

TXM Tutorial for BFM



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

License:

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

Version 3.1

15 May 2019

Table des matières

1 Introduction.....	3
2 General operations of the platform.....	4
3 Access the BFM portal without registering.....	5
4 Sign up for expanded access.....	9
4.1 Signing up.....	9
4.2 Connecting to the BFM.....	12
4.3 Modify your profile.....	12
5 Creating a subcorpus.....	13
5.1 General interface of the <i>Textes</i> option of the <i>Sous-corpus</i> command.....	13
5.1.1 Zone of text descriptors.....	14
5.1.2 Text zone.....	16
5.1.3 The Statistics Zone.....	17
5.2 Subcorpus example.....	17
5.3 Creation of subcorpus.....	19
5.4 Visualizing a subcorpus: the Dimensions and Edition commands.....	20
6 Organize your data: partitioning the subcorpus.....	21
7 Analyze your data.....	23
7.1 Index.....	23
7.1.1 Adjusting the index.....	24
7.2 Lexicon.....	26
7.2.1 Adjusting the Lexicon.....	26
7.3 Concordances.....	27
7.3.1 Creating concordances.....	27
7.3.2 Adjusting the settings for the concordances.....	29
7.4 Context.....	32
7.5 Use word properties to examine the corpus.....	33
7.5.1 Cattex09 Morphosyntatic Tags.....	33
7.5.2 The “Universal dependencies” morphosyntatic tags (UD).....	34
7.5.3 Lemmas.....	34
7.5.4 Available word properties and possible combinations.....	35
7.6 Using structural units to query the corpus.....	35
8 Exporting a document from TXM.....	35
Appendixes.....	41
Appendix 1: List of word properties in the corpora of the BFM.....	41
Appendix 2: List of Cattex09 Tags.....	43
Appendix 3: Conversion Table Cattex09 → UD.....	45
Appendix 4: List of Sources for Lemmas.....	47

1 Introduction

The **BFM Medieval French Corpus Collection**, created in 1989, includes to this day 170 unabridged Old and Middle French texts. By virtue of its size (nearly 4,700,000 word occurrences) and the sheer diversity of texts from which it is composed, the corpus is without parallel in France. It is used by a community of several hundred researchers, teachers, and students from across the globe.

The texts assembled in the BFM cover a sizable geographical area and a long time period, from the ninth century (with the earliest extant text written in French, the *Oaths of Strasbourg*) up until the end of the fifteenth. Both verse and prose texts are represented in the corpus, as well as a wide variety of genres (literature, history, hagiography, law, the sciences, etc.).

The texts of the BFM corpora are tagged in XML in accordance with the recommendations of the TEI (Text Encoding Initiative). The corpus has also been equipped with linguistic annotation (morpho-syntactic tagging, lemmatization) which allows for a richer and more complex research experience.

The BFM portal gives access to several corpora of Medieval French texts, as well as to a corpus of Late Latin texts and a parallel Latin – French corpus prepared as part of the PaLaFra Project¹ in order to study the passage from Latin to French.

The purpose of this tutorial is to enable users to discover the different querying methods of the corpus with the TXM portal software over a typical work session. After a general overview of the interface (section 2), we invite you to discover the features available without prior registration (section 3) before explaining the registration process which enables users to gain access to additional features. We demonstrate how users may create and organize a working corpus (sections 5 and 6), and then show how its data can be analyzed. Finally, readers will learn how to export the results of their queries so that they may be processed with a spreadsheet program or cited in an article (section 8).

To create and organize your working corpus (henceforth called a subcorpus), you will need to make use of the commands `Subcorpus` and `Partition`.

If you wish to work with the entirety of the BFM corpus², there is no need to create a subcorpus. You may therefore proceed to the next step. The `Text Selection` command allows you to build a working subcorpus by selecting from a predefined list of texts already incorporated into the Medieval French Database. The `Partition` command enables users to divide their subcorpus (or their initial corpus, should they have chosen to work with the entirety of texts of which it is composed) into several parts so as to facilitate comparative analyses. Several “prefabricated” subcorpora and partitions are ready for database users to be used directly.

Soon thereafter you will be guided through the use of the `Index` and `Concordance` commands. `Index` generates a list of the vocabulary of a given text so that one may determine different word frequencies. `Concordance` makes it easier to analyze terms of interest in their precise context.

1 <http://palafra.org>.

2 The entire database bibliography is accessible via the command `Bibliographic Records` (please consult the beginning of section 3).

2 General operations of the platform

Section 4.1 of *TXM User Manual*³

The web platform's interface is split into three different zones: the corpus browser (on the left), the results (displayed in the main portion of the window), and messages (situated below along the status line). The icons for each command appear in the corpus browser – which is to say that the browser contains all objects (corpora, subcorpora and partitions) managed by the platform and affected by the commands (create subcorpora and partitions, generate an index, concordances). The different commands are accessed via a contextual menu by right-clicking on an icon or the name of a corpus, subcorpus or partition; these may also be accessed through the toolbar which displays available commands according to the object selected. The partition and (sub-)corpus icons are retained from one to the next session (after log-out); the tabs providing access to search results, however, are not.

A more detailed description of the operations and the link between the browser and results zones will be provided in the following sections. The third zone, which is devoted to messages, contains comments for each action performed during a given session.

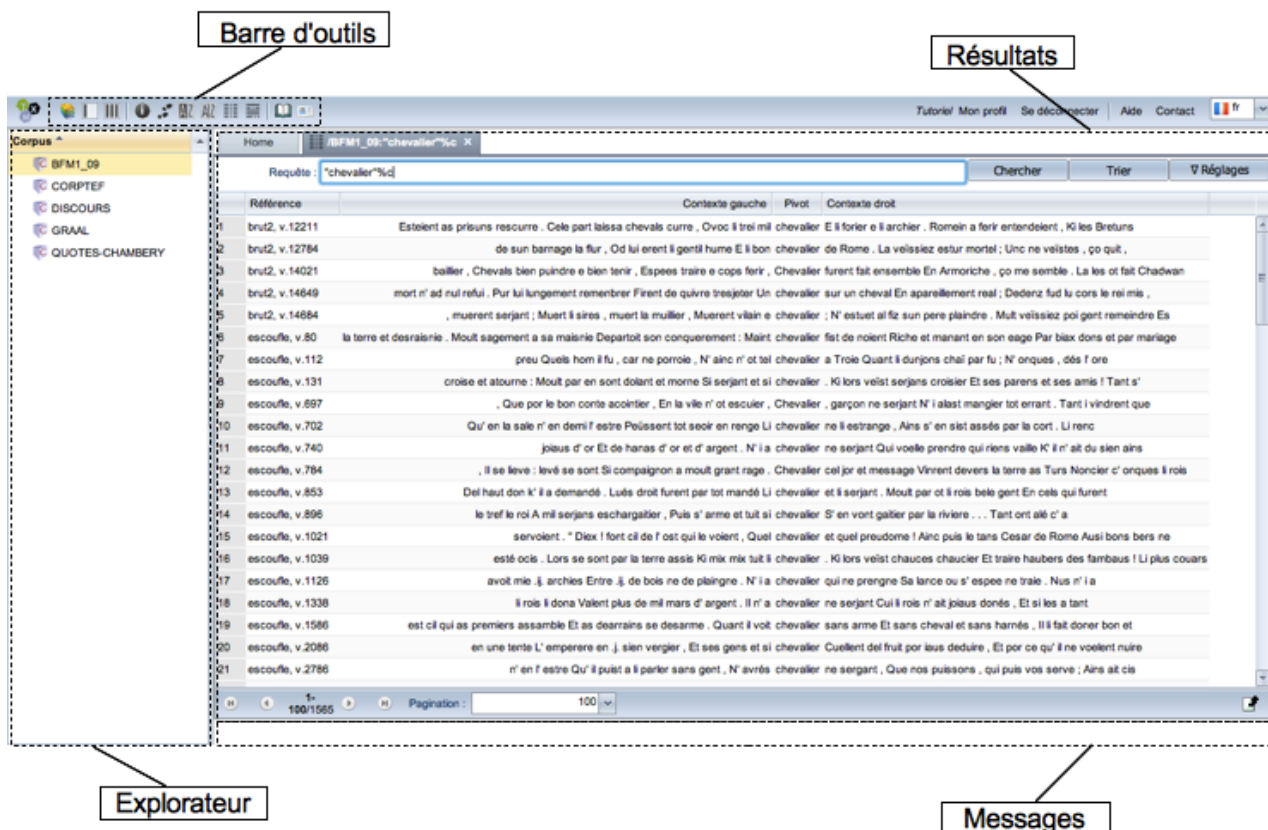


Figure 1: The general interface

³ The *TXM User Manual* is accessible here : <http://textometrie.ens-lyon.fr/files/documentation/TXM%20Manual%200.7.pdf>

3 Access the BFM portal without registering

In a web browser⁴, open the page <http://txm.bfm-corpus.org>.

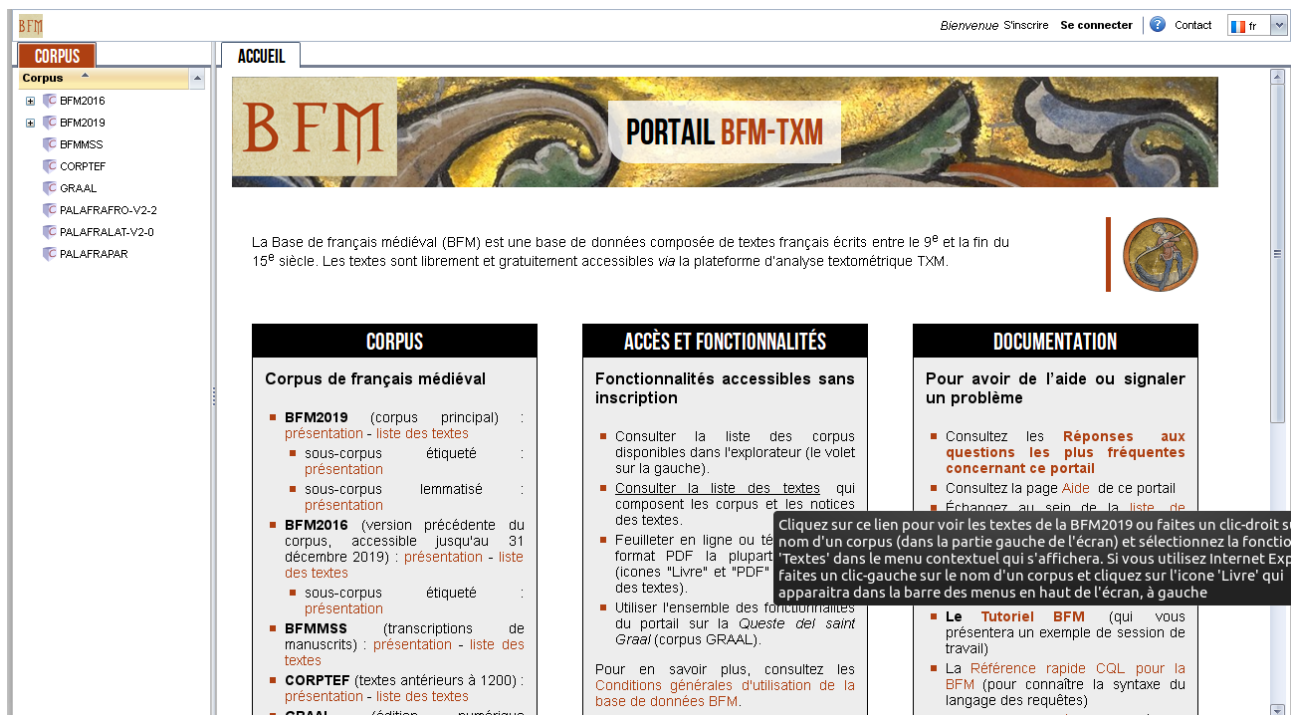


Figure 2: The BFM homepage

If the “Home” tab remains empty and the message « *Aucune donnée à afficher* » appears in the “Corpus” tab on the left, it is necessary to accept cookies from the site <http://txm.ish-lyon.cnrs.fr/bfm> by going to the privacy and security panel under your browser's settings. Otherwise, you may use the alternative address of the portal: <http://portal.textometrie.org/bfm>⁵.

NB: You can switch the language of the portal interface to English by clicking on the flag in the upper-right corner of the window. This will affect the language of TXM commands but not necessarily that of the web pages displayed. In this Tutorial we use English names of TXM commands but most of the screenshots are in French. If you change the portal interface language, you need to confirm closing the current page in the browser.

Some BFM features are accessible without prior registration. For each corpus, you can open a homepage (the ‘Home’ icon) containing a description with useful links. In addition, you can access the list of texts, bibliographic records, and displays of any text (in HTML and PDF format) via the “Corpus” panel on the left or the hypertext link “*Consulter la liste des textes*”.

