

Tutoriel TXM pour la BFM



Copyright © - ENS de Lyon - UMR IHRIM
<http://textometrie.ens-lyon.fr> - <http://txm.bfm-corpus.org>



Ce document est publié sous licence

[Creative Commons BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Version 3.1

15 mai 2019

Table des matières

1	Introduction.....	3
2	Fonctionnement général de la plateforme.....	4
3	J'accède au portail BFM sans inscription.....	5
4	Je m'inscris pour bénéficier d'un accès étendu.....	9
4.1	Inscription.....	9
4.2	Connexion.....	12
4.3	Modifier son profil.....	12
5	Je crée mon sous-corpus.....	13
5.1	Interface générale de la fonctionnalité Textes de la commande Sous-corpus.....	14
5.1.1	Zone des descripteurs.....	15
5.1.2	Zone des textes.....	17
5.1.3	Zone des statistiques.....	17
5.2	Exemple de sous-corpus.....	18
5.3	Création du sous-corpus.....	19
5.4	Visualisation du sous-corpus : commandes Dimensions et Édition.....	20
6	J'organise mes données : partitionner le sous-corpus.....	21
7	J'analyse les données.....	24
7.1	Index.....	24
7.1.1	Réglage de l'index.....	25
7.2	Lexique.....	27
7.2.1	Réglage du Lexique.....	27
7.3	Concordances.....	28
7.3.1	Création des concordances.....	28
7.3.2	Réglage des paramètres des concordances.....	30
7.4	Contexte.....	33
7.5	J'utilise les propriétés de mots pour interroger le corpus.....	34
7.5.1	Étiquettes morphosyntaxiques Cattex09.....	34
7.5.2	Étiquettes morphosyntaxiques <i>Universal dependencies</i> (UD).....	35
7.5.3	Lemmes.....	35
7.5.4	Propriétés de mots disponibles et combinaisons possibles.....	36
7.6	J'utilise des structures pour interroger le corpus.....	36
8	Exporter un document depuis TXM.....	37
	Annexes.....	43
	Annexe 1 : Liste des propriétés de mots dans les corpus de la Base de français médiéval.....	43
	Annexe 2 : Liste des étiquettes Cattex09.....	45
	Annexe 3 : Tableau de conversion Cattex09 → UD.....	47
	Annexe 4 : Liste des sources des lemmes.....	49

1 Introduction

La **Base de français médiéval** (BFM), créée en 1989, comporte à ce jour [170 textes](#) intégraux d'ancien et de moyen français. Par sa taille (près de 4 700 000 d'occurrences-mots) et la diversité des textes qui la composent, cette base est unique en France pour cette période de l'histoire du français. Elle est utilisée par une communauté de plusieurs centaines de chercheurs, enseignants et étudiants du monde entier.

Les textes réunis dans la Base de français médiéval couvrent une aire géographique importante et une longue période temporelle, du IX^e siècle (avec le premier texte écrit en français, les *Serments de Strasbourg*) à la fin du XV^e siècle. Textes en vers et textes en prose sont représentés, ainsi que les différents domaines et genres textuels (littérature, histoire, hagiographie, droit, sciences...).

Les textes de la BFM sont balisés en XML suivant les recommandations de la [TEI](#). Par ailleurs, les textes de la base bénéficient d'un enrichissement linguistique (étiquetage morpho-syntaxique, lemmatisation) qui offre des possibilités d'exploitation plus riches et plus complexes.

Le portail BFM donne accès à plusieurs corpus de textes en français médiéval, ainsi qu'à un corpus de textes de latin tardif et à un corpus parallèle latin – français préparé dans le cadre du projet PaLaFra¹ et destiné à l'étude du passage du latin au français.

Ce tutoriel a pour but de vous faire découvrir les moyens d'interrogation de la BFM à l'aide de la plateforme TXM au fil d'une session de travail type. Après un survol général de l'interface (section [2](#)), nous vous invitons à découvrir les fonctionnalités accessibles sans inscription (section [3](#)) et vous expliquons la procédure d'inscription (section [4](#)) permettant de bénéficier de fonctionnalités supplémentaires. Nous vous proposons de créer et d'organiser votre corpus de travail (sections [5](#) et [6](#)), puis d'en analyser les données (section [7](#)). Vous apprendrez enfin comment vous pouvez exporter les résultats de vos requêtes pour les exploiter avec un logiciel tableur ou les citer dans un article (section [8](#)).

Pour créer et organiser votre corpus de travail (qu'on appellera désormais sous-corpus), vous vous servirez de la commande **Sous-corpus** et de la commande **Partition**.

Si vous souhaitez travailler sur l'ensemble de la BFM², vous n'avez pas besoin de créer un sous-corpus et vous pouvez passer directement à l'étape suivante. La commande **Sous-corpus** vous permet de composer votre sous-corpus de travail en sélectionnant des textes disponibles dans un corpus prédéfini (déjà intégré dans TXM). La commande **Partition** permet de définir des parties à l'intérieur de votre sous-corpus (ou du corpus de départ, si vous décidez de travailler sur l'ensemble des textes qui le composent), en vue d'analyses contrastives permettant de caractériser ces parties les unes par rapport aux autres. Plusieurs sous-corpus et partitions sont « pré-fabriqués » pour les utilisateurs de la BFM. Vous pouvez les interroger directement.

Par la suite, vous serez amené à utiliser les commandes **Index** et **Concordance**. L'**index** vous permet de lister le vocabulaire des textes, puis d'en tirer différentes fréquences. La **concordance** vous aide à analyser les termes qui vous intéressent dans un contexte précis.

1 <http://palafra.org>.

2 L'ensemble de la bibliographie de la BFM est accessible via la commande **Notices Bibliographiques** (voir début de section 3).

2 Fonctionnement général de la plateforme

Section 4.1 du manuel de référence³.

L'interface de la plateforme web se divise en trois zones : l'explorateur (à gauche), les résultats (dans la partie principale), les messages (dans la ligne de statut tout en bas). Les icônes sur lesquelles portent les commandes sont dans l'explorateur, c'est à dire que l'explorateur contient tous les objets (corpus, sous-corpus et partitions) gérés par la plateforme et sur lesquels s'appliquent les commandes (création de sous-corpus, de partitions, génération d'un index, concordances...). On accède aux différentes commandes via un menu contextuel, en faisant un clic-droit⁴ sur une icône, sur le nom d'un corpus ou d'une partition ; on peut également y avoir accès grâce à la barre d'outils qui présente les commandes disponibles en fonction de l'objet sélectionné (ces commandes varient s'il s'agit d'un sous-corpus ou d'une partition par exemple). Les icônes de partitions et de (sous-)corpus sont conservées d'une session à l'autre (après la déconnexion au portail), mais les onglets donnant accès aux résultats ne le sont pas.

La description du fonctionnement et du lien entre la zone de l'explorateur et la zone des résultats sera plus détaillée dans les sections suivantes, au fil des exemples du tutoriel. La troisième zone, celle des messages, contient les commentaires de chaque action exécutée lors d'une session.

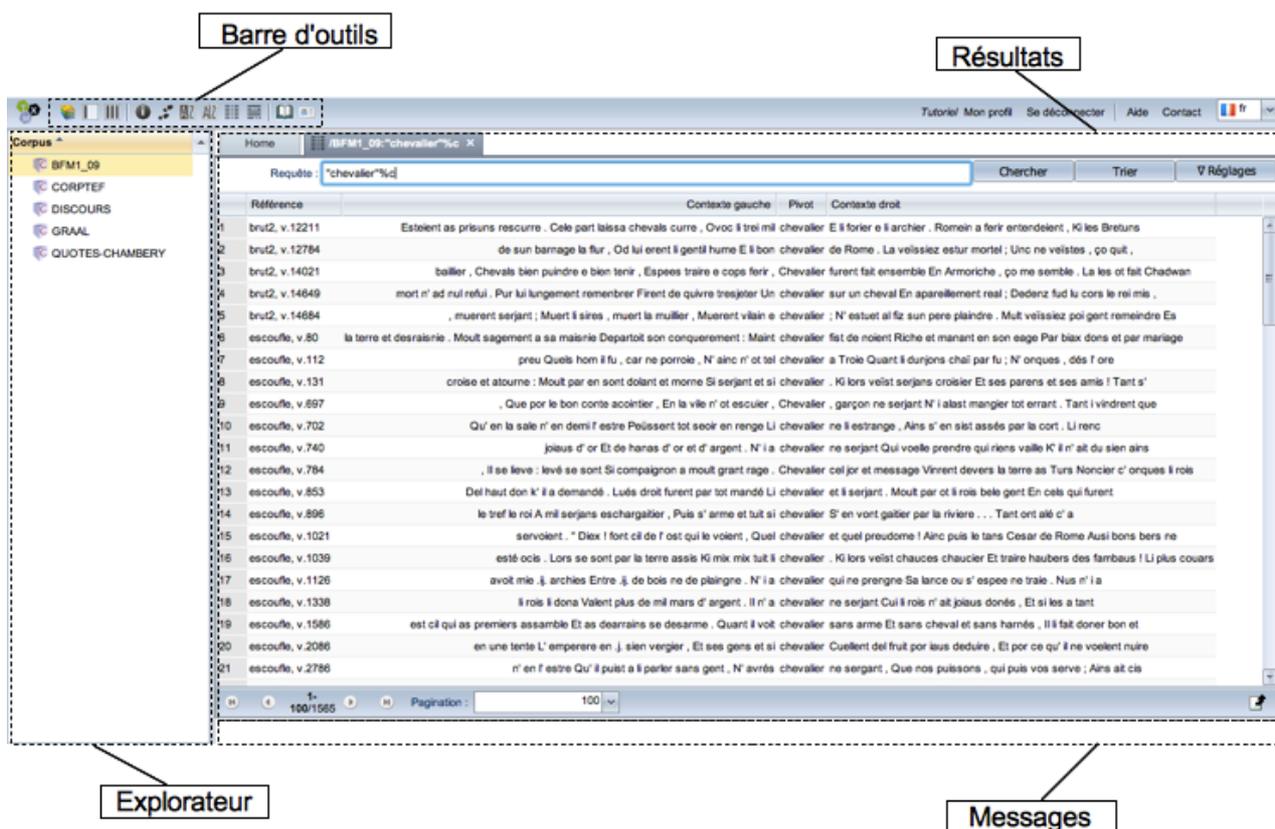


Illustration 1: Interface générale

³ Le manuel de référence en ligne de la plateforme TXM se trouve à l'adresse :

<http://textometrie.ens-lyon.fr/html/doc/manual/0.7.9/fr/manual1.xhtml>

⁴ Le clic-droit fonctionne avec Mozilla Firefox, Chrome et Safari, mais pas sous Internet Explorer. Vous devez utiliser la barre d'outils pour interroger le portail BFM avec ce navigateur.

3 J'accède au portail BFM sans inscription

Dans un navigateur web⁵, j'ouvre l'adresse <http://txm.bfm-corpus.org>.

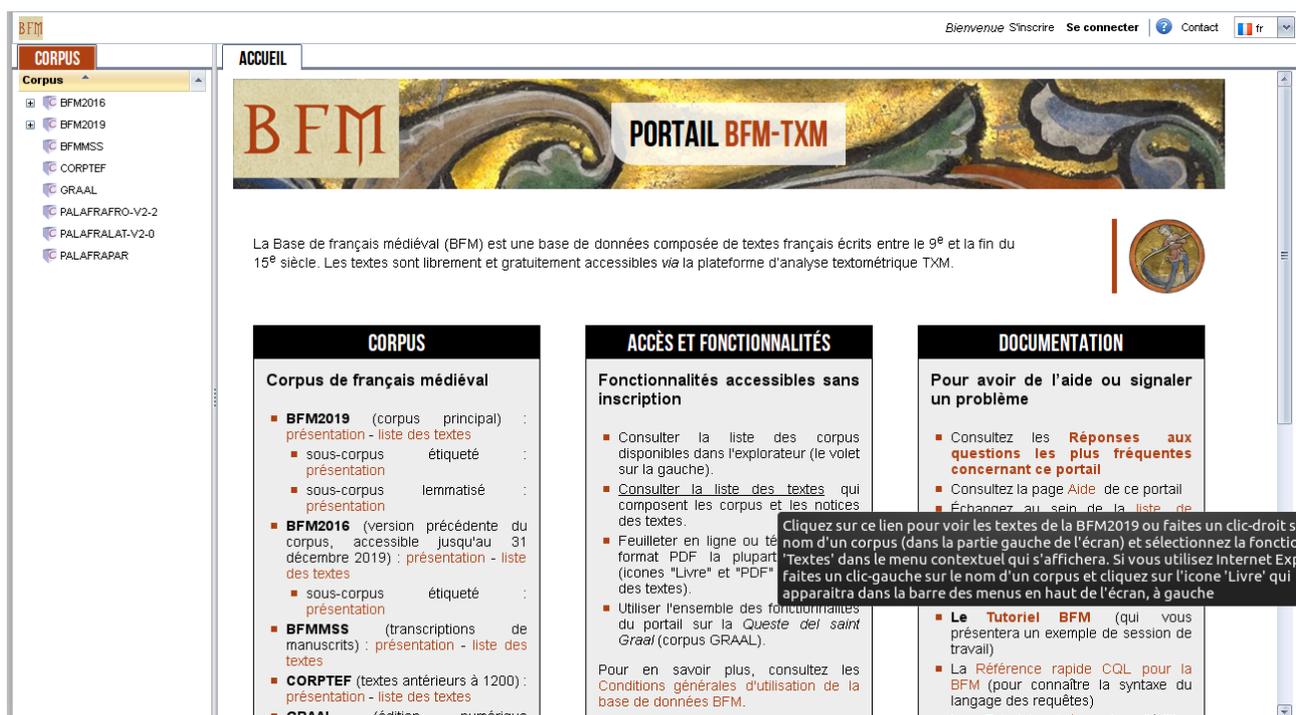


Illustration 2: Écran d'accueil de la BFM

Si l'onglet 'Accueil' reste vide et le message « Aucune donnée à afficher » apparaît dans l'onglet 'Corpus' à gauche, il est nécessaire d'autoriser les cookies provenant du site <http://txm.ish-lyon.cnrs.fr/bfm> dans les paramètres de confidentialité de votre navigateur. Vous pouvez également utiliser l'adresse alternative du portail : <http://portal.textometrie.org/bfm>⁶.

Une partie des fonctionnalités de la BFM sont accessibles sans inscription. Pour chaque corpus, vous pouvez afficher une page d'accueil (icone 'Maison') qui contient sa description et des liens utiles. Vous pouvez également accéder à la liste des textes, aux notices bibliographiques et à la visualisation des textes (formats HTML et PDF) via le panneau 'Corpus' à gauche ou via le lien hypertexte 'Consulter la liste des textes'.

Via le panneau 'Corpus', il faut faire un clic droit sur le corpus désiré, et sélectionner la commande Textes dans le menu qui apparaît.

⁵ Le portail a été testé sous les navigateurs Firefox, Internet Explorer 6, 7, 8 et 11, Microsoft Edge, Chrome et Safari. Sous IE 6, certaines fonctionnalités (clic-droit) ne sont pas disponibles.

⁶ Pour plus de détails, consultez la Foire aux questions de la BFM : <https://groupes.renater.fr/wiki/bfm-liste/faq/aucun-corpus-disponible>.

