique, spatialisée, des proximités des mots (par leur répartition, leur profil temporel) et des proximités des périodes (par leur contenu)
  - Configuration d’ensemble

Fichier Outils Affichage Aide

Accès rapide

Presidents: [frpos="NOM"]:word | Presidents: word | Presidents:word | Presidents:word

Axes : (1,2)



Colonnes	Q12	Q13	Q23	Mass	Dist	Cont
Sarkozy	,18	,18	,00	11,85	0,57	7,2
Pompidou	,19	,52	,56	5,16	1,78	4,0
Mitterrand	,02	,62	,63	20,48	0,34	0,2
Hollande	,09	,10	,02	2,27	0,43	2,0
Giscard	,81	,00	,81	11,55	0,91	0,0
DeGaulle	,94	,78	,16	19,78	0,63	56,2
Chirac	,75	,66	,27	28,91	0,32	30,1

Info lignes | Info colonnes

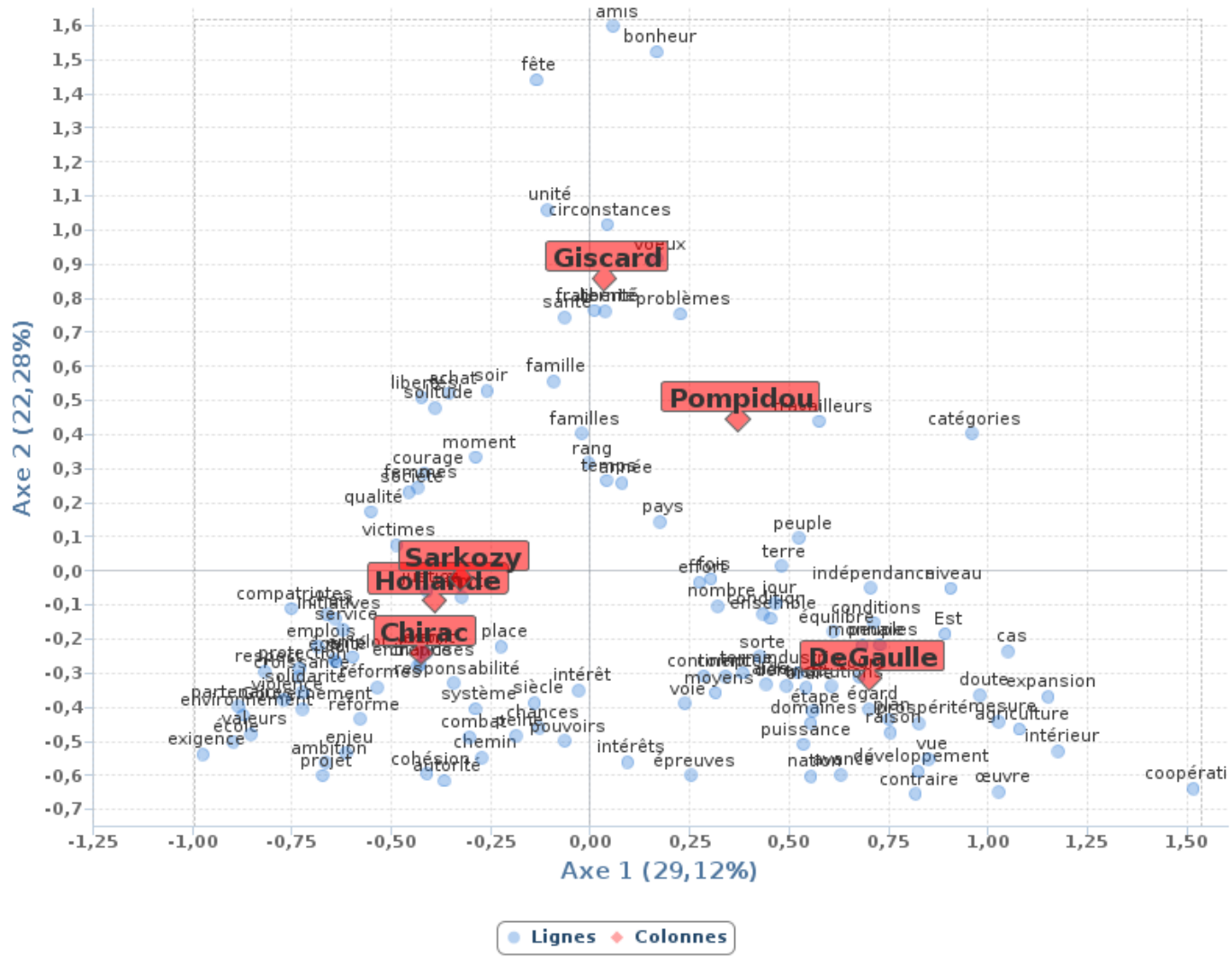
Fichier Outils Affichage Aide

Accès rapide

Presidents: [frpos="NOM"]:word | Presidents: word | Presidents:word

Axes : (1,2)

Plan factoriel de l'analyse des correspondances
sur la partition Presidents du corpus VOEUX



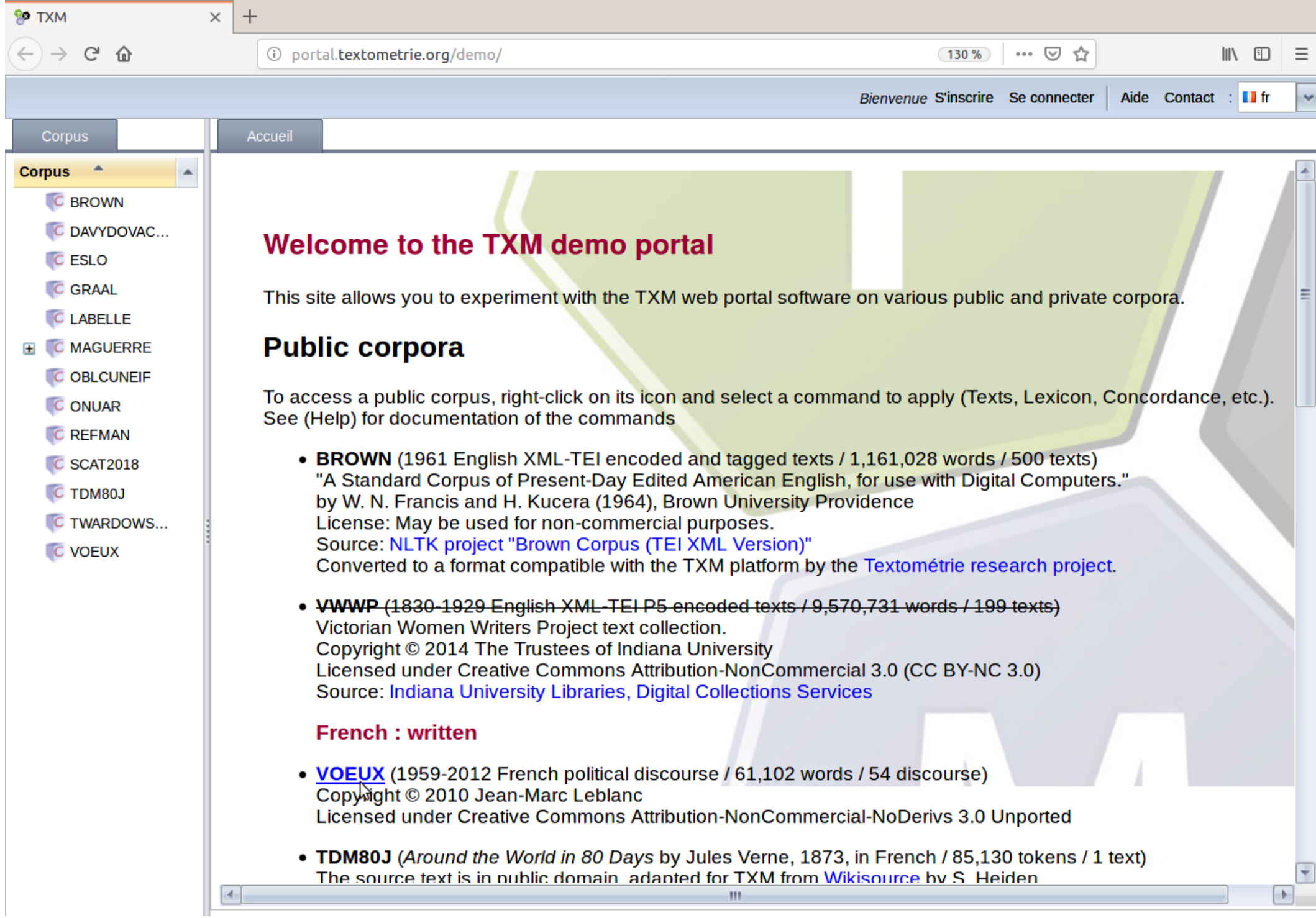
Colonnes	Q12	Q13	Q23	Mass	Dist	Co
DeGaulle	,94	,78	,16	19,78	0,63	56,2
Chirac	,75	,66	,27	28,91	0,32	30,1
Sarkozy	,18	,18	,00	11,85	0,57	7,2
Pompidou	,19	,52	,56	5,16	1,78	4,0
Hollande	,09	,10	,02	2,27	0,43	2,0
Mitterrand	,02	,62	,63	20,48	0,34	0,2
Giscard	,81	,00	,81	11,55	0,91	0,0

