

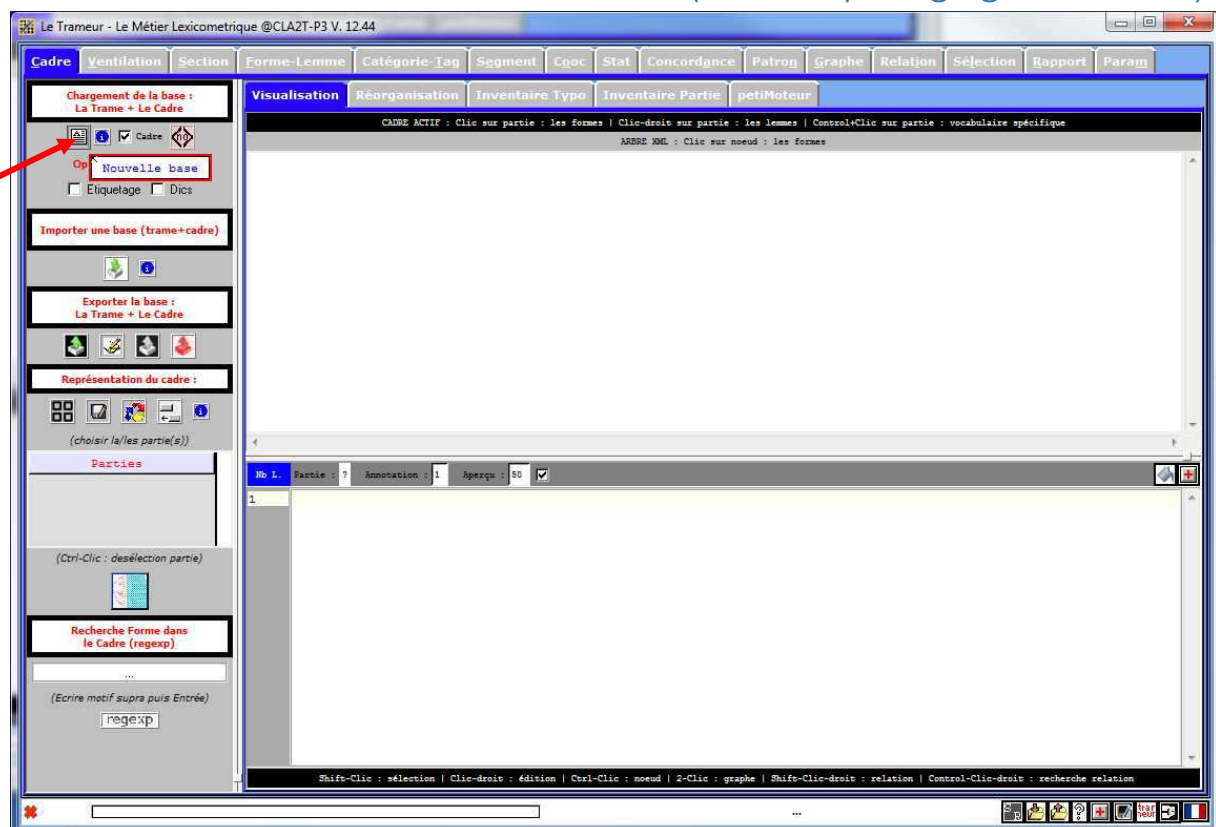
Fiche n°3 :

Le Trameur démarrage

Table des matières

Ouverture d'un fichier dans le Trameur (sans étiquetage grammatical).....	1
Visualisation du corpus	3
Cadre	3
Carte des sections	4
Recherche d'une forme.....	5
Etiquetage d'un corpus	6
Exportation de la base.....	8
Importation d'une base existante	9

Ouverture d'un fichier dans le Trameur (sans étiquetage grammatical)



Un peu de patience...