Via the “Corpus” panel, right-click on your desired corpus and select the command `Textes` in the drop-down menu.

⁴ The portal has been tested with the following web browsers: Firefox, Internet Explorer 6, 7, 8 and 11, Microsoft Edge, Chrome, and Safari. Note that certain operations (namely, right-click) are not available with IE 6.

⁵ For more information, consult the FAQ of the BFM: <https://groupes.renater.fr/wiki/bfm-liste/faq/aucun-corpus-disponible>.

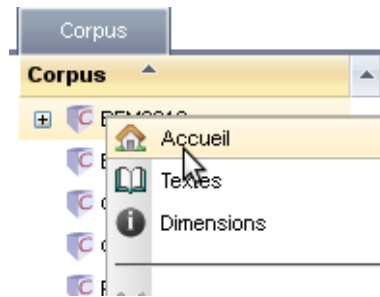



Figure 3: Using the contextual menu

A new tab containing the list of available texts appears.

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
eulial				anonyme	Séquence de saint	peu après 881	ca. 900	vers	religieux	hagiographie	non défini	original
jonas				anonyme	Sermon sur Jonas	prob. entre 938 et f	2e q. 10e s.	prose	religieux	sermon	flandre	indéterminé
passion				anonyme	Passion de Jésus-	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
oxfpe				anonyme	Psauteur 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

Merci de sélectionner un texte pour voir sa fiche bibliographique

Figure 4: Display of the list of texts of a corpus

In this tab, you can access, for all texts, the affiliated metadata, as well as a detailed bibliographic record. To do so, click on the Record button  of the selected text.

identifiant	édition	notice	pdf	auteur	titre	date compo libre	ms date libre	forme	domaine	genre	dialecte	relation
Lapidal				Philippe de Thaon	Lapidaire alphabétique	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 Isembart	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
DescrEngl				anonyme	Description d'Engle peu après 1139		déb. 13e s.	vers	historique	histoire	anglo-normand	adaptation
guill1				anonyme	Chanson de Guilla	ca. 1140	mil. 13e s.	vers	littéraire	epique	anglo-normand	indéterminé
Lapidfp				anonyme	Lapidaire en prose	mil. 12e s.	déb. 13e s.	prose	didactique	lapidaire	anglo-normand	adaptation
ProvSalSan1				Sanson de Nantuil	Les Proverbes de l mil.	12e s.	déb. 13e s.	vers	religieux	proverbes	anglo-normand	original
ProvSalSan2				Sanson de Nantuil	Les Proverbes de l mil.	12e s.	déb. 13e s.	vers	religieux	proverbes	anglo-normand	original
PsOrme				anonyme	Psautier de l'Orme	12e s.	12e s.	prose	religieux	psautier	anglo-normand	traduction
thebes1				anonyme	Roman de Thèbes	ca. 1150	2e t. 13e s.	vers	littéraire	roman	poitevin	indéterminé
thebes2				anonyme	Roman de Thèbes	ca. 1150	2e t. 13e s.	vers	littéraire	roman	poitevin	indéterminé

Gormont et Isembart

Comment citer ce texte :

Gormont et Isembart, 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 Isembart
- 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

Figure 5: Consulting the bibliographic record of a text

For texts in open access, you can access the text itself either in its original HTML form via the Edition  icon or in PDF format via the PDF  icon.

identifiant	édition	notice	pdf	auteur	titre	date compo libre	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-normand	adaptation
Lapidal				Philippe de Tl	Lapidaire alpr	1e tiers 12e s ca.	1200	vers	didactique	lapidaire	anglo-normand	adaptation
oxfps				anonyme	Psautier d'Ox	1e m. 12e s.	mil. 12e s.	prose	religieux	psautier	anglo-normand	traduction

Merci de sélectionner un texte pour voir sa fiche bibliographique

Figure 6: The Edition icon as it appears in the list of texts

Clicking on the Edition icon opens a new tab within the BFM window in which you may view and read the formatted text.

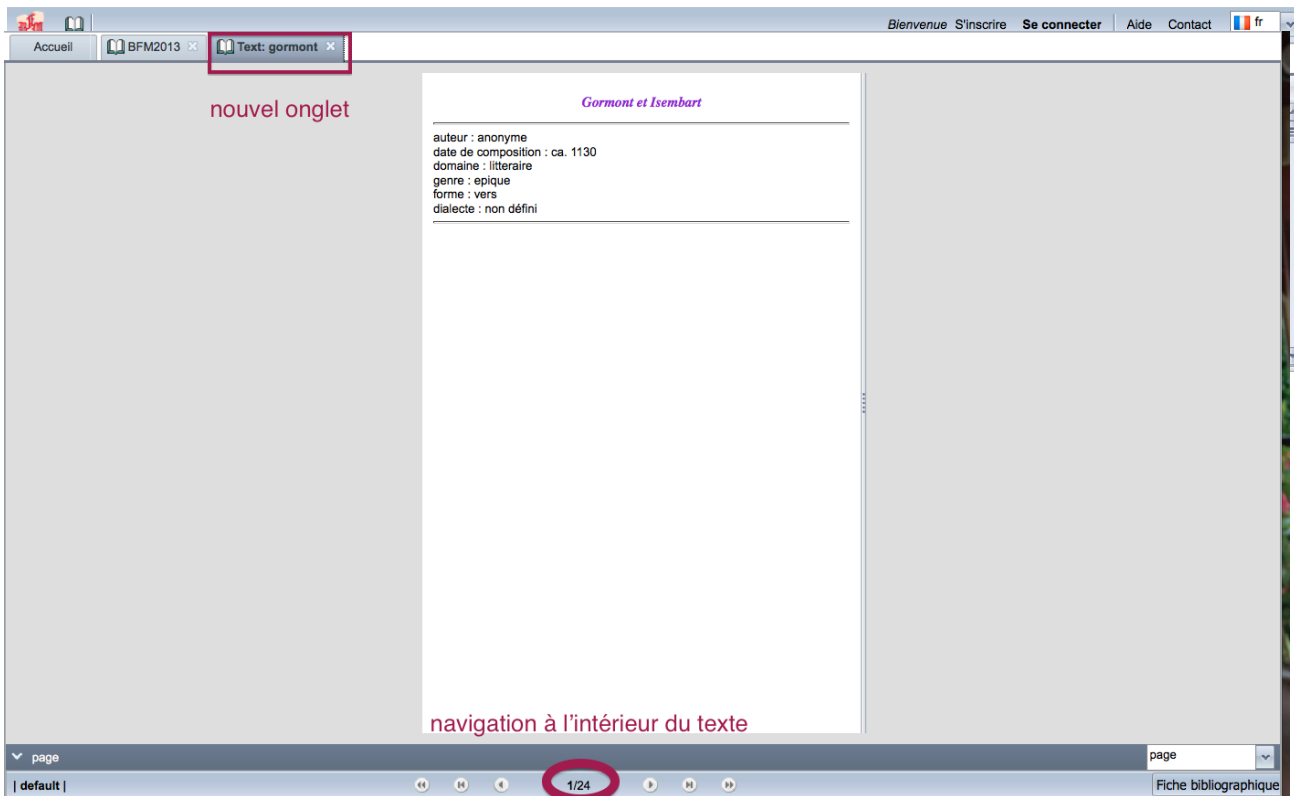


Figure 7: The Edition as it appears in a new tab

Clicking on the PDF button leads to the PDF version of the text, but this time the text will appear in a new tab of your web browser.



Figure 9: Display of an Edition in PDF format in a new browser tab

4 Sign up for expanded access

4.1 Signing up

To have access to more texts and features (lexicon, index, concordances), and to also have the right to download text source files, you must sign up and accept our terms and conditions of use. You can do this by clicking on the hypertext link "S'inscrire" (or "Subscribe"), which appears on the homepage as well as in the toolbar in the upper right corner.



Figure 10: The "Subscribe" button as it appears in the toolbar

The following form appears.

Step 1: Complete the registration form.

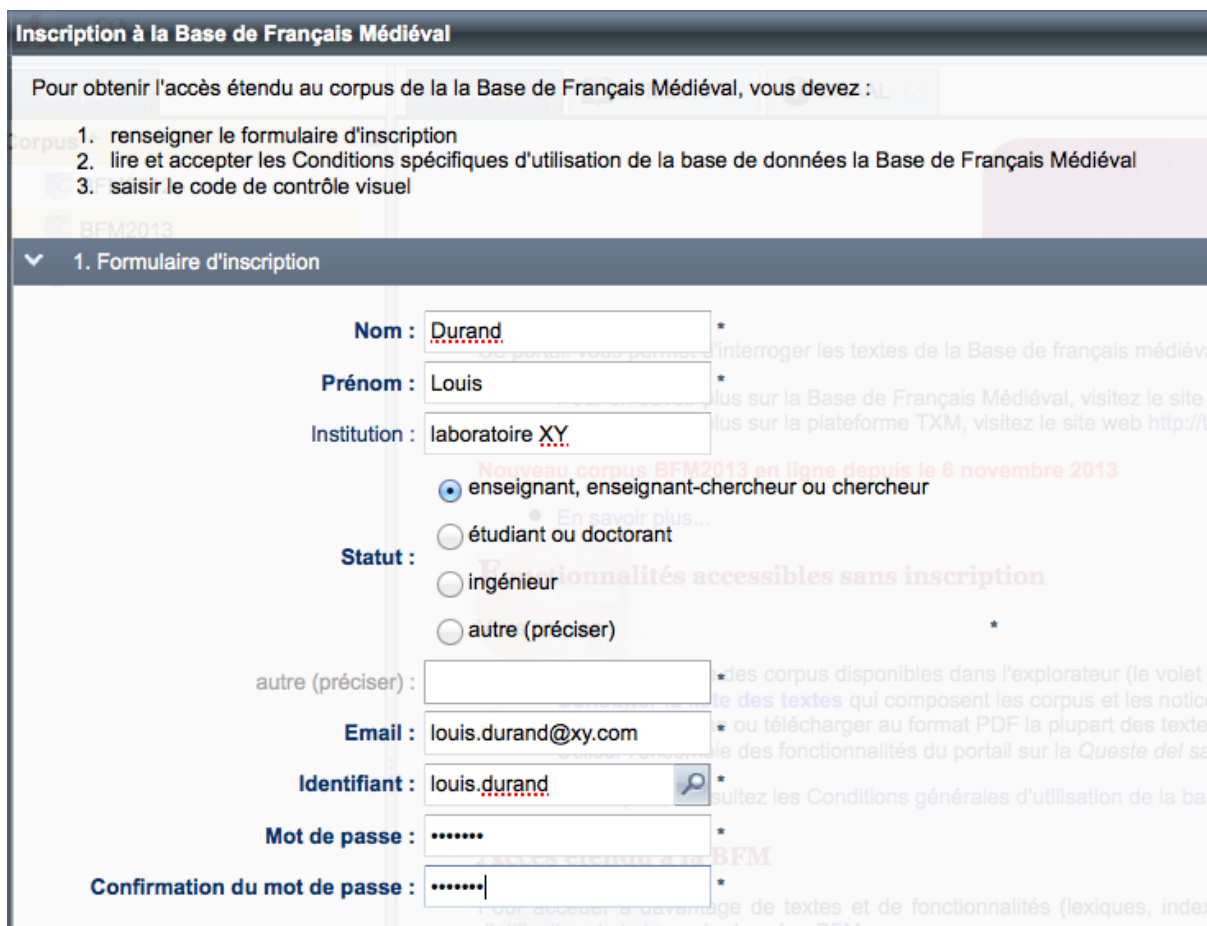
A screenshot of a web registration form titled 'Inscription à la Base de Français Médiéval'. The form is in French and lists three steps: 1. renseigner le formulaire d'inscription, 2. lire et accepter les Conditions spécifiques d'utilisation de la base de données la Base de Français Médiéval, and 3. saisir le code de contrôle visuel. Step 1 is expanded to show the registration form. The form fields include: 'Nom' (Durand), 'Prénom' (Louis), 'Institution' (laboratoire XY), 'Statut' (enseignant, enseignant-chercheur ou chercheur), 'Email' (louis.durand@xy.com), 'Identifiant' (louis.durand), 'Mot de passe' (masked with dots), and 'Confirmation du mot de passe' (masked with dots). There are also radio buttons for 'étudiant ou doctorant', 'ingénieur', and 'autre (préciser)'. A search icon is next to the 'Identifiant' field.

Figure 11: The sign-up form

You must:

- Write your first and last name;
- Indicate the institution with which you are affiliated (university, college...);

- Indicate your professional status (teacher, student);
- Supply your email address;
- Select a username between 6 and 20 characters (to verify the availability of a given username, click on the small magnifying glass);
- Choose a password containing at least 7 characters;
- Confirm your password.

