

iTrameur

Serge Fleury

Le *Trameur* aka Le Métier Textométrique

URL: http://www.tal.univ-paris3.fr/trameur

Téléchargement : http://www.tal.univ-paris3.fr/trameur

Versions exécutables sous Windows/MacOsX disponibles en ligne Documentation en ligne sur la page du projet



URL: http://www.tal.univ-paris3.fr/trameur/iTrameur/

Version õen ligneö disponible directement dans un bon navigateur Cette application met en oeuvre des fonctionnalités disponibles dans *Le Trameur* Documentation dans lognelet « Aide » de iTrameur

Contact: serge.fleury@univ-paris3.fr

Manuel døutilisation

version 12 (et suivantes)

mise à jour : été 2019

CLESTHIA>SYLED>CLA²T

Sommaire

Ta	ble des Figuresble	7
2	Préambule	
	Trame et Cadre textométriques	. 15
3	Présentation générale	. 16
4	Installation.	
	4 .1 Installation du Trameur	
	4.2 Installation du Trameur (version console)	
5	Interface du Trameur (1975) Consore)	
6	Descriptif des icônes	
7	Chargement de la base : mise au jour du Cadre et de la Trame	21
,	7.1 Chargement døune nouvelle base	21
	7.1.1 Paramétrage	
	7.1.2 Chargement	
	7.1.3 Etiquetage du texte	
	7.1.4 Inventaire typographique	
	7.1.5 Inventaire des parties	
	7.2 Importer une base	. 28
8	Les formats døentrée (nouvelle base)	
	8.1 Format texte brut	
	8.2 Format dit « Lexico »	
	8.3 Format XML	. 31
	8.4 Formats pour corpus alignés	. 33
9	Les dictionnaires	. 35
	9.1 Les formes	. 35
	9.2 Les lemmes	. 37
	9.3 Les catégories	
	9.4 Les annotations importées	
10	•	
	10.1 Mise au jour de la Trame	
	10.2 Choix du niveau de løannotation à afficher	41
	10.3 Recherche de chaînes	
	10.4 Recherche « incrémentale » de chaînes	
	10.5 Visualisation des différentes annotations des items de la Trame	
	10.6 Edition et modification des items de la Trame	
11	10.7 Le vocabulaire spécifique doune partie	
11	La carte des sections : représentation cartographique de la <i>Trame</i>	
	11.1 La carte des sections : une vue globale de la Trame	
	11.2 La carte des sections : une vue partielle de la Trame	
	11.3 La carte des sections : masquage de sections	. 53
12	Accès à la <i>Trame</i> par la carte des sections : principales fonctionnalités	
	12.1 Edition de section	
	12.2 Recherche de chaînes	
	12.3 Recherche « incrémentale » de chaînes	
	12.4 Visualisation des différentes annotations des items de la Trame	. 57
	12.5 Edition et modification des items de la Trame	
	12.6 Le vocabulaire spécifique døune section	. 58
	12.7 Le vocabulaire spécifique doune sélection de sections de la carte	. 58
	12.8 Le vocabulaire spécifique des sections de la carte contenant un motif : recherche de cooccurrents	. 60
	12.9 Le vocabulaire spécifique des sections de la carte ne contenant pas un motif	. 62
	12.10 Fonctionnalités complémentaires sur la carte des sections	
13	Concordances	
13	13.1 Modes de déclenchement døune concordance	
	13.2 Paramétrage døune concordance	
	13.3 Editeur de concordance	
1.4	Marquage des annotations de la <i>Trame</i>	
14		
	14.1 Annotation « simple »	
1 ~	14.2 Annotation de type « relation »	
	Vocabulaire spécifique	
16	Graphiques de ventilation	
	16.1 Ventilation par partie døun item	
	16.2 Ventilation par partie døune sélection døtems	
	16.3 Courbe døaccroissement du vocabulaire	
	16.5 Ventilation døune sélection døtems associés à différents niveaux døannotation	
	16.4 Ventilation døune sélection døtems sur une sélection de parties	. 84