Valeurs propres | Info lignes | Info colonnes

Textométrie sur corpus structurés et annotés (1)

Requêtes : du plus simple au plus expressif

- Moteur de recherche CQP (univ. Stuttgart), langage d'interrogation CQL (Corpus Query Language)
  - Exemples d'interrogation sur la page d'accueil du corpus VOEUX sur le portail de démonstration :
    - http://portal.textometrie.org/demo/
    - Un motif linguistique : *Je … souhaiter … année*
      - [frlemma="je"][]*[frlemma="souhaiter"][]*[frlemma="année"] within s
  - Le langage CQL permet :
    - Combinaisons d'informations différentes
    - Discontinuités, longueur variable, alternatives, éléments optionnels
    - Conditions de contextualisation plus globale (en fonction des informations du corpus) : phrase, paragraphe, chapitre, locuteur, année, etc.
      - [word="Algérie" & _.text_loc!="dg"]


TXM
portal.textometrie.org/demo/
130 %
Bienvenue S'inscrire Se connecter Aide Contact : fr
Corpus
Accueil
Corpus
BROWN
DAVYDOVAC...
ESLO
GRAAL
LABELLE
MAGUERRE
OBLCUNEIF
ONUAR
REFMAN
SCAT2018
TDM80J
TWARDOWS...
VOEUX
Welcome to the TXM demo portal
This site allows you to experiment with the TXM web portal software on various public and private corpora.
Public corpora
To access a public corpus, right-click on its icon and select a command to apply (Texts, Lexicon, Concordance, etc.).
See (Help) for documentation of the commands
BROWN (1961 English XML-TEI encoded and tagged texts / 1,161,028 words / 500 texts)
"A Standard Corpus of Present-Day Edited American English, for use with Digital Computers."
by W. N. Francis and H. Kucera (1964), Brown University Providence
License: May be used for non-commercial purposes.
Source: NLTK project "Brown Corpus (TEI XML Version)"
Converted to a format compatible with the TXM platform by the Textométrie research project.
VWWP (1830-1929 English XML-TEI P5 encoded texts / 9,570,731 words / 199 texts)
Victorian Women Writers Project text collection.
Copyright © 2014 The Trustees of Indiana University
Licensed under Creative Commons Attribution-NonCommercial 3.0 (CC BY-NC 3.0)
Source: Indiana University Libraries, Digital Collections Services
French : written
VOEUX (1959-2012 French political discourse / 61,102 words / 54 discourse)
Copyright © 2010 Jean-Marc Leblanc
Licensed under Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported
TDM80J (Around the World in 80 Days by Jules Verne, 1873, in French / 85,130 tokens / 1 text)
The source text is in public domain, adapted for TXM from Wikisource by S. Heiden

TXM

portal.textometrie.org/demo/ 130 %

Bienvenue S'inscrire Se connecter | Aide Contact : fr

Corpus | Accueil | VOEUX

Corpus

- BROWN
- DAVYDOVAC...
- ESLO
- GRAAL
- LABELLE
- MAGUERRE
- OBLCUNEIF
- ONUAR
- REFMAN
- SCAT2018
- TDM80J
- TWARDOWS...
- VOEUX

Lexique

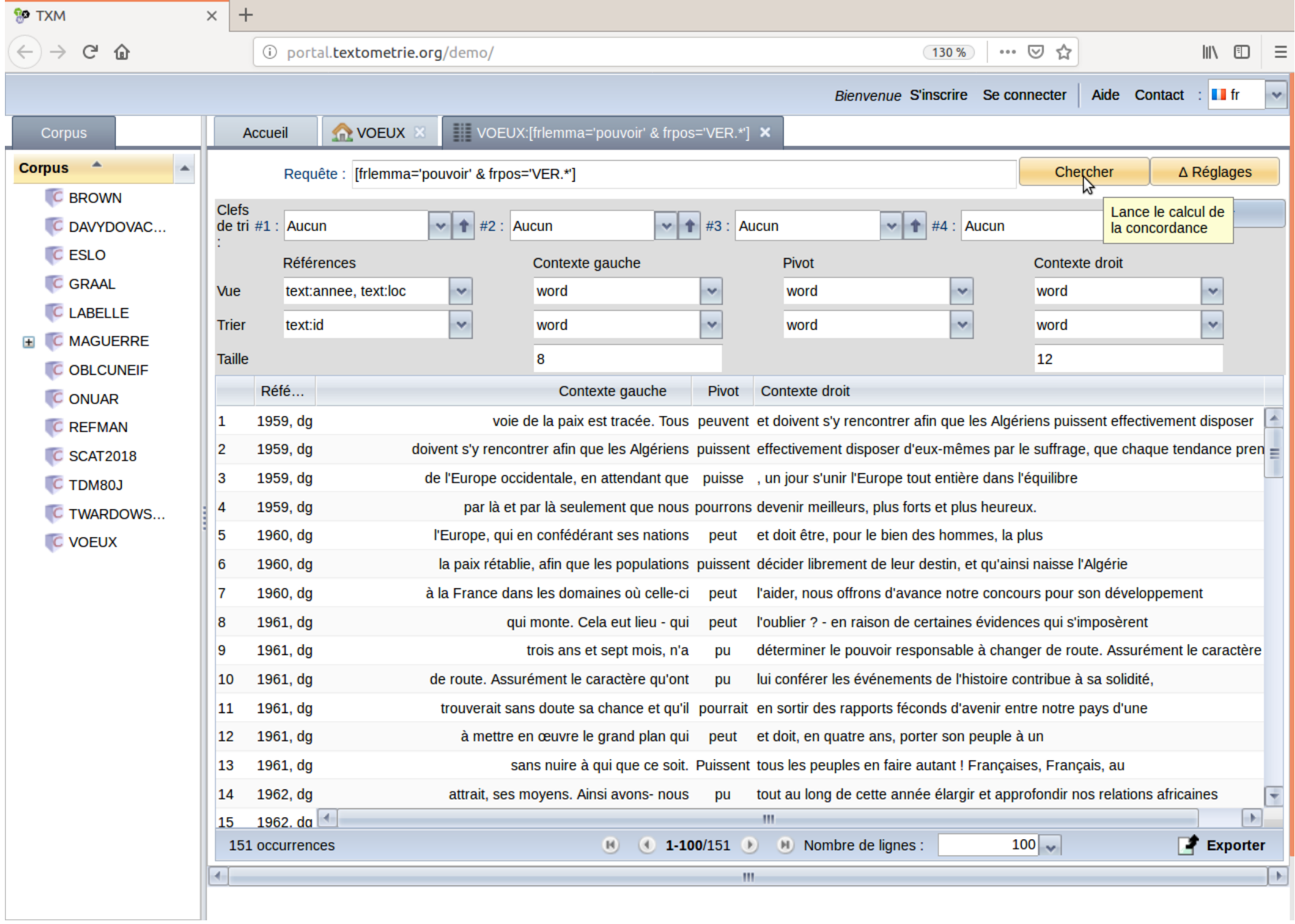
- calculer le lexique complet (double-cliquer sur une ligne pour calculer la concordance du mot) ;
- calculer le lexique des substantifs ;
- calculer le lexique des noms propres ;