If you receive the following message during the inscription process:



Figure 12: 'Username already in use' error message

the username which you have chosen is already in use, and you must therefore opt for another.

Step 2: Read and accept the general terms and conditions of use of the BFM database:



Figure 13: The BFM's general conditions of use

You just need to check the box certifying that you accept the general conditions of use of the BFM.

Step 3: Enter the CAPTCHA code:



Figure 14: Inputting the CAPTCHA code

Finally, click on Submit (Soumettre)

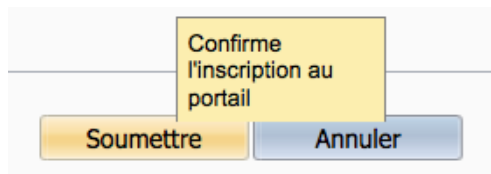


Figure 15: The Submit button

A confirmation window should appear.

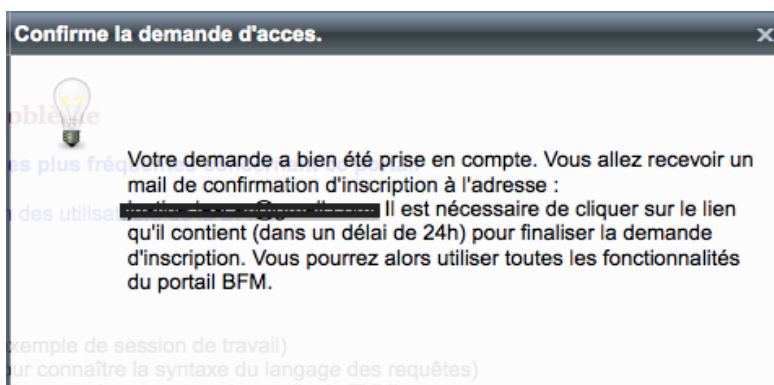


Figure 16: Confirmation of the access request

Step 4: Finalizing your registration.

Simply click on the confirmation link found in the email that you should have received in order to validate the creation of your account.

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.

Figure 17: Confirmation email

Should you not receive the confirmation mail within 5-10 minutes, please do not hesitate to contact the BFM's administrator at the address bfm@ens-lyon.fr.

4.2 Connecting to the BFM

Once your account has been created, you can sign in to the platform by clicking on the Login button in the menu. To do so, indicate your username and password in the following window.

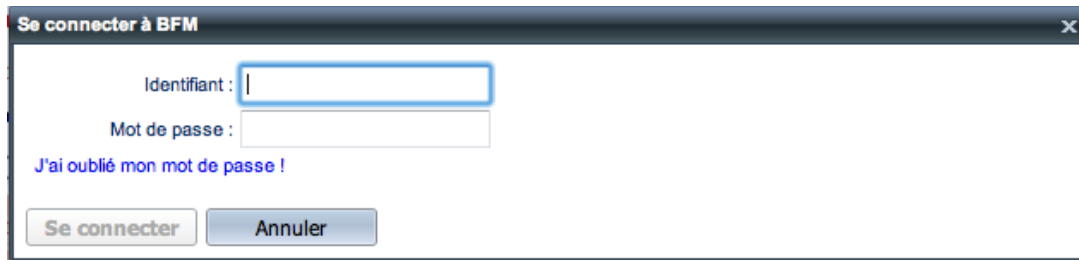


Figure 18: The BFM login window

4.3 Modify your profile

Upon logging in, a new button will appear in the menu which will permit you to access and modify your personal information. Click on My Profile as illustrated in the graphic below.

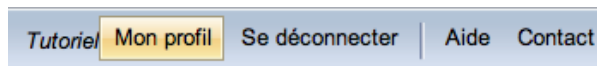


Figure 19: The “My profile” button

You may add or modify the information pertaining to you in the window reproduced hereunder.

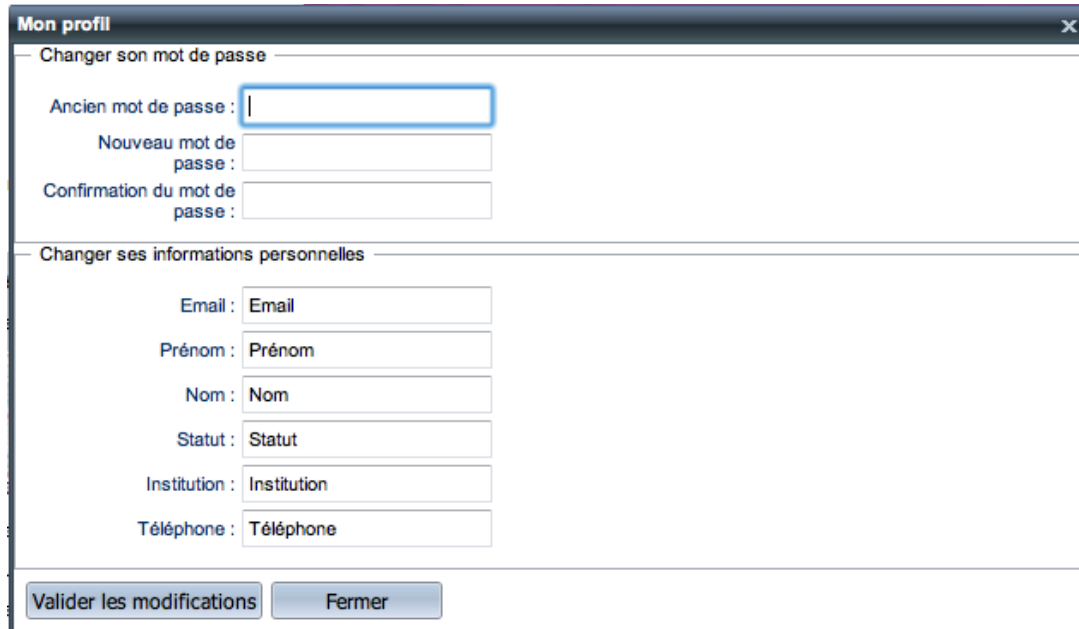


Figure 20: The “My profile” window


At all times you have the ability to:

- Modify your password

- Update your email address
- Update your personal information (Name, Status, Institution, Phone number)

5 Creating a subcorpus

The point of depart will be one of the corpora of the BFM already available on the platform.


If you wish to have access to the bibliographical information concerning the predefined corpora, you must run the command **Texts** accessible from the contextual menu by right-clicking on the corpus that interests you (or by clicking on  in the toolbar). This command will enable you to access the list of texts that constitute the corpus, and to know for each of its texts: the BFM text identifier BFM, the author, the title, the date of composition, the date of the manuscript, the scientific editor, the publisher, the date of publication, the form, the domain, the genre, the dialect, the status of morphosyntactic tagging, and word count. If you would like to carry out a selection within the texts, you should be able to use this information in order to extract the information of interest to you.

By using and combining the texts' descriptors you will be able to create as many subcorpora as you wish. To that end, **select** each time (cf. section 5.2) a group of texts within one of the original corpora: for example, texts from the twelfth and thirteenth centuries that fall under the domains of literature and history, and belong to the BFM2016 corpus. You will also be able to **view** an edition of your choice by means of the **Edition** command.

Next, you will be able to **organize** (section 6) your subcorpus in order to obtain comparative results according to your chosen parameters: for example, you may decide to compare word frequencies in prose and verse texts ('form' parameter), or in different domains. The **Partition** feature assembles the results obtained using the entirety of the subcorpus according to the verse/prose distinction in the first instance, and historical texts/literary texts in the latter. It should be noted that this feature of the platform is used in tandem with the *Index* command.

To create a subcorpus, you will use the command **New subcorpus / Texts** which will enable you to use several descriptors (a chronological descriptor or a descriptor regarding the domain concerned).

5.1 General interface of the *Textes* option of the *Sous-corpus* command

- First issue the command **Text selection** by right-clicking on the name of the corpus (**New Subcorpus/Texts**) or by selecting  from the toolbar. Do not be surprised if nothing occurs at once, the text selection interface sometimes requires some time to load.

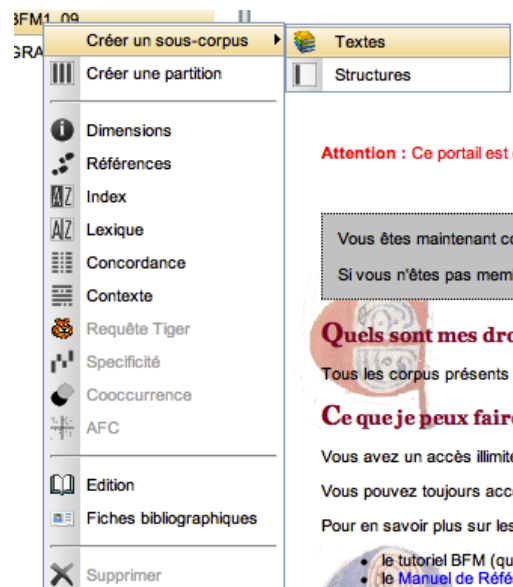


Figure 21: Run the “Text selection” command via the contextual menu on the selected corpus

The results tab for this command is divided into two zones; on the left, dropdown menus list the different descriptors⁶ which categorize the texts; and on the right, a table details the group of texts that comprise the corpus. No text is selected under the default setting, but gradually, as you select your descriptors, checkmarks should appear adjacent to the texts that you have chosen from the list.

While constructing the subcorpus from the available texts, you can add or withdraw texts by using the descriptors (in the left-hand zone). If need be, you may refine the selection text by text (on the right).

For all tables (those included in the dropdown menu of descriptors, just as for those in the detailed list of texts) a left-click on a column header sorts the lines upon the values of the column, and a right-click allows you to choose from a larger selection of formatting operations (including the choice of columns to be displayed).

5.1.1 Zone of text descriptors

A descriptor corresponds to a particular feature of a textual unit. As a whole, the descriptors collect together the metadata that are available and/or useful for the interpretation and exploitation of the texts concerned; they thus depend upon the corpus. Here are some of the useful and available descriptors mobilized by the BFM2019 corpus⁷:

- auteur: author (Bérout, Clari, anonyme⁸...)
- titre: title of work (*Serments de Strasbourg, Queste del saint Graal*...)
- siècle: century during which work was composed (13, 12, 10...)
- date œuvre: date of composition of the work (interval)
- date ms: date of manuscript (interval)

⁶ The BFM terminology differs from that of TXM software. We use the term ‘text descriptor’ which is equivalent to ‘text properties’ for other TXM commands.

⁷ For more information refer to: http://ccfm.ens-lyon.fr/IMG/pdf/Manuel_Descripteurs_BFM.pdf.

⁸ The “anonymous” part of the corpus consists of all works of unknown authors and hence has a different status compared to a collection of works of a certain author.

- **forme**: form of text (verse, prose, mixed)
- **domaine**: domain to which the text belongs (be this literary, historical, didactic, religious or juridical)
- **genre**: text genre (epic, chronicle...)
- **dialecte**: dominant dialect of the work (Picard, Anglo-Norman, undefined...)
- **relation**: relations to other works where applicable (adaptations, translations...)
- **morphosynt**: status of morphosyntactic tagging (verified or unverified)
- **lemmatisation**: status of lemmatization (verified or unverified)

These descriptors may be used to define a subcorpus. To consider a descriptor and thus indicate the values that you would like to retain or exclude, click on the small arrow (^) to the left of the name of the descriptor; a drop-down menu should open out. The purpose of the arrow is to simply show or hide the list of choices for the descriptor. Note that a descriptor may be activated even while the tab remains hidden, and that a descriptor may not be in use even when a table detailing its values appears on the screen. The indicator numbered next to the name of the descriptor, on the blue bar opening each drop-down menu, gives information at all times about the effective use of the descriptor: the figure “(n/N)” indicates that there are *N* different values of the descriptor in the corpus of which *n* have been selected. An unused descriptor therefore displays an indicator of the form “(0/N)”. When opening the drop-down menu, a table should open containing *N* lines corresponding to the *N* available values, with the *n* values you have already selected if any. The situation is different for the date type descriptors, for these operate in terms of time intervals rather than through the selection of values; this difference accounts for the absence of a recapitulatory indicator of the form “(n/N)”.

As seen above, the values of the descriptors are by default unticked (and therefore deactivated), and initially, in the list of texts on the right, no text is selected. To select your texts, you must check the boxes, either by directly clicking on them, by double-clicking on a line in text list, or by selecting one or several lines before, then applying a command to “add” (+) or “remove” (-). Here is a recap of the actions that can be applied to the lines of the table.

