

## 12 questions d'utilisateurs de Frantext

1. Je cherche à afficher les rimes de fins de vers, afin d'étudier la prononciation de certains mots. Pourquoi l'expression CQL (recherche avancée) suivante ne fonctionne pas : `".*" [pos="PONCT"]* </I> ?`
2. Depuis la nouvelle version de Frantext, je suis obligé à chaque fois de retaper mes requêtes. Dans la version précédente, il y avait une fonctionnalité d'historique. Comment puis-je consulter mes requêtes précédentes ?
3. Pourquoi la recherche depuis l'onglet de co-occurrences ne permet pas de chercher plus de 3 séquences, d'afficher un concordancier, d'exporter les résultats ou d'afficher le numéro de page de l'extrait ?
4. J'aurais besoin de sauvegarder et éventuellement d'imprimer une liste de mon corpus. Comment faire ?
5. Dans la recherche assistée, est-il possible de modifier ma recherche sur les lemmes « le » et « chat » ? Je souhaite rechercher un nom commun entre les deux termes.
6. Je cherche dans FRANTEXT intégral des occurrences par exemple de (*sans*) *coup férir* ou (*il faut*) *raison garder*. Avec la formule `[pos="NC"] [pos="VINF"]` j'obtiens des dizaines de milliers de résultats non pertinents, entre les **NC** introduits par une préposition et ceux introduits par un verbe de perception. Je vois mal comment restreindre mon **NC** à un objet direct du **VINF**. Auriez-vous une idée ?
7. Pourquoi est-ce que la recherche de *dent* sur le corpus *Ancien français* avec le lexique *médiéval* retourne également « despit » ?
8. Est-ce que c'est possible d'utiliser la fonction « fréquence » pour chercher « en effet » (et d'autres collocations) pour voir la fréquence par million de mots ? Je n'ai réussi qu'avec « en outre ».
9. Je n'ai pas réussi à trouver comment faire quand, dans une liste de mots, l'un ou plusieurs des éléments de la liste comporte un blanc. Imaginez que je veuille faire une liste du genre : « une fois », « plusieurs fois », « souvent », « rarement ». Comment la formuler ?
10. Comment obtenir une liste alphabétique de graphies, par exemple pour les suffixes en `.+ailles`, afin de consulter une liste comme « *accordailles, antimarmaille, antiquaille* », etc. ?
11. Dans l'ancien Frantext, on pouvait avoir accès aux fréquences absolues et relatives. Comment trouver l'ouvrage de *Duras* qui contient la plus haute concentration des mots de la liste « sommeil » suivante : sommeil, dormir, fatigue, fatigué, endormir, somnoler, léthargique, dormeur et léthargie ?
12. Je suis en train de créer un corpus et ayant défini une période, je voudrais supprimer un seul genre textuel, à savoir la poésie, de la sélection de textes. Comment faire ?

# Expressions FRANTEXT 19.2

## Résumé des expressions de la recherche avancée

### Expressions CQL (Corpus Query Language)

Forme graphique du mot \$	[word="\$ " ] ou "\$ "
Flexions du lemme \$	[lemma="\$ " ]
Catégorie grammaticale \$	[pos="\$ " ]
Joker, n'importe quel mot	[ ]
\$1 suivi de \$2	[word="\$1 " ] [word="\$2 " ]
\$1 et \$2	[word="\$1 " & pos="\$2 " ]
\$1 ou \$2	[word="\$1 "   word="\$2 " ]
\$1 dont la catégorie grammaticale n'est pas \$2	[lemma="\$1 " & pos!="\$2 " ]
\$1 entre les balises <\$2> et </\$2>	<\$2> [lemma="\$1 " ] </\$2>
Recherche sensible à la casse des caractères	[word="\$ " ]
Recherche insensible à la casse des caractères	[word="\$ "%c ]
Recherche insensible aux signes diacritiques	[word="\$ "%d ]
Recherche insensible aux ligatures	[word="\$ "%g ]
Recherche littérale, insensible aux expressions régulières	[word="\$ "%l ]
Cumul de modificateurs	[word="\$ "%cdgl ]

### Expressions de séquence

Liste dont l'identifiant est \$	&liste("\$ " )
Règle \$2 de la grammaire \$1	&règle("\$1 ":: \$2 " )
Flexions du lemme \$1 issues du lexique \$2 (médiéval, préclassique ou moderne). Les catégories grammaticales \$3 et \$4 sont optionnelles	&lexique("\$1 " , "\$2 " , "\$3 " ) ou [word=&lexique("\$1 " , "\$2 " , "\$3 " ) & pos="\$4 " ]
Recherche de multiples unités lexicales \$	&segmenter("\$ " )
Recherche simple de la phrase \$	&développer("\$ " )