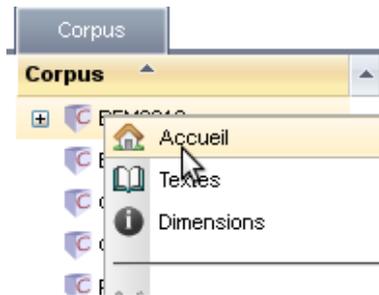


Illustration 3: Utilisation du menu contextuel

Un nouvel onglet s'ouvre contenant la liste des textes disponibles.

identifiant	édition	notice	pdf	auteur	titre	date compo	libre	ms date libre	forme	domaine	genre	dialecte	relation
strasb				anonyme	Serments de Strasl	842		ca. 1000	prose	juridique	serment	non défini	original
eulail				anonyme	Séquence de saint peu après	881		ca. 900	vers	religieux	hagiographie	non défini	original
jonas				anonyme	Sermon sur Jonas	prob. entre 938 et	10e s.		prose	religieux	sermon	flandre	indéterminé
passion				anonyme	Passion de Jésus-c	ca. 1000		ca. 1000	vers	religieux	dramatique	franco-occitan	original
slethgier				anonyme	Vie de saint Léger	ca. 1000		ca. 1000	vers	religieux	hagiographie	franco-occitan	original
alexis				anonyme	Vie de saint Alexis	ca. 1050		ca. 1120	vers	religieux	hagiographie	normand	original
roland				anonyme	Chanson de Rolan	ca. 1100		2e q. 12e s.	vers	littéraire	epique	normand	original
comput				Philippe de Thaon	Comput	1113 ou 1119		3e q. 12e s.	vers	didactique	comput	anglo-normand	adaptation
Lapidal				Philippe de Thaon	Lapidaire alphabéti	1e tiers 12e s.		ca. 1200	vers	didactique	lapidaire	anglo-normand	adaptation
oxfps				anonyme	Psautier d'Oxford	1e m. 12e s.		mil. 12e s.	prose	religieux	psautier	anglo-normand	traduction
bestiaire				Philippe de Thaon	Bestiaire	entre 1121 et 1135		3e q. 12e s.	vers	didactique	bestiaire	anglo-normand	adaptation
gormont				anonyme	Gormont et Isembe	ca. 1130		1e q. 13e s.	vers	littéraire	epique	non défini	original
louis				anonyme	Couronnement de l	ca. 1130		3e q. 13e s.	vers	littéraire	epique	non défini	original
Juise				anonyme	Li ver del juise	2e q. 12e s.		déb. 13e s.	vers	religieux	sermon	liégeois	adaptation

Illustration 4: Affichage de la liste des textes d'un corpus

Dans cet onglet, on peut pour tous les textes accéder aux métadonnées qui décrivent le texte ainsi qu'à une notice bibliographique détaillée. Pour cela on clique sur le bouton Notice du texte choisi.

Gormont et Isebart

Comment citer ce texte :

Gormont et Isebart, édité par A. Bayot, Paris, Champion, 1931. Publié en ligne par la Base de français médiéval, <http://catalog.bfm-corpus.org/gormont>. Dernière révision le 2010-11-19.

Notice détaillée :

- Identifiants :
 - BFM : gormont
 - DEAF : GormB
- Auteur : anonyme
- Titre : Gormont et Isebart
- Date de composition de l'œuvre : ca. 1130
- Date du manuscrit : 1e q. 13e s.
- Éditeur(s) scientifique(s) : A. Bayot
- Édition : 3e éd., Paris, Champion, 1931
- Collection : CFMA, 14
- Forme : vers
- Domaine : littéraire
- Genre : epique
- Dialecte de l'auteur : non défini
- Relation : œuvre originale
- Étiquetage morphosyntaxique : non vérifié

notice du texte choisi

Illustration 5: Consultation de la notice bibliographique d'un texte

Pour les textes en libre accès, on peut également accéder au texte lui-même soit sous la forme de son édition originale HTML via l'icône édition  soit au format PDF via l'icône PDF .

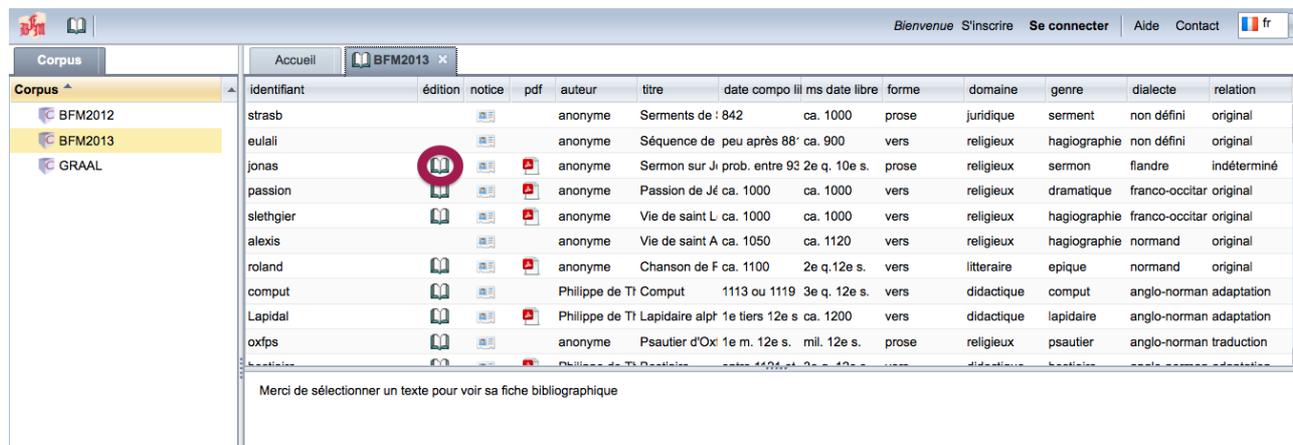


Illustration 6: Icône édition dans la liste des textes

Un clic sur le bouton édition donne accès dans un nouvel onglet de la fenêtre BFM au texte mis en page pour une lecture en ligne.

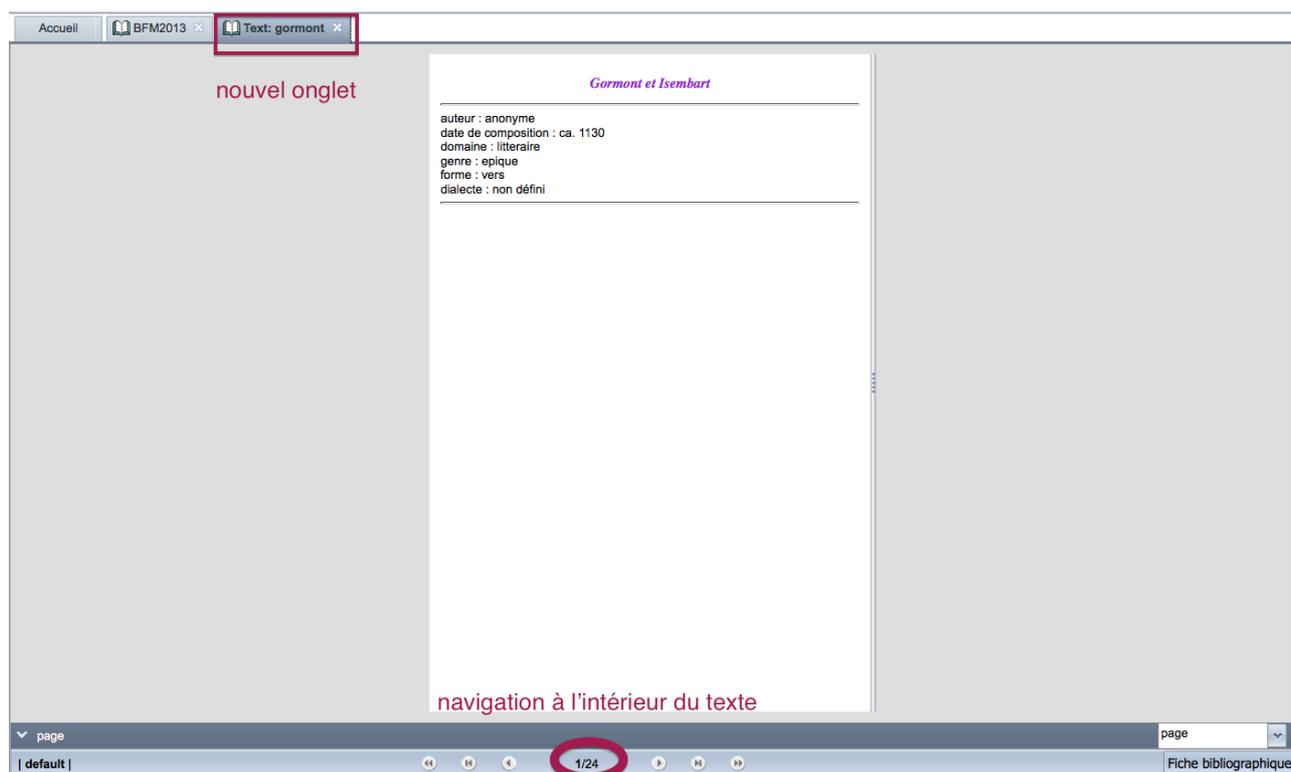


Illustration 7: Affichage de l'édition dans un nouvel onglet

identifiant	édition	notice	pdf	auteur	titre	date compo	ll ms	date libre	forme	domaine	genre	dialecte	relation
strasb				anonyme	Serments de :	842	ca. 1000		prose	juridique	serment	non défini	original
eulali				anonyme	Séquence de	peu après 88'	ca. 900		vers	religieux	hagiographie	non défini	original
jonas				anonyme	Sermon sur J:	prob. entre 93	2e q. 10e s.		prose	religieux	sermon	flandre	indéterminé
passion				anonyme	Passion de Jé	ca. 1000	ca. 1000		vers	religieux	dramatique	franco-occitar	original
slethgier				anonyme	Vie de saint L	ca. 1000	ca. 1000		vers	religieux	hagiographie	franco-occitar	original
alexis				anonyme	Vie de saint A	ca. 1050	ca. 1120		vers	religieux	hagiographie	normand	original
roland				anonyme	Chanson de F	ca. 1100	2e q. 12e s.		vers	littéraire	epique	normand	original
comput				Philippe de Tl	Comput	1113 ou 1119	3e q. 12e s.		vers	didactique	comput	anglo-norman	adaptation
Lapidal				Philippe de Tl	Lapidaire alpr	1e tiers 12e s	ca. 1200		vers	didactique	lapidaire	anglo-norman	adaptation
oxfps				anonyme	Psautier d'Oxi	1e m. 12e s.	mil. 12e s.		prose	religieux	psautier	anglo-norman	traduction

Merci de sélectionner un texte pour voir sa fiche bibliographique

Illustration 8: Icône 'PDF' dans la liste des textes

Un clic sur le bouton PDF donne accès à la version PDF du texte mais cette fois-ci dans un nouvel onglet du navigateur web.

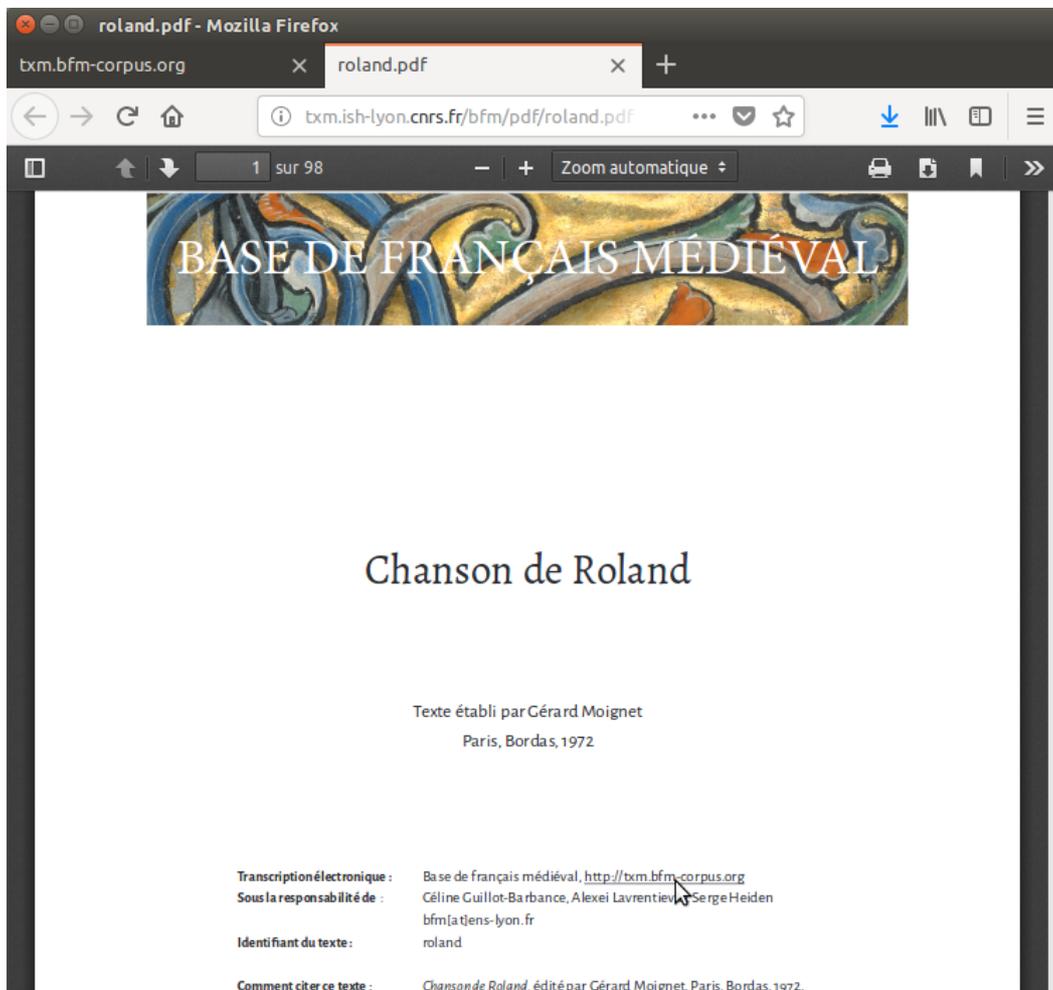


Illustration 9: Affichage d'une édition au format PDF dans un nouvel onglet du navigateur

4 Je m'inscris pour bénéficier d'un accès étendu

4.1 Inscription

Pour avoir accès à davantage de textes et de fonctionnalités (lexiques, index, concordances, téléchargement des sources des textes au format XML-TEI...), vous devez vous inscrire et accepter les conditions d'utilisation de la base de données BFM. Pour cela, vous pouvez cliquer sur le lien hypertexte de la page d'accueil 's'inscrire' ou dans la barre d'outils en haut à droite.



Illustration 10: Bouton 'S'inscrire' dans la barre d'outils

Le formulaire suivant apparaît.

Etape 1 : Renseigner le formulaire d'inscription :

Inscription à la Base de Français Médiéval

Pour obtenir l'accès étendu au corpus de la la Base de Français Médiéval, vous devez :

1. renseigner le formulaire d'inscription
2. lire et accepter les Conditions spécifiques d'utilisation de la base de données la Base de Français Médiéval
3. saisir le code de contrôle visuel

1. Formulaire d'inscription

Nom : Durand *

Prénom : Louis *

Institution : laboratoire XY *

Statut :

- enseignant, enseignant-chercheur ou chercheur
- étudiant ou doctorant
- ingénieur
- autre (préciser) *

autre (préciser) :

Email : louis.durand@xy.com *

Identifiant : louis.durand *

Mot de passe : ***** *

Confirmation du mot de passe : ***** *

Illustration 11: Formulaire d'inscription

Vous devez :

- renseigner vos nom et prénom ;
- indiquer votre institution (ENS, Université...) ;

- indiquer votre statut (enseignant, étudiant...) ;
- renseigner votre adresse mail ;
- choisir un identifiant comportant entre 6 et 20 caractères (vous avez la possibilité de vérifier sa disponibilité en cliquant sur la petite loupe) ;
- choisir un mot de passe à 7 caractères minimum ;
- confirmer votre mot de passe.

Si vous recevez le message suivant lors de votre inscription :



Illustration 12: Message d'erreur 'identifiant déjà utilisé'

cela signifie que l'identifiant que vous avez choisi est déjà utilisé et que vous devez en choisir un autre.

Etape 2 : Lire et accepter les conditions générales d'utilisation de la base de données BFM :



Illustration 13: Conditions générales d'utilisation de la BFM

Il suffit de cocher la case certifiant que l'on accepte les Conditions générales d'utilisation de la BFM.

Etape 3 : Saisir le code de contrôle visuel :

Illustration 14: Saisie du code de contrôle visuel

Enfin cliquer sur Soumettre

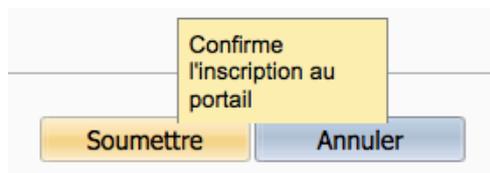


Illustration 15: Bouton 'Soumettre'

Une fenêtre de confirmation apparaît

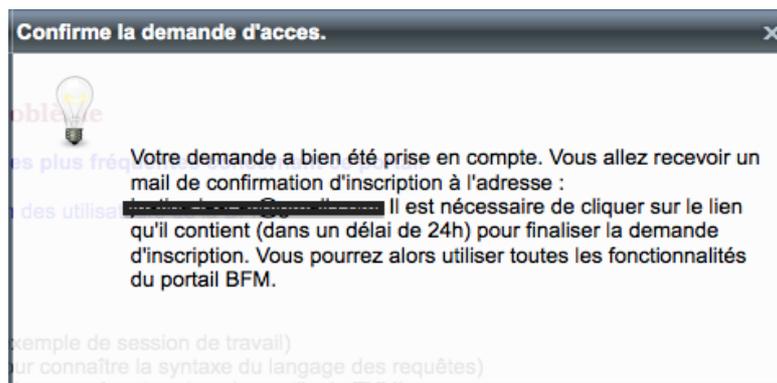


Illustration 16: Confirmation de la demande d'accès

Etape 4 : Finaliser l'inscription

Il suffit de cliquer sur le lien de confirmation d'adresse mail se trouvant dans le mail reçu pour valider la création de compte.