- Double-click on a value to activate or deactivate it
- (+): add the values selected (highlighted by left clicking, or ctrl + left click to select multiple list items)
- (-): deactivate the selected values
- ⇄: reverse the current selection of values

The columns provide quantitative information showing how the available corpus and selected texts are distributed over the different values of the descriptor.

For each of the values, *N* is the number of available texts, *n* the number of texts selected (as no texts are selected at first, you should initially see *n*=0). And the same applies for the word count: for each descriptor value, *T* is the word count of all of what is available, and *t* is the word count of the texts selected.

This table can be sorted by clicking on the header of each column.

You can select values for several different descriptors, which are then criss-crossed (cf. the example

below where representative texts have at once been selected from a certain period and certain genres). To be more precise, the selection of a subcorpus in this interface of the zone of descriptors corresponds to a query expression carried out on the descriptor values of the texts elaborated as follows :

- for each descriptor having activated values, a constraint of the form “OR” is applied to the different activated values;
- a constraint of the form “AND” is applied between the descriptors.

Let there be, therefore, an underlying equation of the form:

descriptor A = value A1 **OR** A2 **OR** A3 **OR** etc.

AND

descriptor B = value B1 **OR** B2 **OR** etc.

AND

descriptor C = value C1 **OR** C2 **OR** etc.

5.1.2 Text zone

The text zone (on the right) is affected by the descriptor zone (on the left). As soon as a descriptor value is checked, the entirety of texts retained in the subcorpus, signaled by green checkmarks, adjusts itself in order to verify the selection criterion indicated. When a single descriptor is considered, each choice of a new descriptor value adds texts, and as a result, green checkmarks appear in the text zone. Then, when one or several other descriptors are implemented, each choice of a new value adds criteria restricting the selection, and the green checkmarks may then be removed.

Once you have chosen your descriptors, you have the option to refine your subcorpus directly in this zone by adjusting its composition text by text. A left-click at the start of a line corresponding to one of the texts modifies the selection for this text. If the text was not already selected, it is added to the subcorpus and a checkmark appears; and if it already figured in the subcorpus of chosen texts, it is removed (represented visually by the disappearance of the green checkmark).

NB: The modifications carried out directly on the table within the text zone are lost if you make a modification to the descriptor zone. In other words, one must begin in practice by making the selection with the aid of the descriptors, and finish by refining the selection (if need be) with the table appearing in the text zone.

In order to visualize more readily the information concerning the text that you wish to select, you can sort in ascending or descending order (by clicking on a column header) or hide/display certain columns (by right-clicking on any column, the menu Columns, or checking/unchecking an element):

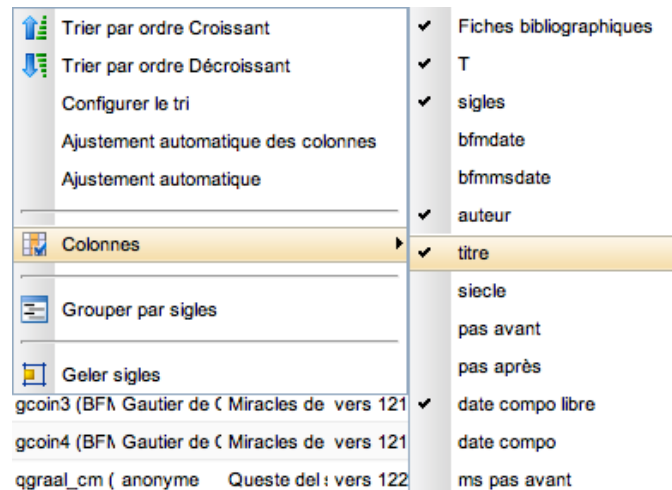


Figure 22: Hiding and displaying the “Title” column

You can also access the bibliographic record of a text by clicking on its icon.

5.1.3 The Statistics Zone

A graphic and more visual representation of the distribution of descriptors in histogram form may be obtained by clicking on the Stats button (above on the left). A third zone then appears on the right. In each tab you will find a histogram of the values of a given descriptor: the length of the bar corresponds to the proportion of the subcorpus for each descriptor value in number of texts (or in number of words according to the option selected). This tool can help you build a more balanced corpus by seeking to obtain a histogram with bars of equal length.

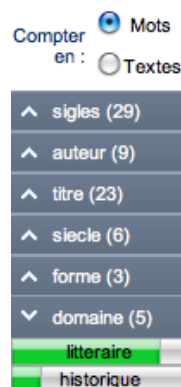


Figure 23: Statistics of the “domain” descriptor in number of words

These statistics are also available in the summary available at the end of the creation of a subcorpus. (cf. Recap).

5.2 Subcorpus example

Say, for example, that we wish to single out literary and historical texts from the twelfth and thirteenth centuries.

- In the lefthand zone, open the pane corresponding to the descriptor `siecle`.
- Select the twelfth and thirteenth centuries (in the pane `siecle`, see figure 24).

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

Figure 24: Selecting the twelfth and thirteenth centuries

As is the case at the start, all centuries are deselected by default. You can either:

- Highlight the lines that you wish to add by holding down the CTRL key while making your selection with the mouse. To confirm your selection, click on .
- Check directly the lines concerned.
- Double-click on these.

The list of texts (the pane on the right) is consequently modified, as well as the distribution histograms (if they are displayed).

We now propose that you apply an additional constraint to a second text descriptor.

- Select the domains `litteraire` and `historique` (see figure 25).

As explained above, add the fields that you wish to conserve. Your list of texts (in the pane on the right) is yet again modified as a result.

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

Figure 25: Selecting the literary and historical domains

For your information, in formal terms, the selection of texts performed corresponds then to the texts that satisfy the following equation.

`siecle= (12 OR 13)`

AND

`domaine= (litteraire OR historique)`

5.3 Creation of subcorpus

- Confirm your selection by clicking OK (above and to the right of the window). You may then name your subcorpus: 12_13_Hist_Litt:

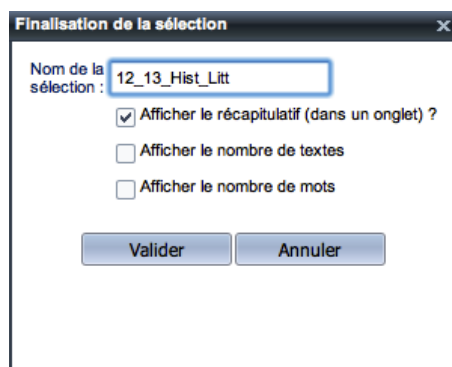


Figure 26: Naming a subcorpus

☞ If you check the option “Display the selection summary in a tab”, a new tab displaying the datum pertaining to your corpus appears. Checking “Show N” and “Show T” adds the columns n , N (words), t , T (texts) and their respective percentages to the summary.

☞ An error message will appear if you attempt to save a subcorpus under a previously used name.

- The new subcorpus now appears in the corpus browser (see figure 27).

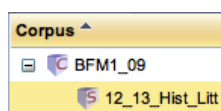


Figure 27: Recently created subcorpus

☞ You have now generated a corpus containing historical and literary texts of the twelfth and thirteenth centuries.

☞ You cannot modify a subcorpus once it has been created. That being said, you do have the option to delete it (right-click, Delete command, or on the toolbar).

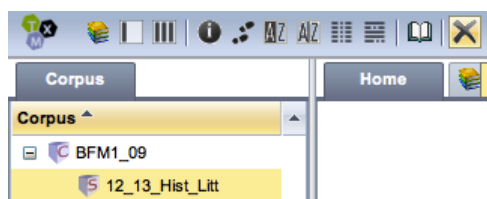


Figure 28: Deleting a subcorpus

Naming rules for subcorpora and partitions.

- The name is case-sensitive.
- The following character is not allowed: forward slash (/).

5.4 Visualizing a subcorpus: the Dimensions and Edition commands

Sections 8.1 and 8.2 of the TXM user Manual

- You can visualize the dimensions of a recently created subcorpus by selecting it and then executing the Dimensions command (in the toolbar):

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

Figure 29: Subcorpus dimensions

This command reports the “size” of your subcorpus by number of occurrences.

☞ When applied to an entire corpus, this command also reports the number of texts.

- You can display an edition of any of the texts contained in the created subcorpus. Select the subcorpus, then execute the Edition command from the toolbar (📖) or contextual menu.

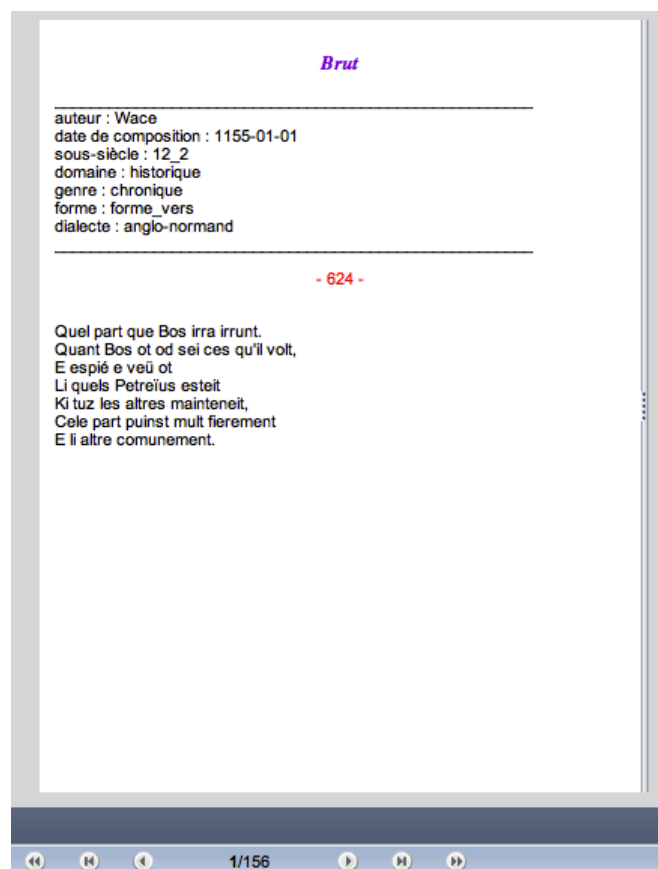


Figure 30: Edition of a text from the 12_13_Hist_Litt subcorpus

You will then be able to navigate through all the accessible texts contained in the subcorpus. On the first page of each text, the principal descriptors are listed. You can navigate through the edition in

this way:

- > : next page
- < : previous page
- >| : last page
- |< : first page
- >> : next text
- << : previous text

☞ The navigation interface is common to all TXM commands.

☞ The commands Edition and Dimensions are available for all corpora: main corpora and subcorpora.

6 Organize your data: partitioning the subcorpus

Section 8.9 of the TXM user Manual

You wish at present to compare the results in terms of text form (verse and prose) but also in terms of the two domains selected (literary and historical texts). The Partition command will allow you to **contrast** certain properties of your corpus. In practice Partition command applies usually to the `text` structural unit.

You need to create two partitions of your new working corpus, one of which will contrast the text forms, and the other the domains.

- Run the Partition command (in its simple mode) on the *12_13 Hist_Litt* subcorpus (see figure 31) from the contextual menu or the tool bar.

Be sure to select the correct icon in the corpus browser. Each created subcorpus can be selected and partitioned. A partition appears in the background as the “descendant” of the original subcorpus.

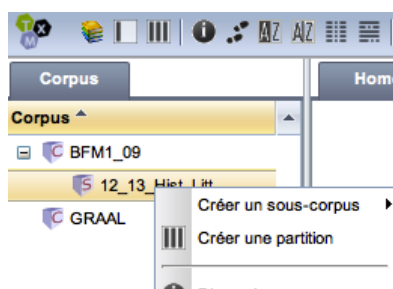


Figure 31: Running the “Partition” command on a selected subcorpus

- Select the structure `text` and the property⁹ `forme`¹⁰. Name your partition, for example `Vers_Prose`¹¹ (see figure 32).

⁹ As indicated above, “property” is equivalent to “descriptor”.

¹⁰ The possible structures and properties correspond to the XML-TEI tags encoded in the text sources.

¹¹ Caution: the forward slash '/' may not be used when naming a partition.



Figure 32: Creating a partition as to the form of the text

- Click on Create, and your partition will appear in the corpus browser :

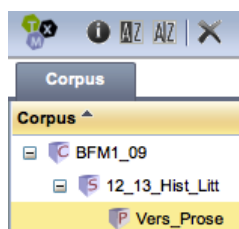


Figure 33: Vers_Prose partition

☞ These actions yield a partition of two parts corresponding to the two text forms defined in the BFM: verse and prose.

- Execute again the Partition command on the 12_13_Hist_Litt subcorpus. Select text again for the structure, but this time choose the property domaine. You will obtain a distribution of the results in terms of the domains represented in the subcorpus, which in the case at hand are literature and history. Name your partition (for example, Hist_Litt).

