

Fiche de travail iTrameur

Faire des analyses textométriques sur les données du corpus

Vous avez créé votre corpus, il est maintenant conforme au format d'entrée attendu par [iTrameur](#) : vous pouvez passer aux analyses textométriques ! Dans cette fiche de travail nous prendrons le thème de la Grèce comme exemple principal (sur un corpus de test de 2015).

SOMMAIRE

Chargement	2
Charger un fichier dans iTrameur (nouvelle base)	2
Visualiser le corpus (vérification)	2
Carte des sections	2
Cadre	3
Exporter une base	4
Importer une base	5
Paramètres	5
Statistiques sur le corpus	6
Principales Caractéristiques Lexicographiques du Corpus (PCLC)	6
Spécificités	7
Spécificités par partie	7
Spécificités totales	8
Vocabulaire spécifique positif par partie	9
Tableau général des items (barycentre temporel, coefficient de Von Neumann)	9
Rechercher une forme	10
Dictionnaire	10
Carte des sections	11
Concordance	12
Ventilation	14
Fréquence absolue	14
Fréquence relative	15
Spécificités	16
Segments répétés	18
Cooccurents	20
Sélection	22

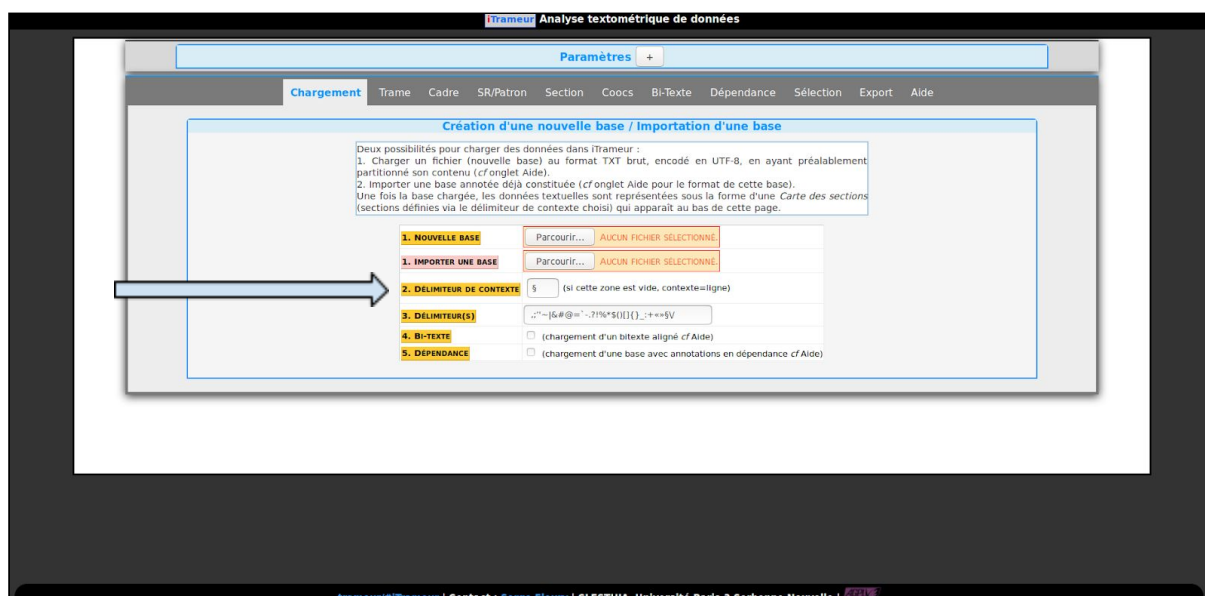
Si vous avez une question, n'hésitez pas à consulter l'onglet "aide" en haut à droite de l'application [iTrameur](#). Vous y trouverez aussi des exemples de fichiers d'exemples prétraités.

Chargement

Charger un fichier dans iTrameur (nouvelle base)

Le texte doit être au format TXT BRUT et être encodé en UTF-8. Il a normalement été préparé en amont pour être segmenté en sections avec le caractère délimiteur §. Il doit apparaître après chaque balise fermante </page>.

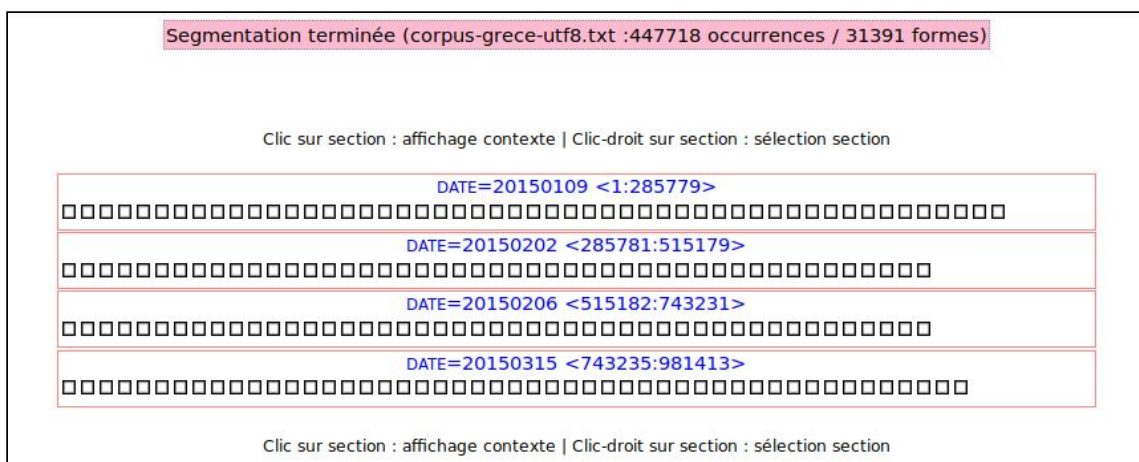
Pour créer une nouvelle base, il faut vérifier que le symbole § est bien le délimiteur de contexte et charger le fichier (1. Nouvelle base) :



Visualiser le corpus (vérification)

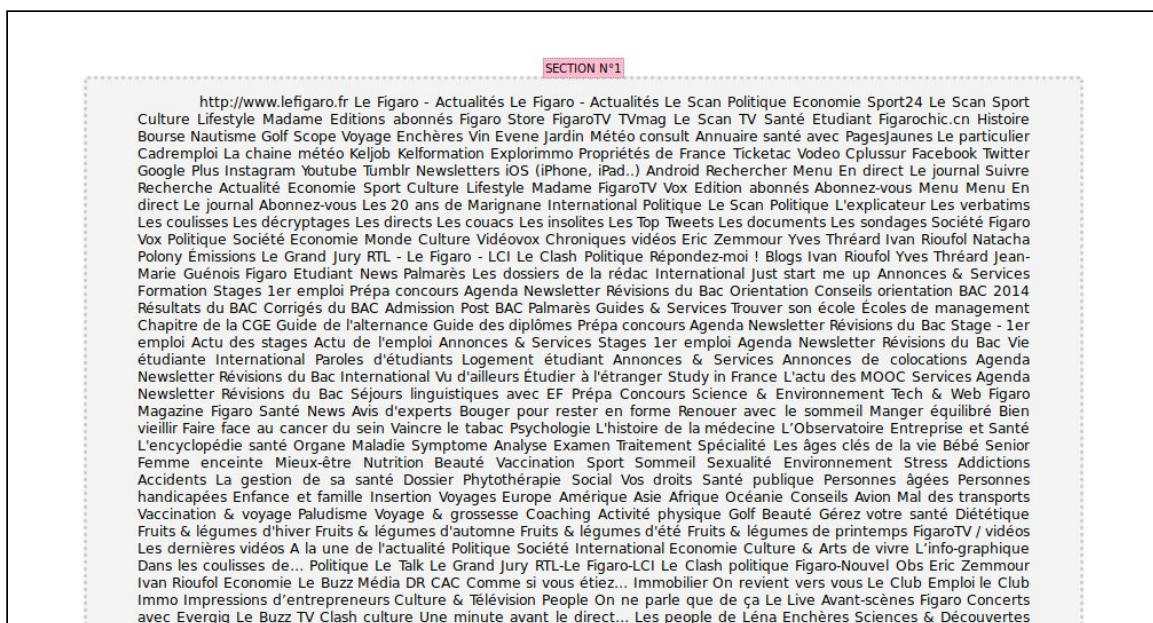
Carte des sections

Une fois le corpus chargé, la carte des sections apparaît automatiquement plus bas :



iTrameur a segmenté le corpus en “mots” en distinguant le nombre d’occurrences total du nombre de formes différentes.

Dans la carte des sections, chaque ligne correspond à une **aspiration** (= un élément **<date>**) et chaque carré correspond à un **journal** (= un élément **<page>**). Quand on clique sur un carré (donc sur une **section**), le texte correspondant à cette section apparaît en dessous :



C'est l'occasion de vérifier que tout semble normal en regardant le contenu de quelques sections, i.e. si la page aspirée n'est pas vide ou ne renvoie pas simplement un message d'erreur.

