

# Fiche n°4 : Exploration textométrique du corpus issu de Gromoteur avec Le Trameur

---

## Sommaire

Préambule : .....	2
Etape n°1 : Le corpus à explorer.....	3
Etape n°2 : Ouvrir/charger le corpus dans le Trameur.....	5
Etape n°3 : Explorer le corpus dans le Trameur .....	8

## Préambule :

A l'issue de l'étape précédente (Fiche n°2), on dispose ici d'un **corpus chronologique** concaténant les contenus des unes de journaux en ligne récupérées à 2 dates différentes.

Les aspirations ont été réalisées: le 8 janvier 2015 et le 9 janvier 2015.

Le corpus a été nettoyé et préparé pour être exploré par le logiciel Le Trameur.

Le Trameur a un site Web : <http://www.tal.univ-paris3.fr/trameur/>

Pour installer Le Trameur sur vos machines personnelles, vous pouvez récupérer le programme d'installation ici :

<http://www.tal.univ-paris3.fr/trameur/setup-trameur-12.exe>

Le programme est installé sur les machines du LaboC : dans le dossier C:\Trameur

La documentation du logiciel est disponible ici :

<http://www.tal.univ-paris3.fr/trameur/leMetierLexicometrique.pdf>

Le **Trameur** : Programme de génération puis de gestion de la *Trame* et du *Cadre* d'un texte (i.e découpage en unité et partitionnement du texte : le métier textométrique) pour construire des opérations lexicométriques / textométriques (ventilation des unités, carte des sections, cooccurrence, spécificité, AFC...).

Le **Trameur** intègre en outre le programme [treetagger](#) : système d'étiquetage automatique des catégories grammaticales des mots avec lemmatisation. Il permet aussi de générer et de gérer des annotations multiples sur les unités du texte (et de traiter les niveaux d'annotations visés)

## Etape n°1 : Le corpus à explorer

Le corpus construit et préparé dans les fiches n°1 et n°2, il est structuré en différentes parties. La représentation graphique suivante (extrait) est produite par Le Trameur : elle donne à voir le système de parties qui sera disponible pour contraster les différentes parties du corpus :

<input type="checkbox"/>	<date(2005-01-08[pos=1])>	...
<input type="checkbox"/>	<page(lefigaro.fr[pos=2])>	...
<input type="checkbox"/>	<page(lemonde.fr[pos=8999])>	...
<input type="checkbox"/>	<page(lopinion.fr[pos=19359])>	...
<input type="checkbox"/>	<page(liberation.fr[pos=21603])>	...
<input type="checkbox"/>	<page(leparisien.fr[pos=29149])>	...
<input type="checkbox"/>	<page(humanite.fr[pos=38890])>	...
<input type="checkbox"/>	<page(la-croix.com[pos=44198])>	...
<input type="checkbox"/>	<page(lejdd.fr[pos=49517])>	...
<input type="checkbox"/>	<page(lecanardenchaine.fr[pos=53039])>	...
<input type="checkbox"/>	<page(nrpyrenees.com[pos=53779])>	...
<input type="checkbox"/>	<page(paris-normandie.com[pos=55692])>	...
<input type="checkbox"/>	<page(lalsace.fr[pos=60200])>	...
<input type="checkbox"/>	<page(petitbleu.fr[pos=63363])>	...
<input type="checkbox"/>	<page(lepopulaire.fr[pos=65806])>	...
<input type="checkbox"/>	<page(lapressedelamanche.fr[pos=73385])>	...
<input type="checkbox"/>	<page(presseocean.fr[pos=73818])>	...
<input type="checkbox"/>	<page(havre-presse.fr[pos=77365])>	...
<input type="checkbox"/>	<page(laprovence.com[pos=79610])>	...
<input type="checkbox"/>	<page(republicain-lorrain.fr[pos=84149])>	...
<input type="checkbox"/>	<page(republique-pyrenees.com[pos=86610])>	...
<input type="checkbox"/>	<page(larep.com[pos=86905])>	...
<input type="checkbox"/>	<page(sudouest.com[pos=93959])>	...
<input type="checkbox"/>	<page(letelegramme.com[pos=100164])>	...
<input type="checkbox"/>	<page(varmatin.fr[pos=104419])>	...
<input type="checkbox"/>	<page(ledauphine.com[pos=105860])>	...
<input type="checkbox"/>	<page(lavoixdunord.fr[pos=107741])>	...
<input type="checkbox"/>	<page(lyonne-republicaine.fr[pos=119625])>	...
<input type="checkbox"/>	<page(leberry.fr[pos=127108])>	...
<input type="checkbox"/>	<page(centre-presse.fr[pos=134792])>	...
<input type="checkbox"/>	<page(charentelibre.com[pos=143177])>	...