Madame, Monsieur,

Une demande d'accès à la Base de Français Médiéval a été formulée sur le portail <http://txm.ish-lyon.cnrs.fr/bfm> avec l'adresse mail [redacted]@gmail.com.

Cette demande d'accès est associée à la création du compte utilisateur [redacted]. Pour pouvoir valider la création de ce compte, merci de cliquer sur le lien suivant [http://txm.ish-lyon.cnrs.fr/bfm?confirmation=89f8e0fb-585c-4d57-80eb-\[redacted\]](http://txm.ish-lyon.cnrs.fr/bfm?confirmation=89f8e0fb-585c-4d57-80eb-[redacted]) (ce lien est valable 24h) pour associer définitivement votre adresse mail avec ce compte. Une fois cette association établie, vous pourrez bénéficier d'un accès étendu à la BFM.

Illustration 17: Mail de confirmation d'adresse mail pour finaliser l'inscription au portail BFM

Si vous ne recevez pas le mail de confirmation au bout de 5-10 minutes après l'inscription, n'hésitez pas à contacter l'administrateur de la BFM à l'adresse bfm@ens-lyon.fr.

4.2 Connexion

Une fois votre compte créé, vous pouvez vous connecter à la plateforme en cliquant sur **Se connecter** dans le menu. Il vous suffit d'indiquer votre identifiant et votre mot de passe dans la fenêtre suivante :

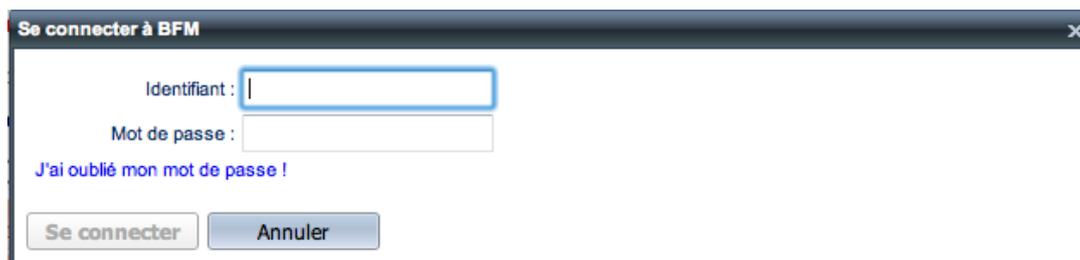


Illustration 18: Fenêtre de connexion

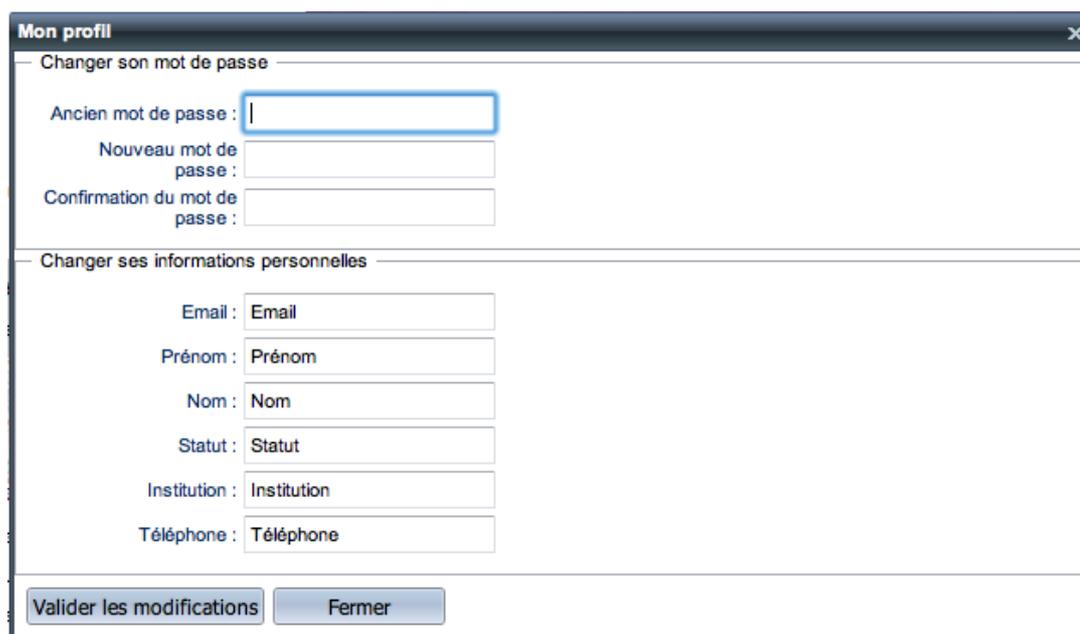
4.3 Modifier son profil

Lorsque vous serez connecté, un nouveau bouton apparaîtra dans le menu, qui vous permettra d'accéder aux informations qui vous concernent et de les modifier. Cliquez sur **Mon profil**, comme indiqué dans l'illustration 19 :



Illustration 19: Bouton 'Mon profil'

Vous pouvez ajouter ou modifier certaines informations à votre profil dans la fenêtre reproduite ci-dessous :



The screenshot shows a window titled "Mon profil" with two main sections. The first section, "Changer son mot de passe", contains three input fields: "Ancien mot de passe", "Nouveau mot de passe", and "Confirmation du mot de passe". The second section, "Changer ses informations personnelles", contains six input fields: "Email", "Prénom", "Nom", "Statut", "Institution", and "Téléphone". At the bottom of the window are two buttons: "Valider les modifications" and "Fermer".

Illustration 20: Fenêtre "Mon profil"

Vous avez la possibilité de :

- Modifier votre mot de passe
- Mettre à jour votre adresse mail
- Mettre à jour vos informations personnelles (Nom, Statut, Institution, Téléphone).

5 Je crée mon sous-corpus

Vous allez partir d'un des corpus de la BFM déjà inclus dans la plateforme.

Si vous souhaitez avoir accès aux informations bibliographiques des corpus prédéfinis, vous devez lancer la commande **Textes** accessible depuis le menu contextuel, en faisant un clic-droit sur le corpus qui vous intéresse (ou en cliquant sur  dans la barre d'outils). Cette commande vous permettra d'accéder à la liste des textes qui constituent le corpus et de connaître pour chacun de ces textes : l'identifiant BFM, l'auteur, le titre, la date de composition, la date du manuscrit, l'éditeur scientifique, la maison d'édition, la date de l'édition, la forme, le domaine, le genre, le dialecte, l'état de l'étiquetage morphosyntaxique et le nombre de mots. Si vous souhaitez opérer une sélection à l'intérieur des textes, vous avez la possibilité d'utiliser ces informations pour extraire uniquement les données qui vous intéressent.

En utilisant et combinant les descripteurs des textes, vous pourrez ainsi créer autant de sous-corpus que vous souhaitez en **sélectionnant** chaque fois (cf. section [5.2](#)) un ensemble de textes à l'intérieur d'un corpus de départ : par exemple, les textes du 12^{ème} et du 13^{ème} siècle qui relèvent des domaines Histoire et Littérature à l'intérieur du corpus BFM2016. Vous pourrez également **visualiser** une édition de votre sélection grâce à la commande **Édition**.

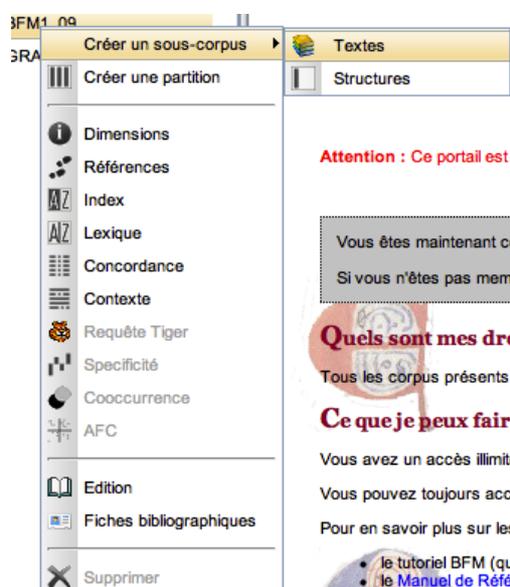
Vous pourrez ensuite **organiser** (section [6](#)) votre sous-corpus afin d'obtenir des résultats comparatifs en fonction des paramètres que vous aurez choisis : par exemple, vous pouvez décider

de répartir les décomptes de mots en fonction de la forme des textes (vers et prose) ou des deux domaines (Histoire et Littérature). La fonctionnalité **Partition** regroupera les résultats obtenus sur l'ensemble du sous-corpus en fonction de la distinction vers/prose dans un cas, et textes historiques/textes littéraires dans l'autre. Cette fonctionnalité de la plateforme s'utilise notamment en association avec la commande **Index** (section 7).

Afin de créer un sous-corpus, vous allez utiliser la commande **Créer un sous-corpus / Textes** qui vous permettra de faire intervenir plusieurs descripteurs (un descripteur chronologique et un descripteur concernant le domaine du texte par exemple) dans la création du sous-corpus.

5.1 Interface générale de la fonctionnalité **Textes** de la commande **Sous-corpus**

- Lancez d'abord la commande **Sous-corpus/Textes** sur le corpus en faisant un clic-droit sur le nom du corpus (**Créer un sous-corpus/Textes**) ou en sélectionnant  depuis la barre d'outils (voir illustration 21). Ne pas s'inquiéter de ne pas voir immédiatement d'effet, le temps d'affichage de cette interface de sélection de textes peut être un peu long.



*Illustration 21: Lancez la commande **Sélection de textes** via le menu contextuel sur le corpus sélectionné*

L'onglet des résultats de cette commande se divise en deux zones : à gauche, des volets déplaçables listent les différents descripteurs⁷ décrivant les textes ; et à droite, un tableau détaille l'ensemble des textes qui composent le corpus. Par défaut, aucun texte n'est sélectionné, et au fur et à mesure de vos choix dans les descripteurs vous verrez apparaître puis s'ajuster des coches à gauche des textes sélectionnés dans la liste.

Pour construire le sous-corpus à partir des textes disponibles, on peut ajouter ou retirer des textes en utilisant les descripteurs (zone de gauche), puis si besoin affiner la sélection texte par texte (zone de droite). Vous avez également la possibilité de faire apparaître sous une forme graphique la répartition des valeurs sélectionnées via le bouton **Stats** (une troisième zone apparaîtra à droite).

Pour tous les tableaux (ceux dans les volets des descripteurs, comme la liste détaillée des textes), un

⁷ La terminologie BFM et TXM est différente, nous parlerons ici de « descripteur de texte » ce qui équivaut à « propriété de texte » pour les autres commandes TXM.

clic-gauche sur un entête de colonne effectue un tri sur les valeurs de cette colonne, et un clic-droit donne accès à un choix plus large d'opérations de mise en forme du tableau (dont le choix des colonnes à afficher).

5.1.1 Zone des descripteurs

Un descripteur correspond à une dimension de description d'une unité textuelle. L'ensemble des descripteurs rassemble les métadonnées disponibles et/ou utiles pour l'interprétation et l'exploitation des textes, ils dépendent donc du corpus. Voici les descripteurs utiles et disponibles dans le corpus BFM2019⁸ :

-
- auteur : Bérout, Clari, anonyme⁹...
- titre : titre de l'œuvre (*Serments de Strasbourg, Queste del saint Graal...*)
- siècle : siècle de composition de l'œuvre (13, 12, 10...)
- date œuvre: date de composition de l'œuvre (intervalle)
- date ms : date du manuscrit (intervalle)
- forme : forme des textes (vers, prose, mixte)
- domaine : domaine fonctionnel du texte (littéraire, historique, didactique, religieux ou juridique.)
- genre : genre discursif (épique, chronique...)
- dialecte : dialecte principal de l'œuvre (picard, anglo-normand, non défini...)
- relation : relation éventuelle à d'autres œuvres (adaptation, traduction...)
- morphosynt : l'état de l'étiquetage morphosyntaxique (vérifié ou non)
- lemmatisation : l'état de la lemmatisation (vérifiée ou non)
-

Ces descripteurs sont utilisables pour définir un sous-corpus. Pour considérer un descripteur et indiquer les valeurs que l'on veut retenir ou exclure, cliquer sur la petite flèche (^) à gauche du nom du descripteur pour déployer le volet correspondant. La flèche a simplement pour effet de montrer ou de masquer le tableau de sélection pour le descripteur, un descripteur peut être actif bien que le détail du tableau soit masqué, et un descripteur peut être inutilisé alors que le tableau détaillant ses valeurs est affiché. L'indicateur chiffré à côté du nom du descripteur, sur la barre bleue qui ouvre chaque volet, renseigne d'ailleurs en permanence sur l'utilisation effective du descripteur : le chiffre « (n/N) » indique qu'il y a N valeurs différentes du descripteur dans le corpus dont n sélectionnées. Un descripteur inutilisé affiche donc un indicateur nul de la forme « (0/N) ». Quand on ouvre le volet, on voit donc s'afficher un tableau comprenant N lignes, correspondant aux N valeurs disponibles, avec les n valeurs déjà sélectionnées le cas échéant. (Le cas des descripteurs de type date est différent, car on raisonne par intervalle et non plus par sélection de valeurs, ce qui explique que l'on n'ait pas d'indicateur récapitulatif « (n/N) ».)

⁸ Plus d'informations : http://ccfm.ens-lyon.fr/IMG/pdf/Manuel_Descripteurs_BFM.pdf

⁹ La partie « anonyme » est composée de l'ensemble des œuvres anonymes et n'a donc pas le même statut que les autres.

Comme on l'a vu, les valeurs des descripteurs sont par défaut décochées, et initialement, dans la liste des textes, à droite, aucun texte n'est sélectionné. Pour constituer une sélection des textes, il faut cocher les cases, soit directement en cliquant sur la case, soit en double-cliquant sur la ligne, soit en sélectionnant une ou plusieurs lignes puis en appliquant une commande pour « ajouter » (+) ou « retirer » (-). Voici le récapitulatif des actions sur les lignes du tableau :

- Double-cliquer sur une valeur pour l'activer ou la désactiver
- + : ajouter les valeurs sélectionnées (surlignées avec clic gauche, ou CTRL+clic gauche pour une sélection multiple)
- - : désactiver les valeurs sélectionnées
- ⇄ : inverser la sélection courante des valeurs

Les colonnes donnent des informations quantitatives en montrant comment le corpus disponible et les textes sélectionnés se répartissent sur les différentes valeurs du descripteur.

Pour chacune des valeurs, N est le nombre de textes disponibles, n le nombre de textes dans la sélection (au départ, comme la sélection ne comprend aucun texte, on a $n=0$). Et même chose pour la taille en nombre de mots : pour chaque valeur de descripteur, T est la taille en nombre de mots de ce qui est disponible, et t est la taille en nombre de mots dans la sélection.

Ce tableau peut être trié en cliquant sur l'entête de chaque colonne.

Remarque : on peut sélectionner ou désélectionner toutes les valeurs (et ainsi changer l'état initial), soit pour *un* descripteur (cf. commandes ci-dessus), soit pour *tous* les descripteurs, en utilisant les boutons généraux en haut à gauche :

Appuyer sur le bouton *Remise à zéro* désactive toutes les valeurs de tous les descripteurs (il n'y a donc initialement plus de contraintes pour la sélection de textes).

- Appuyer sur le bouton *Tout*, coche la totalité des valeurs des descripteurs. Vous pouvez alors effectuer une sélection de textes en décochant des valeurs.

On peut sélectionner des valeurs dans plusieurs descripteurs différents, qui sont alors croisés (cf. exemple ci-après, où l'on sélectionne des textes représentatifs à la fois d'une certaine période et de certains genres). Plus précisément, la sélection du sous-corpus dans cette interface de la zone des descripteurs correspond à une équation de recherche sur les valeurs des descripteurs des textes construite comme suit :

- pour chaque descripteur ayant des valeurs activées, on introduit une contrainte en « OU » sur les différentes valeurs activées ;
- entre les différents descripteurs, on introduit une contrainte en « ET ».

Soit donc, sous-jacente, une équation de la forme :

descripteur A = valeur A1 **OU** A2 **OU** A3 **OU** etc.

ET

descripteur B = valeur B1 **OU** B2 **OU** etc.

ET

descripteur C = valeur C1 **OU** C2 **OU** etc.