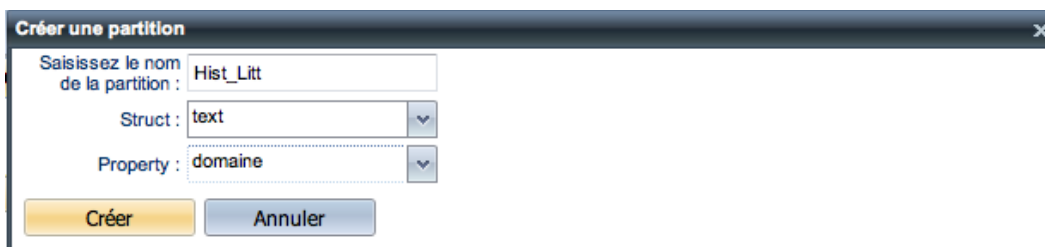


Figure 34: Creating a partition for text domain

Click on Create, and your new partition will appear in the corpus browser.

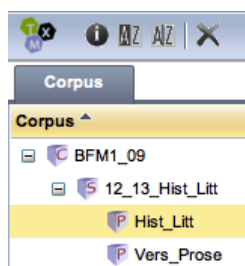


Figure 35: Partition corresponding to the history/literature contrast

☞ You obtain a partition that contains two parts, each corresponding to the two domains of the BFM represented in the subcorpus: history and literature.

You will next be able to use in turns each of these partitions by selecting one of them before executing your commands.

These partitions are but two examples among others. Here is a list of the other available properties:

- auteur (author)
- domaine (domain)
- deaf
 - ID of the source edition in the DEAF bibliography (one part for each edition)
- forme (form)
- genre
- id
 - BFM ID (one part for each textual unit)
- siecle (century)
- titre (title)
 - morphosyntaxe (status of verification of morphosyntactic tagging)
 - lemmatisation (status of verification of lemmatization, not available for the BFM2016 corpus)

7 Analyze your data

You may now analyze the data pertaining to your subcorpus by using TXM commands. The `Index` command will allow you to list the forms (or the tags) of words or constructions corresponding to a certain desired pattern. The index also indicates the **frequency** of these forms in the subcorpus and its partitions. This feature is useful:

- For identifying the forms that interest you: it enables you to find the right formulation for your query by eliminating noise (the irrelevant occurrences). It also assembles all the graphical variants of the word forms with which you are working.
- To get a precise idea of the frequency of these forms in your subcorpus and its potential sections: at this step you will use the sections created by means of the `Partition` feature (section 6).

The `Lexicon` command is similar to the `Index` command, but it will not allow you to enter a precise query. It will simply provide a list of all of the words, including punctuation marks, contained in the working corpus with their respective frequencies.

The `Concordance` command will enable you to **visualize and sort the occurrences** of the desired forms and constructions within a context of adjustable size and form. The `Context` command is similar to concordance but it produces a different visualization in the form of paragraphs rather than in lines of a table.

The texts of the BFM contain morphosyntactic tags and are marked up for direct speech. Certain corpora are also lemmatized. At the end of this section, we will explain how you can use these annotations with the `Index`, `Concordance`, and `Contexte` commands.

7.1 Index

Section 8.3.2 of the TXM User Manual

Let us, for the purpose of illustration, list the occurrences of the word form “chevalier” (case insensitive). To this end, you may use the following CQL query expression: `"chevalier"%c`. (Cf.

the [Référence rapide CQL](#) for more information regarding query syntax).

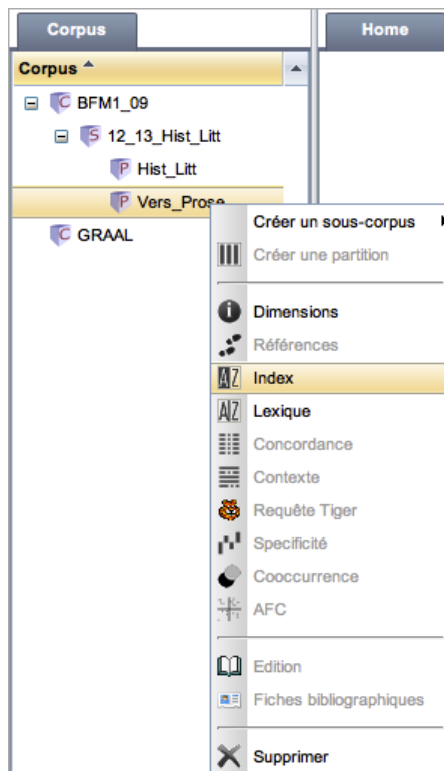


Figure 36: Performing the Index command on a selected partition

Execute the Index command (AZ) on one of the partitions and then enter your CQL expression. Start the search with the Search button.

	word	prose	vers
1	chevalier	829	33
2	Chevalier	10	3

Figure 37: Index of "chevalier"%c with the partition Verse/Prose

The generated table allows you to examine the list of forms corresponding to the CQL expression. What is more, it indicates the frequencies of all forms of the word “chevalier” within the working corpus and the different partitioned sections. In order to analyze more readily these results, you can export the table (see section 8).

7.1.1 Adjusting the index

You may adjust certain index parameters by clicking on the Settings button. Modifications are taken into account only when you click on the Search button.

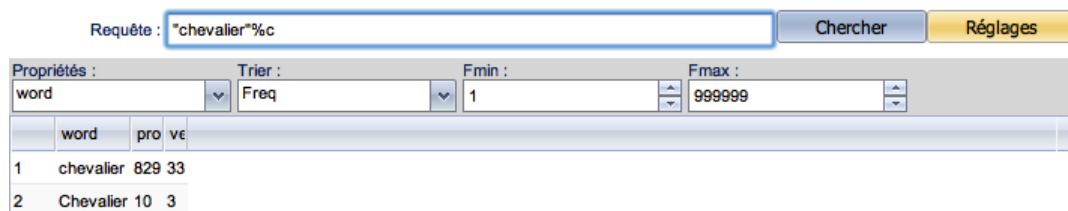


Figure 38: Adjusting the index

- **Properties:** this parameter allows you to choose the target of analysis, namely the representation of the word to be used to construct the list and calculate the frequencies. The list of word properties depends on the corpus (See Appendix 1). It is possible to check multiple boxes in order to display multiple properties. Here, for example, is the list of word properties in the BFM2019 corpus¹².
 - word (graphical form, a default property)
 - cattex-pos (morphological tag in Cattex 09 format)
 - cattex-pos_syn (morphosyntactic tag in Cattex 09 format, see the *Principes d'annotation* in order to familiarize yourself with the difference between morphological and morphosyntactic tagging)
 - etiquetage (verification status of the tagging)
 - fropos (deprecated morphological in Cattex 09 format, identical to *cattex-pos*)
 - id (ID of the occurrence)
 - lang (language of the occurrence)
 - lemma (lemma, see section 7.5 in order to know more about the lemmatization of the BFM)
 - lemma_src (lemma source: DMF, TL, AND, BFM or other)
 - lemmatisation (verification status of the lemmatization)
 - q (level of imbrication in direct discourse: "0" if outside direct discourse... "3" if occurrence is located in direct discourse imbricated in two instances of direct discourse).
 - ref (title of text and the page of the occurrence)
 - sp ("t" if the occurrence is situated in a dialogue, otherwise "f")
 - supplied ("t" if the occurrence is located in a conjecture made by the scientific editor, otherwise "f")
 - ud-pos (part of speech tag according the Universal dependencies tagset)
 - ud-feats (tag indicating the set of morphological features from Universal dependencies tagset)
- **Sort:** allows you to sort your index by frequency or alphabetical order (according to graphical form or any property value) in the left column
- **Fmin:** enables you to choose a minimum threshold frequency
- **Fmax:** enables you to choose a maximum threshold frequency
- **Vmax:** allows you to limit the list to the N most frequent forms (where N is the value selected for Vmax)
- **Pagination:** allows you to change the number of lines per page

You can repeat these steps for each created partition.

- After examining your obtained result, you may want to rework and modify:
 - the query, if it does not retain the desired occurrences: either there are intruders to be excluded, or there are forms not included by the equation;
 - the partition: creation of a new partition (even if you must only make a small adjustment);
 - the subcorpus: you may reduce it (to make it more balanced) or extend it by creating a

¹² In the section 7.5 you will see how to use the word properties for research in the corpus. The full list of word properties available in the BFM corpora is provided in Appendix 1.

new subcorpus.

7.2 Lexicon

Section 6.3.1 of TXM user Manual

- Select the working corpus or partition of your choice and execute the Lexicon command (accessible from the contextual menu as well as from the toolbar).

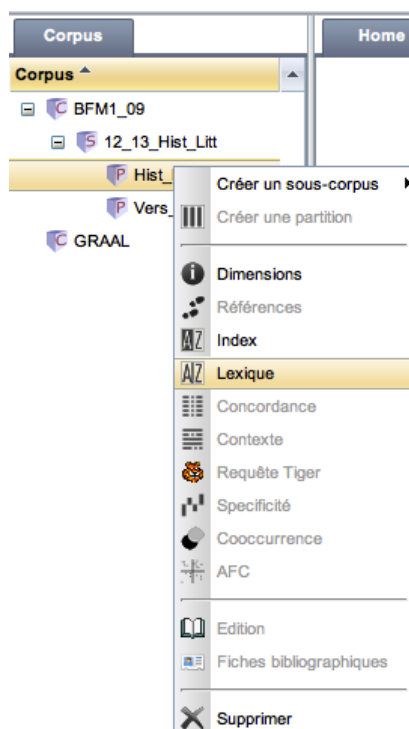


Figure 39: Executing the Lexicon command

This command generates a table representing each graphical form indexed as well as its associated frequency of appearance within the working corpus (when the command is applied to a corpus or a sub-corpus) and each of its sections (when the command is applied to a partition).



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

Figure 40: Lexicon for the partition Hist_Litt

7.2.1 Adjusting the Lexicon

You may adjust certain parameters of the lexicon by clicking on the `Settings` button. Each modification must be confirmed by clicking on the `Process` button.



Figure 41: Adjusting the lexique

- **Properties:** allows you to select a property to display and use for frequency count;
- **Sort:** allows you to sort the *lexique* by frequency or alphabetical order according to graphical form (or another property value) in the left column;
- **Fmin:** enables you to choose a minimum threshold frequency;
- **Fmax:** enables you to choose a maximum threshold frequency;
- **Vmax:** allows you to change the total number of lines (so that only the ten most frequent words appear, for example);
- **Pagination:** allows you to change the number of lines per page.

7.3 Concordances

Section 6.4 of TXM User Manual

Let us now consider how you may recover the context of your studied forms in order to interpret and analyze it.

7.3.1 Creating concordances

You can build a concordance on a selected subcorpus (the command is accessible only for (sub)corpora and not for partitions, for it is not contrastive; while working on a partition, apply the command to the subcorpus from which the partition has been generated). Select the icon of the subcorpus and launch the concordance via the contextual menu (or click on it in the toolbar). Next, simply enter an expression in the query field and submit your concordance with the `Search` button.

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	plaigne. 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

Figure 42: Concordance for the query "chevalier"%c

- Double-clicking on a line of the results enables you to return to the edition. The selected pivot then appears on red background.

- 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
Espurinant, 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 esper.
Cil de Peitou bien les asaillent
E li Bretuns ne se lur failent ;

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

Figure 43: Returning to the edition from a concordance

7.3.2 Adjusting the settings for the concordances

Sections 6.4.4. to 6.4.6 of *TXM User Manual*

You can directly change four parameters: the size of the context(s), the number of results per page, the display properties of the references and results, and the sorting.

1) Size of the context

- You can adjust the size of your different contexts, defined in number of words, by entering the fields `Size left` and `right`, for the lefthand and righthand contexts respectively. In order to confirm your contexts, click on `Search`.



Figure 44: Adjusting context size

2) Number of results per page

- You can indicate the number of lines per page desired in the field titled `Pagination`. To confirm, click on `Search`.

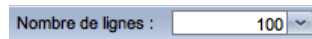


Figure 45: Adjusting the number of results per page

3) View of references, pivot and contexts

- These three adjustments are accessible from the scroll-down menu of the `View` line in the Setting panel.
- The references allow you to label each line of concordance within different textual unities of the subcorpus by displaying the selected properties (the default chosen properties are text ID, page number, and for verse texts, the number of the verse in which the occurrence appears). The form of the references can be defined by checking the reference properties in the menu¹³. For example, you can choose to display the form (text : forme) or the domain (text : domaine)

¹³ The available references are generated from the TEI XML tags used in the corpus source files.