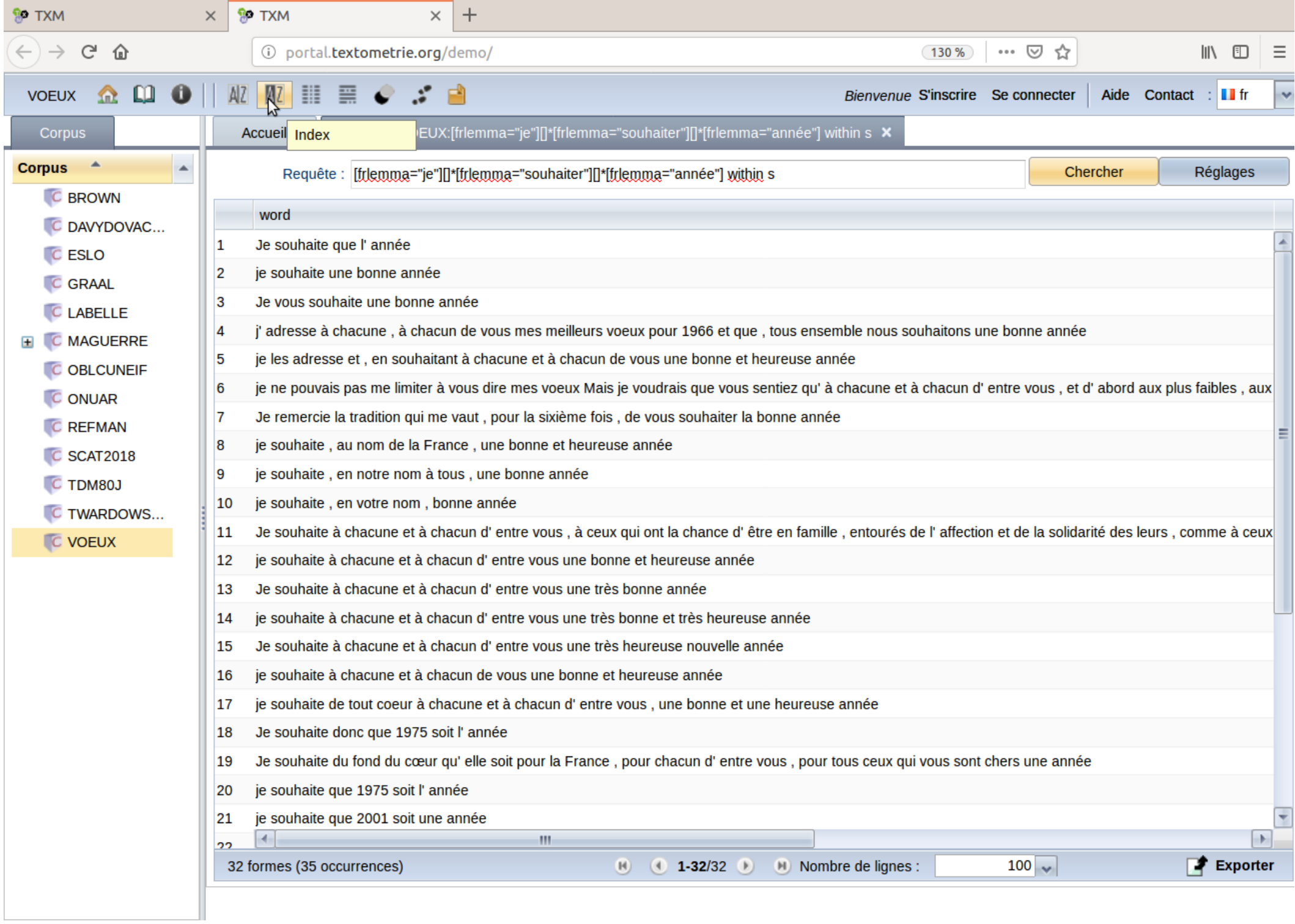
Concordance

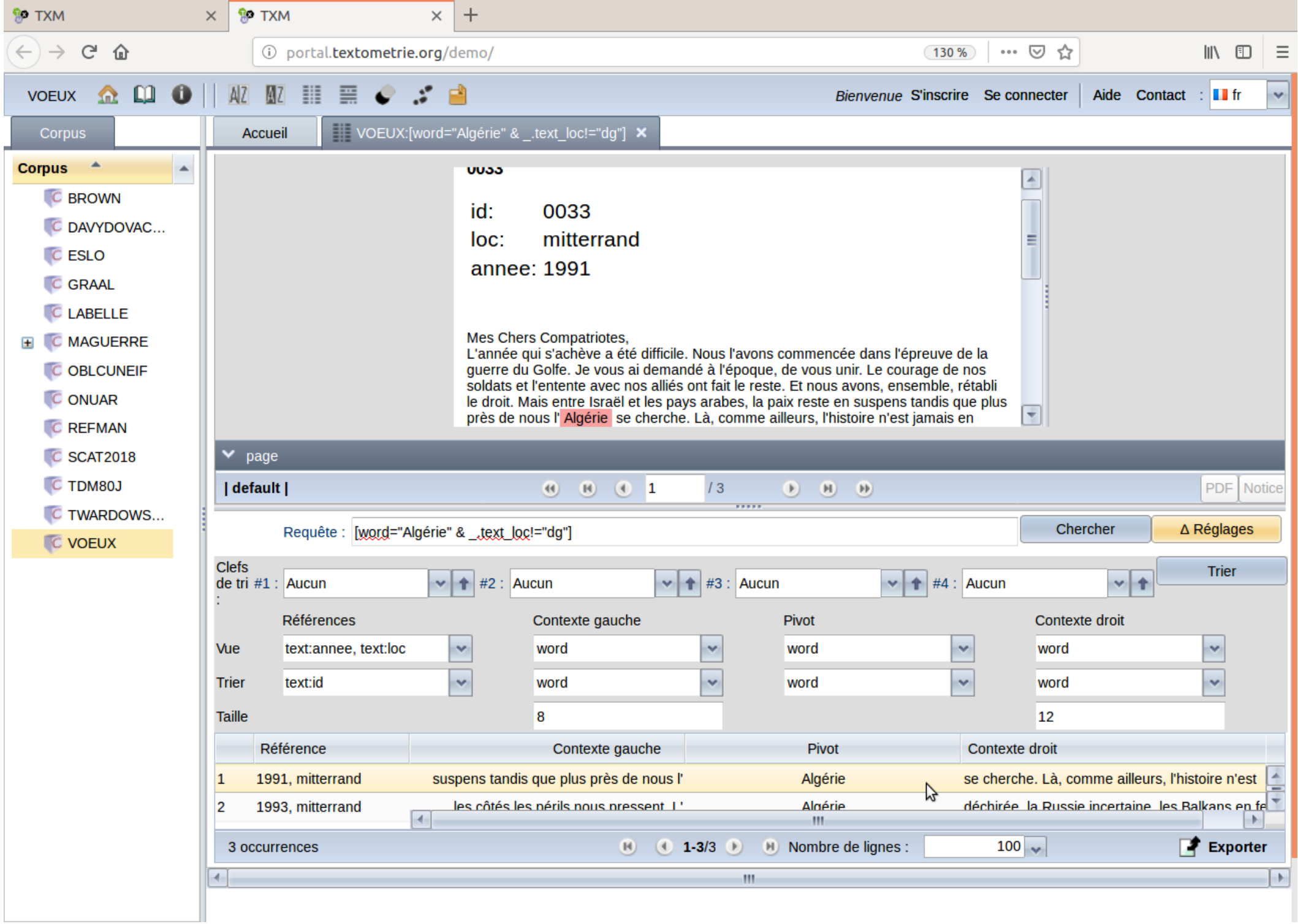
- calculer la concordance :
  - d'un mot : « temps » (double-cliquer sur une ligne de concordance pour lire l'occurrence du pivot en p texte dans l'édition) ;
  - d'un préfixe : « souhait.* » ;
  - d'une disjonction : « européen.*|Europe » ;
  - d'un lemme particulier : « [frlemma='souhaiter'] » ;
  - d'une combinaison de lemme et de partie du discours : « [frlemma='jeune' & frpos='NOM'] » « [frlemma='pouvoir' & frpos='VER.*'] » ;