5.1.2 Zone des textes

La zone des textes (à droite) réagit en fonction de la zone des descripteurs (à gauche). Dès qu'une valeur de descripteur est cochée, l'ensemble des textes retenus dans le sous-corpus, signalés par des coches vertes ✓, s'ajuste pour vérifier la condition de sélection exprimée. Quand un seul descripteur est considéré, chaque sélection d'une nouvelle valeur du descripteur ajoute des textes, si bien que de nouvelles coches vertes apparaissent dans la zone des textes. Puis, quand on fait intervenir un ou plusieurs autres descripteurs, chaque sélection d'une nouvelle valeur ajoute des critères contraignant la sélection, et des coches vertes peuvent alors s'enlever.

Une fois vos descripteurs sélectionnés, vous avez la possibilité d'affiner votre sous-corpus directement dans cette zone, en ajustant la composition de la sélection texte par texte. Un clic-gauche au début de la ligne correspondant à un texte modifie la sélection pour ce texte. Si le texte n'était pas sélectionné, il est ajouté et sera alors coché ; et s'il figurait déjà dans la sélection, il est enlevé (ce qui se traduit visuellement par le retrait de la coche verte)

Remarque : les modifications effectuées directement sur le tableau de la zone des textes sont perdues si l'on fait ensuite une modification dans la zone des descripteurs. Autrement dit, en pratique, il faut commencer par exprimer la sélection à l'aide des descripteurs, et terminer en affinant si besoin la sélection sur le tableau de la zone des textes.

Afin de visualiser plus facilement les informations concernant le texte que vous souhaitez sélectionner, vous pouvez utiliser le tri ascendant ou descendant (un clic sur l'en-tête d'une colonne) ou masquer/afficher certaines colonnes (clic-droit sur n'importe quelle colonne, menu Colonnes, décocher/cocher un élément) :

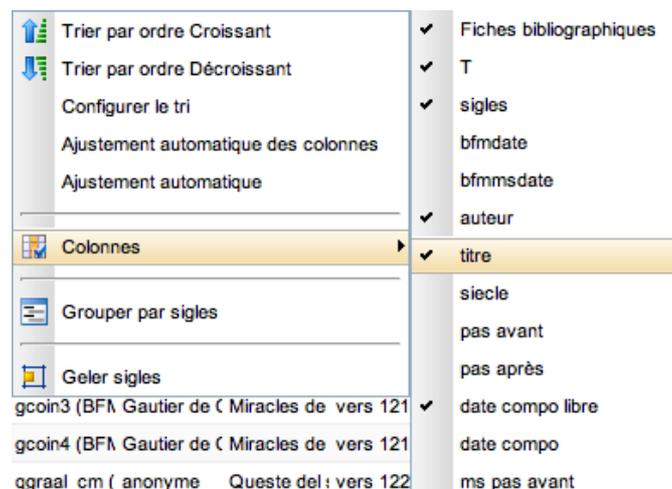


Illustration 22: Afficher/Masquer la colonne "Titre"

Vous pouvez également accéder à la notice bibliographique d'un texte en cliquant sur son icône .

5.1.3 Zone des statistiques

On peut avoir si on le souhaite une représentation graphique, plus visuelle, de la distribution des valeurs de descripteurs, sous la forme d'histogrammes, en cliquant sur le bouton *Stats*, (en haut à gauche). Une troisième zone apparaît alors à droite. Dans chaque volet on trouve un histogramme des valeurs d'un descripteur : la longueur de la barre correspond à la proportion du sous-corpus pour chaque valeur du descripteur, en nombre de textes (ou en nombre de mots en fonction de l'option cochée). Cette visualisation assiste la construction d'un corpus équilibré selon un descripteur en cherchant à obtenir des barres de longueur égale.



Illustration 23: Statistiques du descripteur "domaine" en nombre de mots

Ces statistiques seront également contenues dans le récapitulatif, qu'il est possible d'obtenir à la fin de la création du sous-corpus (cf. récapitulatif)

5.2 Exemple de sous-corpus

Nous voulons isoler les textes du 12^{ème} et 13^{ème} siècles traitant de littérature et d'histoire :

- Dans la zone de gauche, ouvrez le volet correspondant au descripteur `siècle`.
- Sélectionnez les 12^{ème} et 13^{ème} siècles (dans le volet « `siècle` », voir illustration 24).

siècle (1/6)					
valeur ^	n	N	t	T	
<input type="checkbox"/> 11	0	1	0	5806	
<input checked="" type="checkbox"/> 12	4	4	37987	37987	
<input checked="" type="checkbox"/> 13	11	11	854112	854112	
<input type="checkbox"/> 14	0	5	0	374978	
<input type="checkbox"/> 15	0	6	0	411304	
<input type="checkbox"/> 9	0	2	0	344	

Illustration 24: Sélectionner les siècles 12^{ème} et 13^{ème}

Comme au départ, par défaut, tous les siècles sont désélectionnés, vous pouvez soit :

- Surligner les lignes que vous souhaitez ajouter en maintenant la touche `CTRL` lors de votre sélection par la souris, puis valider votre sélection en cliquant sur .
- Cocher directement les lignes concernées.
- Double-cliquer sur ces dernières.

La liste de textes (le volet de droite) se modifie en conséquence, ainsi que les histogrammes de répartition si on a choisi de les afficher.

Nous vous proposons maintenant d'exprimer une contrainte supplémentaire sur un deuxième descripteur de texte.

- Sélectionner les domaines `littéraire` et `historique` (voir illustration 25).

Comme indiqué précédemment, ajoutez les domaines que vous souhaitez conserver. Votre liste de textes (dans le volet de droite) est une nouvelle fois modifiée en conséquence :

valeur ^	n	N	t	T
<input type="checkbox"/> didactique	0	4	0	45280
<input checked="" type="checkbox"/> historique	4	5	77480	321157
<input type="checkbox"/> juridique	0	1	0	129
<input checked="" type="checkbox"/> litteraire	5	11	511573	937970
<input type="checkbox"/> religieux	0	8	0	379995

Illustration 25: Sélectionner les domaines littéraire et historique

Pour information, en termes formels, la sélection des textes opérée correspond alors aux textes vérifiant l'équation suivante :

siècle= (12 OU 13)

ET

domaine= (litteraire OU historique)

5.3 Création du sous-corpus

- Validez votre sélection en cliquant sur Valider (en haut à droite de la fenêtre). Vous pouvez alors nommer votre sous-corpus : 12_13_Hist_Litt :

Finalisation de la sélection

Nom de la sélection : 12_13_Hist_Litt

Afficher le récapitulatif (dans un onglet) ?

Afficher le nombre de textes

Afficher le nombre de mots

Valider Annuler

Illustration 26: Nommer le sous-corpus.

☞ Si vous cochez l'option « Afficher le récapitulatif », un nouvel onglet affiche les données de votre sous-corpus : les descripteurs sélectionnés, les textes retenus et les statistiques si vous avez choisi de les afficher. Vous pouvez également cocher « Afficher le nombre de textes » et « Afficher le nombre de mots » pour ajouter les colonnes n, N (mots), t, T (textes) et leurs pourcentages respectifs au récapitulatif.

☞ Un message d'erreur s'affichera si vous essayez d'enregistrer un sous-corpus sous un nom que vous avez déjà utilisé.

- Le nouveau corpus apparaît maintenant dans l'explorateur (voir illustration 27).



Illustration 27: Sous-corpus nouvellement créé

- ☞ Vous obtenez alors un corpus des textes historiques et littéraires des 12^{ème} et 13^{ème} siècles.
- ☞ Vous ne pouvez pas modifier un sous-corpus déjà créé. Vous avez cependant la possibilité de le supprimer (clic-droit, commande Supprimer ou  depuis la barre d'outils).

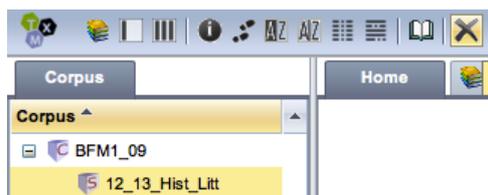


Illustration 28: Supprimer un sous-corpus

Règles pour les noms de sous-corpus et de partition :

- le nom est sensible à la casse (distinction minuscule/majuscule).
- le caractère suivant est interdit : barre oblique (/).

5.4 Visualisation du sous-corpus : commandes Dimensions et Édition

Sections [8.1](#) à [8.2](#) du manuel de référence.

- Vous pouvez visualiser les dimensions du sous-corpus nouvellement créé en le sélectionnant et en lançant la commande Dimensions ( dans la barre d'outils) :

Dimensions de BFM1_09/ :12_13_Hist_Litt
 Nombre d'occurrences (T): 589062

Illustration 29: Dimensions du sous-corpus

Cette commande vous donne la « taille » de votre sous-corpus en nombres d'occurrences.

- ☞ Appliquée à un corpus, cette commande donne également le nombre de textes.
 - Vous avez la possibilité d'accéder à une édition des textes contenus dans le sous-corpus créé. Sélectionnez le sous-corpus, puis dans la barre d'outils () ou dans le menu contextuel, lancez la commande Édition.

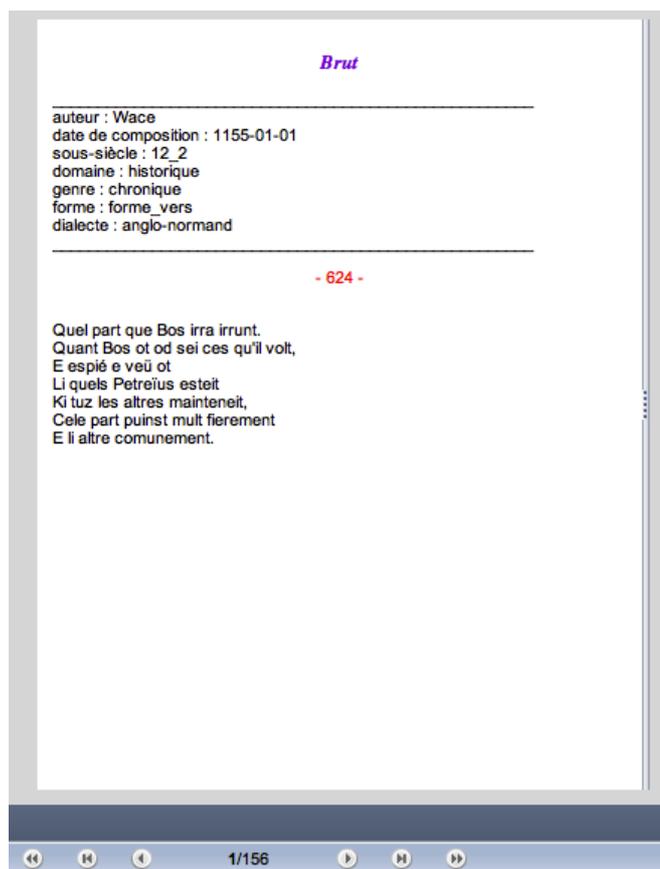


Illustration 30: Édition d'un texte du sous-corpus 12_13_Hist_Litt

Vous pourrez alors naviguer dans tous les textes accessibles du sous-corpus. Chaque texte présente sur sa première page ses principaux descripteurs. Vous pouvez naviguer dans l'édition ainsi :

- > : page suivante
- < : page précédente
- >| : dernière page
- |< : première page
- >> : texte suivant
- << : texte précédent

☞ L'interface de navigation est commune à toutes les commandes de TXM.

☞ Les commandes Édition et Dimensions sont disponibles pour tous les corpus : corpus principaux et sous-corpus.

6 J'organise mes données : partitionner le sous-corpus

Section [8.9](#) du manuel de référence.

Vous souhaitez à présent comparer les résultats en fonction de la forme des textes (vers et prose) mais aussi en fonction des deux domaines sélectionnés (littéraire et historique). La commande Partition vous permettra de **mettre en contraste** certaines propriétés de votre corpus. En pratique, la

commande `Partition` s'applique le plus souvent à la structure `text`.

Nous vous proposons donc de créer deux partitions sur votre nouveau corpus de travail, l'une contrastant les formes, l'autre les domaines.

- Lancez la commande `Partition` en mode simple sur le sous-corpus `12_13_Hist_Litt` (voir illustration 31) depuis le menu contextuel ou la barre d'outils (▮▮▮).

Nous vous conseillons, pour cela, de veiller attentivement à sélectionner la bonne icône dans l'explorateur. Chaque sous-corpus créé peut être sélectionné et partitionné. Une partition se présente comme « le descendant » de son sous-corpus d'origine (elle est en retrait sur la droite). Vos partitions seront donc d'abord définies par leur sous-corpus d'origine.

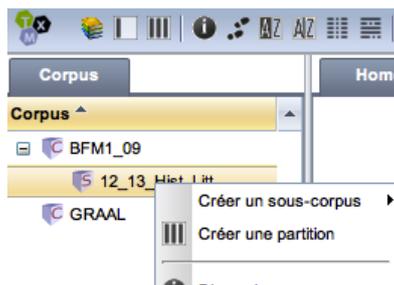


Illustration 31: Lancer la commande `Partition` sur un sous-corpus sélectionné

- Sélectionnez la structure `text` et la propriété¹⁰ `forme`¹¹. Nommez votre partition, par exemple `Vers_Prose`¹² (voir illustration 32).



Illustration 32: Création d'une partition portant sur la forme du texte

- Cliquez sur `Créer` et votre partition apparaîtra dans l'explorateur :

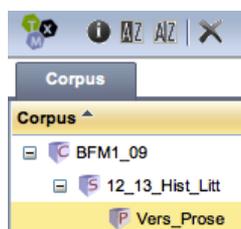


Illustration 33: Partition `Vers_Prose`

10 Comme souligné précédemment, « propriété » équivaut à « descripteur ».

11 Les structures et les propriétés possibles correspondent aux balises XML-TEI encodées dans les sources des textes.

12 Attention : l'usage du caractère « / » est interdit dans les noms de partitions.

☞ Vous obtenez une partition contenant deux parties correspondant aux deux formes définies dans la BFM : vers et prose.

- Lancez à nouveau la commande **Partition**, toujours sur le sous-corpus 12_13 Hist_Litt. Sélectionnez encore la structure `text` mais cette fois-ci avec la propriété `domaine`. Vous obtiendrez une répartition des résultats en fonction des domaines représentés dans le sous-corpus, dans le cas présent les domaines littéraire et historique. Nommez votre partition, par exemple `Hist_Litt`.



Illustration 34: Création d'une partition sur le domaine

Cliquez sur **Créer** et votre nouvelle partition apparaîtra dans l'explorateur :

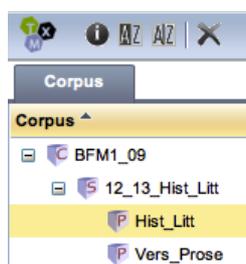


Illustration 35: Partition sur l'opposition des deux domaines, histoire et littérature

☞ Vous obtenez une partition contenant deux parties correspondant aux deux domaines de la BFM représentés dans le sous-corpus : histoire et littérature.

Par la suite, vous pourrez utiliser alternativement chacune de ces partitions en sélectionnant l'une d'entre elles puis en lançant vos commandes.

Ces partitions ne sont que deux exemples parmi d'autres, voici la liste des propriétés utiles et disponibles :

- auteur
- domaine
- deaf
 - identifiant d'édition dans la bibliographie du DEAF (une partie par édition)
- forme
- genre

- id
 - identifiant BFM (une partie par unité textuelle)
- siecle
- titre
- morphosyntaxe
- lemmatisation (non disponible dans la BFM2016)

7 J'analyse les données

Vous pouvez à présent analyser les données de votre sous-corpus en utilisant les commandes de TXM. La commande `Index` vous permettra de **lister les formes** (ou les étiquettes) de mots ou de constructions correspondant à un certain motif recherché. L'index indique aussi la **fréquence** de ces formes dans le sous-corpus et ses partitions. Cette fonctionnalité est utile :

- pour identifier les formes qui vous intéressent : elle vous permet de trouver la bonne expression de requête de motif qui élimine le bruit (les occurrences non souhaitées) et intègre toutes les variantes graphiques des formes sur lesquelles vous travaillez ;
- pour se faire une idée précise de la fréquence de ces formes dans votre sous-corpus et ses éventuelles parties : c'est à ce stade que vous utiliserez les parties créées grâce à la fonctionnalité `Partition` (section [6](#)).

La commande `Lexique` est similaire à la commande `Index` mais ne vous permettra pas d'entrer une requête précise. Elle vous donnera simplement la liste de tous les mots du corpus de travail, ponctuations incluses, avec leurs fréquences respectives.

La commande `Concordance` vous permettra de **visualiser et de trier les occurrences** des formes et constructions recherchées dans un contexte de taille et de forme modulables. La commande `Contexte` est similaire à cette dernière mais permet une visualisation différente, sous forme de paragraphes plutôt que de lignes.