Cadre

Pour vérifier que le corpus est bien formé, il faut regarder le cadre :



Dans l'onglet "cadre", en cliquant sur le bouton "cadre", les parties vont apparaître. Elles sont définies par les balises du corpus. Il faut vérifier que les parties "**page**" sont bien imbriquées dans les parties "**date**" et que les bordures des rectangles qui représentent les parties ne se croisent pas.

À titre indicatif, la fin du cadre doit être similaire à :

page=nicematin.fr <971774:977833>

url=url <971775:971787>

title=title <971788:971808>

text=text <971809:977832>

page=lanouvellerepublique.fr <977839:981404>

url=url <977840:977852>

title=title <977853:977877>

text=text <977878:981403>

Si vous n'avez pas un affichage similaire, regardez ce qui s'affiche dans le menu de "partition" dans l'onglet "paramètres" en haut de la page. Il pourrait y avoir une balise mal formée, ex. < {{ latestCtrl.page }} / {{ latestCtrl.lastPage }} >.

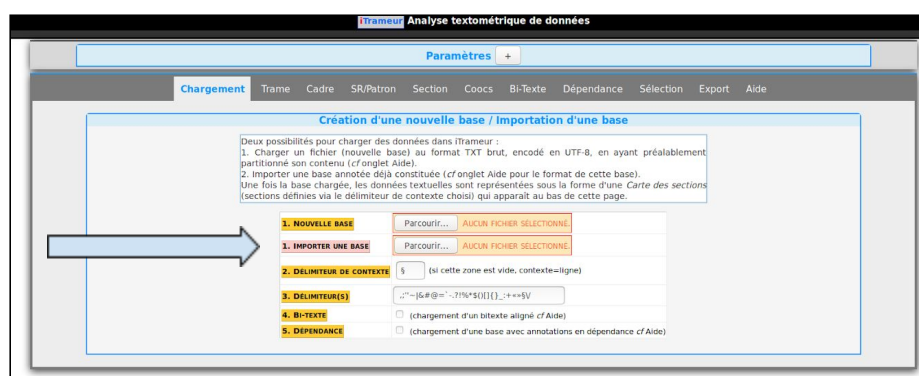
Exporter une base

Exporter une base permet de sauvegarder localement la base chargée dans un fichier au format TXT BRUT (encodé en UTF-8). Ce fichier contient **la trame** (système de **coordonnées** sur le texte : chaque item est repéré par son numéro d'ordre) et **le cadre** (information de **structure**) du corpus. Le fichier obtenu est un fichier tabulé (comme un tableau : les colonnes sont séparées par une tabulation et les lignes par un saut de ligne) avec un "mot" par ligne. Ce fichier peut être ouvert dans un éditeur de texte brut (ex. Notepad++) ou un tableur (ex. Excel). Il peut être corrigé et enrichi avec des annotations (ex : étiquetage morpho-syntactique) puis réimporté.



Importer une base

Il est possible d'importer une base déjà existante si elle a au moins 3 couches d'annotation (forme, lemme, pos). Pour cela, il suffit de charger le fichier dans "importation d'une base".



Paramètres

En haut de la fenêtre de l'application, il y a un onglet qui permet de définir des paramètres. Pour y accéder, cliquer sur la croix...



... permet de dérouler le menu :

Paramètres			
PARTITION	date	PARTIES	20150109
GRAPHE H	400	SEUIL	5
GRAPHE L	800	SR LGMAX	12
ANNOTATION	1	ANNOTATION SORTIE	1
		LG CONTEXTE	10
		CO-FREQ	2
		SR FQMIN	10
		RELATION	REL
		PÔLE SOURCE	Charlie
		FQ MAX	5
		INDSPMIN	5
		NB SÉLECTION SECTION	1
		ANNOTATION RELATION	4

Une **partition** est un critère utilisé pour diviser un corpus. Les **parties** sont les morceaux de corpus obtenues en appliquant une partition. Par exemple, la partition en dates permet d'avoir une partie pour chaque aspiration. Mais choisir une partition en pages permet d'avoir une partie pour chaque journal, quelle que soit la date. Ce sont des paramètres qu'il est possible de changer selon la recherche souhaitée.

Les informations concernant les paramétrages se trouvent à la fin de l'onglet "aide" :

Paramétrage

DELIMITEUR DE CONTEXTE : Cette zone de saisie doit contenir le caractère utilisé pour la construction de la *Carte des Sections* (et aussi pour déterminer les contextes utilisés pour le calcul des cooccurrences via la *Carte des Sections*).

DELIMITEUR(S) : Cette zone de saisie contient la liste des caractères délimiteurs utilisés pour segmenter le texte en formes graphiques.

BI-TEXTE : Cette case à cocher permet de charger un bi-texte aligné (et les fonctionnalités associées).

DÉPENDANCE : Cette case à cocher permet de charger les fonctionnalités associées aux traitements sur une base contenant des annotations en dépendance.

PARTITION : Liste permettant de sélectionner une partition.

PARTIE : Liste permettant de sélectionner une partie de la partition choisie.

ANNOTATION : Cette liste, mise à jour à l'issue du chargement d'une base annotée, permet de sélectionner une annotation pour réaliser le calcul visé.

ANNOTATION SORTIE : Cette liste, mise à jour à l'issue du chargement d'une base annotée, permet de sélectionner l'annotation à utiliser pour afficher les zones textuelles (concordance, section, contexte cooccurrence) quelle que soit l'annotation utilisée. Par exemple : concordance du pôle NOM (annotation n°3) et affichage en sortie via les formes graphiques (annotation n°1) des contextes visés. Par défaut l'annotation en sortie a la même valeur que l'annotation sélectionnée pour les calculs.

SEUIL : Par défaut, l'indice de spécificité est calculé avec un seuil de probabilité fixé à 5 %.

CO-FREQ : Par défaut, le calcul de cooccurrence est calculé en ne retenant que les candidats cooccurents dont la co-fréquence est supérieure à la valeur donnée.

INDSPMIN : Par défaut, le calcul de cooccurrence est calculé en ne retenant que les candidats cooccurents dont l'indice de spécificité est supérieur à la valeur donnée.

FQ MAX : Par défaut, le calcul des spécificités totales est calculé en ne retenant que les formes dont la fréquence est supérieure à la valeur donnée. Idem pour le calcul du "Réseau de cooccurents".

LG CONTEXTE : Longueur du contexte pour l'affichage d'une concordance.

GRAPHE H : Par défaut, les graphiques construits ont une hauteur correspondant à la valeur donnée.

GRAPHE L : Par défaut, les graphiques construits ont une largeur correspondant à la valeur donnée.

PÔLE : Zone de saisie utilisée pour définir le pôle visé (remplissage par auto-complétion).

NB SÉLECTION SECTION : Zone de saisie permettant de définir le nombre de sections à sélectionner simultanément (via le clic-droit).

RELATION : Zone de saisie permettant de saisir le nom d'une relation.

ANNOTATION RELATION : Zone de saisie permettant le numéro d'annotation portant le nom de la relation visée.

Statistiques sur le corpus

Principales Caractéristiques Lexicographiques du Corpus (PCLC)

Le calcul des **Principales Caractéristiques Lexicographiques du Corpus** (PCLC) permet d'obtenir respectivement pour chaque partie (en fonction de la partition choisie) le nombre de formes totales, le nombre de formes différentes, la forme la plus fréquente et le nombre d'occurrences de cette forme.

P.C.L.C

Recherche :

Partie	F occ	f typ	Forme max	Max
20150109	129503	15055	de	4921
20150202	105074	14905	de	4001
20150206	104333	15249	de	4156
20150315	108808	15224	de	4285

Affichage de 1 à 4 des 4 items

Préc. 1 Suiv.

Cela peut être intéressant de regarder les PCLC pour la partition en “pages” afin de vérifier qu’il n’y ait pas de résultats étranges. Dans la plupart des cas, la forme la plus fréquente est un “mot-outil” (ex : “de”, “à”, “le”). Ce n’est pas étonnant car il s’agit des mots les plus fréquents en français même si ce ne sont pas réellement eux qui sont vecteurs de sens. Dans le corpus d’exemple, dans la page du journal *Var Matin* c’est la forme “chaussure” qui est la plus fréquente. En suivant le lien de l’url qui a servi pour la création du corpus, il s’est avéré qu’il s’agissait d’un site de magasin de chaussures de sport. Le bon url pour ce journal aurait été www.varmatin.com et non www.varmatin.fr. Explorer son corpus avec ce type de fonctionnalité permet de détecter de potentielles erreurs et de les corriger (par exemple enlever les parties “page=varmatin” dans le corpus d’exemple).