Index

- calculer la fréquence du mot « bonheur » ;
- lister les réalisations du radical .*patri.* ;
- lister comment est qualifié le pays (index de [word='pays|nation'] [frpos='ADJ']) ;
- lister ce qui est français (index de [frpos='NOM'] [word='français.*']) ;
- lister les différentes politiques (index de [frlemma='politique' & frpos='NOM'] [frpos='ADJ']) ;


TXM
portal.textometrie.org/demo/
130 %
Bienvenue S'inscrire Se connecter Aide Contact : fr
Corpus
Accueil
VOEUX
VOEUX:[frlemma='pouvoir' & frpos='VER.*']
Corpus
BROWN
DAVYDOVAC...
ESLO
GRAAL
LABELLE
MAGUERRE
OBLCUNEIF
ONUAR
REFMAN
SCAT2018
TDM80J
TWARDOWS...
VOEUX
Requête : [frlemma='pouvoir' & frpos='VER.*']
Chercher
Δ Réglages
Lance le calcul de la concordance
Clefs de tri : #1 : Aucun #2 : Aucun #3 : Aucun #4 : Aucun
Références
Contexte gauche
Pivot
Contexte droit
Vue text:annee, text:loc word word word
Trier text:id word word word
Taille 8 12
Réfé... Contexte gauche Pivot Contexte droit
1 1959, dg voie de la paix est tracée. Tous peuvent et doivent s'y rencontrer afin que les Algériens puissent effectivement disposer
2 1959, dg doivent s'y rencontrer afin que les Algériens puissent effectivement disposer d'eux-mêmes par le suffrage, que chaque tendance pren
3 1959, dg de l'Europe occidentale, en attendant que puisse , un jour s'unir l'Europe tout entière dans l'équilibre
4 1959, dg par là et par là seulement que nous pourrons devenir meilleurs, plus forts et plus heureux.
5 1960, dg l'Europe, qui en confédérant ses nations peut et doit être, pour le bien des hommes, la plus
6 1960, dg la paix rétablie, afin que les populations puissent décider librement de leur destin, et qu'ainsi naisse l'Algérie
7 1960, dg à la France dans les domaines où celle-ci peut l'aider, nous offrons d'avance notre concours pour son développement
8 1961, dg qui monte. Cela eut lieu - qui peut l'oublier ? - en raison de certaines évidences qui s'imposèrent
9 1961, dg trois ans et sept mois, n'a pu déterminer le pouvoir responsable à changer de route. Assurément le caractère
10 1961, dg de route. Assurément le caractère qu'ont pu lui conférer les événements de l'histoire contribue à sa solidité,
11 1961, dg trouverait sans doute sa chance et qu'il pourrait en sortir des rapports féconds d'avenir entre notre pays d'une
12 1961, dg à mettre en œuvre le grand plan qui peut et doit, en quatre ans, porter son peuple à un
13 1961, dg sans nuire à qui que ce soit. Puissent tous les peuples en faire autant ! Françaises, Français, au
14 1962, dg attrait, ses moyens. Ainsi avons- nous pu tout au long de cette année élargir et approfondir nos relations africaines
15 1962, dg
151 occurrences 1-100/151 Nombre de lignes : 100 Exporter


TXM
TXM
portal.textometrie.org/demo/
130 %
VOEUX
Bienvenue S'inscrire Se connecter Aide Contact : fr
Corpus
Accueil
Index
EUX:[frlemma="je"][]*[frlemma="souhaiter"][]*[frlemma="année"] within s
Corpus
BROWN
DAVYDOVAC...
ESLO
GRAAL
LABELLE
MAGUERRE
OBLCUNEIF
ONUAR
REFMAN
SCAT2018
TDM80J
TWARDOWS...
VOEUX
Requête : [frlemma="je"][]*[frlemma="souhaiter"][]*[frlemma="année"] within s
Chercher
Réglages
word
1 Je souhaite que l' année
2 je souhaite une bonne année
3 Je vous souhaite une bonne année
4 j' adresse à chacune , à chacun de vous mes meilleurs voeux pour 1966 et que , tous ensemble nous souhaitons une bonne année
5 je les adresse et , en souhaitant à chacune et à chacun de vous une bonne et heureuse année
6 je ne pouvais pas me limiter à vous dire mes voeux Mais je voudrais que vous sentiez qu' à chacune et à chacun d' entre vous , et d' abord aux plus faibles , aux
7 Je remercie la tradition qui me vaut , pour la sixième fois , de vous souhaiter la bonne année
8 je souhaite , au nom de la France , une bonne et heureuse année
9 je souhaite , en notre nom à tous , une bonne année
10 je souhaite , en votre nom , bonne année
11 Je souhaite à chacune et à chacun d' entre vous , à ceux qui ont la chance d' être en famille , entourés de l' affection et de la solidarité des leurs , comme à ceux
12 je souhaite à chacune et à chacun d' entre vous une bonne et heureuse année
13 Je souhaite à chacune et à chacun d' entre vous une très bonne année
14 je souhaite à chacune et à chacun d' entre vous une très bonne et très heureuse année
15 Je souhaite à chacune et à chacun d' entre vous une très heureuse nouvelle année
16 je souhaite à chacune et à chacun de vous une bonne et heureuse année
17 je souhaite de tout coeur à chacune et à chacun d' entre vous , une bonne et une heureuse année
18 Je souhaite donc que 1975 soit l' année
19 Je souhaite du fond du cœur qu' elle soit pour la France , pour chacun d' entre vous , pour tous ceux qui vous sont chers une année
20 je souhaite que 1975 soit l' année
21 je souhaite que 2001 soit une année
32 formes (35 occurrences)
1-32/32
Nombre de lignes : 100
Exporter


TXM
TXM
portal.textometrie.org/demo/
130 %
VOEUX
Bienvenue S'inscrire Se connecter Aide Contact : fr
Corpus
Accueil
VOEUX:[word="Algérie" & _.text_loc!="dg"]
Corpus
BROWN
DAVYDOVAC...
ESLO
GRAAL
LABELLE
MAGUERRE
OBLCUNEIF
ONUAR
REFMAN
SCAT2018
TDM80J
TWARDOWS...
VOEUX
0033
id: 0033
loc: mitterrand
annee: 1991
Mes Chers Compatriotes,
L'année qui s'achève a été difficile. Nous l'avons commencée dans l'épreuve de la guerre du Golfe. Je vous ai demandé à l'époque, de vous unir. Le courage de nos soldats et l'entente avec nos alliés ont fait le reste. Et nous avons, ensemble, rétabli le droit. Mais entre Israël et les pays arabes, la paix reste en suspens tandis que plus près de nous l' Algérie se cherche. Là, comme ailleurs, l'histoire n'est jamais en
page
| default |
1 / 3
PDF
Notice
Requête : [word="Algérie" & _.text_loc!="dg"]
Chercher
Δ Réglages
Clefs de tri :
#1 : Aucun
#2 : Aucun
#3 : Aucun
#4 : Aucun
Trier
Références
Contexte gauche
Pivot
Contexte droit
Vue
text:annee, text:loc
word
word
word
Trier
text:id
word
word
word
Taille
8
12
Référence
Contexte gauche
Pivot
Contexte droit
1 1991, mitterrand suspens tandis que plus près de nous l' Algérie se cherche. Là, comme ailleurs, l'histoire n'est
2 1993, mitterrand les côtés les périls nous pressent. L' Algérie déchirée, la Russie incertaine, les Balkans en fe
3 occurrences
1-3/3
Nombre de lignes : 100
Exporter

Textométrie sur corpus structurés et annotés (2)

Intégration fine des annotations aux calculs

- Dans les différentes fonctionnalités, on peut distinguer
  - Ce que l'on cherche,
  - Ce que l'on groupe pour l'analyse, l'angle sous lequel on regarde le texte, ce qui « type » les unités (formes lexicales, lemmes, catégories, etc.)
  - Ce que l'on voit, affiché à l'écran

(cf. Pincemin JADT 2004)

- ex. concordance du verbe « aller » (*lemme*) triée sur l'*étiquette morphosyntaxique* et visualisée en *graphie*.