On vérifie l'encodage, on en profite pour regarder le nombre de mot :

The screenshot shows the 'Param' window of the 'Le Trameur' software. The window title is 'Le Trameur - Le Métier Lexicométrique @CLA2T-P3 V. 12.44'. The interface includes a menu bar with options like 'Cadre', 'Ventilation', 'Section', 'Forme-Lemme', 'Catégorie-Tag', 'Segment', 'Cooc', 'Stat', 'Concordance', 'Patron', 'Graphe', 'Relation', 'Sélection', 'Rapport', and 'Param'. The main area displays a list of parameters and their values:

Fichier.....	corpus_JanvFev15.txt
Délimiteurs utilisés pour segmenter les formes du texte de travail.....	.,!/?_-'"0123456789;<+*~> segmentation en caractères
Nombre d'items (longueur de la trame : formes + délimiteurs).....	1018182
Nombre de délimiteurs.....	571505
Nombre d'occurrences de forme.....	448577
Nombre de formes.....	30797
Nombre d'hapax.....	10059
Fréquence maximale.....	17357
Forme maximale.....	de
Nombre d'annotations par item.....	1
Nombre de délimiteurs de partie.....	1386
Longueur Segment (max) pour SR.....	12
Fréquence Segment (min) pour SR.....	10
Seuil pour calcul des SPECIFS.....	10
Codage de la base.....	UTF-8 (Unicode)
Codage source (fichier importé).....	utf-8
Langue pour treetagger.....	(fr) french (utf-8)
Taille Police Edition (cadre, section, concordance...).....	8
Police d'édition.....	15 Courier
À propos Le Trameur.....	Help

At the bottom of the window, a status bar indicates 'Chargement de la base à partir du fichier corpus_JanvFev15.txt OK'.

Visualisation du corpus

Cadre

On sélectionne les parties pertinentes (1), celles que l'on va étudier ; elles sont délimitées par des balises dans notre corpus. On vérifie que les parties « page » sont bien emboîtées dans les parties « date » et que les bordures des rectangles qui représentent les parties ne se croisent pas (3)...

Carte des sections

On sélectionne les parties pertinentes (1), puis on indique au logiciel que l'unité de base de notre corpus est la page et que celle-ci correspond à une partie (2). Dans la représentation obtenue, chacun des carrés est une page de notre corpus (4).

The screenshot shows the 'Le Trameur - Le Métier Lexicométrique @CLA2T-P3 V. 12.44' application window. The 'Section' menu is active, displaying options for loading and configuring the section map. Red annotations highlight the following elements:

- 1:** The 'page' option selected in the 'Parties' list.
- 2:** The 'page' option selected in the 'Chargement de la Carte des sections' dropdown menu.
- 3:** The 'Chargement de la Carte des sections' button.
- 4:** The grid of small squares representing the corpus pages.

The interface also shows a search bar for 'Recherche Forme sur la carte', a 'Spécificités sur Sections' section with 'BI-TEXT' options, and a 'Sélection Annotation' section with 'Forme' selected. The main window displays a grid of sections with dates (20150108, 20150109, 20150206, 20150208) and a status bar at the bottom indicating 'Nb. l. sections sélectionnées: 0', 'Nb. Sect.: 2', 'Annotation: 1', and 'Aperçu: 50'.

Recherche d'une forme

Si l'on se demande où est « Charlie », on peut le chercher dans le dictionnaire des formes (1), sélectionner la forme (2) et projeter la forme sur la carte des sections (3).

The screenshot shows the 'Forme-Lemme' window with the search results for 'Charlie'. The search term 'Charlie' is entered in the 'Recherche Forme(s) (Dico Forme):' field. The results table shows the following entries:

Fq	Forme	Delim	Num
48492	\n	+	1
46134	\$	+	1
17631	.	+	1
17357	de	+	1
16588	-	+	1
13213	,	+	1
12520	:	+	1
11081	'	+	1
9994	/	+	1
8798	à	+	1
8646	la	-	1
6141	le	-	1
5306	les	-	1
5228	et	-	1
5154	en	-	1
4678	des	-	1
4430	An	-	1

The 'Projection sur la carte des sections des formes sélectionnées' window shows the following results:

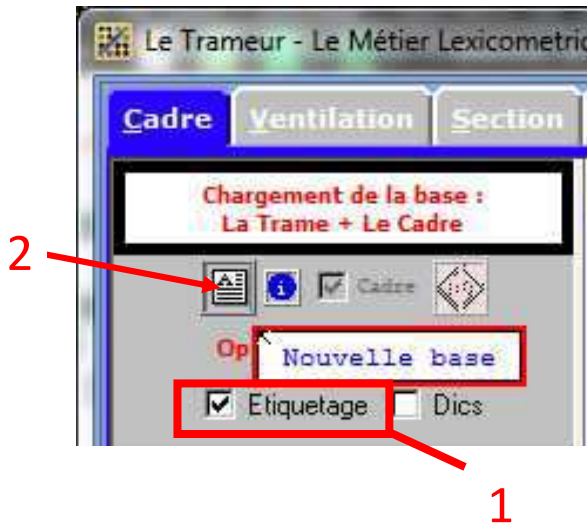
Fq	Forme	Delim	Num
14	Charlie	-	1
8	Charlie Hebdo	-	1
1	#JeSuisCharlie	-	1
9	#JeSuisCharlie	-	1
2	Charlie Coulibaly	-	1
1	satirique Charlie	-	1
2	Charlie's	-	1
1	#LesEtudiantsSontCharlie	-	1
2	#JeSuisCharliePO	-	1
58	Charlie	-	1
3	#JeSuisCharlie	-	1
48	#CharlieHebdo	-	1
3	Charlie...	-	1
1	"Charlie"	-	1
3	Charl...aHebdo	-	1

The 'Carte des sections' window shows a grid of sections with green squares indicating the presence of the form 'Charlie'. The search results are displayed in the bottom panel, showing the form 'Charlie' with a frequency of 3217.

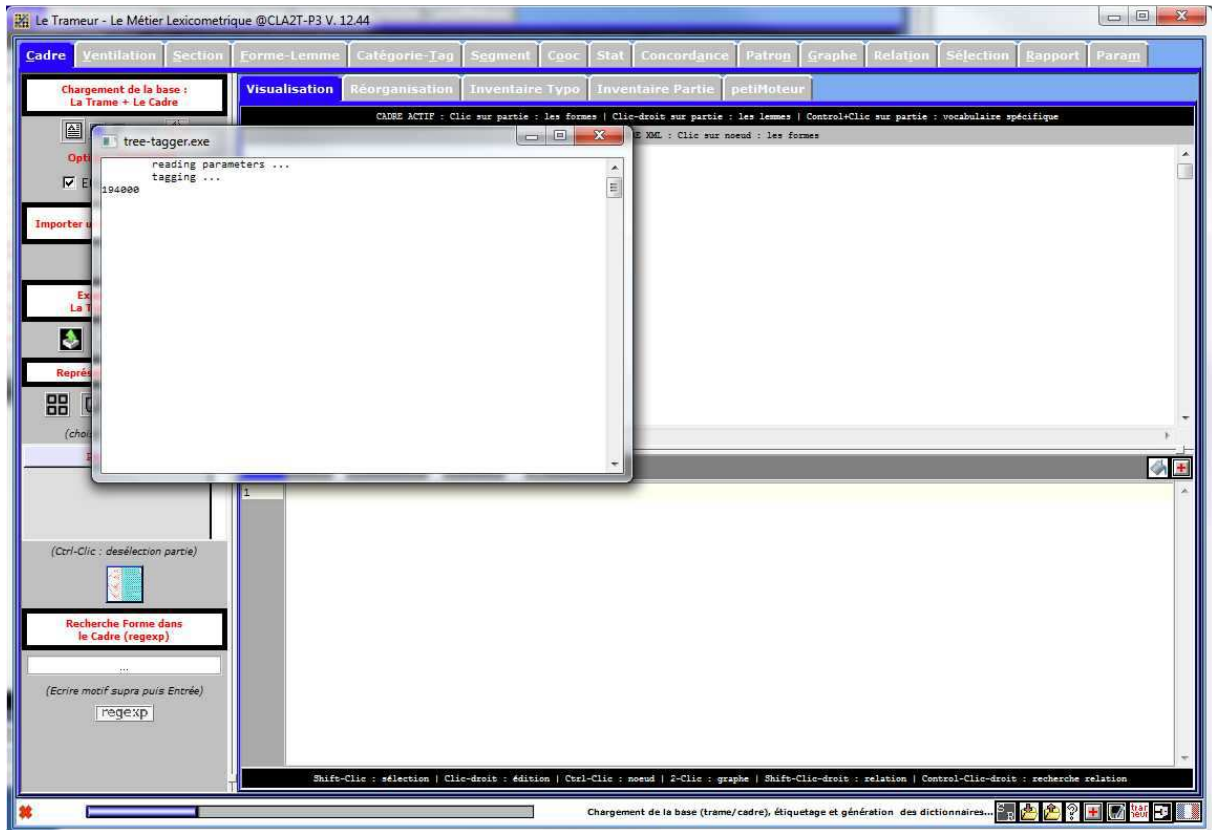
Les carrés verts représentent les pages qui contiennent la forme « Charlie ».