P.C.L.C

Recherche :

Partie	F occ	f typ	Forme max	Max
nordlittoral.fr	3668	1195	de	138
nrpyrenees.com	3346	1170	de	133
paris-normandie.com	8785	2400	de	429
petitbleu.fr	4136	1381	de	156
presseeocean.fr	6350	1493	Nantes	239
pyrenees.com	5510	1316	de	246
republicain-lorrain.fr	5686	1625	de	247
republique-pyrenees.com	520	93	de	32
sudouest.com	11589	2499	de	435
varmatin.fr	2090	308	Chaussure	194

Affichage de 41 à 50 des 50 items

Préc. 1 2 3 4 5 Suiv.

Spécificités

Dans l’onglet “cadre”, on peut avoir accès à des statistiques sur les spécificités du corpus. La **spécificité** est une métrique qui permet de donner un aperçu global du vocabulaire surreprésenté ou sous-représenté dans un ensemble d’items. Une valeur de spécificité positive indique une **surreprésentation**. Il est donc statistiquement anormal que cet item soit aussi fréquent dans cet ensemble. Au contraire, une valeur négative indique une **sous-représentation**. Il est donc statistiquement étonnant que cet item soit aussi peu fréquent dans cet ensemble.

Par exemple, il peut y avoir un vocabulaire spécifique à une date en particulier par rapport aux autres dates, tous journaux confondus, en fonction de l'actualité (ex. attentats, jeux olympiques...). Il peut aussi y avoir un vocabulaire spécifique à un journal par rapport aux autres, toutes dates confondues, en fonction de sa ligne directrice (ex : *La Croix*, journal chrétien et catholique). Parallèlement, dans ces cas-là il peut aussi y avoir un vocabulaire sous-représenté.

Vous pouvez vous servir de cette fonctionnalité comme un tri sur le contenu. Il se peut que des éléments indésirables apparaissent alors qu'ils ne sont pas du "contenu textuel". Par exemple, si les robots de plusieurs sites ont bloqué Gromoteur, vous trouverez de nombreuses occurrences du mot "blocked". Le calcul de spécificités vous permet de faire émerger ces cas et de corriger le corpus en conséquence. De même, certains caractères spéciaux sont saillants mais n'apportent aucune information et ne sont pas présents dans tous les journaux. Vous pouvez les supprimer pour éviter un biais supplémentaire. Il n'est cependant pas nécessaire de supprimer tout ce qui n'entre pas dans le contenu des articles. Toutes les pages d'accueil proposent de se connecter et de s'abonner au journal. Ces éléments sont inhérents aux différentes parties du corpus et n'influenceront pas a priori l'analyse.

Spécificités par partie

Le bouton "spécifs-partie*" permet de calculer le **vocabulaire spécifique de la partie visée** par rapport aux autres parties. La partie en question est celle qui est entrée dans les [paramètres](#) (attention à bien choisir la partition en fonction de la partie souhaitée : il faut choisir une partition en pages pour pouvoir sélectionner la partie correspondant au journal *Le Monde* par exemple).

Le tableau donne les informations de fréquence totale et dans la partie sélectionnée ainsi que la spécificité des items présents dans la partie. Si un item est **surreprésenté**, la forme est dite "caractéristique" de la partie car sa fréquence est anormalement élevée dans cette partie par rapport au reste du corpus. Il est possible de faire des recherches d'item dans le tableau ou encore de le trier en fonction des fréquences ou par ordre alphabétique des items.

Spécificités Partie : la-croix.com (Partition : page)

Recherche :

Item	FQ	fq	Sp
Croix	86	71	102
http	280	84	69
com	282	80	64
www	381	84	57
	136	57	57
partenaire	34	30	46
PARTENARIATS	26	26	44
Ethique	24	24	41
Urbi	24	24	41
Orbi	24	24	41

Affichage de 1 à 10 des 1,514 items Préc. 1 2 3 4 5 ... 152 Suiv.

Spécificités totales

Le bouton “spécif totales*” permet de calculer **les spécificités pour chaque partie en fonction de la partition sélectionnée** dans les [paramètres](#). Le tableau de résultats affiche pour chaque item sa fréquence totale dans le corpus puis sa fréquence et son indice de spécificité dans chaque partie. Il est possible de faire des recherches d’item dans le tableau ou encore de le trier en fonction des fréquences, spécificités ou par ordre alphabétique des items.

Tableau Général des Items (FQ > 5 | annotation:1)
Partition : page

Copy CSV Excel PDF Print

Recherche :

Item	FQ	lefigaro.fr / fq	lefigaro.fr / sp	lemonde.fr / fq	lemonde.fr / sp	lopinion.fr / fq	lopinion.fr / sp	libera /
d	4446	170	0	209	0	49	0	1
Le	3599	238	14	226	9	60	5	1
2015	3590	196	6	21	-42	7	-9	1
un	3027	121	0	130	0	43	3	1
pour	2722	109	0	89	-3	26	0	9
une	2638	95	0	98	0	23	0	9
dans	2551	96	0	79	-4	19	0	8
La	2466	55	-7	121	0	35	3	1
Les	2446	253	42	183	14	17	0	6
sur	2412	72	-3	102	0	36	3	9

Affichage de 11 à 20 des 7,585 items Préc. 1 2 3 4 5 ... 759 Suiv.

Vocabulaire spécifique positif par partie

Le bouton “Mots Spécifs+” permet de trouver **les mots de spécificité positive d’une partition**. Les résultats sont présentés sous forme de tableau mais aussi de graphique associant pour chaque partie ses mots spécifiques. **Il est parfois utile d’augmenter le seuil de spécificité positive retenu (IndSPmin) pour ce calcul afin de limiter le nombre d’unités à afficher comme le montre le graphique suivant.**

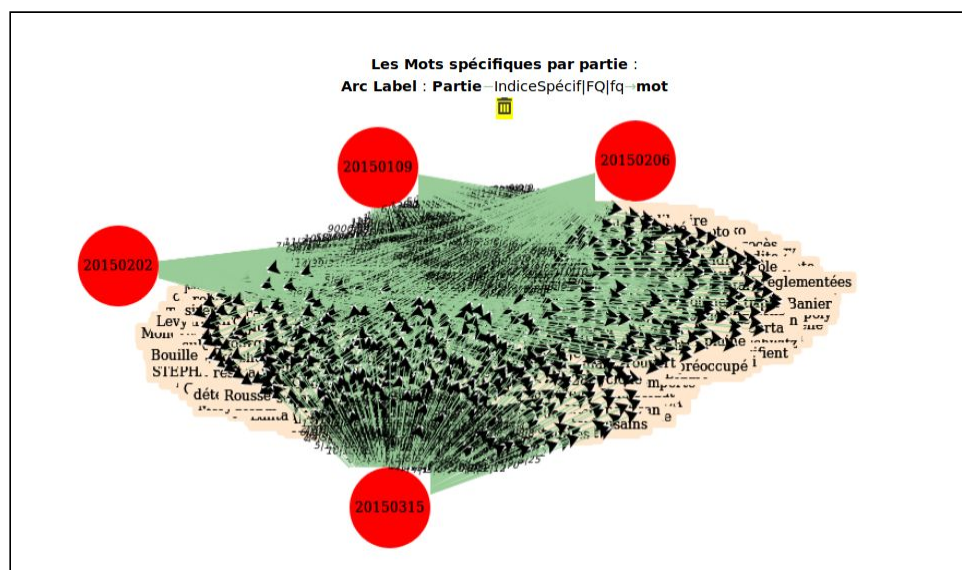


Tableau général des items (barycentre temporel, coefficient de Von Neumann)

Les outils mis en oeuvre ici sont utiles pour mettre au jour des termes évolutifs dans un corpus chronologique.

Le bouton "TGF+BT+VN" permet d'obtenir le **tableau général des items sur la partition**, le barycentre temporel et le coefficient VN. Le tableau de résultats présente pour chaque item sa **fréquence absolue** sur toutes les parties de la partition choisie.

Le tableau indique aussi pour ces items le **barycentre temporel (BT)**, un indice de la concentration de l'item dans les différentes périodes du corpus. Plus les occurrences du terme seront concentrées dans les premières périodes du corpus, plus l'indice sera proche de 1. Plus elles seront concentrées dans les dernières périodes, plus l'indice sera proche de n (n = nombre de période : ici, n = nombre d'aspirations de votre corpus).

Le tableau présente aussi enfin le **coefficient de Von Neumann (VN)**. Si ce coefficient est proche de 1, la suite des items ne représente pas d'auto-corrélation (l'item n'est pas corrélé aux précédents / suivants). L'auto-corrélation est dite positive lorsque l'indice est plus petit. On dit alors que le terme "participe à l'évolution chronologique"¹. Si les **segments répétés** ont été calculés au préalable, le tableau les intégrera (mais attention, le temps de calcul sera plus long...).