TXM | TXM

portal.textometrie.org/demo/ 130 %

VOEUX | Bienvenue S'inscrire Se connecter | Aide Contact : fr

Corpus | Accueil | VOEUX:[frlemma="aller"]

Corpus
- BROWN
- DAVYDOVAC...
- ESLO
- GRAAL
- LABELLE
- MAGUERRE
- OBLCUNEIF
- ONUAR
- REFMAN
- SCAT2018
- TDM80J
- TWARDOWS...
- VOEUX

Requête : [frlemma="aller"] Chercher Δ Réglages

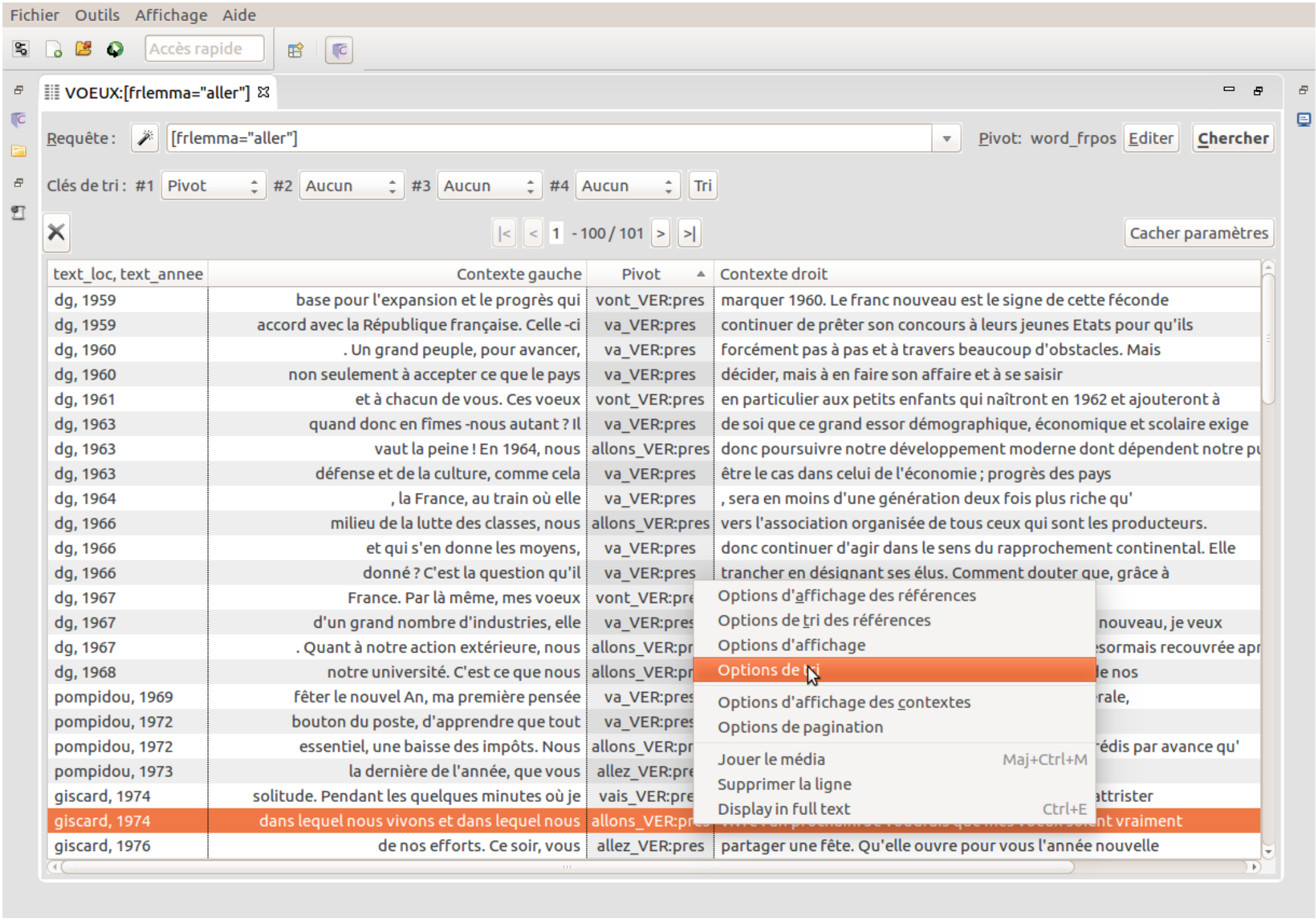
Clefs de tri : #1 : Pivot #2 : Référence #3 : Aucun #4 : Aucun Trier

Trie les résultats

	Références	Contexte gauche	Pivot	Contexte droit
Vue	text:annee, text:loc	word	word, frpos	word
Trier	text:annee	word	frpos	word
Taille		8		12