Les textes de la BFM sont étiquetés en morphosyntaxe et balisés pour le discours direct. Certains corpus sont également lemmatisés. À la fin de cette section, nous vous expliquerons comment vous pouvez utiliser ces annotations avec les commandes `Index`, `Concordance` et `Contexte`.

7.1 Index

Section [8.3.2](#) du manuel de référence.

Nous vous proposons de lister les occurrences de la forme « chevalier » (avec ou sans majuscule). Pour cela, on utilise le motif exprimé par la requête CQL suivante : `"chevalier"%c`. (Cf. la [Référence rapide CQL](#) pour plus d'informations sur la syntaxe des requêtes)

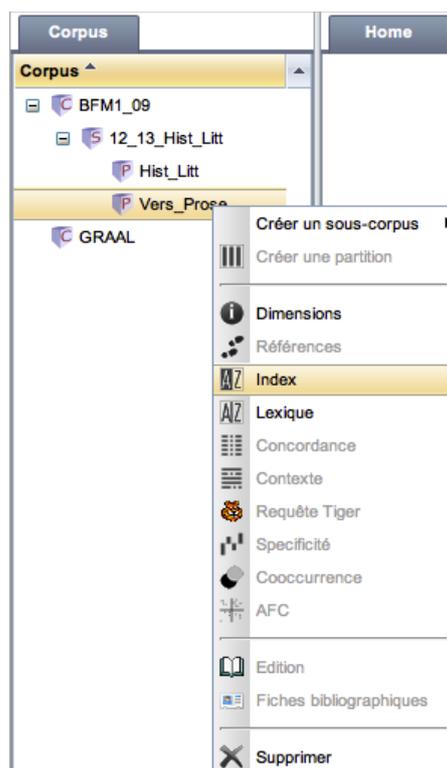


Illustration 36: Lancer la commande Index sur une partition sélectionnée

Lancez la commande Index (AZ) sur une des partitions puis saisissez votre expression CQL. Validez ensuite la recherche avec le bouton Chercher

Requête : "chevalier"%c			
	word	prose	vers
1	chevalier	829	33
2	Chevalier	10	3

Illustration 37: Résultat de l'Index de "chevalier"%c sur la partition Vers/Prose

Le tableau obtenu vous permet d'examiner la liste des formes correspondant à l'expression CQL. Il indique également les fréquences de toutes les formes du mot « chevalier » dans le corpus de travail et dans les différentes parties de la partition. Afin d'analyser plus facilement ces résultats, vous pouvez exporter le tableau obtenu (voir la section 8).

7.1.1 Réglage de l'index

Vous pouvez régler certains paramètres de l'index en cliquant sur le bouton Réglages. Les changements ne sont pris en compte que lorsque l'on appuie sur le bouton Chercher :

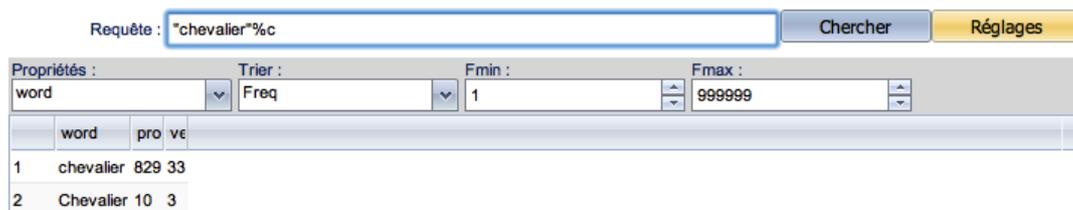


Illustration 38: Réglages de l'index

- Propriétés : permet de sélectionner la propriété d'analyse, à savoir la représentation du mot à utiliser pour construire la liste et calculer les fréquences. La liste des propriétés dépend du corpus (voir Annexe 1). Il est possible de cocher plusieurs cases pour afficher plusieurs propriétés. Voici par exemple la liste des propriétés de mots dans le corpus BFM2019¹³ :
 - word (forme graphique, propriété par défaut)
 - cattex-pos (étiquette morphologique au format [Cattex 09](#))
 - cattex-pos_syn (étiquette morphosyntaxique au format [Cattex 09](#), voir les [Principes d'annotation](#) pour connaître la différence entre l'étiquetage morphologique et morphosyntaxique)
 - etiquetage (état de vérification de l'étiquetage)
 - fropos (étiquette morphologique au format [Cattex 09](#), conservée pour la compatibilité avec les versions plus anciennes de la BFM, identique à *cattex-pos*)
 - id (identifiant de l'occurrence)
 - lang (langue de l'occurrence)
 - lemma (lemme, voir la section 7.5 pour en savoir plus sur la lemmatisation de la BFM)
 - lemma_src (source du lemme : DMF, TL, AND, BFM ou autre)
 - lemmatisation (état de vérification de la lemmatisation)
 - q (niveau d'imbrication dans discours direct : « 0 » si hors discours direct... « 3 » si l'occurrence se trouve dans un discours direct imbriqué dans deux autres discours directs)
 - ref (titre du texte et page où se trouve l'occurrence)
 - sp (« t » si l'occurrence se trouve dans une prise de parole dans un texte dramatique ou dans un dialogue, sinon « f »)
 - supplied (« t » si l'occurrence se trouve dans une conjecture de l'éditeur scientifique, sinon « f »)
 - ud-pos (étiquette de partie de discours du jeu [Universal dependencies](#))
 - ud-feats (étiquette qui regroupe les traits morphologiques du jeu [Universal dependencies](#))
- Trier : permet de trier votre index par fréquence ou par ordre alphabétique des mots (ou valeurs de propriété) en colonne gauche
- Fmin : permet de choisir un seuil de fréquence minimum
- Fmax : permet de choisir un seuil de fréquence maximum
- Vmax : Vmax : permet de limiter la liste aux N formes les plus fréquentes (N étant la valeur choisie pour Vmax)
- Nombre de lignes : permet de changer le nombre de lignes par page

Vous pourrez renouveler ces étapes sur chaque partition créée.

- L'observation du résultat peut amener à remanier :
 - la requête, si elle ne sélectionne pas les occurrences voulues : soit qu'il y ait des intrus à exclure, soit qu'il y ait des formes non couvertes par l'équation ;
 - la partition : la partition : création d'une nouvelle partition (même pour un petit ajustement) ;

¹³ Dans la section [7.5](#) vous verrez comment utiliser les propriétés pour interroger le corpus. La liste complète des propriétés de mots disponibles dans les corpus de la BFM est donnée dans l'Annexe 1.

- le sous-corpus : le réduire (par exemple pour gagner en équilibre) ou l'étendre en créant un nouveau sous-corpus.

7.2 Lexique

Section 6.3.1 du manuel de référence.

- Sélectionnez le corpus de travail ou la partition de votre choix et lancez la commande Lexique (accessible depuis le menu contextuel ou depuis la barre d'outils )

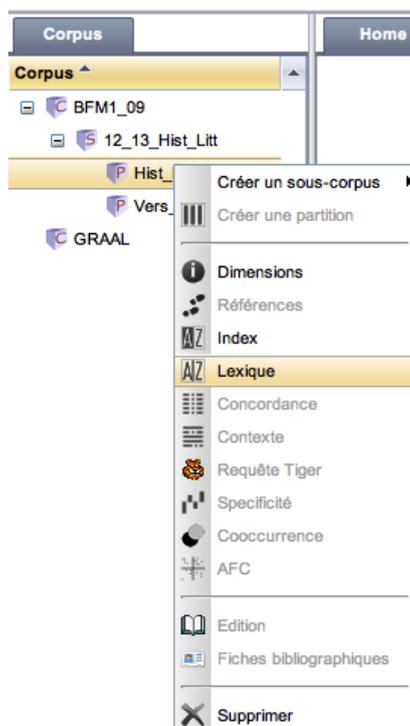


Illustration 39: Lancer la commande Lexique

Cette commande génère un tableau présentant chaque forme graphique recensée et sa fréquence d'apparition au sein du corpus de travail (lorsque la commande est lancée sur un corpus ou sous-corpus) ou au sein de chacune des parties (lorsque la commande est lancée sur une partition) :

	word	histo	litterai
1	,	5668	27077
2	et	3347	16306
3	.	2361	14602
4	il	1363	15595
5	que	897	11077

Illustration 40: Lexique de la partition Hist_Litt

7.2.1 Réglage du Lexique

Vous pouvez régler certains paramètres du lexique en cliquant sur le bouton Réglages. Chaque changement doit être validé en appuyant sur le bouton Calculer :



Illustration 41: Réglages du Lexique

- Propriétés : permet de sélectionner une propriété d'affichage pour la faire apparaître au niveau du pivot.
- Trier : permet de trier votre lexique par fréquence ou par ordre alphabétique des mots (ou valeur de propriété) en colonne gauche
- Fmin : permet de choisir un seuil de fréquence minimum
- Fmax : permet de choisir un seuil de fréquence maximum
- Vmax : permet de changer le nombre total de lignes (pour n'afficher par exemple que les 10 mots les plus fréquents)
- Nombre de lignes : permet de changer le nombre de lignes par page.

7.3 Concordances

Section [6.4](#) du manuel de référence.

Il s'agit à présent de retrouver le contexte des formes que vous étudiez, pour les interpréter et les analyser.

7.3.1 Création des concordances

Vous pouvez lancer une concordance sur un sous-corpus sélectionné (la commande n'est accessible que pour les corpus et non pour les partitions, car elle n'est pas contrastive ; quand on travaille sur une partition, appliquer la commande au sous-corpus dont la partition est issue). Sélectionnez l'icône du sous-corpus et lancez la concordance via le menu contextuel (ou cliquez sur  dans la barre d'outils). Il vous suffit ensuite de saisir une expression dans le champ requête et de valider votre concordance avec le bouton « Chercher ».

Requête : "chevalier"%c Chercher ∇ Réglages

Référence	Contexte gauche	Pivot	Contexte droit
1	brut2, v.12211	, Ovoc li trei mil chevalier	E li forier e li
2	brut2, v.12784	gentil hume E li bon chevalier	de Rome . La veïssiez
3	brut2, v.14021	traire e cops ferir, Chevalier furent fait ensemble	En Armoriche
4	brut2, v.14649	Firent de quivre tresjeter Un chevalier	sur un cheval En apareillement
5	brut2, v.14684	mullier, Muerent vilain e chevalier ; N' estuet al fiz	
6	escoufle, v.80	Deparitoit son conquerement : Maint chevalier	fist de noient Riche et
7	escoufle, v.112	N'ainc n'ot tel chevalier	a Troie Quant li dunjons
8	escoufle, v.131	morne Si serjant et si chevalier .	Ki lors veïst serjans
9	escoufle, v.697	vile n'ot escuier, Chevalier , garçon	ne serjant N'
10	escoufle, v.702	tot seoir en renga Li chevalier	ne li estrange , Ains
11	escoufle, v.740	argent. N'i a chevalier	ne serjant Qui voelle prendre
12	escoufle, v.784	a moult grant rage. Chevalier	cel jor et message Vinrent
13	escoufle, v.853	furent par tot mandé Li chevalier	et li serjant . Moult
14	escoufle, v.896	s'arme et tuit si chevalier	S' en vont gaitier par
15	escoufle, v.1021	qui le voient, Quel chevalier	et quel preudome ! Ainc
16	escoufle, v.1039	Ki mix mix tuit li chevalier .	Ki lors veïst chauces
17	escoufle, v.1126	plaingne. N'i a chevalier	qui ne prengne Sa lance
18	escoufle, v.1338	argent. Il n'a chevalier	ne serjant Cui li rois
19	escoufle, v.1586	desarme. Quant il voit chevalier	sans arme Et sans cheval
20	escoufle, v.2086	Et ses gens et si chevalier	Cuellent del fruit por laus

875 occurrences 1-50/875 Nombre de lignes : 50 Exporter

Illustration 42: Concordance sur la requête "chevalier"%c

- Double-cliquer sur une ligne de résultat vous permet de retourner à l'édition. Le pivot sélectionné apparaît alors en rouge.

- 636 -

Ovoc li trei mil **chevalier**
E li forier e li archier.
Romein a ferir entendeient,
Ki les Bretuns descunfiseient,
Quant Guitart vint od ses meisnees
Espurunant, lances beissees ;
Plus de cent en deschevalchierent
Ki unques puis ne redrescierent.
Es vus Romeins tuz esbaiz
E tuz se tindrent pur huniz ;
Quidoent que Artur venist
E sa gent tute le siewist.
Quant tant virent des lur chaeir,
De guarisun n'orent espeir.
Cil de Peitou bien les asaillent
E li Bretuns ne se lur faillet :

brut2 13/156 Fiche bibliographique

Requête : "chevalier"%c Chercher Trier ∇ Réglages

Référence	Contexte gauche	Pivot	Contexte droit
1	brut2, v.12211	, Ovoc li trei mil chevalier	E li forier e li
2	brut2, v.12784	gentil hume E li bon chevalier	de Rome . La veïssiez
3	brut2, v.14021	traire e cops ferir , Chevalier furent fait ensemble	En Armoriche

Illustration 43: Retour à l'édition

7.3.2 Réglage des paramètres des concordances

Sections [6.4.4](#) à [6.4.6](#) du manuel de référence.

Vous pouvez intervenir directement sur quatre paramètres : la taille des contextes, le nombre de résultats par page, les propriétés d'affichage des références et des résultats et le tri des concordances. Vous ferez apparaître ces paramètres en cliquant sur le bouton Réglages.

1) Taille des contextes

- Vous pouvez régler la taille de vos différents contextes, définie en nombre de mots, en renseignant les champs `Taille` gauche et droit, respectivement pour le contexte gauche et le contexte droit. Pour valider vos contextes, cliquez sur `Chercher`.

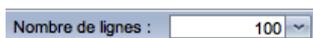


The image shows a horizontal bar with a light gray background. On the left, the word "Taille" is written in a small font. To its right, there are two white input boxes. The first box contains the number "5". To the right of the second box, there is another "5".

Illustration 44: Réglage de la taille des contextes

2) Nombre de résultats

- Vous pouvez indiquer le nombre de lignes par page souhaité dans le champ `Nombre de lignes`. Pour valider, cliquez sur `Chercher`.



The image shows a small rectangular box with a light blue border. Inside, the text "Nombre de lignes :" is followed by a white input field containing the number "100". To the right of the input field is a small downward-pointing arrow.

Illustration 45: Réglage du nombre de résultats par page

3) Vue des références, du pivot et des contextes

- Ces trois réglages sont accessibles dans les menus déroulants de la ligne `Vue`.
- Les références permettent de repérer chaque ligne de concordance à l'intérieur des différentes unités textuelles du sous-corpus en affichant les propriétés sélectionnées (par défaut, le système propose l'identifiant du texte, le numéro de la page et, pour les textes en vers, le numéro du vers où se trouve l'occurrence). La forme des références peut être définie en cochant les propriétés des références dans le menu¹⁴. Vous pouvez par exemple choisir d'afficher la forme (`text:forme`) ou le domaine (`text:domaine`).

¹⁴ Les références disponibles correspondent aux balises XML-TEI encodées au sein des sources du corpus.

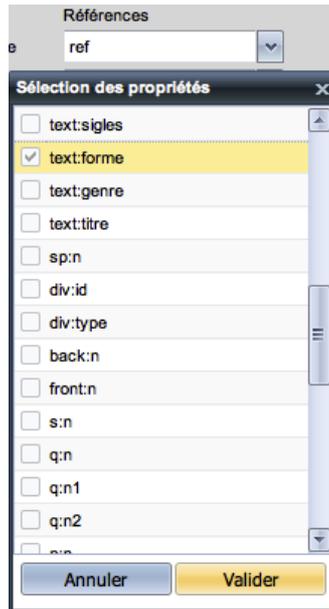


Illustration 46: Définir les références

	Référence	Contexte gauche	Pivot	Contexte droit
1	vers, brut2, v.12211		, Ovoc li trei mil chevalier	E li forier e li
2	vers, brut2, v.12784		gentil hume E li bon chevalier	de Rome. La veissiez
3	vers, brut2, v.14021		traire e cops ferir, Chevalier	furent fait ensemble En Armoriche
4	vers, brut2, v.14649		Firent de quivre tresjeter Un chevalier	sur un cheval En apareillement
5	vers, brut2, v.14684		mullier, Muerent vilain e chevalier ;	N'estuet al fiz
6	vers, escoufle, v.80	Departoit son conquerement :	Maint chevalier fist de noient	Riche et

Illustration 47: Résultat des réglages ci-dessus

- Vous pouvez également choisir les informations (ou propriétés) que vous voulez afficher au niveau du pivot et des contextes. Par exemple, si vous souhaitez savoir si vos occurrences sont dans le discours direct ou la narration, choisissez comme propriété d'affichage du pivot « word » (qui correspond à la forme du mot, propriété qui s'affiche par défaut) et « q » (le texte indiquera _1 (premier niveau d'imbrication), _2 (deuxième niveau d'imbrication) ou _3 (troisième niveau d'imbrication) pour signifier que l'occurrence se trouve dans la discours direct, _0 pour indiquer la narration).