Tableau Général des Items (Fq > 5 | annotation:1)
Partition : date

Copy CSV Excel PDF Print

Recherche :

Item	FQ	BT	VN	20150109 / fq	20150202 / fq	20150206 / fq	20150315 / fq
de	17363	2.4	0.91	4921	4001	4156	4285
la	8585	2.5	0.89	2414	1994	2057	2120
à	8475	2.4	0.89	2684	1792	1814	2185
le	6404	2.5	0.49	1680	1651	1526	1547
l	6142	2.4	0.83	1800	1410	1486	1446
les	5432	2.4	0.42	1564	1389	1221	1258
et	5167	2.4	0.65	1506	1224	1221	1216
en	5010	2.4	0.8	1540	1127	1204	1139
du	4692	2.5	1.01	1050	1263	1211	1168
des	4513	2.4	0.94	1319	1020	1108	1066

Affichage de 1 à 10 des 7,585 items Préc. 1 2 3 4 5 ... 759 Suiv.

¹ [Les séries textuelles chronologiques](#)", A. Salem 1991

Rechercher une forme

Dictionnaire

Pour avoir une idée des mots qui se trouvent dans le corpus, il faut regarder **la trame**. **Le dictionnaire des formes** affiche les items du corpus selon un classement du plus fréquent au moins fréquent. Ce dictionnaire peut être exporté selon différents formats (csv, excel, pdf). Il est possible de sélectionner ou projeter chaque item sur une **concordance**, une **ventilation** ou une **carte des sections**.

Chargement

Trame

Cadre

SR/Pattern

Section

Coccs

Bi-Texte

Dépendance

Sélection

Export

Aide

Opérations sur la Trame

Dictionnaire*

Concordance*

Fusion Annotation

ANNOTATIONS À FUSIONNER

2/3

Dictionnaire

Copy

CSV

Excel

PDF

Print

Recherche :

Item	Fq	Concordance	Ventilation	Carte	Sélection
de	17363				
la	8585				
à	8475				
le	6404				
i	6142				
les	5432				
et	5167				
en	5010				

Supposons que nous voulions étudier le traitement de la Grèce dans la presse écrite française. Il faut tout d'abord vérifier si le mot-clé est bien présent dans le corpus en le cherchant dans le dictionnaire :

Dictionnaire

Copy

CSV

Excel

PDF

Print

Recherche : Grèce

Item	Fq	Concordance	Ventilation	Carte	Sélection
Grèce	104				

Affichage de 1 à 1 des 1 items (filtrage à partir des 31,391 items)

Préc.

1

Suiv.

C'est le cas, il est donc raisonnable de continuer notre étude sur ce thème.

Carte des sections

Une fois la forme trouvée dans le dictionnaire, on peut la sélectionner et la projeter sur la carte des sections en cliquant sur le bouton dans la colonne "carte". Les carrés en rouge sur la carte des sections correspondent aux pages qui contiennent la forme "Grèce".

Il suffit de cliquer sur un carré pour avoir le texte de la section correspondante dans lequel est surlignée en rouge la forme recherchée :

Clic sur section : affichage contexte | Clic-droit sur section : sélection section

DATE=20150109 <1:285779>

DATE=20150202 <285781:515179>

DATE=20150206 <515182:743231>

DATE=20150315 <743235:981413>

Clic sur section : affichage contexte | Clic-droit sur section : sélection section

SECTION N°53

http://www.lemonde.fr Le Monde.fr - Actualité à la Une Le Monde.fr - Actualité à la Une En poursuivant votre navigation sur ce site, vous acceptez l'utilisation de cookies pour vous proposer des contenus et services adaptés à vos centres d'intérêts. En savoir plus et gérer ces paramètres . Le Monde Télérama Le Monde diplomatique Le Huffington Post Courrier international La Vie au Jardin S'abonner au Monde à partir de 1 € Services Le Monde Boutique Le Monde Accueil Hors-séries Livres DVD CD Unes du Monde Loisirs & papeterie Partenaires Le Monde Annonces auto Annonces emploi Annonces immo Comparateur de banques Comparateur de crédit Cours d'anglais Modèles de lettres Prix de l'immobilier Shopping Édition globale Édition afrique Rechercher : Facebook Twitter Suivre le Monde sur Twitter @lemondefr Suivre @lemondefr Google+ Emploi Newsletters Affichez ici la météo de votre ville S'inscrire Connexion Le Monde.fr Édition globale Édition afrique Mise à jour à 20h26 Accueil International Politique Société Éco Culture Idées Planète Sport Sciences Techno Campus Le Mag Édition Abonnés En ce moment Affaire Bettencourt Grèce Ukraine Politique Procès du Carlton Handball Attentats en région parisienne Les décodeurs Grèce : les jours de la troïka sont comptés 6 Les Européens savent qu'au moins en termes d'affichage, il faut qu'ils donnent ce gage à Athènes : faire disparaître cette structure, qui réunit les créanciers de la Grèce Grèce : compte à rebours enclenché pour éviter un défaut sur la dette 11 Après la victoire de Syriza, Athènes veut « arrêter d'imiter Sisyphe » Obama appelle à lâcher du lest sur la Grèce, Renzi veut de la « prudence » 24 Athènes demande du temps pour faire ses propositions sur sa dette 6 Le trésor des pirates : bataille autour de l'héritage de The Pirate Bay Du site « officiel » aux clones multiples, plusieurs groupes se disputent l'héritage de « l'esprit »

Concordance

Pour avoir une idée des contextes d'apparition, il est intéressant de regarder la forme recherchée sous forme de concordance. Cela peut se faire en projetant une forme depuis le dictionnaire (comme nous venons de le faire pour la carte des sections mais en cliquant sur le bouton de la colonne “Concordance”) ou en utilisant directement le bouton “Concordance” dans l’onglet “Trame” mais une “Pôle source” doit être entrée dans les paramètres :

Afficher 10 items

Recherche :

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
1	20150109	la justice, Eric Holder, devait retrouver dimanche ses homologues européens. En	Grèce	, les grands chantiers de Syriza En Turquie, des revues satiriques menacées
2	20150109	et Coralie Schaub Economie / Ecofutur Abandon de l'euro par la	Grèce	: SOS fantômes Décryptage Par Jean Quatremer Thierry Lepaon, une démission qui
3	20150109	Les Grecs de France se cotisent pour pouvoir voter L'association	Grèce	-France Résistance a lancé une souscription pour permettre aux électeurs
4	20150109	INFOGRAPHIE INTERACTIVE Libye, la guerre civile autour du pétrole Portfolio En	Grèce	, les centres de santé débordés 30 ans de prévention AXA Prévention
5	20150109	économiste iconoclaste et vulgarisateur hors pair 07/01, 18 h 07	Grèce	: contre-offensive de Syriza face aux scénarios catastrophe 07/01, 17
6	20150109	regagner sa place en Amérique latine 07/01, 12 h 11	Grèce	: Schulz critique "les spéculations irresponsables" de Berlin 07/01, 12 h
7	20150109	48 heures 06/01, 19 h 14 Une sortie de la	Grèce	de la zone euro pourrait coûter cher aux Français 06/01
8	20150109	économiste iconoclaste et vulgarisateur hors pair 07/01, 18 h 07	Grèce	: contre-offensive de Syriza face aux scénarios catastrophe 07/01, 17
9	20150109	regagner sa place en Amérique latine 07/01, 12 h 11	Grèce	: Schulz critique "les spéculations irresponsables" de Berlin 07/01, 12 h

Il est aussi possible de réaliser une concordance de plusieurs items simultanément : il suffit de séparer deux pôles par une espace dans les [paramètres](#). Si on souhaite chercher le bigramme “grèce antique” par exemple, il faudra le faire avec les [segments répétés](#).

The screenshot shows the concordance tool interface. At the top, there are dropdown menus for 'ANNOTATION' (set to 1), 'ANNOTATION SORTIE' (set to 1), 'RELATION' (set to REL), and 'ANNOTATION RELATION' (set to 4). Below these is a search bar containing 'Grèce Syriza'. A navigation bar includes links like 'Chargement', 'Trame', 'Cadre', 'SR/Paron', 'Section', 'Coocs', 'Bi-Texte', 'Dependance', 'Sélection', 'Export', and 'Aide'. A section titled 'Opérations sur la Trame' contains buttons for 'Dictionnaire*', 'Concordance*', 'Fusion Annotation', and 'ANNOTATIONS À FUSIONNER' (with a count of 263). Below this, there's a table with columns: 'N°', 'Partie', 'Contexte Gauche', 'Pôle', and 'Contexte Droit'. The table lists 10 items, all with 'Partie' 20150109 and 'Pôle' Grèce or Syriza. The search bar at the bottom right is empty.