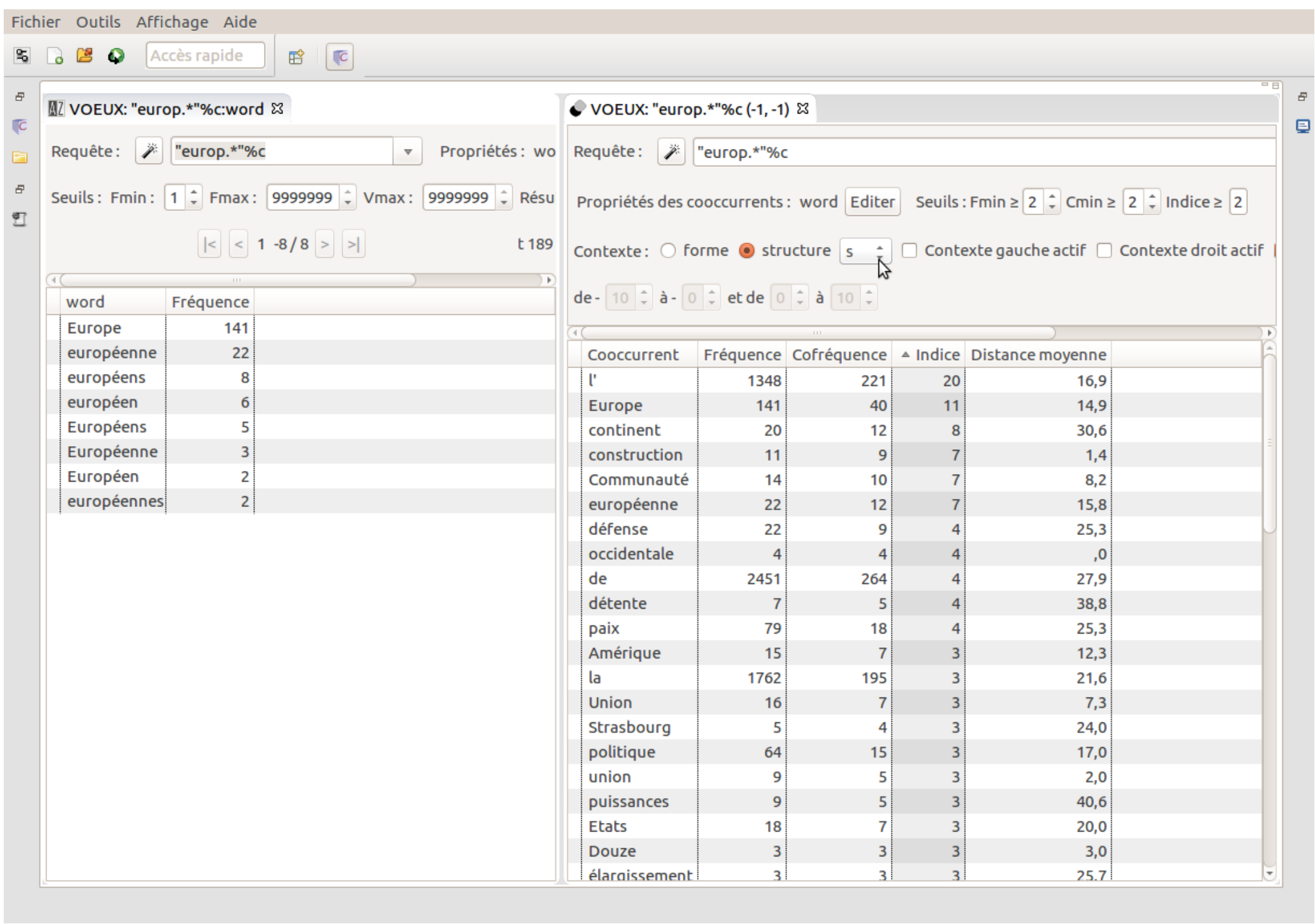
	Référence	Contexte gauche	Pivot	Contexte droit
1	1959, dg	base pour l'expansion et le progrès qui	vont_VER:pres	marquer 1960. Le franc nouveau est le signe de cette féconde
2	1959, dg	en accord avec la République française. Celle-ci	va_VER:pres	continuer de prêter son concours à leurs jeunes Etats pour qu'ils
3	1960, dg	. Un grand peuple, pour avancer,	va_VER:pres	forcément pas à pas et à travers beaucoup d'obstacles. Mais
4	1960, dg	non seulement à accepter ce que le pays	va_VER:pres	décider, mais à en faire son affaire et à se saisir
5	1961, dg	et à chacun de vous. Ces voeux	vont_VER:pres	en particulier aux petits enfants qui naîtront en 1962 et ajouteront à
6	1963, dg	, quand donc en fîmes-nous autant ? Il	va_VER:pres	de soi que ce grand essor démographique, économique et scolaire exige
7	1963, dg	vaut la peine ! En 1964, nous	allons_VER:pres	donc poursuivre notre développement moderne dont dépendent notre puiss
8	1963, dg	défense et de la culture, comme cela	va_VER:pres	être le cas dans celui de l'économie ; progrès des pays
9	1964, dg	, la France, au train où elle	va_VER:pres	, sera en moins d'une génération deux fois plus riche qu'
10	1966, dg	milieu de la lutte des classes, nous	allons_VER:pres	vers l'association organisée de tous ceux qui sont les producteurs.
11	1966, dg	et qui s'en donne les moyens,	va_VER:pres	donc continuer d'agir dans le sens du rapprochement continental. Elle
12	1966, dg	donné ? C'est la question qu'il	va_VER:pres	trancher en désignant ses élus. Comment douter que, grâce à
13	1967, dg	France. Par là même, mes voeux	vont_VER:pres	à chacune et à chacun de vous. Car, quand la
14	1967, dg	d'un grand nombre d'industries, elle	va_VER:pres	marquer une importante étape vers un ordre social nouveau, je veux
15	1967, dg			

101 occurrences 1-100/101 Nombre de lignes : 100 Exporter


Fichier Outils Affichage Aide
Accès rapide
VOEUX:[frlemma="aller"]
Requête : [frlemma="aller"]
Pivot: word_frpos
Editer
Chercher
Clés de tri : #1 Pivot #2 Aucun #3 Aucun #4 Aucun Tri
|< < 1 - 100 / 101 > >|
Cacher paramètres
text_loc, text_annee
Contexte gauche
Pivot
Contexte droit
dg, 1959 base pour l'expansion et le progrès qui vont_VER:pres marquer 1960. Le franc nouveau est le signe de cette féconde
dg, 1959 accord avec la République française. Celle -ci va_VER:pres continuer de prêter son concours à leurs jeunes Etats pour qu'ils
dg, 1960 . Un grand peuple, pour avancer, va_VER:pres forcément pas à pas et à travers beaucoup d'obstacles. Mais
dg, 1960 non seulement à accepter ce que le pays va_VER:pres décider, mais à en faire son affaire et à se saisir
dg, 1961 et à chacun de vous. Ces voeux vont_VER:pres en particulier aux petits enfants qui naîtront en 1962 et ajouteront à
dg, 1963 quand donc en fîmes -nous autant ? Il va_VER:pres de soi que ce grand essor démographique, économique et scolaire exige
dg, 1963 vaut la peine ! En 1964, nous allons_VER:pres donc poursuivre notre développement moderne dont dépendent notre pu
dg, 1963 défense et de la culture, comme cela va_VER:pres être le cas dans celui de l'économie ; progrès des pays
dg, 1964 , la France, au train où elle va_VER:pres , sera en moins d'une génération deux fois plus riche qu'
dg, 1966 milieu de la lutte des classes, nous allons_VER:pres vers l'association organisée de tous ceux qui sont les producteurs.
dg, 1966 et qui s'en donne les moyens, va_VER:pres donc continuer d'agir dans le sens du rapprochement continental. Elle
dg, 1966 donné ? C'est la question qu'il va_VER:pres trancher en désignant ses élus. Comment douter que, grâce à
dg, 1967 France. Par là même, mes voeux vont_VER:pre
dg, 1967 d'un grand nombre d'industries, elle va_VER:pres nouveau, je veux
dg, 1967 . Quant à notre action extérieure, nous allons_VER:pr sormais recouvrée apr
dg, 1968 notre université. C'est ce que nous allons_VER:pr le nos
pompidou, 1969 fêter le nouvel An, ma première pensée va_VER:pres rale,
pompidou, 1972 bouton du poste, d'apprendre que tout va_VER:pres
pompidou, 1972 essentiel, une baisse des impôts. Nous allons_VER:pr rédis par avance qu'
pompidou, 1973 la dernière de l'année, que vous allez_VER:pre
giscard, 1974 solitude. Pendant les quelques minutes où je vais_VER:pre attrister
giscard, 1974 dans lequel nous vivons et dans lequel nous allons_VER:pr nt vraiment
giscard, 1976 de nos efforts. Ce soir, vous allez_VER:pres partager une fête. Qu'elle ouvre pour vous l'année nouvelle
Options d'affichage des références
Options de tri des références
Options d'affichage
Options de tri
Options d'affichage des contextes
Options de pagination
Jouer le média Maj+Ctrl+M
Supprimer la ligne
Display in full text Ctrl+E

Import : s’adapter aux données

- Idem : on va du plus simple (copier/coller) à des imports enrichis
  - Métadonnées sur les textes
    - Par ex. dans VOEUX : le président, l’année
  - Étiquetage des unités lexicales (et délimitation)
    - À la volée (couplage avec TreeTagger), ou
    - Prise en compte d’une annotation existante (manuelle, autre étiqueteur, étiquetage vérifié, etc.)
  - Structures internes → pour spécifier la portée des recherches et contextualiser les résultats
    - Phrases, paragraphes…
      - ex. dans VOEUX, cooccurrence de "europ.*"%c dans les phrases (s)
    - Parties, listes, catégorisations, entités nommées…
      - ex. dans les Notices des Actualités françaises, listes les descripteurs de lieu : <item_type="DEL">[descripteursaffcol]+</item>
    - Tours de parole (locuteurs, interviewer/interviewé)