☞ NA signifie que la propriété choisie n'a pas de valeur renseignée.

- Si vous souhaitez voir la catégorie morphosyntaxique associée à chaque occurrence du pivot, ajoutez « cattex-pos » (étiquette morphosyntaxique Cattex 09) ou « ud-pos » (partie du discours selon le jeu *Universal dependencies*) à la liste des propriétés d'affichage.

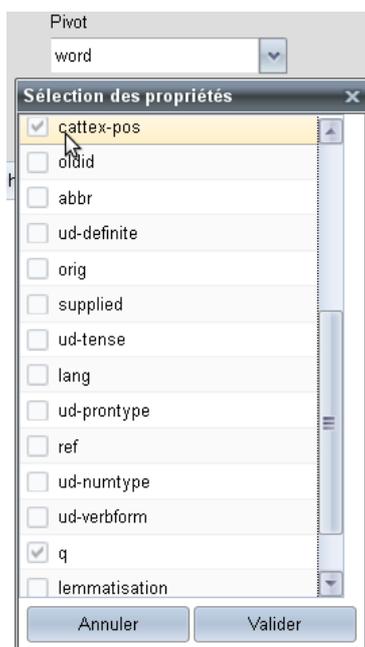


Illustration 48: Ajouter les propriétés « q » et « cattex-pos » au pivot.

	Référence	Contexte gauche	Pivot	Contexte droit
1	vers, brut2, v.12211		, Ovoc li trei mil chevalier_0_NOMcom	E li forier e li
2	vers, brut2, v.12784	gentil hume E li bon chevalier_0_NOMcom		de Rome. La veïssiez
3	vers, brut2, v.14021		traire e cops ferir, Chevalier_0_NOMcom	furent fait ensemble En Armoriche
4	vers, brut2, v.14649	Firent de quivre tresjeter Un chevalier_0_NOMcom		sur un cheval En aparellement
5	vers, brut2, v.14684		muillier, Muerent vilain e chevalier_0_NOMcom	; N'estuet al fiz
6	vers, escoufle, v.80	Deparitoit son conquerement : Maint chevalier_0_NOMcom		fist de noient Riche et

Illustration 49: Nouvel affichage des propriétés

Les propriétés disponibles sont identiques à celles de la commande Index.

4) Tri des concordances

- Ces trois réglages sont accessibles dans les menus déroulants de la ligne `Trier`.
- Vous avez la possibilité d'indiquer pour chacune des colonnes de la concordance sur quelle(s) propriété(s) s'opère le tri : ce n'est donc pas forcément la forme graphique (`word`) ni la propriété choisie pour l'affichage. Par exemple pour les références on peut trier sur une information qu'on indique en 2^e ou 3^e position dans l'affichage (une référence `ref+text:siècle` peut ainsi être triée sur le siècle).



Illustration 50: Choisir la propriété de tri (fropos) pour le contexte droit

- Vous avez aussi la possibilité d'opérer un tri multiple sur vos résultats, grâce aux « clés de tri » (précédées du signe #). Elles vous permettent de combiner et de hiérarchiser vos critères de tri : par exemple, vous pouvez trier vos résultats sur la forme du pivot en premier, puis sur le contexte droit en second. Validez votre sélection en cliquant sur `Trier`. Vous pouvez aussi choisir un tri ascendant ou descendant en cliquant sur la flèche à droite de chaque champ.



Illustration 51: Clés de tri sur le pivot et le contexte droit.

7.4 Contexte

Comme la commande `Concordance`, la commande `Contexte` permet d'observer toutes les occurrences d'un motif dans leurs contextes de réalisation. C'est la présentation qui diffère :

- la `Concordance`, en présentant les contextes superposés dans un tableau triable, est plus efficace pour analyser le contexte immédiat du pivot et ses récurrences (constructions, figements) ;
- le `Contexte`, qui se présente comme une suite d'extraits avec mise en valeur typographique du motif cherché, facilite quant à lui la lecture de chacun des contextes.

Cliquer sur une ligne de résultat vous permet de faire un retour à l'édition.

- Lancez la commande `Contexte` sur votre sous-corpus :

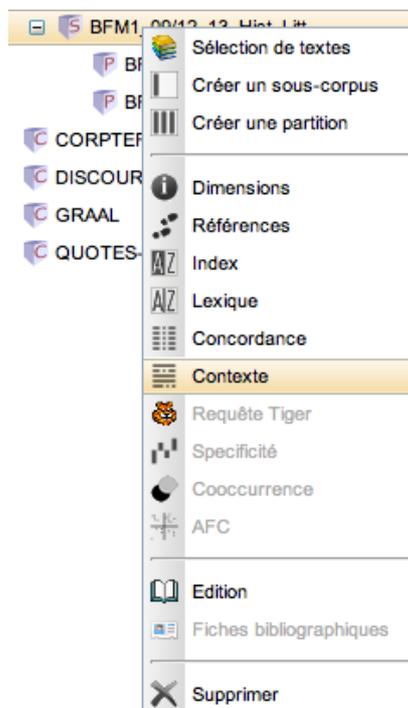


Illustration 52: Lancer la commande Contexte

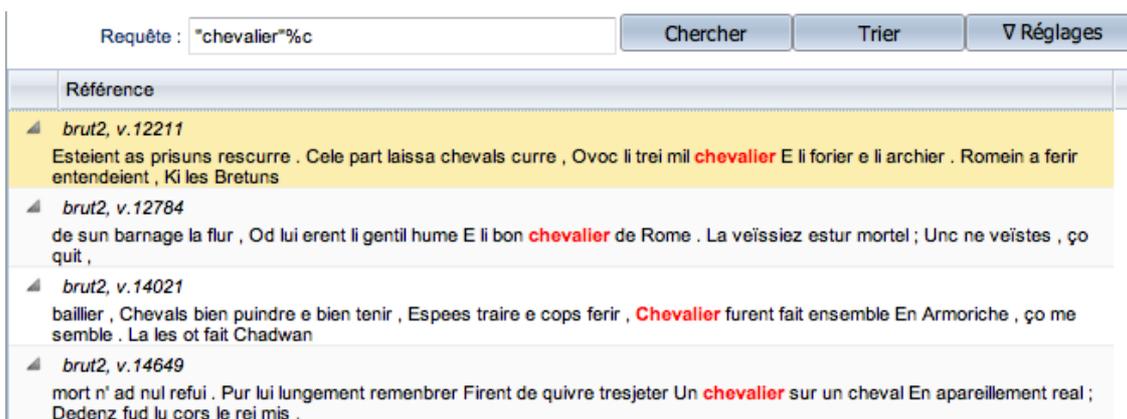


Illustration 53: Contexte sur la requête "Chevalier"%c

Les réglages proposés sont identiques à ceux des concordances (vous reporter à la section [7.3.2](#)).

7.5 J'utilise les propriétés de mots pour interroger le corpus

7.5.1 Étiquettes morphosyntaxiques Cattex09

Tous les textes de tous les corpus de la BFM sont étiquetés en morphosyntaxe. Cet étiquetage est vérifié dans une [quarantaine de textes](#), et dans le reste du corpus il est automatique. Dans les corpus BFM2016 et CORPTEF, les étiquettes vérifiées sont disponibles grâce à la propriété de mot « pos » et les étiquettes automatiques, par la propriété « fropos ».

Dans les corpus BFM2019, PALAFRAFRO-V2-2 et PALAFRAPAR, une seule propriété « cattex-pos » donne accès aux étiquettes morphologiques Cattex, vérifiées ou non, et la propriété

« étiquetage » permet de restreindre la requête aux étiquettes vérifiées.

Il est également possible d'utiliser le sous-corpus 'étiqueté' de la BFM2019 pour travailler uniquement sur textes dont l'étiquetage a été vérifié.

Ce nouveau système sera désormais utilisé dans les nouveaux corpus de la BFM.

La propriété « cattex-pos_syn » contient l'étiquette morphosyntaxique qui peut être différente de l'étiquette morphologique dans certains cas précis (infinitifs ou participes substantivés, par exemple). Le document [Principes d'annotation Cattex 09](#) accessible sur le site de la BFM explique cet usage plus en détail. Ce principe n'est cependant appliqué que dans certains textes vérifiés, il convient donc d'utiliser la distinction morphologique/morphosyntaxique avec précaution. Dans les textes non vérifiés, l'étiquette « cattex-pos_syn » est toujours identique à « cattex-pos ».

Voici un exemple de requête à utiliser avec les commandes Index, Concordance ou Contexte pour retrouver toutes les formes verbales :

```
[cattex-pos="VER.*"]
```

Pour restreindre la requête aux étiquettes vérifiées, il faut soit travailler sur un sous-corpus de textes vérifiés, soit utiliser la requête suivante :

```
[cattex-pos="VER.*" & etiquetage="vérifié"]
```

La liste des étiquettes Cattex09 se trouve dans l'Annexe 2 et dans la [rubrique « Documentation »](#) du site <http://bfm.ens-lyon.fr>. Pour en savoir plus sur la formulation des requêtes, consultez la [Référence rapide CQL pour la BFM](#).

7.5.2 Étiquettes morphosyntaxiques *Universal dependencies* (UD)

Dans les corpus BFM2019, PALAFRAFRO-V2-2 et PALAFRAPAR, ainsi que dans le corpus latin PALAFRALAT-V2-0, il est possible d'utiliser les étiquettes du jeu universel défini par le projet [Universal dependencies](#). Pour les textes français ces étiquettes ont été obtenues par conversion automatique à partir des étiquettes morphologiques Cattex 09. Des propriétés différentes sont utilisées pour indiquer la partie du discours et les différents traits morphologiques :

- **ud-pos** (partie du discours)
- **ud-feats** (ensemble de traits morphologiques)

Voici un exemple de requête permettant de retrouver les participes présents :

```
[ud-pos="VERB" & ud-feats contains "VerbForm=Part\|Tense=Pres"]
```

La documentation complète du jeu UD est consultable sur le site <http://universaldependencies.org>.

Le tableau de conversion des étiquettes Cattex09 → UD est donné dans l'Annexe 3 et dans la [rubrique « Documentation »](#) du site <http://bfm.ens-lyon.fr>. Pour en savoir plus sur la formulation des requêtes, consultez la [Référence rapide CQL pour la BFM](#).

7.5.3 Lemmes

La lemmatisation est actuellement disponible uniquement dans les corpus BFM2019, PALAFRAFRO-V2-2 et PALAFRAPAR. Elle est vérifiée dans [un nombre restreint de textes](#) et automatique dans le reste du corpus. Il est possible de travailler sur le sous-corpus 'lemmatisé' dans le BFM2019 pour se restreindre aux textes dont la lemmatisation et l'étiquetage ont été vérifiés (lorsque les lemmes sont vérifiés, les étiquettes morphosyntaxiques le sont aussi).

En cas de lemmatisation automatique, plusieurs hypothèses peuvent être proposées. Le format des lemmes a été optimisé pour faciliter la recherche parmi les hypothèses multiples. Ainsi, pour rechercher toutes les formes du verbe *être* :

```
[lemma contains "être"]
```

La majorité des lemmes sont issus du [Dictionnaire du moyen français](#) (DMF), mais dans certains cas les lemmes proviennent du dictionnaire Tobler-Lommatzsch ou d'autres sources. Il est donc utile d'inclure les formes des lemmes de différentes sources :

```
[lemma contains "être|estre"]
```

Notez que les chiffres utilisés dans le DMF pour distinguer les homonymes ne sont pas inclus dans les lemmes. Une requête combinée sur le lemme et l'étiquette morphosyntaxique peut aider à lever certaines ambiguïtés :

```
[lemma contains "être|estre" & ud-pos="VERB"]
```

 (permet d'exclure les occurrences des noms *fust* et *esté* par exemple).

Pour restreindre la recherche aux lemmes vérifiés, il faut utiliser la propriété « lemmatisation » :

```
[lemma contains "être|estre" & lemmatisation="vérifiée"]
```

Enfin, pour les lemmes vérifiés, il est possible de connaître la source du lemme grâce à la propriété « lemma_src ». Vous pouvez ajouter cette propriété dans les réglages d'affichage de l'index (cf. [7.1.1](#)) ou de la concordance (cf. [7.3.2](#)). La liste des sources des lemmes est donnée en Annexe 4 et dans la [rubrique « Documentation »](#) du site <http://bfm.ens-lyon.fr>. Pour en savoir plus sur la formulation des requêtes, consultez la [Référence rapide CQL pour la BFM](#).

7.5.4 Propriétés de mots disponibles et combinaisons possibles

Toutes les propriétés associées aux mots peuvent être visualisées dans les index et les concordances grâce au bouton « Réglages ». Elles peuvent être interrogées et associées les unes aux autres grâce au langage CQL (voir la [Référence rapide CQL pour la BFM](#)).

Par exemple, pour chercher les occurrences de la forme *a* dont l'étiquette Cattex « VERcjcjg » (verbe conjugué) a été vérifiée, il faut écrire la requête suivante :

```
[word="a" & cattex-pos="VERcjcjg" & etiquetage="vérifié"]
```

Si la requête concerne un lemme, il convient d'utiliser l'opérateur *contains* :

```
[lemma contains "par" & ud-pos="ADV"]
```

La liste complète des propriétés de mots se trouve dans l'Annexe 1 et dans la [rubrique « Documentation »](#) du site <http://bfm.ens-lyon.fr> (document *Liste des propriétés de mots disponibles dans les corpus de la BFM*).

7.6 J'utilise des structures pour interroger le corpus

En plus des propriétés de mots, il est possible d'utiliser des éléments de structure de textes pour travailler avec le corpus. Ces structures sont encodés par des balises XML-TEI, et il est possible d'utiliser ces balises dans les requêtes.

Les passages au discours direct ont ainsi été encodés dans la BFM par l'élément TEI <q>. La requête suivante permet de construire un index ou une concordance des passages au discours direct :

```
<q> []+ </q>
```

Pour rechercher un mot ou une expression se trouvant dans le discours direct le plus simple est cependant d'utiliser la propriété de mot *q* :

```
[word="fortune"%c & q!="0"]
```

Quand la propriété *q* a la valeur 0, cela signifie que l'occurrence ne se trouve pas dans le discours direct. Quand elle a la valeur 1, l'occurrence se trouve dans un discours direct de niveau 1 (les niveaux d'imbrication du discours dans le discours direct sont rendus par des chiffres allant de 1 à 3)

Pour afficher le discours direct entier contenant le mot recherché, il suffit d'ajouter l'opérateur *expand to* à la requête :

```
[word="fortune"%c & q!="0"] expand to q
```

8 Exporter un document depuis TXM

Après avoir généré un index ou une concordance, vous avez la possibilité de conserver les résultats obtenus en cliquant sur , en bas à droite de la fenêtre : vous pouvez ainsi les retravailler dans un autre logiciel ou les inclure dans une publication.

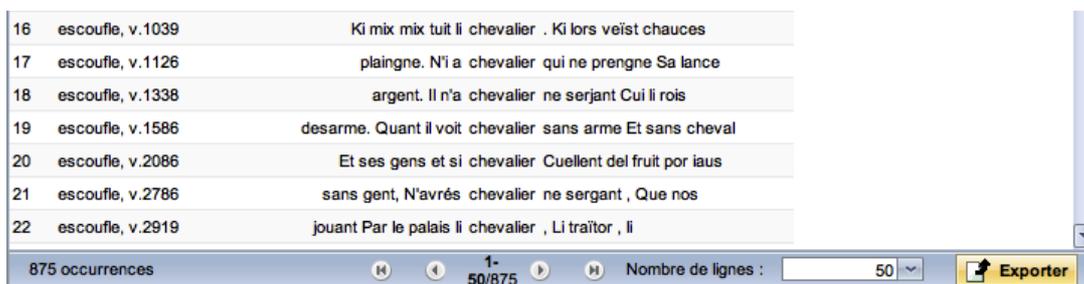


Illustration 54: Exporter le résultat sélectionné, via le bouton Exporter situé en bas à droite du volet des concordances

Le résultat de l'export sera affiché dans un nouvel onglet ou une nouvelle fenêtre de votre navigateur. Il est possible que votre navigateur bloque l'affichage de fenêtres « popup » (comme les « fenêtres publicitaires »). Dans ce cas, vous devez d'abord autoriser l'ouverture de ces fenêtres par votre navigateur pour le site <http://txm.bfm-corpus.org>, puis relancer la commande d'export.