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
1	20150109	la justice, Eric Holder, devait retrouver dimanche ses homologues européens. En	Grèce	, les grands chantiers de Syriza En Turquie, des revues satiriques menacées
2	20150109	retrouver dimanche ses homologues européens. En Grèce, les grands chantiers de	Syriza	En Turquie, des revues satiriques menacées A Madrid, chez Goldman Sachs
3	20150109	et Coralie Schaub Economie / Ecofutur Abandon de l'euro par la	Grèce	: SOS fantasmes Décryptage Par Jean Quatremer Thierry Lepaon, une démission qui
4	20150109	Les Grecs de France se cotisent pour pouvoir voter L'association	Grèce	- France Résistance a lancé une souscription pour permettre aux électeurs
5	20150109	INFOGRAPHIE INTERACTIVE Libye, la guerre civile autour du pétrole Portfolio En	Grèce	, les centres de santé débordés 30 ans de prévention AXA Prévention
6	20150109	économiste iconoclaste et vulgarisateur hors pair 07/01, 18 h 07	Grèce	: contre-offensive de Syriza face aux scénarios catastrophe 07/01, 17
7	20150109	hors pair 07/01, 18 h 07 Grèce: contre-offensive de	Syriza	face aux scénarios catastrophe 07/01, 17 h 06 Wall Street
8	20150109	regagner sa place en Amérique latine 07/01, 12 h 11	Grèce	: Schulz critique "les spéculations irresponsables" de Berlin 07/01, 12 h
9	20150109	48 heures 06/01, 19 h 14 Une sortie de la	Grèce	de la zone euro pourrait coûter cher aux Français 06/01
10	20150109	économiste iconoclaste et vulgarisateur hors pair 07/01, 18 h 07	Grèce	: contre-offensive de Syriza face aux scénarios catastrophe 07/01, 17

La concordance des deux pôles “Grèce” et “Syriza”²

Il est possible de chercher un autre item (ou plusieurs) dans la concordance, “BCE”, par exemple :

The screenshot shows the concordance tool interface with a search for 'BCE'. The search bar at the bottom right now contains 'BCE'. The table below shows results for items 19 through 111, all with 'Partie' 20150202 or 20150206 and 'Pôle' Grèce or Syriza. The search results are filtered to show items 1 to 9 out of 9.

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
19	20150202	le 02/02/2015 à 19:30 Le sort de la	Grèce	entre les mains de la BCE Les gouverneurs de la BCE
20	20150202	échéance fin février. Partager cet article Recommander Le sort de la	Grèce	entre les mains de la BCE Les gouverneurs de la BCE
46	20150202	de la BCE n'est pas qu'une bataille grecque » La	Grèce	veut en finir avec la troïka Francis Wurtz : "Syriza ou la
88	20150206	700.000 enfants harcelés à l'école 226 46% 5 La	Grèce	choquée dénonce le «chantage» de la BCE 225 46% 1 II
101	20150206	et brûlé la grande mosquée» «La BCE met le gouvernement de	Syriza	dans une position impossible» Interview Grèce : à Berlin, Schäuble et Varoufakis
102	20150206	BCE met le gouvernement de Syriza dans une position impossible» Interview	Grèce	: à Berlin, Schäuble et Varoufakis même pas d'accord pour ne
105	20150206	Par Cordélia Bonal Economie / Ecofutur «La BCE met le gouvernement de	Syriza	dans une position impossible» Interview Par Christian Losson Les espoirs de
109	20150206	d'Espagne, du Portugal, d'Irlande et d'ailleurs. Monde BCE /	Grèce	: il faut au moins écouter Platon !! Allez les Grecs ! Philippe Martínez
111	20150206	Économistes atterrés : "La BCE tente un coup d'État contre la	Grèce	" Voir toutes les brèves Inscription à la newsletter Abonnez-vous » Discount

Affichage de 1 à 9 des 9 items (filtrage à partir des 139 items) Préc. 1 Suiv.

À l'aide des flèches, il est possible de trier les résultats en fonction du contexte gauche ou droit par exemple.

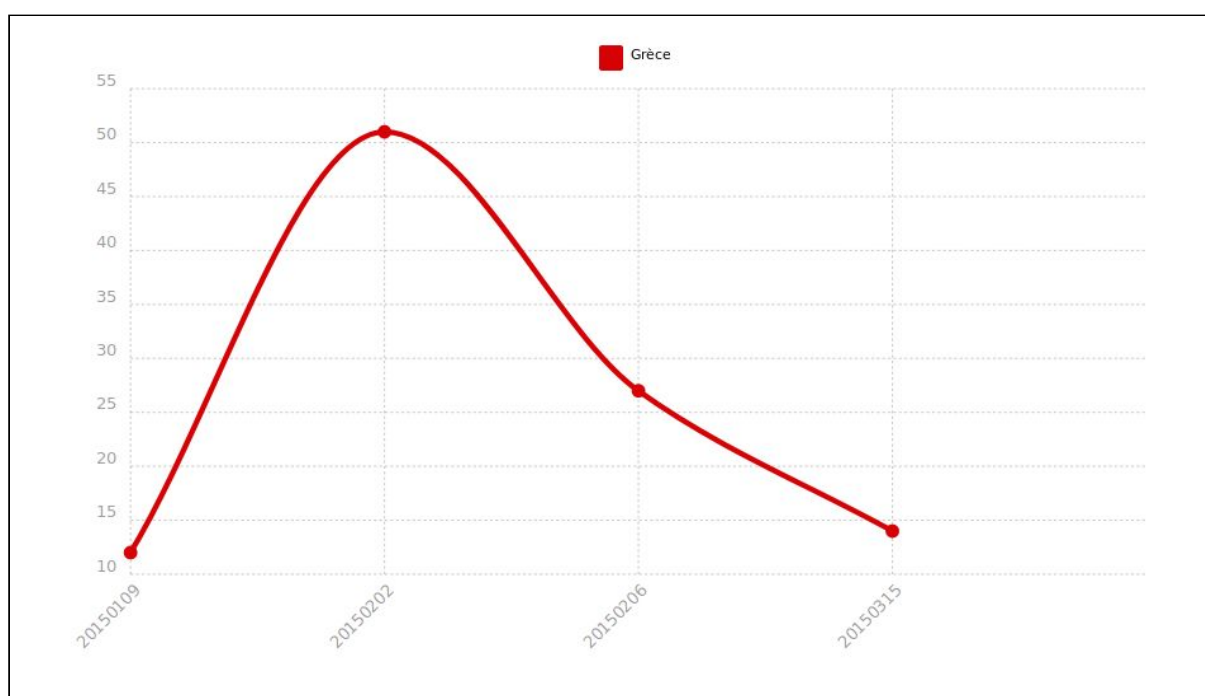
² Parti arrivé au pouvoir en Grèce en 2015, considéré comme étant de gauche radicale.

Ventilation

La fréquence d'un pôle dans le corpus peut être visualisée avec un graphique de ventilation. Comme pour la [concordance](#), cela peut se faire depuis le dictionnaire (onglet "Trame") en projetant une forme sur le graphique de ventilation ou depuis le bouton "ventilation" de l'onglet "Cadre" en fonction du champ "pôle source" entré dans les paramètres. iTremeur produit automatiquement 3 graphiques sur la partition sélectionnée dans les paramètres.

Fréquence absolue

Le premier graphique de ventilation (haut) est réalisé en fonction de la **fréquence absolue** du pôle recherché. C'est-à-dire en fonction du nombre d'occurrences de cette forme dans le corpus. Il est possible de choisir la partition (pages ou dates) dans les [paramètres](#).

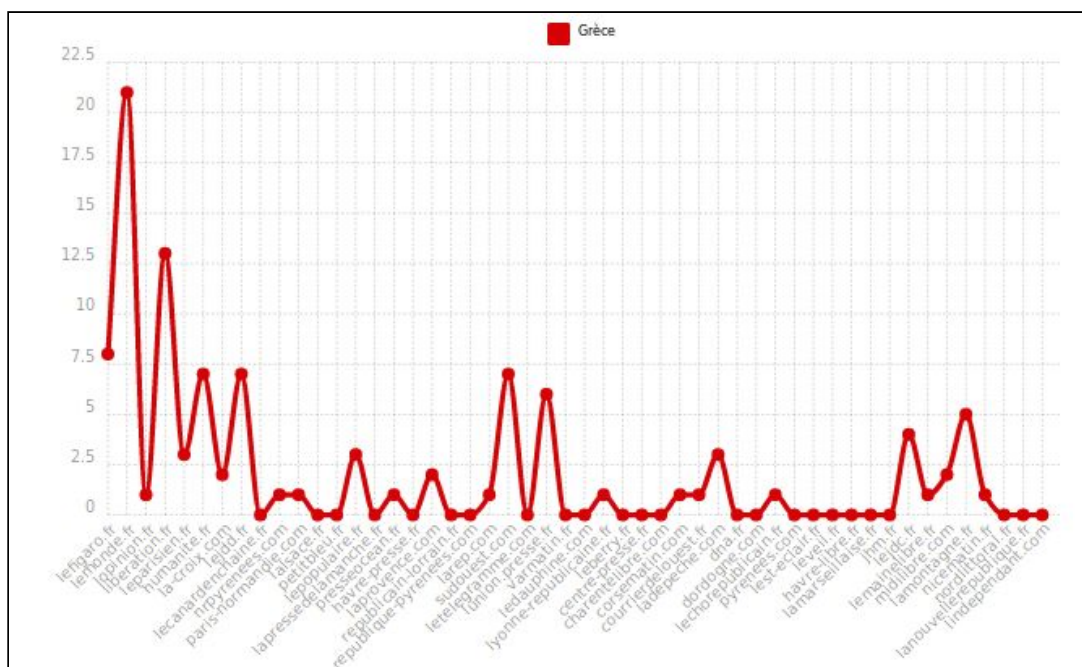


Ventilation de la forme "Grèce" - partition en dates - fréquences absolues

Avec la partition en dates, on constate que les mentions de la forme "Grèce" sont plus fréquentes le 5 février qu'aux autres dates. Il faut faire des recherches pour savoir pourquoi elle est plus élevée à cette date. Par exemple, ce pic peut être lié à la victoire du parti Syriza lors des élections du 25 janvier, qui a eu des effets sur les négociations concernant, entre autres, la dette grecque³.