Fichier Outils Affichage Aide
Accès rapide
VOEUX: "europ.*"%c:word
Requête : "europ.*"%c
Propriétés : wo
Seuils : Fmin : 1 Fmax : 9999999 Vmax : 9999999 Résu
1 -8 / 8
t 189
word Fréquence
Europe 141
européenne 22
européens 8
européen 6
Européens 5
Européenne 3
Européen 2
européennes 2
VOEUX: "europ.*"%c (-1, -1)
Requête : "europ.*"%c
Propriétés des cooccurrents : word Editer
Seuils : Fmin ≥ 2 Cmin ≥ 2 Indice ≥ 2
Contexte : forme structure s Contexte gauche actif Contexte droit actif
de - 10 à - 0 et de 0 à 10
Cooccurrent Fréquence Cofréquence Indice Distance moyenne
l' 1348 221 20 16,9
Europe 141 40 11 14,9
continent 20 12 8 30,6
construction 11 9 7 1,4
Communauté 14 10 7 8,2
européenne 22 12 7 15,8
défense 22 9 4 25,3
occidentale 4 4 4 ,0
de 2451 264 4 27,9
détente 7 5 4 38,8
paix 79 18 4 25,3
Amérique 15 7 3 12,3
la 1762 195 3 21,6
Union 16 7 3 7,3
Strasbourg 5 4 3 24,0
politique 64 15 3 17,0
union 9 5 3 2,0
puissances 9 5 3 40,6
Etats 18 7 3 20,0
Douze 3 3 3 3,0
élargissement 3 3 3 25,7

Fichier Outils Affichage Aide

Accès rapide

1945 - 6

Titre propre

La chasse aux V1

Résumé

Dans le Nord de la France, des centaines de points de départ de V1 ont été neutralisés, comme au village de Lumbres où une rampe de lancement a été retrouvée. Cependant des V1 continuent à être lancés par les allemands vers les côtes anglaises. Les chasseurs britanniques tentent de les détruire au dessus de la Manche, souvent avec succès.

Séquences

- Vue générale LUMBRES, au crépuscule, avec chevalets de mine

- 2 Plans d'une rampe de lancement de V1 examinée par des curieux

- Enfants courant sur la rampe

- Divers plans des éléments d'un V1 en partie détruit

- Plusieurs plans d'un V1, au départ et en vol

- Avion de chasse anglais, au décollage

- Chasseur anglais poursuivant un V1 et ce dernier piquant vers le sol

Descripteurs (Aff. Col.)

- DET Seconde Guerre mondiale
- DEI arme (v1)
- DEI avion de chasse (anglais)
- DEI rampe de lancement
- DEL France
- DEL Pas de Calais

default 6 / 553

AFNOTICES: <item_type="DEL">[descripteursaffcol]+</item>:word

Requête : <item_type="DEL">[descripteursaffcol]+</item> Propriété

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

|< < 1 -100 / 3264 > >| t 3440

word	Fréquence
France	6753
Paris	3304
Etats Unis	880
Belgique	773
Algérie	714
Grande Bretagne	499
Allemagne	458
Région parisienne	442
Angleterre	425
Maroc	393
Alger	389
Bruxelles	365
Tunisie	278
Italie	256
Londres	238
Autriche	224
Allemagne RF	214
URSS	180
New York	176
Indochine-Asie	161
République fédérale d' Allemagne	161
Colombes	160
Casablanca	138

Fichier Corpus Outils Affichage Aide

Accès rapide

P1S8 30 avril 2014 - 2

E: 0:08:06 *[bruit de fond]*

P: 0:08:06 heu 0:08:07 effectivement on peut difficilement faire mieux que voir ça *[montre le spectre de lumière blanche]* c' est-à-dire que 0:08:14 ils sont pas assez dispersifs ils dispersent pas la lumière assez précisément ces spectroscopes pour pouvoir voir les petites bandes noires 0:08:23 donc là il faut prendre un spectroscope beaucoup plus perfectionné d' accord le vieux spectroscope là très 19ème siècle que vous avez utilisé mardi à la limite celui-ci 0:08:34 si on le pointait vers le ciel on pourrait commencer à voir les raies noires 0:08:38 il faut déjà prendre un spectroscope de très bonne qualité donc c' est pour ça que je vous propose là je vous l' ai mis en français dans l' activité il y a des raies noires 0:08:46

[Jeu 3: Comprendre ce qui a été fait, récupérer les bons documents et répondre aux questions sur le spectre de raies du soleil]

report: , id: 4, theme: Thème 3 : Reconnaissance des spectres d'absorption, organisation: class

0:08:46 la question c' est maintenant comment à l' aide de l' ensemble du modèle que vous avez sous les yeux vous pouvez avec une phrase ou deux pas plus interpréter l' existence de ces raies noires c' était la première question de euh cette activité 5 0:09:03 *[voir texte du chapitre 4: Activité 5 : Détermination de la composition d'une étoile à l'aide de son spectre]* donc c' est parti je vous laisse échanger sur la première et sur la deuxième question et puis sur le petit exercice à faire à la maison 0:09:09 même si vous du coup je pensais avoir plus de temps que ça mardi vous avez tout fait à la maison les l' exercice et les deux premières questions 0:09:15

E: 0:09:15 *[inaud.]*

P: 0:09:20 alors pendant que vous commencez à réfléchir à ça à deux ou à trois je donne ce qu' il manque *[le professeur commence à distribuer]* 0:09:25

report: , id: 5, theme: none, organisation: group

[Les élèves travaillent en groupe] [le professeur donne des documents]

0:09:25 tu l' as le document

E: 0:09:27 oui oui 0:09:30

default 2 / 12

Retour au document source

- Edition synoptique
  - TDM80J : Jules Verne, *Le tour du monde en 80 jours*
    - Import XTZ (« XML TEI Zero »)
    - Images en ligne (Wikisource)
  - GRAAL : La *Queste del saint Graal*
    - Alignement avec un manuscrit
- Multimédia
  - VOIXOFF : projet ANR ANTRACT, corpus de transcriptions automatiques alignées sur les vidéos des *Actualités françaises* (1945-1969)
    - Le retour à la vidéo utilise l'extension MediaPlayer (niveau beta)

TXM | TXM

portal.textometrie.org/demo/ 130 %

TDM80J | Bienvenue S'inscrire Se connecter | Aide Contact : fr

Corpus | Accueil | Text: tdm80j | TDM80J:maître | Lexique /TDM80J

Corpus

- BROWN
- DAVYDOVAC…
- ESLO
- GRAAL
- LABELLE
- MAGUERRE
- OBLCUNEIF
- ONUAR
- REFMAN
- SCAT2018
- TDM80J
- TWARDOWS…
- VOEUX

Chapitre I.
Dans lequel Phileas Fogg et Passepartout s'acceptent réciproquement l'un comme maître, l'autre comme domestique

En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens – maison dans laquelle Sheridan mourut en 1814 –, était

PHILEAS FOGG.

I

DANS LEQUEL PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT, L'UN COMME MAITRE, L'AUTRE COMME DOMESTIQUE.

En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens, — maison dans laquelle Shéridan mourut en 1814, — était habitée par Phileas Fogg, esq., l'un des membres les plus singuliers et les plus remar-

page | TDM80J - default, TD …

| default | facs | 2 / 219 | PDF | Notice

Requête : maître | Chercher | ∇ Réglages

	Ré…	Contexte gauche	Pivot	Contexte droit
1	tdm80j	et Passepartout s'acceptent réciproquement l'un comme	maître	, l'autre comme domestique En l'année 1872, la maison
2	tdm80j	une première fois : c'était son nouveau	maître	qui sortait ; puis une seconde fois : c'était son prédécesseur
3	tdm80j	Tussaud des bonshommes aussi vivants que mon nouveau	maître	! » Il convient de dire ici que les « bonshommes »

120 occurrences | 1-100/120 | Nombre de lignes : 100 | Exporter

TXM | TXM

portal.textometrie.org/demo/ 130 %

GRAAL | Bienvenue S'inscrire Se connecter | Aide Contact : fr

Corpus | Accueil | GRAAL | Text: qgraal_cm

Corpus

- BROWN
- DAVYDOVAC...
- ESLO
- GRAAL
- LABELLE
- MAGUERRE
- OBLCUNEIF
- ONUAR
- REFMAN
- SCAT2018
- TDM80J
- TWARDOWS...
- VOEUX

<160a>

Ici commence la version de la Queste del saint Graal donnée par le manuscrit K (Bibliothèque Municipale de Lyon, Palais des Arts n° 77), folios 160 recto à 224 verso. Tout le début du texte a été mutilé : la première grande lettre a été découpée, comme on le voit sur la reproduction du manuscrit, et quelques lignes du texte manquent, que nous donnons ici entre crochets, en bleu, d'après le manuscrit Z (Paris, BNF n. acq. fr. 1119, folio 138 recto, colonne a) qui est un manuscrit proche de celui que nous éditons ici.