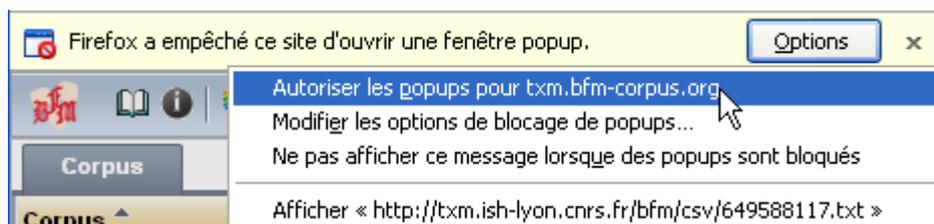


Illustration 55: Autoriser les fenêtres « popup » pour le portail BFM

Si le mot 'Référence' au début de la concordance exportée ne s'affiche pas correctement, vous devez sélectionner l'encodage des caractères Unicode (UTF-8) dans le menu de votre navigateur :

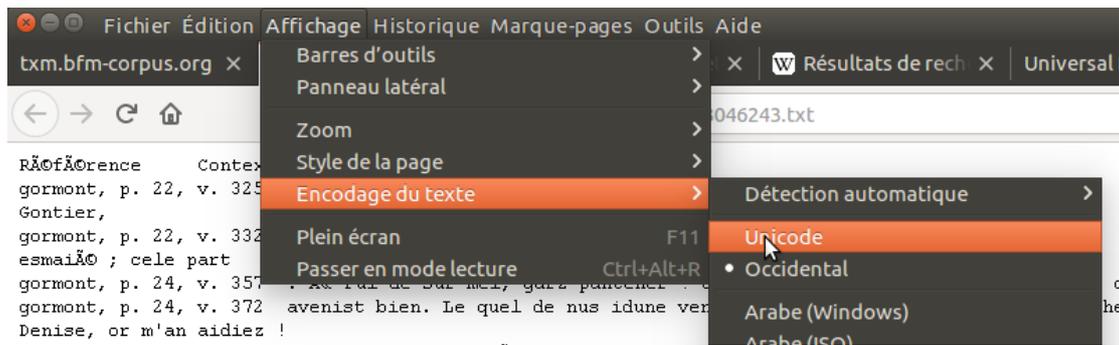


Illustration 56: Modification de l'encodage des caractères (Firefox 59)

Pour ouvrir la concordance ou l'index exporté dans un logiciel tableur (comme Excel ou Calc), vous pouvez utiliser l'une des deux méthodes suivantes. Si votre navigateur ne permet pas de choisir manuellement l'encodage du texte, vous devrez utiliser la méthode 2.

Méthode 1 : Copier – coller

- Ouvrez Calc ou Excel
- Sélectionnez et copiez tout dans l'onglet des résultats de l'export (touches Ctrl + A ou Pomme/Commande + A, puis Ctrl +C ou Pomme/Commande + C)
- **Dans Excel**
 - utilisez le menu Edition → Collage spécial → Texte Unicode
 - vérifiez si les colonnes de certaines lignes ne sont pas décalées par rapport aux autres. Cela peut se produire si le caractère guillemet droit (") se trouve au début d'une case. Dans ce cas, vous devez utiliser la méthode 2.
- **Dans Calc**
 - utilisez le menu Edition → Collage spécial → Texte non formaté
 - dans la fenêtre de paramètres d'import de texte, vérifiez et corrigez si nécessaire les options suivantes :
 - Jeu de caractères : Unicode (UTF-8)
 - Séparé par : cochez « Tabulation », décochez les autres cases
 - Séparateur de texte : supprimez le guillemet
 - Cliquez sur OK

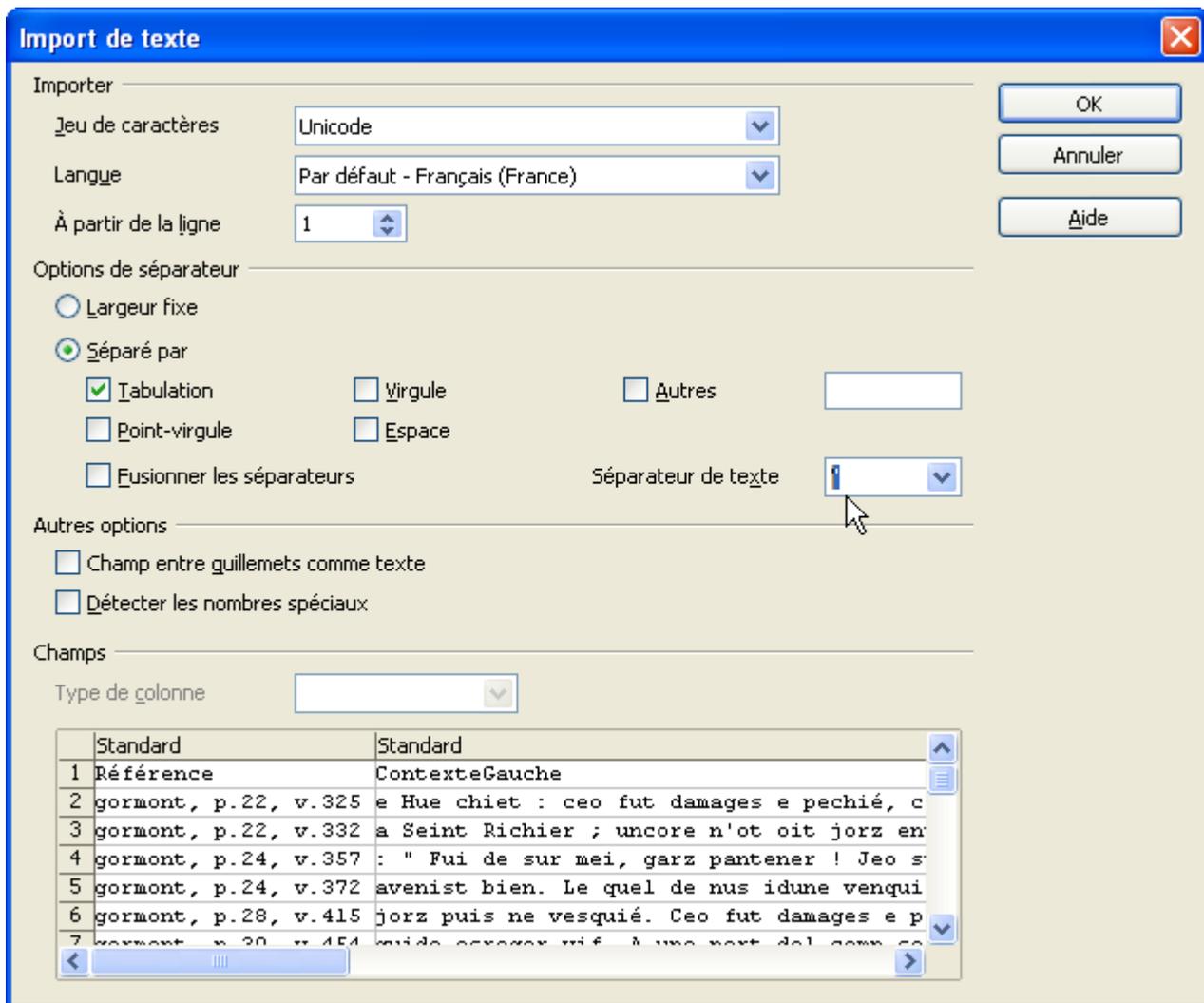


Illustration 57: Paramètres d'import de texte dans Libre Office Calc

Méthode 2 : Ouvrir un fichier texte

- Enregistrez la concordance ou l'index exporté sur votre ordinateur (Menu Fichier → Enregistrer sous...)
 - n'hésitez pas à remplacer le nom automatique du fichier
- Ouvrez Calc ou Excel
- **Dans Calc**
 - Utilisez le menu Fichier → Ouvrir
 - Ouvrez le dossier où vous avez enregistré votre export et sélectionnez le fichier correspondant
 - Configurez les paramètres d'import de texte (illustration 57) comme indiqué ci-dessus dans la méthode « copier - coller ».
- **Dans Excel**
 - Utilisez le menu Données → À partir du texte

- Ouvrez le dossier où vous avez enregistré votre export
- Sélectionnez votre fichier exporté et cliquez sur « Importer » :

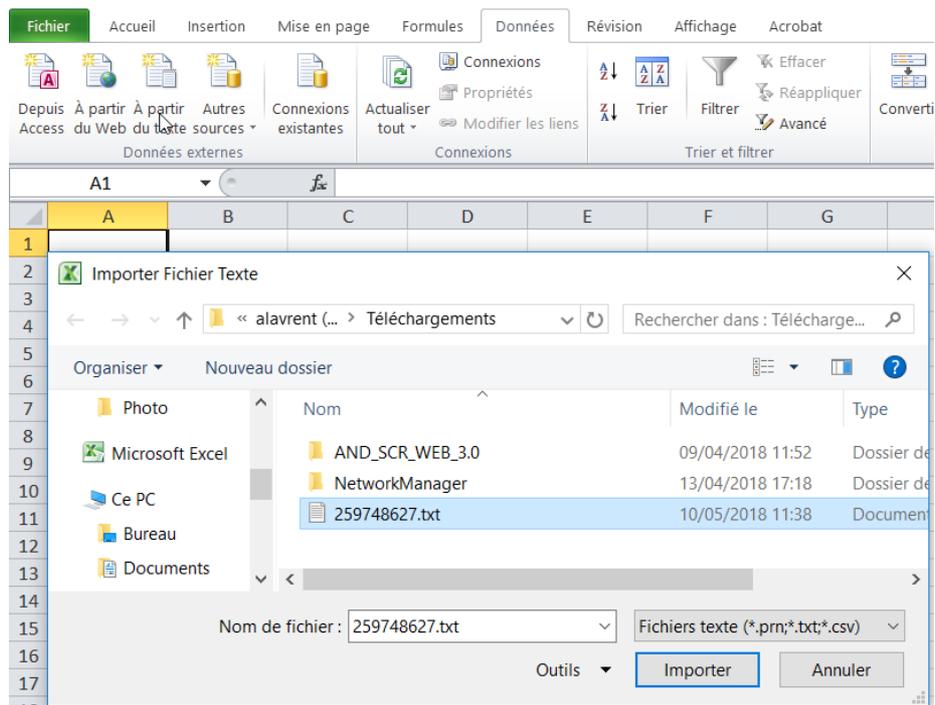


Illustration 58: Ouverture de fichier texte (Excel 2010)

- Dans l'assistant Importation de texte, choisissez les options suivantes :
 - Types de données d'origine : Délimité
 - Origine du fichier : Unicode (UTF-8)
- Cliquez sur « Suivant »

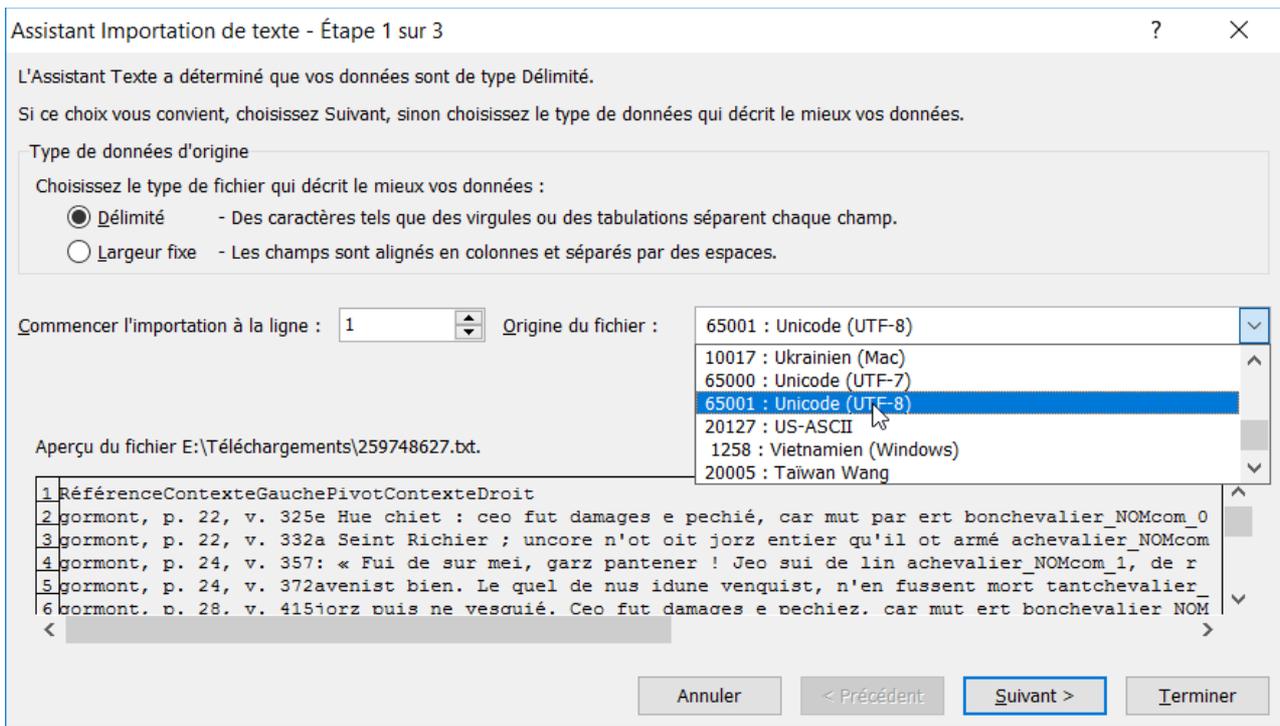


Illustration 59: Assistant d'importation de texte, étape 1 (Excel 2010)

- Dans la fenêtre suivante :
 - cochez la case Tabulation et décochez les autres
 - sélectionnez {aucun} dans la liste déroulante « Identificateur de texte »
- Cliquez sur « Terminer » (l'étape 3 est optionnelle)

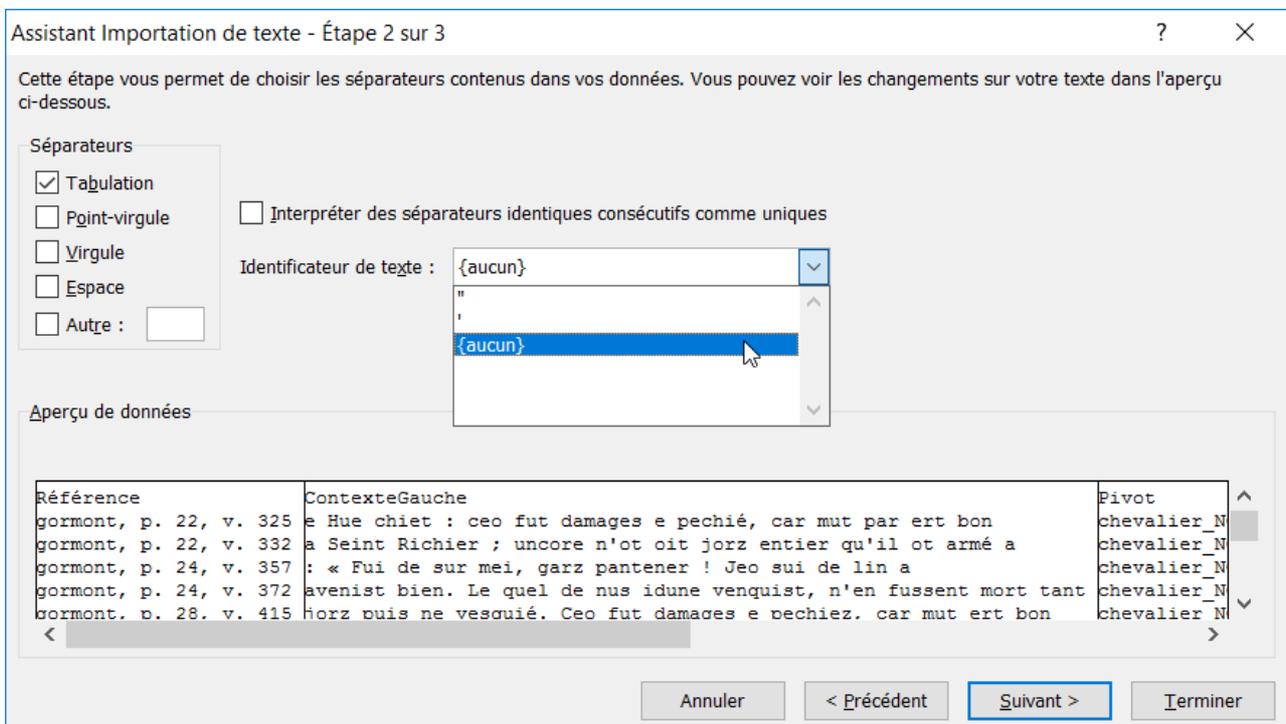


Illustration 60: Assistant d'importation de texte, étape 2 (Excel 2010)

Ainsi, vous obtenez ce tableau, que vous avez la possibilité de modifier (pour faciliter la lecture, pensez à aligner à droite la colonne qui se trouve à gauche du pivot) :

	A	B	C	D
17	escoufle, v.1039	Ki mix mix tuit li	chevalier	. Ki lors veïst chauces
18	escoufle, v.1126	plaigne. N'i a	chevalier	qui ne prengne Sa lance
19	escoufle, v.1338	argent. Il n'a	chevalier	ne serjant Cui li rois
20	escoufle, v.1586	desarme. Quant il voit	chevalier	sans arme Et sans cheval
21	escoufle, v.2086	Et ses gens et si	chevalier	Cuellent del fruit por iaus
22	escoufle, v.2786	sans gent, N'avrés	chevalier	ne sergant, Que nos
23	escoufle, v.2919	jouant Par le palais li	chevalier	, Li traïtor, li

Illustration 61: Affichage d'une concordance dans un tableur

Remarque : le tableur n'a plus les possibilités de tri adaptées aux données textuelles :

- pour le contexte gauche, le tri du mot le plus proche du pivot au mot le plus éloigné (tri de droite à gauche au niveau des mots, et de gauche à droite au niveau des caractères de chaque mot) ;
- tri sur des propriétés que l'on n'affiche pas forcément.