<input type="checkbox"/>	<page(corsematin.com[pos=149431])>	...
<input type="checkbox"/>	<page(courrierdelouest.fr[pos=155497])>	...
<input type="checkbox"/>	<page(ladepeche.com[pos=159145])>	...
<input type="checkbox"/>	<page(dna.fr[pos=165503])>	...
<input type="checkbox"/>	<page(dordogne.com[pos=168884])>	...
<input type="checkbox"/>	<page(lechorepublicain.fr[pos=172013])>	...
<input type="checkbox"/>	<page(pyrenees.com[pos=179540])>	...
<input type="checkbox"/>	<page(veilleil.fr[pos=182619])>	...
<input type="checkbox"/>	<page(havre-libre.fr[pos=186304])>	...
<input type="checkbox"/>	<page(havre-presse.fr[pos=188549])>	...
<input type="checkbox"/>	<page(lamarseillaise.fr[pos=190794])>	...
<input type="checkbox"/>	<page(lindependant.com[pos=196080])>	...
<input type="checkbox"/>	<page(jhm.fr[pos=200138])>	...
<input type="checkbox"/>	<page(lejdc.fr[pos=203250])>	...
<input type="checkbox"/>	<page(lemainelibre.fr[pos=210059])>	...
<input type="checkbox"/>	<page(midilibre.com[pos=213469])>	...
<input type="checkbox"/>	<page(lamontagne.fr[pos=217631])>	...
<input type="checkbox"/>	<page(nicematin.fr[pos=226471])>	...
<input type="checkbox"/>	<page(nordeclair.fr[pos=232376])>	...
<input type="checkbox"/>	<page(nordlittoral.fr[pos=236468])>	...
<input type="checkbox"/>	<page(lanouvellerepublique.fr[pos=244136])>	...

On ne voit ici que la structuration sur la première date d'aspiration, contenant toutes les pages aspirées. Une structuration similaire est disponible pour les autres dates d'aspiration.

Deux partitions seront utiles pour contraster les parties :

- la partition **DATE** qui va permettre d'analyser les données dans la chronologie des données aspirées
- la partition **PAGE** qui va permettre de contraster les données par journal (le journal *lefigaro* versus le journal *lemonde*)