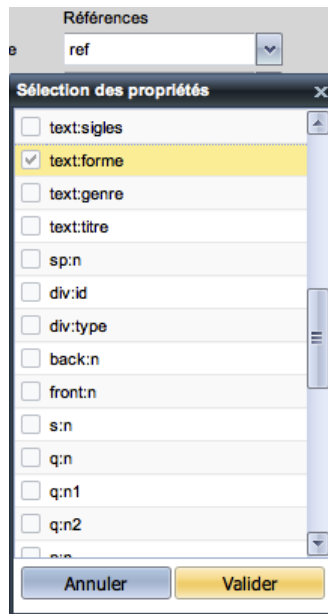


Figure 46: Defining the references

	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 vivre 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	Deparloit son conquerement :	Maint chevalier fist de noient	Riche et

Figure 47: Result of the above adjustments

- You can also choose the information (or properties) that you would like to have displayed in the pivot and context columns. For example, if you wish to know whether your occurrences appear in direct discourse or narration, choose “word” and “q” as displayed properties for the pivot (“word” corresponds to the form of the word, a property displayed by default, and “q” indicates the level of imbrication in the direct discourse: “_1” for the first level of imbrication, “_2” for the second or “_3” for the third, whereas “_0” indicates narration).

☞ NA signifies that the chosen property has no provided value or is not applicable.

- If you wish to see the morphosyntactic tag associated with each occurrence of the pivot, add “cattex-pos” (morphosyntactic tag Cattex 09) or “ud-pos” (part of speech according to the *Universal dependencies* framework) to the list of display properties.

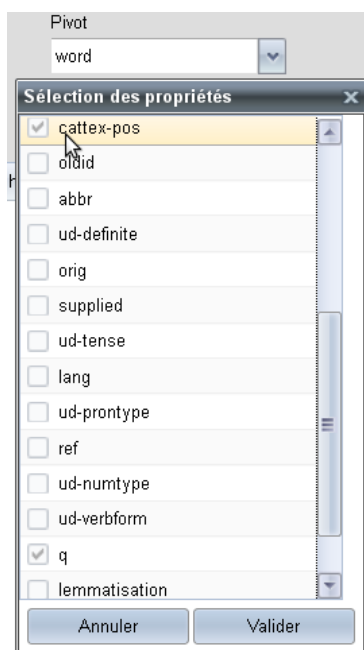


Figure 48: Add the properties “q” and “cattex-pos” to the pivot

	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 apareillement	
5	vers, brut2, v.14684	muillier, Muerent vilain e chevalier_0_NOMcom	; N'estuet al fiz	
6	vers, escoufle, v.80	Deparitoit son conqerement : Maint chevalier_0_NOMcom	fist de noient Riche et	

Figure 49: Display of pivot including “q” and “cattex-pos” properties

The available properties are identical to those of the Index command.

4) Sorting of concordances

- These three adjustments are accessible from the drop-down menus of the `Sort` line in the `Settings` panel.
- For each column of the concordance, you have the option to indicate the property in fonction of which the sorting should be performed. This property is not necessarily the graphical form (`word`) nor the property chosen for the display. For example, the references can be sorted in terms of information placed in second or third position in the display (a reference `ref+text:siecle` can thus be sorted by century).



Figure 50: Choosing the sorting property (`cattex-pos`) for the righthand context

- You also can sort your results in a more complex manner thanks to the “Sort keys” (preceded by the number sign: #). These keys allow you to combine and hierarchize your

sorting criteria: for example, you may sort your results in the first instance according to the form of the pivot, and then according to the righthand context in the second. Confirm your choice by clicking `Sort`. You may also choose whether to sort in ascending or descending order by clicking on the arrow to the right of each field.



Figure 51: The “Sort keys” for the pivot and right-hand context

7.4 Context

Like the Concordance command, the Context command enables you to observe all the occurrences of a search pattern in their actual contexts. The two differ, however, when it comes to presentation:

- Since it presents the overlapping contexts within a sortable table, the Concordance feature proves to be more efficient for analyzing the immediate context of the pivot and its recurrence (constructions, idiomatic structures, etc.);
- As for the Context, which appears as a sequence of text excerpts with highlighted occurrences of the sought pattern, it makes it easier to read each context.

Double-clicking on an excerpt allows you to display the whole page of the edition.

- Perform the Context command on your subcorpus:

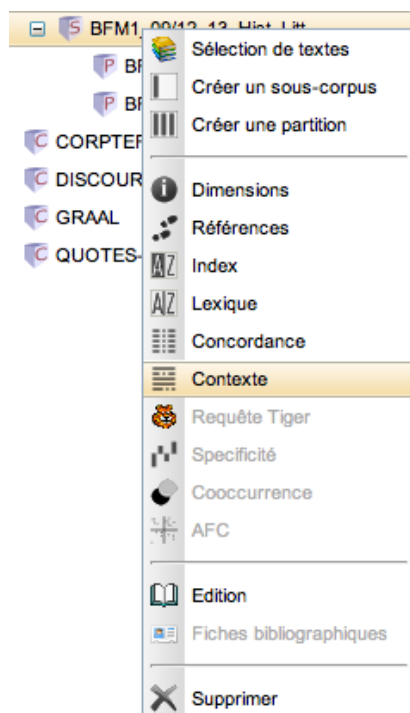


Figure 52: Performing the Context command

Requête : "chevalier"%c		Chercher	Trier	∇ Réglages
Référence				
▲	brut2, v.12211	Esteient as prisuns rescurre . Cele part laissa chevals curre , Ovoc li trei mil chevalier E li forier e li archier . Romein a ferir entendeient , Ki les Bretuns		
▲	brut2, v.12784	de sun barnage la flur , Od lui erent li gentil hume E li bon chevalier de Rome . La veïssiez estur mortel ; Unc ne veïstes , ço quit ,		
▲	brut2, v.14021	baillier , Chevals bien puindre e bien tenir , Espees traire e cops ferir , Chevalier furent fait ensemble En Armoriche , ço me semble . La les ot fait Chadwan		
▲	brut2, v.14649	mort n' ad nul refui . Pur lui lungement remembrer Firent de quivre tresjeter Un chevalier sur un cheval En apareillement real ; Dedenz fud lu cors le rei mis ,		

Figure 53: Context for the query "chevalier"%c

The adjustments proposed for Contexts are identical to those for the concordances (refer to section 7.3.2).

7.5 Use word properties to examine the corpus

7.5.1 Cattex09 Morphosyntactic Tags

All texts contained in the BFM corpora have been tagged for morphosyntax. This tagging has been verified for [some forty texts](#), and is automatic for the rest of the corpus. In the BFM2016 and CORPTEF corpora the verified tags are available thanks to the word properties “pos” and the automatic tags through the property “fropos”.

In the BFM2019, PALAFRAFRO-V2-2 and PALAFRAPAR corpora, a single property “cattex-pos” gives access to the Cattex morphological tags, be they verified or not, and the “etiquetage” property makes it possible to restrict your queries to include only verified tags.

It is also possible to use the ‘étiqueté’ subcorpus of the BFM2019 to work with verified tags only.

From this point forward, this new system will be used with all new BFM corpora.

In certain precise cases, the morphosyntactic tag contained by the property “cattex-pos_syn” may differ from the morphological tag (such is the case with substantivized infinitives and participles, for example). The document [Principes d’annotation Cattex 09](#) accessible from the BFM site explains this usage in more detail. Even so, because this principle has only been applied to certain verified texts, we recommend that you make use of the morphological/morphosyntactic distinction with caution. In non-verified texts, the “cattex-pos_syn” tag is always identical with “cattex-pos”.

Here is an example query that could be used with the Index, Concordance or Contexte commands to find all verbal forms.

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

To restrict the search to only verified tags, you must either work with a subcorpus of verified texts, or use the following query:

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

A list of Cattex09 tags appears in the second appendix and under the [rubrique « Documentation »](#) of the site <http://bfm.ens-lyon.fr>. In order to learn more about how you should enter your query, consult the following link: [Référence rapide CQL pour la BFM](#).

7.5.2 The “Universal dependencies” morphosyntactic tags (UD)

When using the BFM2019, PALAFRAFRO-V2-2 and PALAFRAPAR corpora, as well as the PALAFRALAT-V2-0 Latin corpus, it is possible to use the tags defined by the [Universal dependencies](#) project. For French texts these tags have been obtained through automatic conversion of the Cattex 09 morphological tags. Different properties are used to indicate part of speech and various morphological traits.

- **ud-pos (part of speech)**
- **ud-feats** (a set of morphological features)

Here is a sample query that you could use to find occurrences of the present participle:

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

More information regarding the UD tagset can be consulted by going to the site <http://universaldependencies.org>.

The conversion table for transforming Cattex09 tags into UD tagging is given in Appendix 3 and under “[Documentation](#)” at the site <http://bfm.ens-lyon.fr>. In order to learn more about how you should enter your query, consult the following link: [Référence rapide CQL pour la BFM](#).

7.5.3 Lemmas

Lemmatization is currently only available for the BFM2019, PALAFRAFRO-V2-2 and PALAFRAPAR corpora. The accuracy of the process has been verified manually for a [restricted number of texts](#), and is computer-generated for the rest of the corpus. You may use the “lemmatisé” sub-corpus of the BFM2019 to work only with texts with verified lemmas and morphological tags (whereas the lemmas are verified, the morphological tags are verified, too).

With automatic lemmatization, several hypotheses may be provided. The format of the lemmas has been optimized to facilitate searches through multiple hypotheses. Thus, to search all forms of the verb ‘être’, use the following query:

```
[lemma contains "être"]
```

The majority of lemmas have been borrowed from the [Dictionnaire du moyen français](#) (DMF), but in certain instances they come from the Tobler-Lommatzsch dictionary or other sources. In consequence, it is useful to include the lemma forms from different sources:

```
[lemma contains "être|estre"]
```

Note that the figures used in the DMF to distinguish homonyms have not been included within the lemmas. A query that takes into account both the lemma and the morphosyntactic tag can help you clarify and resolve certain ambiguities:

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

(allows you to exclude the occurrences of the nouns *fust* and *esté* for instance)

To restrict the query to verified lemmas only, you must use the property “lemmatisation”:

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

Finally, for verified lemmas, it is possible to know their source dictionary or lexicon through the “lemma_src” property. You may add this property to the technical settings for the Index (cf. 7.1.1) or the concordance (cf. 7.3.2) commands. The list of sources is given in Appendix 4 and also appears under [rubrique « Documentation »](#) of the website <http://bfm.ens-lyon.fr>. In order to learn more about how you should enter your query, consult the following link: [Référence rapide CQL pour la BFM](#).

7.5.4 Available word properties and possible combinations

All word properties can be visualized in the indexes and the concordances through the `Settings` button. They can also be queried and linked with one another by means of CQL language (see [Référence rapide CQL pour la BFM](#)).

For example, to search the occurrences of the form *a* for which the Cattex tag “VERcjg” (conjugated verb) has been verified, you must write and submit the following query:

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

If the query concerns a lemma, you must use the operator `contains`:

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

The exhaustive list of word properties is located in Appendix 1 and also appears under [rubrique « Documentation »](#) of the website <http://bfm.ens-lyon.fr> (*Liste des propriétés de mots disponibles dans les corpus de la BFM* document).

7.6 Using structural units to query the corpus

In addition to word properties, you may use the units of text structure to work with a corpus. These units are encoded with TEI XML tags that can be used in queries.

For example, the instances of direct speech have been marked up with the TEI `<q>` tag in the BFM texts. The following query allows you to build an index or a concordance of the instances of direct speech:

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

To search for a word or a phrase located within a direct speech it is however easier to use the “q” word property:


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

When the value of the “q” property is “0”, this means the occurrence is not located inside a direct speech. When the value is “1”, the occurrence is located in the first level of a direct speech (the levels of nesting of direct speech are expressed with numbers from 1 to 3).

To display the whole direct speech containing the word in the query, you may just add the “expand to q” operator to your query:

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

8 Exporting a document from TXM

Upon generating an index or a concordance, you may save your obtained results by clicking on  `Export` button in the right bottom corner of the window: you can thus rework them in another software program or include them in a publication.

16	escoufle, v.1039	Ki mix mix tuit li chevalier . Ki lors veïst chauces
17	escoufle, v.1126	plaigne. 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 iaus
21	escoufle, v.2786	sans gent, N'avrés chevalier ne sergant , Que nos
22	escoufle, v.2919	jouant Par le palais li chevalier , Li traïtor , li

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

Figure 54: Exporting the selected result, via the “Exporter” button at the bottom righthand corner of the “concordances” tab

The export result will appear in a new tab or a new window of your browser. It may be that your browser blocks pop-up windows from showing up on your screen. Should such be the case, you will first need to configure your browser to authorize popup windows from the site <http://txm.bfm-corpus.org>, then relaunch the Export command.