(p. 1)

[§ 1]

[A la veille de la Pente-
coste quant li compai-
gnon de la table re-
onde furent venu
a Kamaalot et il o-
rent oï le servisse et
l'en voloit metre les
tables a heure de]
nonne.´ lors en[tra] [a cheval en la][1]
damoisele, et fu venue si grant oirre
l'en veoir car ses chevaux en fu enc

<160a>

Ici commence la version de la Queste del saint Graal donnée par le manuscrit K (Bibliothèque Municipale de Lyon, Palais des Arts n° 77), folios 160 recto à 224 verso. Tout le début du texte a été mutilé : la première grande lettre a été découpée, comme on le voit sur la reproduction du manuscrit, et quelques lignes du texte manquent, que nous donnons ici entre crochets, en bleu, d'après le manuscrit Z (Paris, BNF n. acq. fr. 1119, folio 138 recto, colonne a) qui est un manuscrit proche de celui que nous éditons ici.

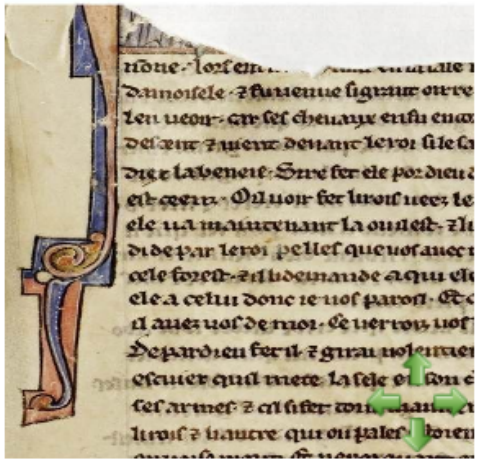
(p. 1)

[§ 1]

A [la ueille delapente
coſte qñt li compaí
gnon de latable re
onde furent uenu
a kamaalot ⁊ il o
rent oı leſeruiſſe ⁊
len uoloit metre leſ
tableſ a heure de]
nō̄e ·ꝛ· loꝛſ en[tra] [acheual en la][1] ſale
damoiſele . ꝛ fuuenue ſigꝛant oirre que b
len ueour car ſeſ cheuaux enfu encoꝛe te

Fragment du ms. Z (B
à la place d'un fragment manquant

Retour au manuscrit K (Lyon, BM, P.A.



page | GRAAL - courante, GR...

| courante | fac-similaire | ms-colonne | 160a / 268 PDF Notice

Fichier Outils Affichage Aide

Accès rapide

AFE02014770 - 1

S 3: 0:00:02 l'algérie récompense les ouvriers qui sont toujours restés fidèles à l'antique tradition de l'artisanat algérien

S 2: 0:00:08 pour faux

S 3: 0:00:17 c'est au palais consulaire en présence d'une foule nombreuse que monsieur berniole qu'accompagnait monsieur tom secrétaire général du gouvernement générale à distribuer aux meilleurs ouvriers les récompenses de leurs efforts et de leur valeur

default 1 / 1

AFE02014770.mp4

REPRE 00:20111 Répéter Taux: Vol :

AFVOIXOFFSUJ:[frplemma="foule"]

Requête : [frplemma="foule"] Pivot: word Editer **Chercher**

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

|< < 1 - 100 / 1134 > >| Cacher paramètres

ref	Contexte gauche	Pivot	Contexte droit
AFE00001865, S21, 0:02:14	reine de pornichet dans radio sourire saluer la	foule	s'excuser tout à la fois de est lui karen suspecte non
AFE02000035, S1, 0:00:10	chirurgiens américains y réussir à la tauber une	foule	anxieuse et ému en empathie composé des médecins qui l'a
AFE02014622, S1, 0:00:08	elle devait le départ donné à cuba de	foule	nombreuse coloré et curieux regarde les géant de la route a
AFE02014740, S1, 0:00:02	tous les ans sur les quais algérois une	foule	pittoresques et variées derrière lui la lutte pour < unk > de
AFE02014770, S3, 0:00:17	est au palais consulaire en présence d'une	foule	nombreuse que monsieur berniole qu'accompagnait monsi
AFE02014774, S0, 0:00:01	sur la piste du stade municipal ou une	foule	énorme et puis d'admirer à loisir les prescrire mais les engin
AFE02014803, S3, 0:00:27	incessamment poursuivis pour l'industrialisation du maroc la	foule	

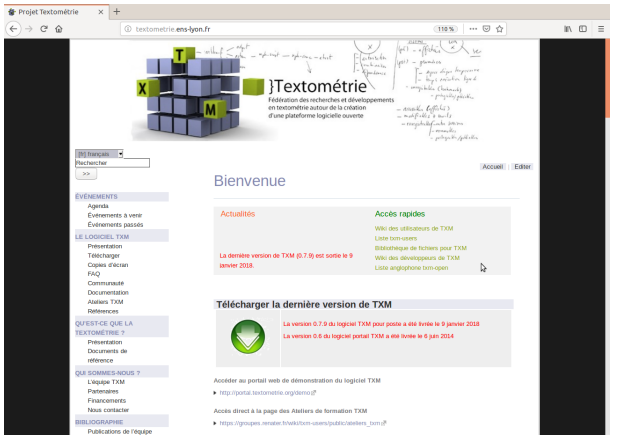
Ouverture des traitements (1)

- Macros et scripts
  - Pour augmenter ou personnaliser les traitements sans rentrer dans le code « noyau » compilé du logiciel
    - Fonctionnalités non encore intégrées, ex. :
      - Vocabulaire de base (BasicVocabulary)
      - Filtrage d'un plan d'AFC en fonction des cos2 et contributions (CAFilter – prototype non encore diffusé)
      - Construction d'une liste utilisable dans les requêtes (CreateCQPList)
      - Calcul direct d'un score de spécificités par ses quatre paramètres (PlotSpecif)
    - Préparation de corpus, ex. :
      - Chercher/remplacer avancé et sur plusieurs fichiers (SearchReplaceinDirectory)
      - Changer l'encodage des caractères (ChangeEncoding)
      - Adapter un script d'import à ses données

Ouverture des traitements (2)

- Articulation avec R
  - Utilisation de R pour l'implémentation des calculs statistiques (AFC, CAH, …)
  - Contribution à R : spécificités implémentées dans le package Textometry
  - Vue R avec console R et objets R dans TXM
- Articulation avec d'autres logiciels
  - Passerelles d'import dans les deux sens avec Iramuteq
  - Exports de tableaux de données et de résultats en format tableur générique

Ressources en ligne



- Site du projet Textométrie
  - http://textometrie.ens-lyon.fr
  - Point d'accès général,
    cf. rubriques en marge et encadré en haut à droite
- Communauté utilisateurs (liste et wiki)
- Contact : équipe TXM (ENS Lyon et Univ. Besançon)
  - textometrie *à* groupes.renater.fr
- Un exemple de publication récente avec une problématique historique et utilisant TXM (et Iramuteq) :
  - Mayaffre Damon, Pincemin Bénédicte, Heiden Serge, Weyl Philippe (2018). « L'évolution de la mémoire de la Shoah au prisme de la statistique textuelle », *in* Denis Peschanski & Brigitte Sion (dir.), *La vérité du témoin. Mémoire et mémorialisation, vol. 2*, collection Mémoire(s), Hermann Éditeurs, Paris & Institut National de l'Audiovisuel, Bry-sur-Marne, chapitre VI, p. 93-124. https://halshs.archives-ouvertes.fr/hal-01890536v1