³

https://fr.wikipedia.org/wiki/Crise_de_la_dette_publique_grecque#Troisi%C3%A8me_phase_critique_d%C3%A9but_2015



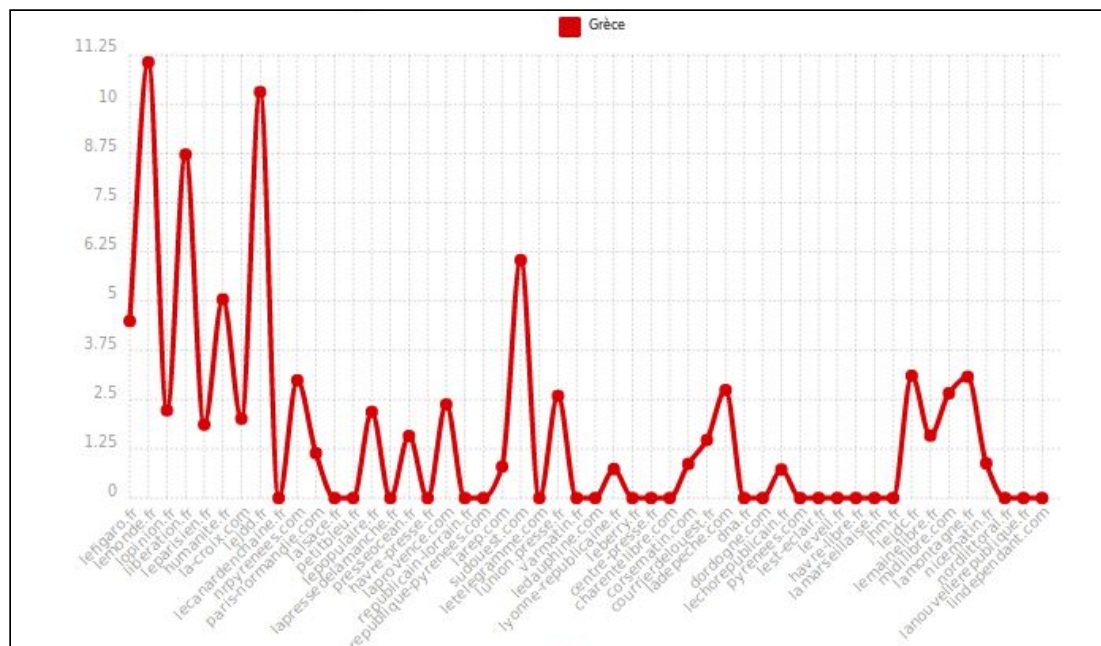
Avec la partition en pages, on constate de grandes différences de fréquence absolue selon les journaux. *Le Monde* est de loin le journal qui mentionne le plus la forme “Grèce”, avec plus de 20 mentions. *Libération* vient en second avec 13 mentions, puis *Le Figaro* avec 8 mentions et plusieurs journaux comme *L’Humanité* mentionnent cette forme 7 fois. Pourquoi ces différences ? Il faut peut être regarder l’orientation politique de ces journaux ou regarder s’il s’agit de journaux nationaux ou régionaux.

Fréquence relative

Le second graphique de ventilation (milieu) est réalisé en fonction de la **fréquence relative** du pôle recherché. Le nombre d'occurrences de l'item dans la partie est mis en rapport avec la taille (le nombre d'items) de la partie. Il s'agit d'une manière plus juste de comparer des textes de taille différente. Pour avoir une idée de son utilité, calculer les **PCLC** sur la partition en pages et trier le tableau selon les **F occ** (nombre total de mots dans le texte pour chaque journal aux 4 dates confondues) à l'aide de la petite flèche. Ces valeurs varient de 520 à plus de 23 000 !

Partie	F occ	f typ	Forme max	Max
republique-pyrenees.com	520	93	de	32
lapressedelamanche.fr	930	258	de	29
lecanardenchaine.fr	1394	201	du	72
varmatin.fr	2090	308	Chaussure	194
ledauphine.com	3339	933	à	152
nrpyrenees.com	3346	1170	de	133
nordlittoral.fr	3668	1195	de	138
petitbleu.fr	4136	1381	de	156
lopinion.fr	4489	1300	de	128
havre-libre.fr	4578	1290	de	214

Partie	F occ	f typ	Forme max	Max
lunion.presse.fr	23126	2777	de	896
lemonde.fr	18977	4505	de	804
lefigaro.fr	17801	3226	de	806
lamontagne.fr	16243	3371	de	576
leparisien.fr	16098	3639	de	535
centre-presse.fr	15930	2354	de	580
liberation.fr	14904	3892	de	568
leberry.fr	14410	3097	de	523
humanite.fr	13878	3749	de	689
lechorepublicain.fr	13852	2947	de	526



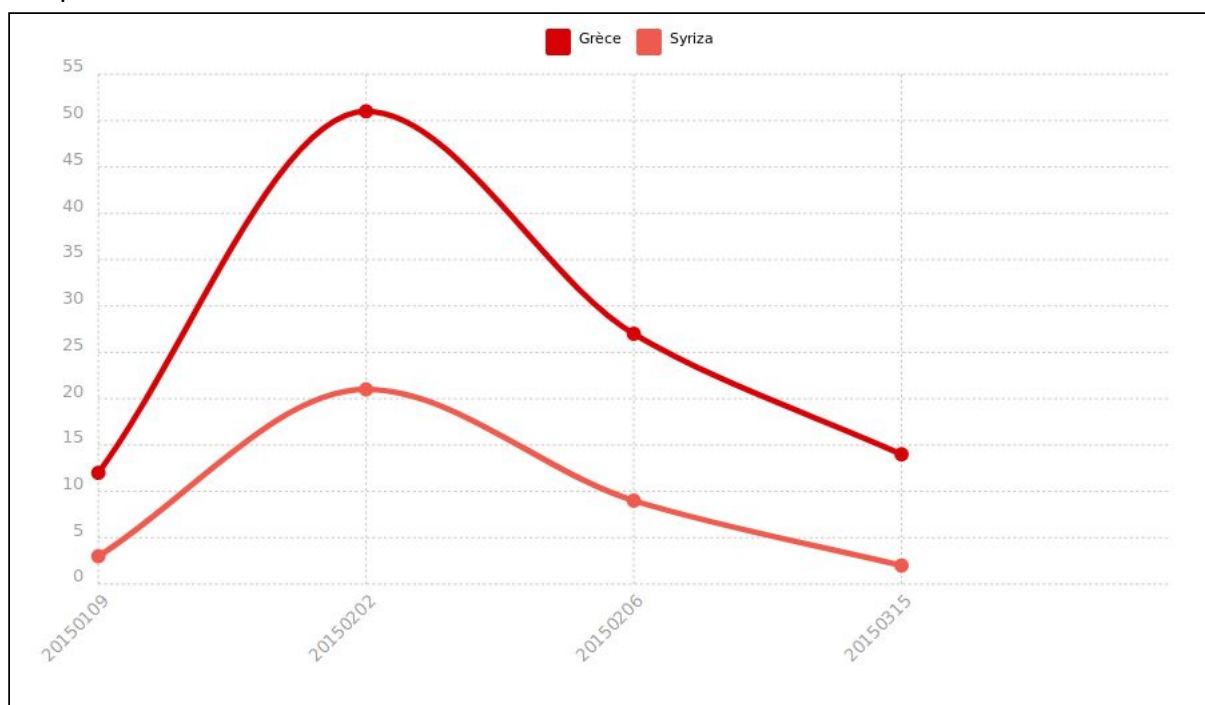
Avec la fréquence relative, on remarque que les différences entre *Le Monde* et *Libération* vs. *Le Figaro* sont plus claires qu'avec la fréquence absolue. On voit aussi que la fréquence relative pour *Le JDD* explose par rapport à la fréquence absolue (7 occurrences / 1977 items au total).