### Catégories grammaticales

Adjectif	ADJ
Adverbe	ADV
Conjonction de coordination	CC
Conjonction de subordination	CS
Clitique objet	CLO
Clitique sujet	CLS
Déterminant	DET
Mot étranger	ET
Interjection	I
Nom commun	NC
Nom propre	NP
Préposition + déterminant	P\+D
Ponctuation	PONCT
Pronom	PRO
Pronom relatif	PROREL
Pronom interrogatif	PROWH
Préposition	P
Verbe conjugué	V
Verbe à l'infinitif	VINF
Verbe participe passé	VPP
Verbe participe présent	VPR
Mot non traité	X
Adverbe + clitique objet	ADV\+CLO
Adverbe + déterminant	ADV\+DET
Clitique sujet + clitique objet	CLS\+CLO
Clitique sujet + verbe	CLS\+V
Conjonction de subordination + clitique	CS\+CLO
Conj. de subordination + clitique sujet	CS\+CLS
Verbe + clitique sujet	V\+CLS

### Expressions régulières

Groupe	( \$ )
Alternative (\$1 ou \$2)	\$1   \$2
N'importe quel caractère	.
Interpréter littéralement un opérateur	\
Un des caractères entre crochets	[ \$ ]
Tout caractère hormis ceux entre crochets	[ ^ \$ ]
Intervalle alphabétiques	[ a - z ]
Intervalle numérique	[ 0 - 9 ]
0 ou 1 fois	\$ ?
0 fois ou plus	\$ *
1 fois ou plus	\$ +
Exactement n fois	\$ { n }
Entre n et m fois	\$ { n , m }
Au moins n fois	\$ { n , }
Un chiffre	\d
Tout sauf un chiffre	\D
Nombre hexadécimal sur 2 chiffres	\x
Nombre hexadécimal sur 4 chiffres	\u
Tout sauf un espace	\S
Un caractère alphanumérique sans diacritiques	\w
Tout sauf un caractère alphanumérique	\W

Un caractère alphanumérique (1)	[[:alnum:]]
Un caractère alphabétique (1)	[[:alpha:]]
Un caractère numérique	[[:digit:]]
Un caractère alphabétique en minuscules (1)	[[:lower:]]
Un caractère alphabétique en majuscules (1)	[[:upper:]]
Un caractère alphanumérique ou un tiret bas	[[:word:]]
Un nombre hexadécimal	[[:xdigit:]]

(1) Sans les caractères diacritiques

## 12 réponses aux utilisateurs de Frantext

1. Frantext considère que l'expression CQL est trop longue ou trop complexe. Il est possible de contourner le problème en modifiant les quantificateurs des expressions régulières. Exemple : `[pos!="PONCT"] [pos="PONCT"]{0,3}</I>`. Attention toutefois car les quantificateurs sont limités à 31 caractères consécutifs, et la recherche de balises XML n'est pas totalement opérationnelle.
2. Il n'y a pas d'historique des recherches. Vous pouvez toutefois modifier à tout moment la recherche en cours en cliquant sur la zone grise d'édition située au dessus des résultats. Vous pouvez également enregistrer explicitement vos recherches à l'aide du bouton **SAUVEGARDER** situé à droite de l'interface, afin de les réutiliser par la suite.
3. Étant donné le nombre considérable de traitements informatiques qui interviennent à chaque recherche, vous disposez de 2 manière d'effectuer des recherches de co-occurrences. L'onglet **Co-occurrences** dispose de peu de paramètres, mais ne limite pas le nombre de résultats. Les modes de recherche simple, assistée et avancée sont limités à 100 000 résultats, mais permettent d'ajouter autant d'**Actions de Co-occurrence** que vous le souhaitez à partir d'une première recherche, ainsi que d'exporter les résultats. Enfin, concernant le numéro de page en mode **Co-occurrences**, il est bien disponible, mais seulement lorsque vous utilisez la fonction de Zoom.
4. Il n'est pas possible d'exporter ou d'imprimer la liste d'un corpus mais vous pouvez télécharger la liste des ouvrages du corpus intégral au format PDF disponible dans les informations de mise à jour sur le site de documentation <https://wiki.frantext.fr>. De plus, lorsque vous vous connectez depuis un espace numérique de travail vous avez la possibilité d'exporter vos corpus personnels depuis l'onglet **Paramètres > Paramètres utilisateur > Import / Export > Exporter**. Le fichier exporté contient alors les numéros de cotes des ouvrages du corpus (en cherchant un peu), vous pouvez alors comparer cette liste avec la bibliographie où chaque référence est suivie d'informations entre accolades : la cote, la date et le nombre de mots.
5. Cliquez dans la zone grise qui permet d'éditer la recherche au dessus des résultats, puis ajoutez la recherche du nom commun en cliquant sur le bouton **+** (Astuce : Dans le menu **Catégorie**, cliquez sur la touche **n** jusqu'à atteindre la ligne nom commun). Puis survolez la partie supérieure du bloc jusqu'à ce que le curseur prenne la forme d'une croix. Il vous suffit alors de saisir le bloc et de le déplacer.
6. Pour enlever un peu de bruit, vous pouvez essayer d'identifier les catégories qui ajoutent du bruit (par exemple ici les prépositions et verbes placés avant les noms communs) et effectuer une recherche avancée qui exclut ces catégories : `[pos!="P|V"] [pos="NC"] [pos="VINF"]`. Appliquez ensuite une Action **Modifier la taille du pivot** en choisissant **Mot avant** : -1 afin de récupérer une vedette sur *Nom commun + Verbe à l'infinitif*.