## Etape n°2 : Ouvrir/charger le corpus dans le Trameur

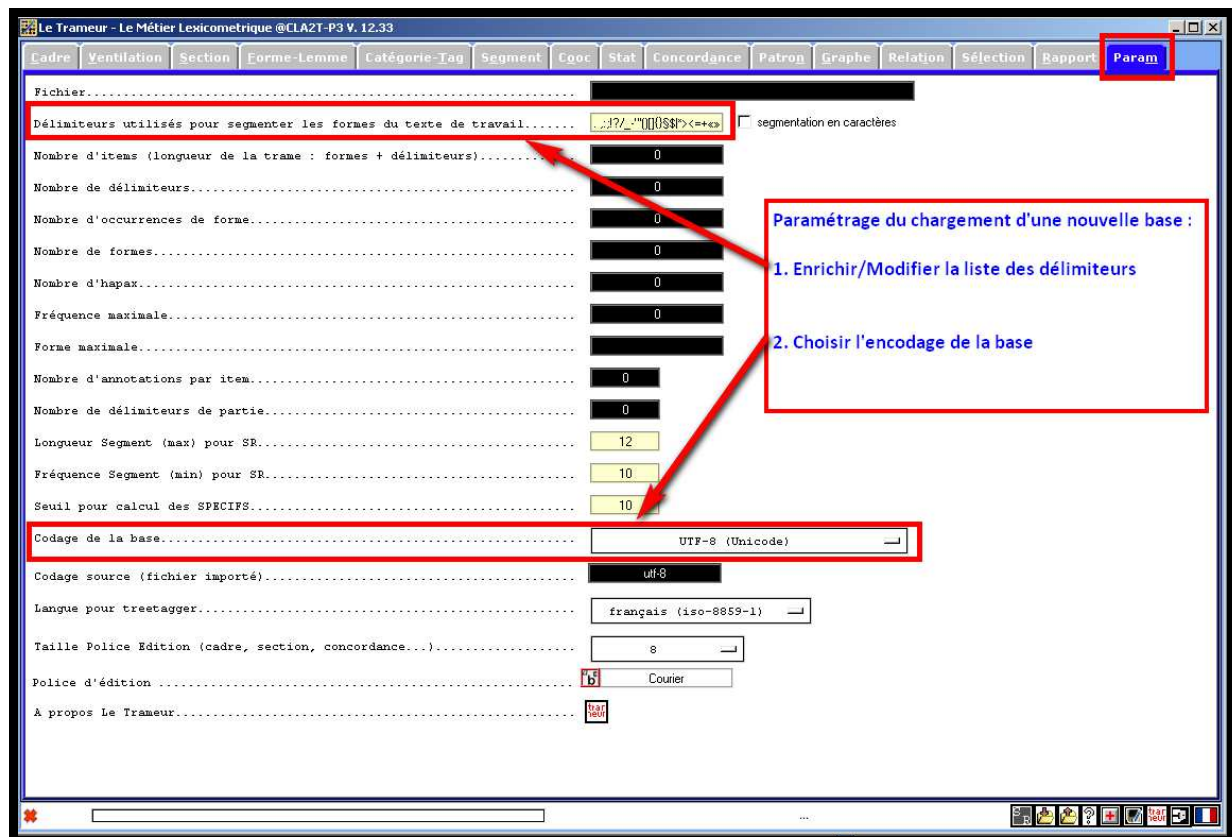
Pour charger un corpus dans le Trameur, il faut procéder en 2 temps :

1. Paramétrer le chargement
2. Réaliser le chargement

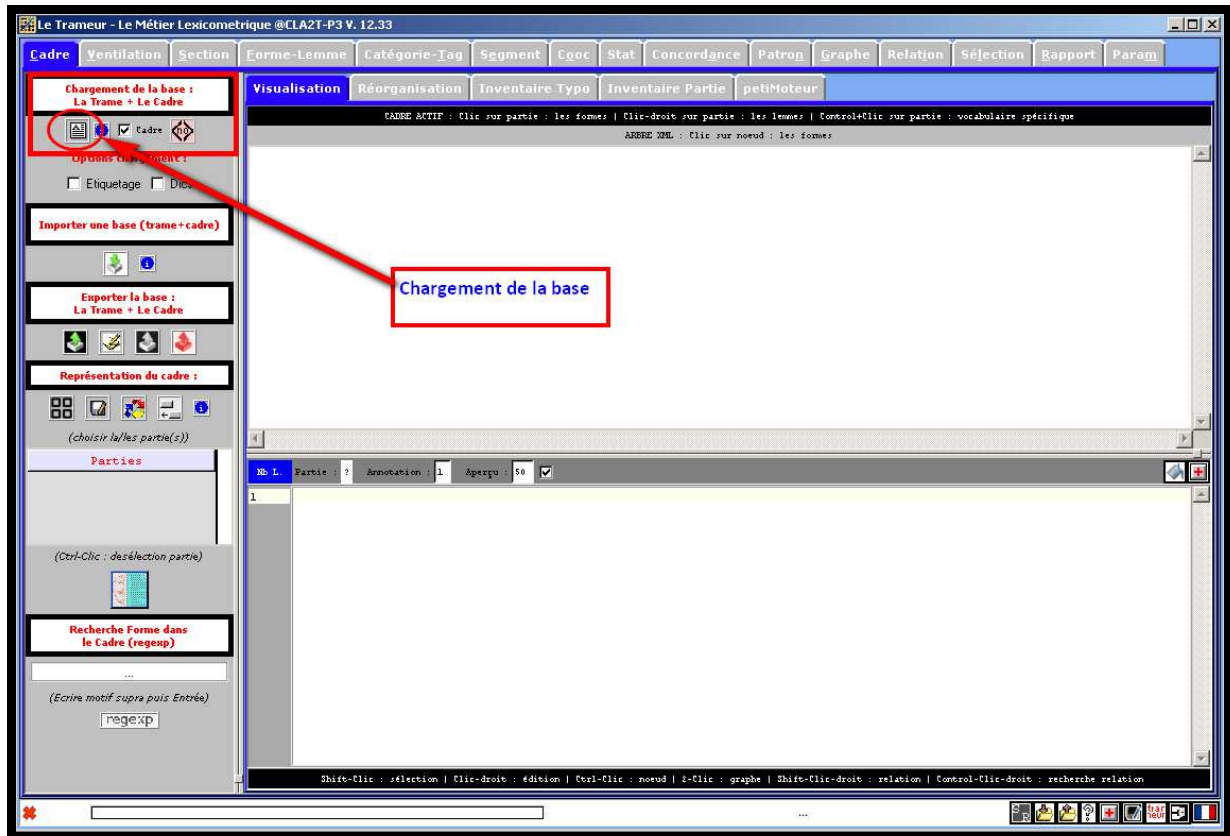
Le paramétrage consiste à spécifier notamment :