Pensez à enregistrer le tableau au format de votre logiciel (.xls, .xlsx ou .ods).

Annexes

Annexe 1 : Liste des propriétés de mots dans les corpus de la Base de français médiéval

Le tableau ci-dessous donne la liste des propriétés de mots disponibles et utiles pour l'interrogation des différents corpus de la BFM. La colonne « Corpus » indique les corpus dans lesquels les propriétés sont disponibles :

- BFM2019 : corpus BFM2019, PALAFRAFRO-V2-2 et PALAFRAPAR
 - BFM2016 : BFM2016 et CORPTEF (ces propriétés disparaîtront dans les nouvelles versions des corpus)
- tous : tous les corpus français

Nom de la propriété	Explication	Exemple de requête	Corpus
cattex-pos	Étiquette morphologique Cattex 09	[cattex-pos="ADJ.*"]	BFM2019
cattex-pos_syn	Étiquette morphosyntaxique Cattex 09, identique à cattex-pos, sauf dans les cas de substantivation, etc., identifiés dans certains textes vérifiés	[cattex-pos_syn="NOMcom" & cattex-pos="VER.*"]	BFM2019
etiquetage	Indique si l'étiquetage morphosyntaxique a été vérifié pour une occurrence donnée	[etiquetage="vérifié"]	BFM2019
frolemma	Identique à la forme (peu utile pour les requêtes)		BFM2016
fropos	Étiquette morphosyntaxique automatique non vérifiée (Cattex 09)	[fropos="ADJ.*"]	BFM2016 ¹⁵
id	Identifiant du mot dans le corpus (peu utile pour les requêtes)		tous
lang	Langue du mot étranger	[lang="la.*"]	tous
lemma	Lemme de l'occurrence	[lemma contains "avoir"]	BFM2019
lemma_src	Source du lemme, voir la section 3	[lemma_src="BFM"]	BFM2019
lemmatisation	Indique si le lemme est vérifié ou non	[lemmatisation="vérifié"]	BFM2019
orig	Indique si le mot est balisé <orig> (graphie non régularisée) dans le document XML-TEI, utilisée pour	[orig!="NA"]	tous

¹⁵ La propriété `fropos` est également disponible dans le corpus BFM2019 pour le confort des utilisateurs, sa valeur est identique à `cattex-pos` (étiquette Cattex 09 vérifiée ou non vérifiée).

	certaines ponctuations dans la <i>Queste del saint Graal</i>		
pos	Étiquette morphosyntaxique Cattex 09 vérifiée	[pos="ADJ.*"]	BFM2016
q	Niveau d'imbrication de l'occurrence dans un discours direct (« 0 » si hors du discours direct, « 1 » si dans un discours direct non imbriqué, « 2 » ou « 3 » si dans un discours direct imbriqué dans un ou deux autres discours directs)	[q="[123]"]	tous
ref	Référence par défaut pour la concordance (peu utile pour les requêtes)		tous
sic	Indique si le mot est balisé <sic> (graphie visiblement erronée) dans le document XML-TEI,	[sic!="NA"]	tous
sp	Indique si le mot se trouve dans une prise de parole d'un texte dramatique ou d'un dialogue (balise TEI <sp>), valeurs possibles : « t » (vrai) ou « f » (faux)	[sp="t"] ou [sp="f"]	tous
supplied	Indique si le mot se trouve dans un passage balisé <supplied> (conjecture éditoriale) dans le document XML-TEI, valeurs possibles : « t » (vrai) ou « f » (faux)	[supplied="f"]	tous
ud-feats	Ensemble de traits morphologiques UD	[ud-feats contains "PronType=Det"]	BFM2019
ud-pos	Étiquette « partie du discours » UD	[ud-pos="ADJ"]	BFM2019
word	La forme du mot, propriété par défaut. Une syntaxe raccourcie peut être utilisée si elle n'est pas combinée avec d'autres propriétés	a, "a" ou [word="a"] [word="a" & cattex-pos="VERcjk"]	tous

Annexe 2 : Liste des étiquettes Cattex09

Étiquette	Interprétation
ABR	ABRÉVIATION
ADJcar	ADJECTIF cardinal
ADJind	ADJECTIF indéfini
ADJord	ADJECTIF ordinal
ADJpos	ADJECTIF possessif
ADJqua	ADJECTIF qualificatif
ADVgen	ADVERBE (général)
ADVgen.PROper	contraction ADVERBE + PRONOM personnel
ADVint	ADVERBE interrogatif
ADVneg	ADVERBE de négation
ADVneg.PROper	contraction ADVERBE de négation + PRONOM personnel
ADVsub	ADVERBE subordonnant
CONcoo	CONJONCTION de coordination
CONsub	CONJONCTION de subordination
DETCar	DÉTERMINANT cardinal
DETCom	DÉTERMINANT défini composé
DEDef	DÉTERMINANT défini
DEDem	DÉTERMINANT démonstratif
DEInd	DÉTERMINANT indéfini
DEInt	DÉTERMINANT interrogatif
DEndf	DÉTERMINANT non défini
DEPos	DÉTERMINANT possessif
DERel	DÉTERMINANT relatif
ETR	MOT ÉTRANGER
INJ	INTERJECTION
NOMcom	NOM commun
NOMpro	NOM propre
OUT	ERREUR
PONfbl	PONCTUATION faible
PONft	PONCTUATION forte
PONpdr	PONCTUATION parenthèse droite
PONpga	PONCTUATION parenthèse gauche

PONpxx	PONCTUATION parenthèse (gauche ou droite)
PRE	PRÉPOSITION
PRE.DETdef	contraction PRÉPOSITION + DÉTERMINANT défini
PRE.DETrel	contraction PRÉPOSITION + DÉTERMINANT relatif
PRE.PROper	contraction PRÉPOSITION + PRONOM personnel
PROadv	PRONOM adverbial
PROcar	PRONOM cardinal
PROdem	PRONOM démonstratif
PROimp	PRONOM impersonnel
PROind	PRONOM indéfini
PROint	PRONOM interrogatif
PROord	PRONOM ordinal
PROper	PRONOM personnel
PROper.PROper	contraction PRONOM personnel + PRONOM personnel
PROpos	PRONOM possessif
PROrel	PRONOM relatif
RED	REDONDANT
RES	RÉSIDU
VERcjcjg	VERBE conjugué
VERinf	VERBE infinitif
VERppa	VERBE participe présent
VERppe	VERBE participe passé

Annexe 3 : Tableau de conversion Cattex09 → UD

L'organisation des jeux CATTEX2009-min et UD et la répartition entre catégories (*pos*) et traits morphologiques (*feats*) sont assez différentes. Le tableau qui suit établit la correspondance entre les étiquettes CATTEX2009-min (propriété *cattex-pos*) et les étiquettes UD (propriété *ud-pos*) avec les traits associés. Tous les traits morphologiques UD sont encodés dans une seule propriété de mot *ud-feats*, sa valeur est présentée dans un format similaire à CONLL-U (<http://universaldependencies.org/format.html>).

Dans quelques cas, l'étiquette CATTEX2009-min n'a pas d'équivalent direct UD et l'étiquette UD correspondante est moins précise :

- ADJqua (*cattex-pos*) correspond à ADJ (*ud-pos*)
- ADVing (*cattex-pos*) correspond à ADV (*ud-pos*)
- ADVsub (*cattex-pos*) correspond à ADV (*ud-pos*)
- PROadv (*cattex-pos*) correspond à ADV (*ud-pos*)
- PROimp (*cattex-pos*) correspond à PRON (*ud-pos*)
- PONfbl (*cattex-pos*) correspond à PUNCT (*ud-pos*)
- PONfirt (*cattex-pos*) correspond à PUNCT (*ud-pos*)
- PONpdr (*cattex-pos*) correspond à PUNCT (*ud-pos*)
- PONpga (*cattex-pos*) correspond à PUNCT (*ud-pos*)
- PONpxx (*cattex-pos*) correspond à PUNCT (*ud-pos*)

Le trait *Definite* du jeu UD a été utilisé dans quelques cas très limités. Cela concerne :

- l'article défini, qui combine les propriétés *ud-pos* (« Det ») et *PronType* (« Art ») avec la propriété *Definite* (« Def ») [ud-pos="Det" & ud-feats contains "PronType=Art\|Definite=Def"]
- l'article indéfini, qui combine les propriétés *ud-pos* (« Det ») et *PronType* (« Art ») avec la propriété *Definite* (« Ind ») [ud-pos="Det" & ud-feats contains "PronType=Art\|Definite=Ind"]
- le déterminant et le pronom *ledit*, qui combinent la propriété *ud-pos* (« Det » ou « Pron ») avec la propriété *Definite* (« Com ») ; on utilise ici de manière détournée la valeur « Com » (pour *complex*) prévue dans le jeu UD : [ud-pos="Det|Pron" & ud-feats contains "Definite=Com"]

Selon la spécification UD, les formes contractées comme le français *del* devraient être étiquetées comme deux mots distincts. Dans Cattex, les formes contractées sont traitées comme des occurrences uniques portant une double étiquette (par exemple, « PRE.DETdef »). Afin d'assurer la continuité d'architecture de corpus, le système de doubles étiquettes a été transposé au jeu UD de la manière suivante :

- *del* [ud-pos="PREP.DET"]

L'ensemble des règles de conversion est présenté dans le tableau suivant :

cattex-pos	ud-pos	ud-feats
ABR	X	Abbr=Yes
ADJcar	ADJ	NumType=Card
ADJind	ADJ	PronType=Ind
ADJord	ADJ	NumType=Ord
ADJpos	ADJ	Poss=Yes
ADJqua	ADJ	_
ADVgen	ADV	_
ADVgen.PROper	ADV.PRON	_ . PronType=Prs
ADVing	ADV	PronType=Int
ADVint	ADV	PronType=Int
ADVneg	ADV	Prontype=Neg
ADVneg.PROper	ADV.PRON	PronType=Neg . PronType=Prs
ADVsub	ADV	_
CONcoo	CCONJ	_
CONsub	SCONJ	_
DETCar	DET	NumType=Card
DETCom	DET	Definite=Com
DETdef	DET	PronType=Art Definite=Def
DETDem	DET	PronType=Dem
DETind	DET	PronType=Ind
DETint	DET	PronType=Int
DETndf	DET	PronType=Art Definite=Ind
DETord	DET	NumType=Ord
DETpos	DET	Poss=Yes
DETrel	DET	PronType=Rel
ETR	X	Foreign=Yes
INJ	INTJ	_
NOMcom	NOUN	_
NOMpro	PROPN	_
OUT	X	_
PONfbl	PUNCT	_
PONfrit	PUNCT	_
PONpdr	PUNCT	_
PONpga	PUNCT	_
PONpxx	PUNCT	_
PRE	ADP	_
PRE.DETdef	ADP.DET	_ . PronType=Art Definite=Def
PRE.DETrel	ADP.DET	_ . PronType=Rel

PRE.PROper	ADP.PRON	_ . PronType=Prs
PROadv	ADV	_
PROcar	PRON	NumType=Card
PROcom	PRON	Definite=Com
PROdem	PRON	PronType=Dem
PROimp	PRON	PronType=Prs
PROind	PRON	PronType=Ind
PROint	PRON	PronType=Int
PROord	PRON	NumType=Ord
PROper	PRON	PronType=Prs
PROper.PROper	PRON.PRON	PronType=Prs . PronType=Prs
PROpos	PRON	Poss=Yes
PROrel	PRON	PronType=Rel
RED	X	_
RES	X	_
VERcjk	VERB	VerbForm=Fin
VERinf	VERB	VerbForm=Inf
VERppa	VERB	VerbForm=Part Tense=Pres
VERppe	VERB	VerbForm=Part Tense=Past

Le format de la propriété *ud-feats* a été optimisé pour les requêtes CQL avec l'opérateur *contains*. Par exemple, la requête suivante permet de retrouver tous les pronoms personnels (y compris dans les contractions) :

```
[ud-feats contains "PronType=Prs"]
```

Pour trouver les articles indéfinis, vous pouvez utiliser la requête suivante :

```
[ud-pos="DET" & ud-feats contains "PronType=Art\|Definite=Ind"]
```

Annexe 4 : Liste des sources des lemmes

AND	<i>Anglo-Norman Dictionary</i> , http://www.anglo-norman.net/
BFM	Base de français médiéval (création)
DECT	<i>Dictionnaire Électronique de Chrétien de Troyes</i> , http://www.atilf.fr/dect
DMF	<i>Dictionnaire du Moyen Français</i> , http://www.atilf.fr/dmf
FEW	<i>Französisches Etymologisches Wörterbuch</i> de Walther von Wartburg, http://www.atilf.fr/few
GDF	<i>Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle</i> de Frédéric Godefroy
TL	<i>Tobler-Lommatzsch Altfranzösisches Wörterbuch</i>

Table des révisions

Date	Nom	Modification
26/10/10	Lauranne Bertrand (LB)	Création du document
27/10/10	B. Pincemin (BP) et A. Lavrentiev (AL)	Enrichissement de certaines sections
27/10/10	LB	Ajout des images
02/11/10	Céline Guillot (CG)	Compléments sur le déroulement de la session
08/11/10	LB	Ajout d'une section détaillée sur CQL
06/12/10	LB	Refonte du tutoriel pour la version Web
15/12/10	LB	Ajout de la commande Sélection de textes
24/01/11	CG, BP, Serge Heiden (SH) & LB	Ajustement de la commande Sélection de textes, de l'enchaînement des étapes du tutoriel et de certaines illustrations
21/06/11	LB	Mise à jour de la section concernant l'inscription et la connexion
08/08/11	LB	Mise à jour de la section concernant la sélection de textes, les partitions, l'index et les réglages
28/08/11	CG, AL & LB	Corrections et ajouts après relecture
04/10/11	CG, AL, SH, BP & LB	Mise à jour de l'ensemble du tutoriel pour la release de la plateforme
14/10/11	LB	Ajout d'un glossaire
24/10/11	LB	Ajout de sections sur les autres commandes disponibles
11/12/11	CG, BP & AL	Corrections et ajouts après relecture
13/12/11	LB	Ajout d'une table commentée
14/01/12	LB	Révision de la totalité du tutoriel et mise à jour des copies d'écran
12/02/12	LB	Nettoyage pour la sortie de la plateforme le 1er Mars
23/02/12	LB	Mise à jour des copies d'écran suite aux résolutions de bugs
09/03/12	BP, SH & LB	Corrections mineures après test en masse
11/06/12	AL, CG	Corrections mineures, déplacement de la table
13/01/14	JL (Justine Lascar)	Relecture et mise à jour pour la BFM2013
31/01/14	AL	Validation des modifications et révision générale
12/02/14	SLH, LB, CG, BP	relecture et corrections ponctuelles
01/06/18	AL, CG	Relecture et mise à jour (corpus PALAFRA, étiquettes UD, lemmatisation)

17/07/18	AL	Corpus PALFRAFRO-V2-2, synchronisation avec la traduction anglaise
15/05/19	AL, CG	Corpus BFM2019, corrections ponctuelles suite au remarques de BP et SLH.