Spécificités

Le dernier graphique de ventilation (bas) est réalisé en fonction des **spécificités** (cf) du pôle pour chaque partie. Cela permet de voir si cet item est surreprésenté ou sous-représenté dans chaque partie par rapport à la norme que constitue la fréquence totale de cet item dans le corpus. Si la spécificité est positive, la forme est dite “caractéristique” de la partie car sa fréquence est anormalement élevée dans cette partie par rapport au reste du corpus. Lorsque la spécificité est négative, cela revient à dire que la forme apparaît moins souvent que ce à quoi on peut s’attendre statistiquement.

Sur la partition en dates, la spécificité est positive pour le 2 février, cela confirme les observations selon les [fréquences absolues](#).

Comme pour la concordance, il est possible de chercher plusieurs pôles afin de les comparer :



Ventilation des formes "Grèce" et "Syriza" - partition en dates - fréquences absolues

Segments répétés

Un **segment répété** (ou "polyforme répétée") est une suite de formes dont la fréquence est supérieure ou égale à 2 dans le corpus (le paramétrage par défaut est **SR LGMAX** : 12 et **SR FQMIN** : 10).

Dans l'onglet "SR/Patron", le bouton "Segments répétés*" permet d'obtenir les segments répétés du corpus triés par fréquence absolue (mais il y a aussi l'information de longueur du segment). Les résultats permettent d'explorer le contenu du tableau et de lancer des opérations sur chacun des items du tableau (concordance, ventilation, carte des sections).

Copy	CSV	Excel	PDF	Print	Recherche :
SR	Fq	Lg	Concordance	Ventilation	Carte
de la	2419	2			
de l	1821	2			
Charlie Hebdo	1511	2			
à la	965	2			
à l	794	2			
2015 à	684	2			
dans le	645	2			
y a	483	2			
de France	478	2			
sur le	416	2			

Affichage de 1 à 10 des 12,359 items

Préc. 1 2 3 4 5 ... 1236 Suiv.

Il est possible de chercher des segments répétés de votre choix dans la case **SR** et de projeter les résultats sur la [carte des sections](#), une [ventilation](#) ou une [concordance](#) :

Opérations sur les *Segments Répétés*

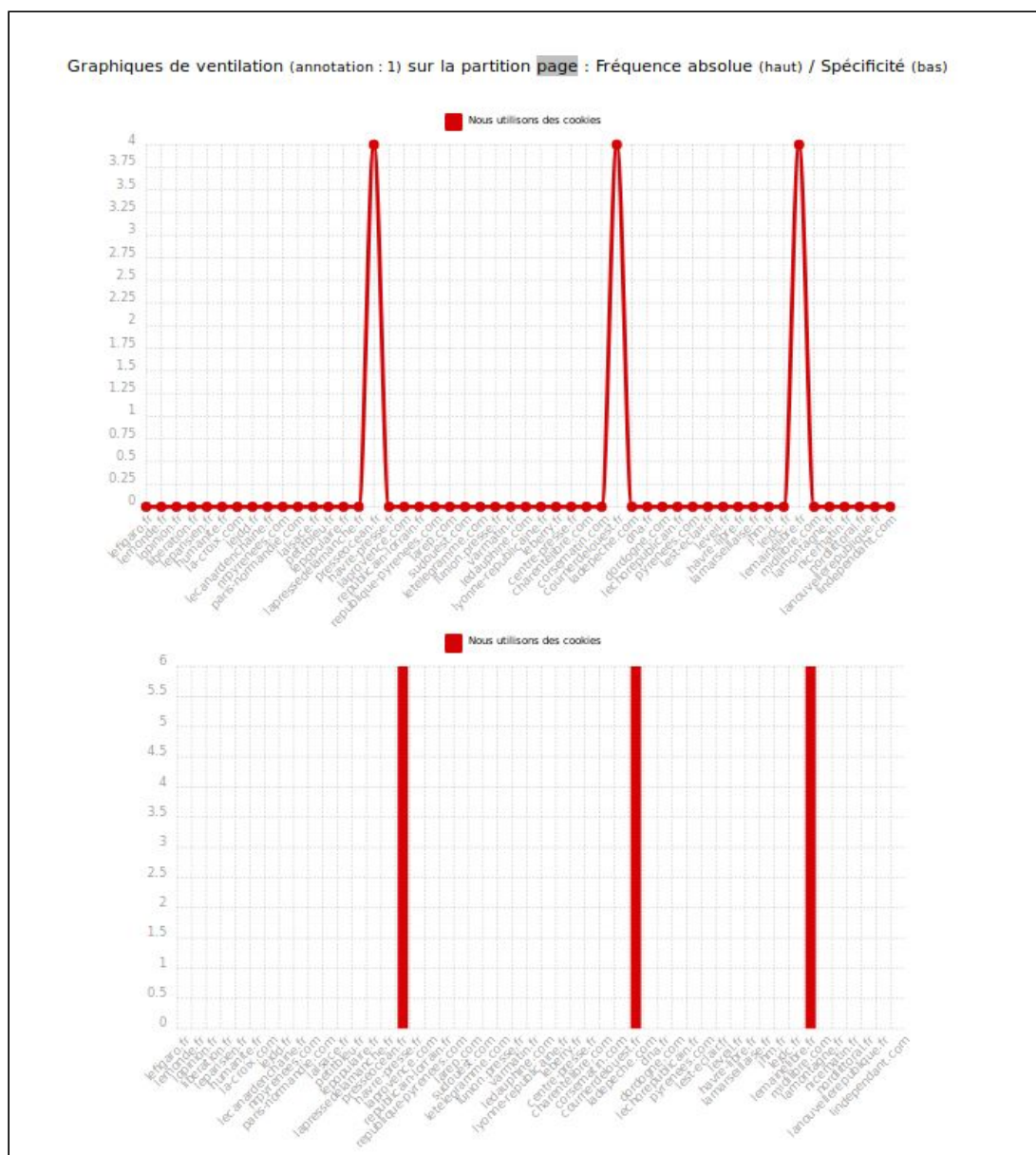
SR Nous utilisons des

Segments répétés*

Carte Sections(SR)*

Ventilation(SR)*

Concordance(SR)*



Ventilation du SR “Nous utilisons des cookies” - partition en pages

Presse Océan, Courrier de l’Ouest et Le Maine Libre contiennent chacun 4 fois le segment “Nous utilisons des cookies”.

Cooccurents

La **cooccurrence** est la présence simultanée (mais non forcément contiguë) dans un fragment de texte des occurrences de deux formes données. L'onglet "Coocs" permet de rechercher les cooccurents du pôle source entré dans les [paramètres](#) en cliquant sur le bouton "cooccurents*". Le calcul des cooccurents d'un terme se fait en mettant au jour les termes co-fréquents surreprésentés dans les contextes de ce terme. Il est possible de préciser le contexte dans lequel on veut chercher ces cooccurents (NB Terme Gauche / NB Terme Droite). Par défaut, ce calcul de cooccurrence est réalisé en ne retenant que les candidats cooccurents dont l'indice de spécificité est supérieur à la valeur IndSpMin (cf [paramètres](#)). Pour le moment nous avons cherché le terme "Grèce" seul mais il peut être intéressant de regarder quels sont les mots qui apparaissent régulièrement autour.

Paramètres

PARTITION: date
PARTIES: 20150109
LG CONTEXTE: 10
FO MAX: 5

GRAPHE H: 400
SEUIL: 5
CO-FREQ: 2
INDSPMIN: 5

GRAPHE L: 800
SR LGMAX: 12
SR FQMIN: 10
NB SÉLECTION SECTION: 1

ANNOTATION*: 1
ANNOTATION SORTIE: 1
RELATION: REL
ANNOTATION RELATION: 4

POLE SOURCE: Grèce

Chargement
Trame
Cadre
SR/Patron
Section
Coocs
Bi-Texte
Dépendance
Sélection
Export
Aide

Calcul de cooccurents

NB TERME GAUCHE: 10
NB TERME DROITE: 10

Cooccurents*
Cooccurents* sur partie sélectionnée

Réseau Cooccurents*
Réseau Cooccurents* sur partie sélectionnée

☐ STOPLISTE="GESTIONNAIRE DE SÉLECTION"

Les résultats apparaissent sous forme de tableau :

Cooc	FqCooc	CoFreq	IndSP
rangent	11	11	29
derrière	31	11	21
Syriza	35	10	18
Dilemme	13	8	18
économie	91	11	16
la	8585	68	15
réfléchit	11	6	14
créanciers	3	4	12
insignifiance	4	4	11
stratégie	23	6	11
Démystifier	7	4	10
Sarkozy	115	8	10
Varoufakis	17	4	8

Cliquer sur un terme (ici “Syriza”) dans le tableau donne accès à son contexte (plus bas) :