- L'encodage du fichier à charger : ici ce sera UTF-8
- Préciser ou enrichir la liste des caractères délimiteurs utilisés pour segmenter le texte en mots. Par défaut, le Trameur propose une liste de caractères délimiteur, il faut parfois en ajouter voire en enlever.

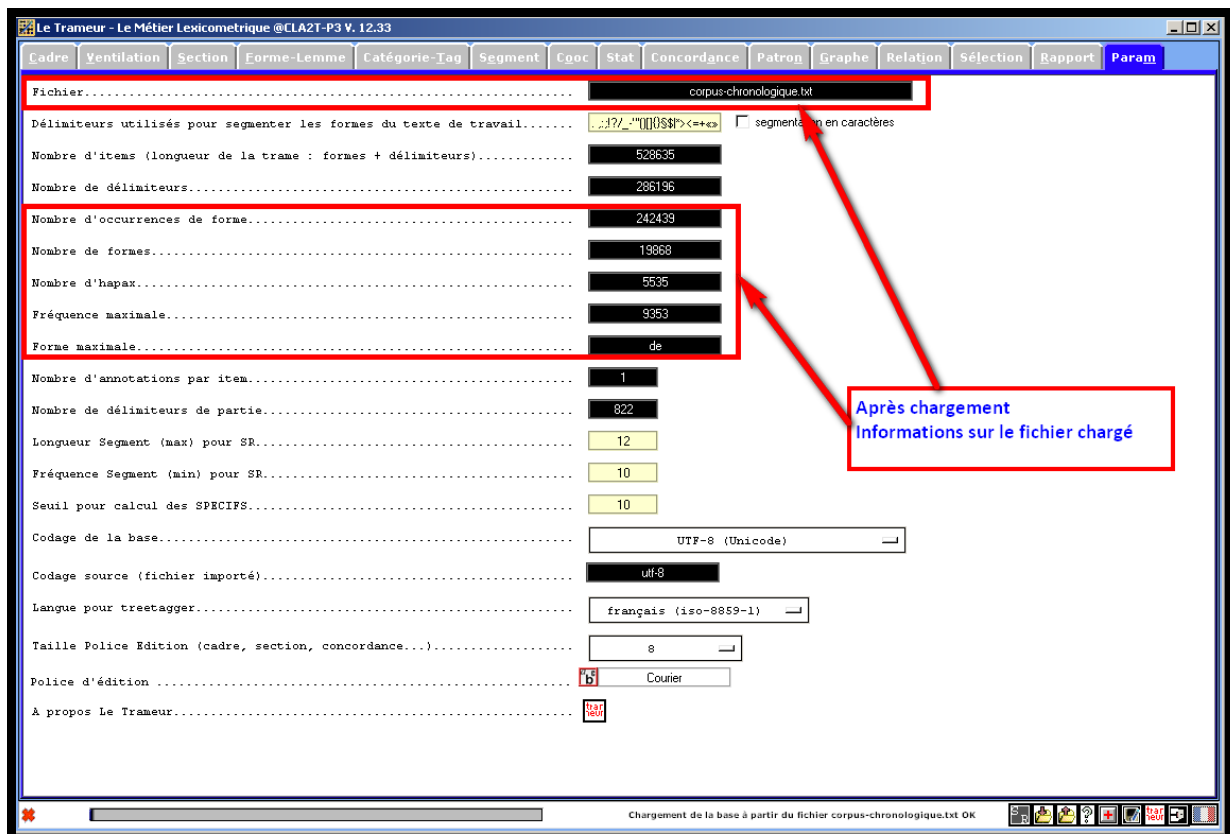
Ce paramétrage se fait dans l'onglet **PARAM** :