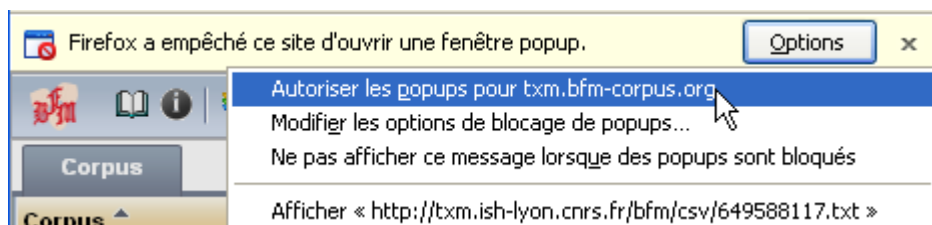


Figure 55: Authorizing popup windows from the BFM portal

If the word “Référence” at the beginning of the exported concordance does not appear correctly, you must manually select the Unicode (UTF-8) character encoding in your browser menu.

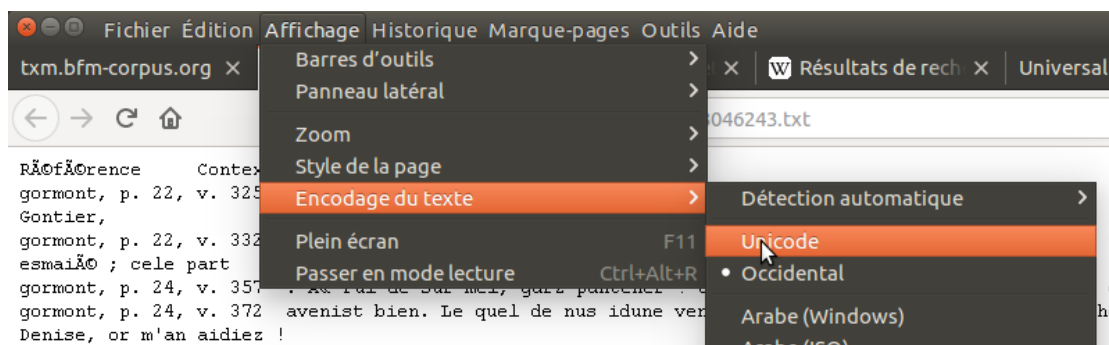


Figure 56: Modifying character encoding (Firefox 59)

To open the exported concordance or index in a spreadsheet software program (such as Excel or Calc), you may use one of the two following methods. If your browser does not enable you to manually choose the character encoding of the text, you will need to use the second method.

Method 1: Copy and paste

Open Calc or Excel

- Select all the content of the tab with the export results and copy it to the clipboard (Ctrl + A or Apple/Command key + A, then Ctrl + C or Apple/Command + C)

With Excel

- Go to the Editing menu → Paste special → Unicode Text
- Verify to make sure that the columns of certain lines have not shifted in relation to the others. This may occur if the right double quotation mark character (") lies at the start of a cell. In such a case, you must use second method.

With Calc

- Go to the Editing menu → Paste special → unformatted text
- In the settings window for Text Import, confirm and correct if necessary the following options:
 - Character set: Unicode (UTF-8)
 - Separated by: check “Tab”, uncheck the boxes for the other options
 - Text delimiter: remove the quotation mark
 - Click OK

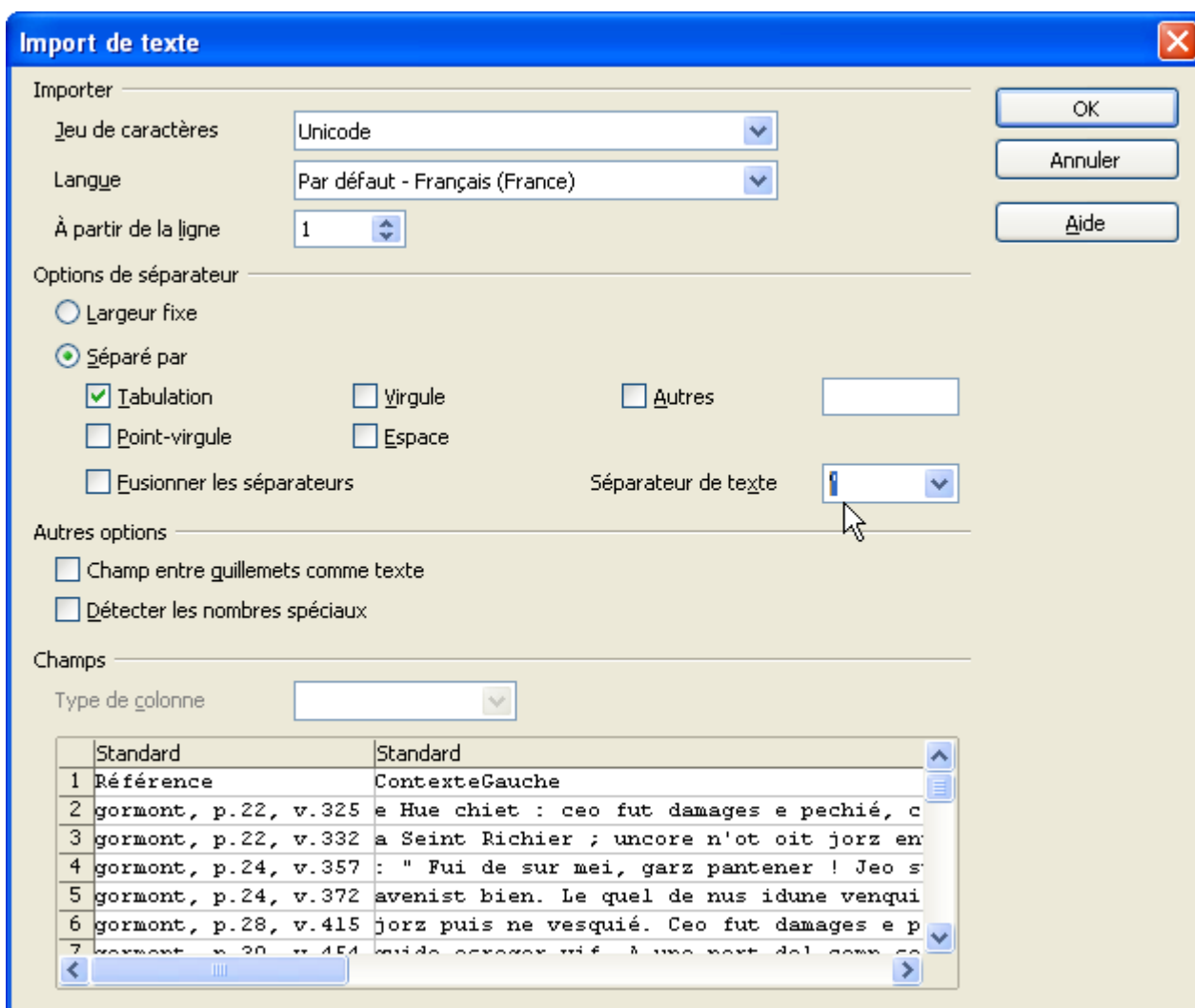


Figure 57: Text Import settings in OpenOffice Calc

Method 2: Opening a text file

- Save the exported concordance or index to your computer (File Menu → Save as)

Do not hesitate to replace the automatic name of the file

- Open Calc or Excel
- **With Calc**
 - Go to the File menu → Open
 - Open the folder in which you have saved your export and select the corresponding file
 - Configure the import text settings (figure 57) as indicated here above for the Copy and Paste method
- **With Excel**
 - Go to the Data menu → From Text
 - Open the folder in which you have saved your export
 - Select the exported file and click “Import”

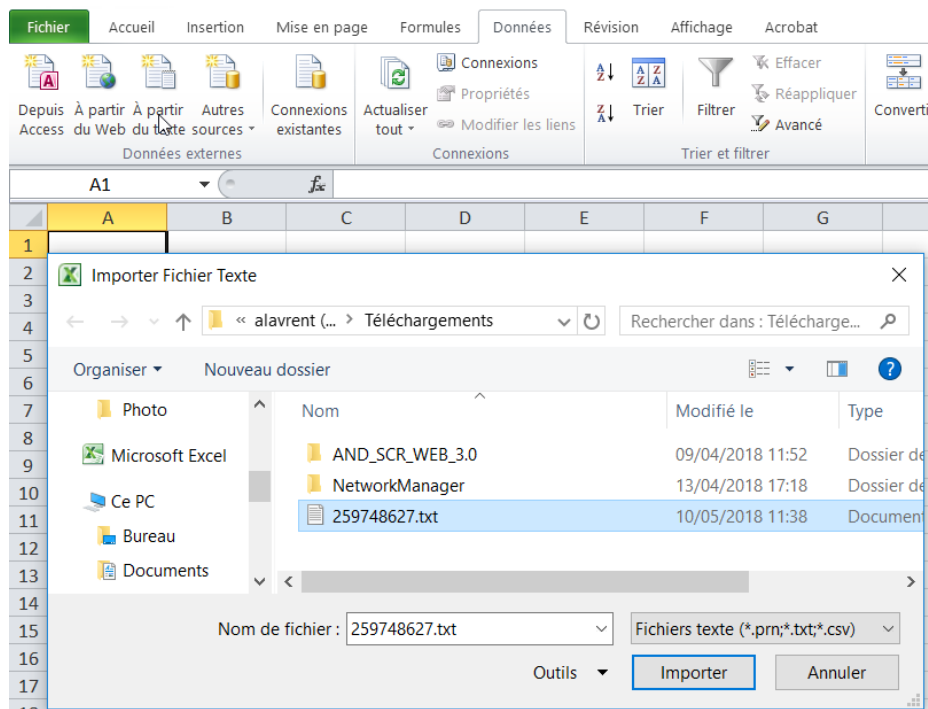


Figure 58: Opening a text file (Excel 2010)

- In the Text Import Wizard, choose the following options:
 - Original data type: “delimited”
 - File origin: Unicode (UTF-8)
- Click “Next”

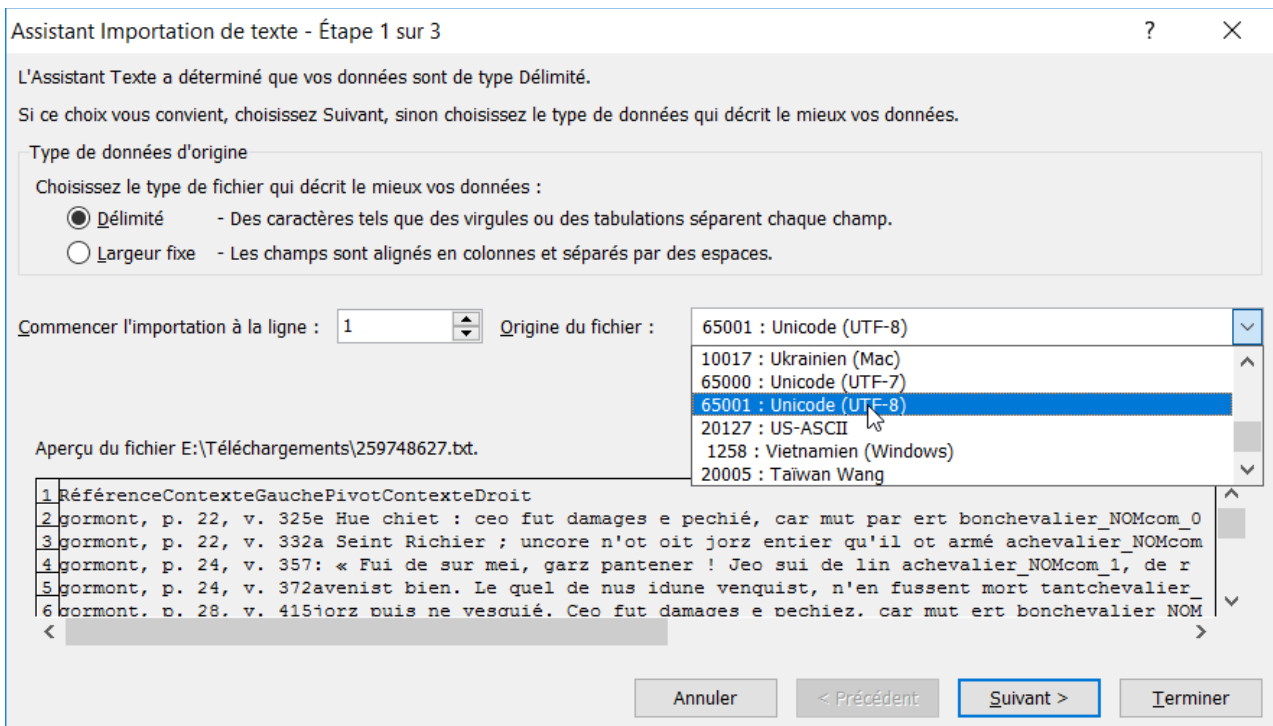


Figure 59: Text Import Wizard, Step 1 (Excel 2010)