7. Il s'agit d'une erreur dans le lexique. Dans votre cas, l'utilisation d'une action de **Fréquence** sur le *Pivot* permet de consulter les 10 entrées présentes dans le lexique. En attendant que l'entrée soit corrigée, vous pouvez essayer de copier les résultats de fréquence dans un traitement de texte et constituer une règle de grammaire comme `Règle dent = "dant|denz|dans|dens|dan|dent|adens|den|dentz"` en supprimant l'entrée qui pose problème. Vous pouvez également créer une requête CQL du type `[word=&lexique("médiéval","dent") & lemma!="dépit|despit"]`.
8. Du fait de la segmentation des mots, vous devez utiliser des expressions CQL pour parvenir à vos fins, ce que ne permet pas l'onglet **Fréquences**. Effectuez une recherche simple, assistée ou avancée, puis utilisez une action de **Fréquence** sur le *Pivot*. Pour calculer la fréquence par million de mots équivalente à celle de l'onglet **Fréquences**, vous devez diviser le **Nombre d'occurrences** par le **Nombre de tokens dans le corpus**, puis multiplier le tout par 1 000 000. Le **Nombre de tokens dans le corpus** est disponible sous l'onglet **Corpus > Voir le corpus > Informations**. Attention, contrairement à l'onglet **Fréquences**, le nombre de résultats dans la recherche simple, assistée et avancée est limité à 100 000.
9. Lorsque vous effectuez une recherche simple de l'expression **Mot ou suite de mots**, le champ de recherche situé au dessus des résultats vous indique que votre requête correspond à l'expression CQL `&développer("Mot ou suite de mots"%cdl)`. C'est cette expression CQL que vous devez ajouter dans votre liste de mots. Cochez ainsi la case **Expression CQL**, puis saisissez l'expression contenant votre séquence de mots, par exemple `&développer("une fois"%cdl)`.
10. Ce n'est pas possible dans Frantext. L'action de **Fréquence** ne permet pas le tri alphabétique. L'interface permet toutefois de consulter une liste alphabétique de flexions à partir d'une expression régulière lorsque vous constituez une liste, par exemple en ajoutant les flexions du lemme `.*ailles?` depuis le lexique **Moderne**. Mais la meilleure solution consiste à effectuer une recherche avancée de `[word=".+ailles?"%cd]` puis à exporter le concordancier au format CSV pour l'exploiter dans Excel. Les dernières versions permettent en effet d'effectuer un *tri alphabétique sans doublons* depuis **Données > Options avancées** (de Trier/Filtrer) > **Extraction sans doublons**.
11. Il n'existe pas de fonction pour cela, mais vous pouvez effectuer le calcul à la main. Commencez par créer un corpus de textes de *Duras*. Effectuez ensuite une recherche sur ce corpus avec votre liste ou une expression CQL du type `[word="sommeil|(en)?dormir|fatigu(é)|somnoler|léthargi(qu)?e|dormeur"%cd]`, puis sous l'onglet **Actions**, choisissez **Fréquence par document** pour obtenir le nombre d'occurrences par ouvrages. Depuis la liste de fréquences, vous pouvez retrouver le nombre de mots de chaque ouvrage en cliquant sur sa *Cote* à gauche. Pour chaque ouvrage, vous pourrez ainsi diviser le nombre d'occurrences par le nombre total de mots afin d'obtenir la fréquence relative.
12. À partir de votre corpus, vous pouvez ajouter ou supprimer des textes en utilisant les filtres. Après chaque étape de filtrage, vous ne pouvez sélectionner ou désélectionner que le contenu filtré. Il est donc important de réinitialiser les filtres à chaque étapes. Vous pouvez également filtrer les résultats d'une recherche simple, assistée ou avancée selon le même principe, à l'aide de l'Action **Filtrer par métadonnées** ou de l'onglet **Métadonnées** situé à gauche des résultats.