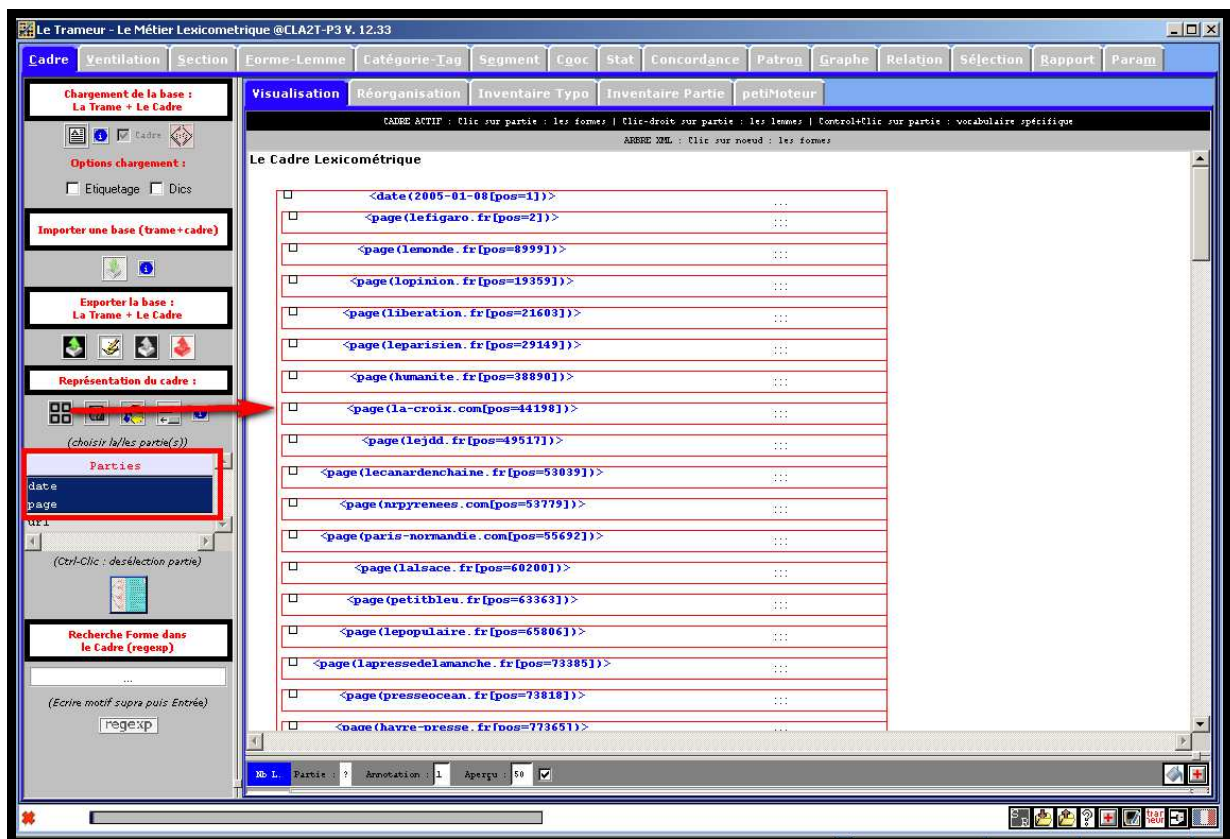
Une fois le paramétrage réalisé on peut charger le fichier. On se place dans l'onglet **CADRE** et on déclenche la sélection du fichier à charger :



Une fois le chargement effectué, le Trameur se positionne sur l'onglet **PARAM** et donne quelques informations quantitatives sur le fichier chargé : nombre de mots par exemple.



Pour vérifier la bonne structuration du fichier en parties, on peut construire une représentation graphique de la structure des parties disponibles, on se place sur l'onglet CADRE et on active le bouton construisant la représentation du **CADRE** de la base textométrique chargée :



## Etape n°3 : Explorer le corpus dans le Trameur

L'exploration à réaliser consiste à utiliser différents outils disponibles dans le Trameur. Ces opérations seront décrites en détail pendant le cours :

- Explorer le dictionnaire
- Distribution de mot(s) sur un graphique de ventilation (par parties)
- Distribution de mots sur une carte des sections
- Concordance de mot(s)
- Calcul de cooccurrence