- In the following window:
 - Check the box for “Tab” and deselect the others
 - Select {none} from the drop-down list “Text Identifier”
- Click “Finish” (step 3 is optional)

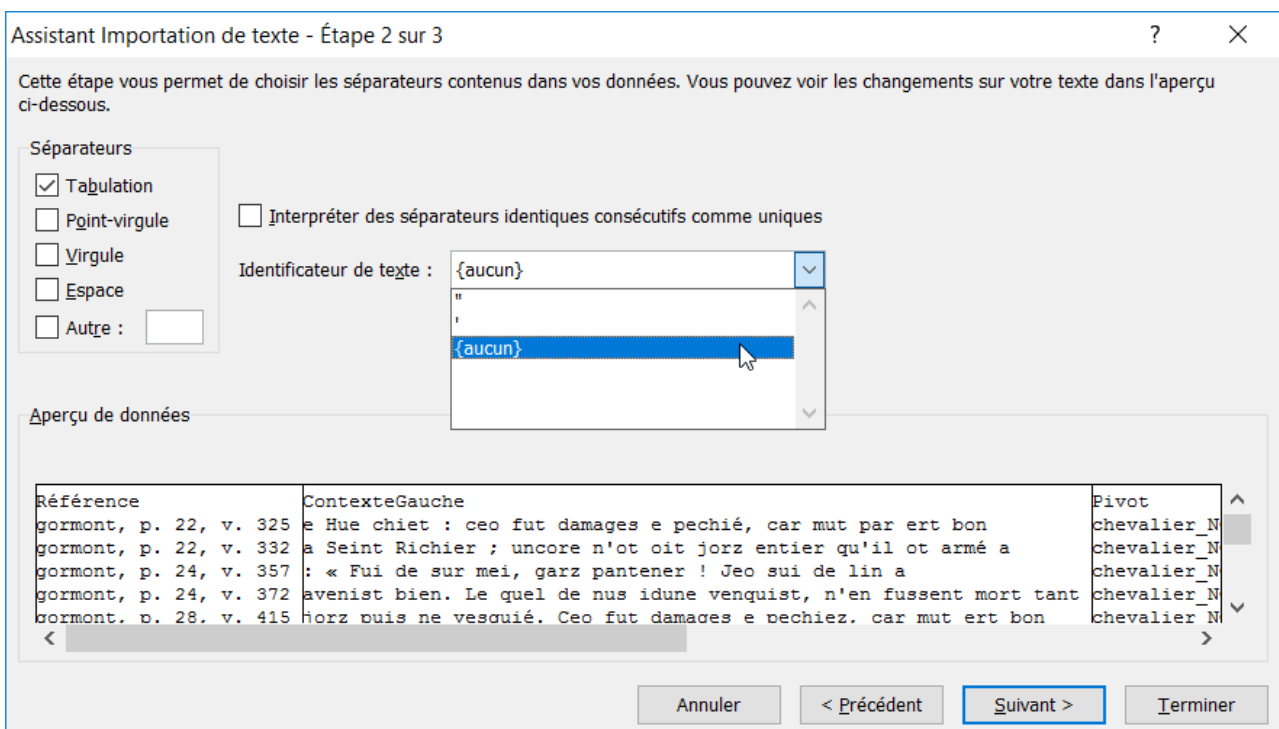


Figure 60: Text Import Wizard, step 2 (Excel 2010)

You thereby obtain the following table, which you may modify (for reading purposes, think to right-align the column to the left of the pivot).

	A	B	C	D
17	escoufle, v.1039	Ki mix mix tuit li	chevalier	. Ki lors veïst chauce
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

Figure 61: A Concordance Viewed in a Spreadsheet

NB: The spreadsheet does not offer sorting possibilities designed for textual data. Therefore you will not be able:

- For the left-hand context, to sort from the word closest to the pivot to the one which is farthest away (sorting from right to left, at the level of words, and from left to right at the level of word characters)
- To use for sorting the properties that are not actually displayed.

Think to save the table in the format of your spreadsheet software (.xls, .xlsx or .ods)

Appendixes

Appendix 1: List of word properties in the corpora of the BFM

The table produced below provides a list of available word properties that are useful for examining the different corpora of the BFM. The “Corpus” column indicates the corpora for which the properties are available.

- BFM2019: BFM2019, PALAFRAFRO-V2-2 and PALAFRAPAR corpora
- BFM2016: BFM2016 and CORPTEF (these properties will disappear in all new versions of the corpora)
- all: all French corpora

Property name	Explanation	Sample Query	Corpora concerned
cattex-pos	Cattex 09 morphological tag	[cattex-pos="ADJ.*"]	BFM2019
cattex-pos_syn	Cattex 09 morphosyntactic tag, identical to cattex-pos, save for exceptional cases of substantivation, etc., identified in certain verified texts	[cattex-pos_syn="NOMcom" & cattex-pos="VER.*"]	BFM2019
etiquetage	Indicates if the morphosyntactic tagging has been verified for a given occurrence	[etiquetage="vérifié"]	BFM2019
frolemma	Identical to the form (of little use for your queries)		BFM2016
fropos	Automatic, unverified morphosyntactic tag (Cattex 09)	[fropos="ADJ.*"]	BFM2016 ¹⁴
id	Word ID within the corpus (of little use for queries)		all
lang	Language of foreign (not Old French) word	[lang="la.*"]	all
lemma	Lemma of the occurrence	[lemma contains "avoir"]	BFM2019
lemma_src	Lemma source, see section 3	[lemma_src="BFM"]	BFM2019
lemmatisation	Indicates whether or not the lemma has been verified	[lemmatisation="vérifié"]	BFM2019
orig	Indicates whether the word has been tagged <orig> (nonregularized	[orig!="NA"]	all

¹⁴ The property `fropos` is also available in the BFM2019 corpus for the users' convenience, its value is identical to `cattex-pos` (Cattex 09 tag, verified or unverified).

	graphical form) in the TEI XML source file, used for some punctuation marks in the <i>Queste del saint Graal</i>		
pos	Verified Cattex 09 morphosyntactic tag	[pos="ADJ.*"]	BFM2016
q	Level of imbrication of occurrence within direct discourse ("0" if outside direct discourse, "1" if within direct discourse that is not imbricated, "2" or "3" if within one or two other instances of direct discourse)	[q="[123]"]	all
ref	Default reference for the concordance (of little use for queries)		all
sic	Indicates whether a word has been tagged <sic> (conspicuously erroneous graphical form) in the XML-TEI document	[sic!="NA"]	all
sp	Indicates whether a word appears within a dialogue (TEI tag <sp>), possible values: "t" (true) or "f" (false)	[sp="t"] ou [sp="f"]	all
supplied	Indicates whether a word appears in a passage tagged as <supplied> (editorial conjecture) in the XML-TEI document, possible values: "t" (true) or "f" (false)	[supplied="f"]	all
ud-feats	Tag for all UD features	[ud-feats contains "PronType=Det"]	PALAFRA2-2
ud-pos	UD tag for "part of speech"	[ud-pos="ADJ"]	PALAFRA
word	Word form, a default property. An abbreviated syntax may be used if the query does not combine several properties	a, "a" ou [word="a"] [word="a" & cattex-pos="VERcjg"]	all

Appendix 2: List of Cattex09 Tags

Tag	Interpretation
ABR	ABBREVIATION
ADJcar	Cardinal ADJECTIVE
ADJind	Indefinite ADJECTIVE
ADJord	Ordinal ADJECTIVE
ADJpos	Possessive ADJECTIVE
ADJqua	Qualitative ADJECTIVE
ADVgen	ADVERB (general)
ADVgen.PROper	Contraction ADVERB + Personal PRONOUN
ADVint	Interrogative ADVERB
ADVneg	Negative ADVERB
ADVneg.PROper	Contraction Negative ADVERB + Personal PRONOUN
ADVsub	Subordinating ADVERB
CONcoo	Coordinating CONJUNCTION
CONsub	Subordinating CONJUNCTION
DETCar	Cardinal DETERMINER
DETCom	Compound Definite DETERMINER
DETdef	Definite DETERMINER
DETdem	Demonstrative DETERMINER
DETind	Indefinite DETERMINER
DETint	Interrogative DETERMINER
DETndf	Non-definite DETERMINER
DETpos	Possessive DETERMINER
DETrel	Relative DETERMINER
ETR	Foreign word
INJ	INTERJECTION
NOMcom	Common NOUN
NOMpro	Proper NOUN
OUT	ERROR
PONfbl	Weak PUNCTUATION
PONft	Strong PUNCTUATION
PONpdr	PUNCTUATION Close Parenthesis ‘)’
PONpga	PUNCTUATION Open Parenthesis ‘(’

PONpxx	PUNCTUATION Parentheses (be them opening or closing)
PRE	PREPOSITION
PRE.DETdef	Contraction PREPOSITION + definite DETERMINER
PRE.DETrel	Contraction PREPOSITION + Relative DETERMINER
PRE.PROper	Contraction PREPOSITION + Personal PRONOUN
PROadv	Adverbial PRONOUN
PROcar	Cardinal PRONOUN
PROdem	Demonstrative PRONOUN
PROimp	Impersonal PRONOUN
PROind	Indefinite PRONOUN
PROint	Interrogative PRONOUN
PROord	Ordinal PRONOUN
PROper	Personal PRONOUN
PROper.PROper	Contraction Personal PRONOUN + Personal PRONOUN
PROpos	Possessive PRONOUN
PROrel	Relative PRONOUN
RED	REDUNDANT
RES	RESIDUE
VERcjk	Conjugated VERB
VERinf	VERB in its Infinitive Form
VERppa	VERB in its Present Participle Form
VERppe	VERB in its Past Participle Form

Appendix 3: Conversion Table Cattex09 → UD

The organization of the CATTEX2009-min and UD tagsets, and the breakdown between part of speech categories and morphological feats are rather different from one another. The following table illustrates the corresponding feats for both the CATTEX2009-min (cattex-pos property) and UD tags. Each feat is a property (just like *ud-pos*) whose possible values are detailed in the table below. All UD feats are encoded in a single word property “ud-feats”, its value is quite similar to that of CONLL-U format (<http://universaldependencies.org/format.html>).

In certain instances, the CATTEX2009-min tag does not have a UD equivalent and so the corresponding UD tag is less precise:

- ADJqua (cattex-pos) corresponds to ADJ (ud-pos)
- ADVing (cattex-pos) corresponds to ADV (ud-pos)
- ADVsub (cattex-pos) corresponds to ADV (ud-pos)
- PROadv (cattex-pos) corresponds to ADV (ud-pos)
- PROimp (cattex-pos) corresponds to PRON (ud-pos)
- PONfbl (cattex-pos) corresponds to PUNCT (ud-pos)
- PONfrit (cattex-pos) corresponds to PUNCT (ud-pos)
- PONpdr (cattex-pos) corresponds to PUNCT (ud-pos)
- PONpga (cattex-pos) corresponds to PUNCT (ud-pos)
- PONpxx (cattex-pos) corresponds to PUNCT (ud-pos)

The “Definite” feature of the UD tagset has been used in a very limited number of cases. This use concerns:

- The definite article, which combines the properties *ud-pos* (“Det”) and *PronType* (“Art”) with the *Definite* property (“Def”) [ud-pos="Det" & ud-feats contains "PronType=Art\|Definite=Def"]
- The indefinite article, which combines the properties *ud-pos* (“Det”) and *PronType* (“Art”) with the *Definite* property (“Ind”) [ud-pos="Det" & ud-feats contains "PronType=Art\|Definite=Ind"]
- The determiner and pronoun *ledit*, which combines the properties *ud-pos* (“Det” or “Pron”) with the *Definite* property (“Com”); the value “Com” (for complex) is here used in a roundabout manner: [ud-pos="Det|Pron" & ud-feats contains "Definite=Com"].

The UD specification requires that contracted forms, such as French *del*, be tagged as two separate words. During the conversion, these forms were duplicated, the second being prefixed with an asterisk:

- *del* (cattex-pos="PRE.DETdef") is converted to *del* (ud-pos="PREP") **del* (ud-pos="DET")

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

The format of ud-feats word property has been optimized for CQL queries with the operator *contains*. For instance, the following query will find all the personal pronouns (including contracted forms):

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

To find indefinite articles, use the following query:

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

Appendix 4: List of Sources for Lemmas

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>

Revision Table

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)

25/06/18	Ian Byrd (IB)	Traduction en anglais
05/07/18	AL, CG	Révision de la traduction
15/05/19	AL	Synchronisation avec la version française 3.1