The screenshot shows the 'Carte des sections' window with a grid of sections. The search results are displayed in the bottom panel, showing the form 'Charlie' with a frequency of 3217. The search results are displayed in the bottom panel, showing the form 'Charlie' with a frequency of 3217.

Etiquetage d'un corpus



Le chargement peut être un peu long et une nouvelle fenêtre doit s'afficher temporairement.



À présent, on a aussi le dictionnaire des lemmes en plus du dictionnaire des formes.

The screenshot displays a software interface with two main dictionary panels. The left panel, titled 'Dictionnaire des formes', contains a table with columns: Fq, Forme, Delim, Num. The right panel, titled 'Dictionnaire des lemmes', contains a table with columns: Fq, Lemme, Num. Below these are sections for 'Formes avec motif' and 'Lemmes avec motif', each with a table of Fq, Forme, Delim, Num. The interface also includes search filters and a 'TagCloud' section on the left.

On a aussi la liste des étiquettes de catégories grammaticales utilisées. Notre corpus a donc maintenant 3 « couches » pour chaque mot :

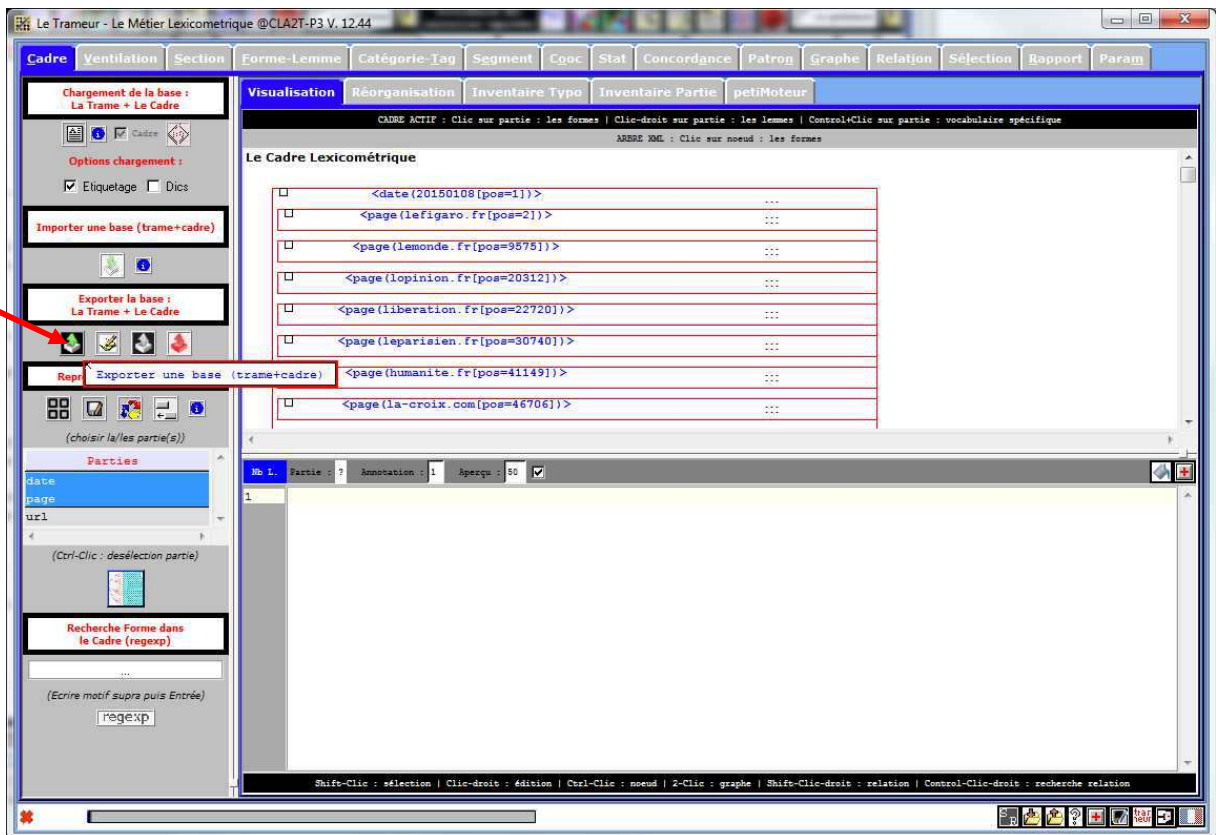
1. La forme : le mot tel qu'il apparaît dans le texte,
2. le lemme correspondant à la forme,
3. la catégorie grammaticale de la forme.