Contextes

Afficher 10 items

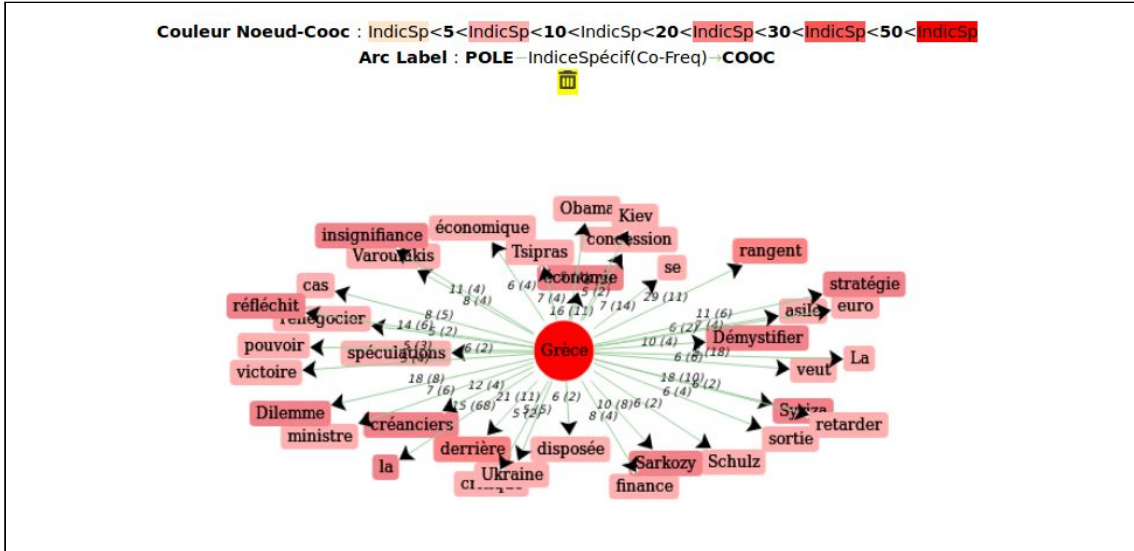
Recherche :

N°	Contexte
1	/01, 18 h 07 crise contre-offensive de Syriza
2	/01, 18 h 07 crise contre-offensive de Syriza
3	la victoire de Syriza en crise , le secrétaire national du
4	la bourse Syriza transforme la crise en "place to be"
5	vient de se passer en crise avec la victoire de Syriza
6	heures Syriza au pouvoir en crise : Obama est contre l
7	heures Syriza au pouvoir en crise : Obama est contre l
8	vient de se passer en crise avec la victoire de Syriza
9	de l'affaire. Vidéos crise : Syriza , le jour
10	victoire de la Syriza en crise est l'occasion d'un

Affichage de 1 à 10 des 10 items

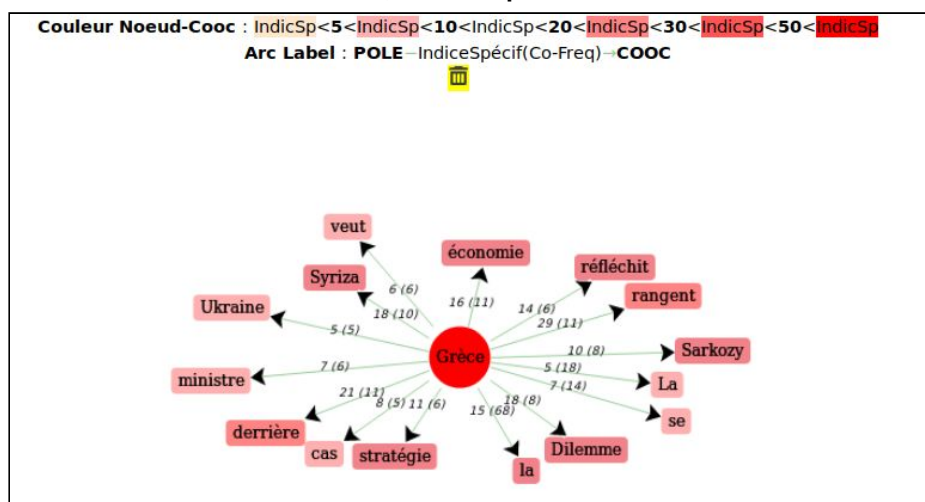
Préc. 1 Suiv.

Les résultats sont aussi donnés sous forme de graphique :



S'il y a trop ou pas assez de candidats dans votre graphique, il suffit de changer la valeur Co-Freq (par défaut, le calcul de cooccurrence est calculé en ne retenant que les candidats cooccurents dont la co-fréquence est supérieure à la valeur donnée). Il est aussi possible de déplacer les flèches à l'aide de la souris pour que le graphique soit plus lisible (ou cliquer dessus et attendre que les cooccurents se déplacent).

Ici, Co-Freq = 5 :



Le bouton “Cooccurrents sur partie sélectionnée” permet de faire le calcul des cooccurrents sur la partie sélectionnée dans les [paramètres](#) (par exemple sur un journal en particulier).

Le bouton “Réseau Cooccurrents” effectue le calcul des cooccurrents sur l’ensemble des items de fréquence supérieure à **FQ Max** (cf [paramètres](#)). Il est aussi possible de faire ce calcul sur la partie sélectionnée.

Pour les calculs de cooccurrents, il est possible d'utiliser le Gestionnaire de Sélections comme une stop-liste (une liste de mots que l’on ne souhaite pas prendre en compte : par exemple les mots vides). Pour cela, il faut cocher la case **StopListe="Gestionnaire de Sélection"** et sélectionner les items à excludre de ces calculs (par exemple à partir du [dictionnaire](#)).

Il est important de toujours prendre du recul lorsque l’on fait des statistiques textuelles pour se demander pourquoi on obtient ces résultats-là. Par exemple, il faut faire attention aux “cooccurrences accidentelles” causées par un problème de formatage dans le texte. Par exemple, un point manquant dans le texte peut induire iTrameur en erreur, le laissant croire que deux mots sont dans la même phrase alors que ce n’est pas exact. Il peut être intéressant de regarder la concordance d’un cooccurrent pour vérifier s’il apparaît dans le contexte proche du pôle.

Sélection

Il est possible de sélectionner plusieurs formes du [dictionnaire](#) et de les projeter sur la carte des sections ou une concordance grâce à l’onglet “sélection” :



Restons sur le thème de la Grèce : on pourrait sélectionner les mots “BCE” et “Troïka”⁴ car ils représentent de nouveaux acteurs dans la crise de la dette grecque. Cette sélection pourrait représenter le point de vue non-grec de cette crise.

Dictionnaire

Copy CSV Excel PDF Print

Recherche : Troïka

Item	Fq	Concordance	Ventilation	Carte	Sélection
troïka	12				

Affichage de 1 à 1 des 1 items (filtrage à partir des 31,391 items) Préc. 1 Suiv.

Une fois la sélection opérée, on peut afficher les termes qu'elle contient dans une concordance et rechercher des mots dans les contextes :

Afficher 10 items Recherche : Grèce

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
12	20150202	30 Le sort de la Grèce entre les mains de la	BCE	Les gouverneurs de la BCE ont rappelé que leurs financements accordés
13	20150202	Grèce entre les mains de la BCE Les gouverneurs de la	BCE	ont rappelé que leurs financements accordés aux banques sont liés au
15	20150202	Recommander Le sort de la Grèce entre les mains de la	BCE	Les gouverneurs de la BCE ont rappelé que leurs financements accordés
16	20150202	Grèce entre les mains de la BCE Les gouverneurs de la	BCE	ont rappelé que leurs financements accordés aux banques sont liés au
18	20150202	Attentats en région parisienne Les décodeurs Grèce : les jours de la	troïka	sont comptés 6 Les Européens savent qu'au moins en termes
19	20150202	du Sud en vigilance 16:36 Grèce : jours comptés pour la	troïka	16:28 Procès de l'affaire du Carlton 16:16 Tempêtes
20	20150202	banques échouent à trouver un accord Grèce : les jours de la	troïka	sont comptés 6 Assurance-vie : les taux 2014 des fonds en
21	20150202	et défiscalisez Pourquoi la Grèce et l'Europe veulent enterrer la	troïka	Par Isabelle Marchais Jean-Claude Juncker propose d'instituer à terme
22	20150202	Abonnés Sur le même sujet: Frédéric Boccarda « L'argent de la	BCE	n'est pas qu'une bataille grecque » La Grèce veut en
23	20150202	qu'une bataille grecque » La Grèce veut en finir avec la	troïka	Francis Wurtz : "Syriza ou la brèche dans la forteresse des puissants

Affichage de 1 à 10 des 15 items (filtrage à partir des 35 items) Préc. 1 2 Suiv.

Il est important de vérifier que dans les contextes où ils apparaissent il est effectivement question de la Grèce, par exemple pour pouvoir tirer des conclusions à des hypothèses liées au thème de départ.

Pour répondre aux questions posées au préalable, il est possible aussi de projeter la sélection sur la carte des sections afin d'avoir une idée de la répartition de ces mots dans les journaux et les dates. Il suffit de cliquer sur un carré pour avoir le texte du journal dans lequel les mots de la sélection sont surlignés.

⁴ BCE + FMI + Commission européenne