The screenshot displays a software interface with a 'Dictionnaire des catégories' table. The table has columns: Fq, Catégorie (tree-tagger), Num. The interface also includes an 'Importation d'une annotation' section and a 'Visualisation d'annotation' section.

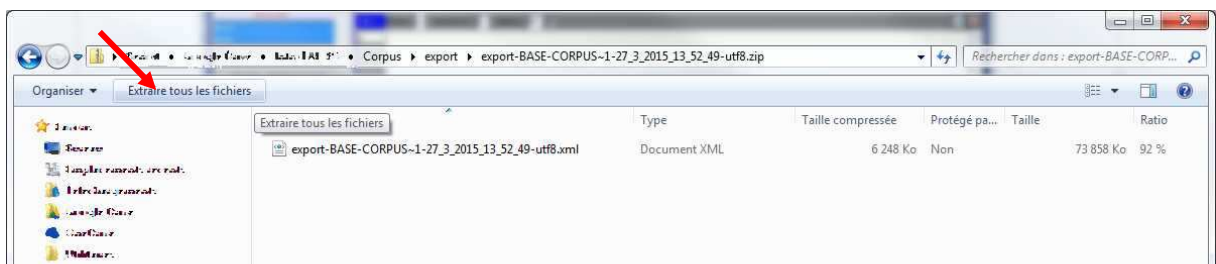
Fq	Catégorie (tree-tagger)	Num
123469	NAM	3
68794	NOM	3
48492	RETURN	3
47726	NUM	3
45398	PRP	3
33465	DET_ART	3
24621	ADJ	3
22706	VER_pres	3
12414	ABR	3
12300	PRP_det	3
10766	VER_pper	3
9768	VER_futu	3
7239	KON	3
5818	ADV	3
5298	PRO_PER	3
4400	VER_infi	3
3736	DET_POS	3
1775	VER_simp	3
1682	PRO_DEM	3
1350	PRO_REL	3
1070	PRO_IND	3
1021	VER_impf	3
732	VER_ppre	3
566	SYM	3
291	VER_cond	3
135	VER_subp	3
97	PUN	3
24	INT	3
14	PRO_POS	3
5	PUN_cit	3
3	VER_subi	3

Exportation de la base

Pour recharger plus rapidement la base la fois suivante, on peut exporter la base.



Un fichier au format XML est enregistré dans le sous-dossier « export » du dossier dans lequel se trouvait le corpus. Attention, le fichier est compressé (zippé) lors de l'enregistrement, il faudra l'extraire pour le réutiliser.



Importation d'une base existante