Le *Trameur* aka Le métier textométrique

17	Segments répétés	
18	Cooccurrences ó Poly-Cooccurrences	88
	18.1 Paramétrage et lancement des modules	
	18.2 Mise en ò uvre du calcul	
	18.3 Cooccurrents et segments répétés	
	18.4 Sélection des chemins de cooccurrences à examiner	
	18.5 Fonctionnalités complémentaires sur les résultats produits	94
10	Extraction de patrons et graphes de mots	
19		
	19.1 Extraction de patrons (recherche sur un niveau déannotation)	
	19.2 Extraction de patrons (avec croisement døannotations)	
	19.3 Actions sur les séquences extraites à partir døun patron donné	
	19.4 Graphes de mots avec Le Trameur	
	19.5 Graphes de mots avec Pajek	
20	Calculs textométriques complémentaires	
	20.1 Tableau lexical døune partition	
	20.2 AFC : Analyse Factorielle des Correspondances	108
	20.3 Intégration de modules R (phase 1)	111
	Mise en ò uvre des modules R dans le Trameur	112
	Illustration sur la base « Etats finaux BROUILLONS »	113
	20.4 Corpus chronologiques : barycentre temporel, coefficient de Von Neumann	
21	Sélection. Gestionnaire de sélection	
	21.1 Sélection de positions sur la Trame	
	21.1.1 Sélection directe à partir de la <i>Trame</i>	
	21.1.2 Sélection directe à partir de la <i>Trame</i> 21.1.2 Sélectionner dans les dictionnaires	
	21.1.2 Selectionner dans les dictionnaires 21.2 Traitements døune sélection dans le Gestionnaire de sélection	
	21.3 Opérations sur sélections : concordance, ventilation, modification annotation, nouvelle annotation	
	21.3.1 Concordance	
	21.3.2 Ventilation	
	21.3.3 Modification des annotations sur les items sélectionnés	
	21.3.4 Ajout døune nouvelle annotation sur les items sélectionnés	
	21.4 Méthodes de génération de sélection via le Gestionnaire	
	21.4.1 Génération et sélection døune liste de formes	
	21.4.2 Génération et sélection døune liste de segments	134
	21.4.3 Génération et sélection multicritère	
	21.4.4 Génération døune liste døtems spécifiques	135
	21.5 Sélection de parties (comme ensembles de sélection de positions sur la Trame)	
	21.5.1 Calcul de segments	
	21.5.2 Calcul de spécificité	
	21.5.3 Extraction de patron	
	21.5.4 Calcul de cooccurrents	
	21.5.5 Calcul de cooccurrents sur corpus chronologique	
	21.6 Gestion des sélections : sauvegarde et chargement	150
	21.6.1 Sauvegarde døune Sélection	151
22	21.6.2 Chargement døune Sélection	
22	Importation døannotation sur une Trame	15/
23	Ajout/modification døannotation sur une <i>Trame</i>	
	23.1 Ajout/modification døannotation par sélection døtems dans un dictionnaire	
	23.1.1 Ajout døannotation	
	23.1.2 Modification døannotation	
	23.2 Ajout/fusion de couche d\(\phi\)annotation	
	23.3 Ajout/modification dønnotation dans une base exportée	
	23.4 Annotation de sections	
24	Des nuages de d	172
	Relations de dépendance entre les items de Trame (via leurs annotations)	174
	25.1 Extrait de la Trame : « un jour, un petit ange de sapin de í »	175
	25.2 Recherche de relations	
	25.2.1 Sélection des relations à afficher	
	25.2.2 Graphe des relations	
	25.3 Sélection des relations avec filtrage sur le nò ud source de la relation.	
	25.4 Sélection des relations avec filtrage sur la position du dépendant	
	25.5 Actions sur les nò uds du graphe : retour en contexte	
	25.6 Rechercher dans un graphe de dépendance	
	25.7 Affichage de relation en contexte	
	25.8 Recherche de relation en contexte	
	25.9 Graphe de relations en contexte	191
	25.10 Visualisation de toutes les relations dans un contexte donné	
	25.11 Recherche de collocation : spécificités sur relation	194

CLESTHIA>SYLED>CLA²T

	25.12 Recherche source-cible døune relation.	196
26	Le Cadre: visualisation, réorganisation	199
27	Exporter une base textométrique	202
28	Exportation du Cadre et de la Trame	208
	28.Î Fichier de travail	208
	28.2 Fichiers døexport	208
29	Le rapport : enregistrement des résultats produits	215
	29.1 Ajouter un élément au rapport	
	29.2 Ajouter un élément externe au rapport	216
	29.2 Enregistrer le rapport	
30	Fonctionnalités complémentaires	220
	30.1 Graphes døannotation	220
	30.1.1 Intégration dans le <i>Trameur</i> døun module døimportation døun graphe døannotations sur la <i>Trame</i>	220
	30.1.1 Génération de graphes døannotations	223
	30.1.3 Exploitation des graphes déannotation	227
	30.2 Base textométrique de textes alignés	228
	30.2.1 Mise au jour døalignement dans la carte des sections	228
	30.2.2 Exploration doune base textométrique de textes alignés	231
	30.2.3 Outils et méthodes dœxploration døun alignement	249
	30.3 Annotation de données orales	255
	30.4 Module petiMoteur	257
	30.5 MOTIFS : repérage et sélection de motifs multi-annotés	258
	Définitions	
	Mise en ò uvre dans le Trameur	258
	Ecriture des motifs dans un fichier	259
	Chargement du fichier contenant les motifs	260
	Affichage du résultat dans le Gestionnaire de Sélection	
	Visualisation en contexte du "motif"	262
31	Le Trameur en mode console	263
	31.1 Fichier de travail:	263
	31.2 Le Cadre :	264
	31.3 La Trame :	265
	31.4 La Trame « annotée » (si treetagger est invoqué)	
32	Bibliographie	269
33	Glossaire	271

Table des Figures

Figure 1 : <i>Le Trameur</i> Fenêtre principale	. 19
Figure 2 : Le Trameur Onglet PARAM	. 22
Figure 3 : Le Trameur Onglet CADRE	. 23
Figure 4 : Le Trameur Onglet PARAM : choix de langue pour Treetagger	. 23
Figure 5a : <i>Le Trameur</i> Inventaire Typographique	. 24
Figure 5b : Le Trameur Inventaire Typographiquesur petitMoteur (1)	
Figure 5c : <i>Le Trameur</i> Inventaire Typographiquesur petitMoteur (2)	
Figure 5d : Le Trameur Inventaire des parties	
Figure 5e : Le Trameur Inventaire des parties sur Gromoteur	
Figure 6 : Le Trameur Onglet FORME	
Figure 7 : Le Trameur Actions sur Items sélectionnés à partir des dictionnaires	. 36
Figure 8 : Le Trameur Onglet FORME	
Figure 9 : Le Trameur Onglet CADRE, base chargée	
Figure 10 : <i>Le Trameur</i> Onglet CADRE, <i>Cadre</i> généré	
Figure 11 : Le Trameur Onglet CADRE, contenu textuel d'une partie (les mots)	
Figure 12 : Le Trameur Onglet CADRE, contenu textuel d'une partie (les lemmes)	
Figure 13 : Le Trameur Onglet CADRE, chargement d'une représentation arborescente d'une base au format	
XML (1)	
Figure 14 : Le Trameur Onglet CADRE, chargement d'une représentation arborescente d'une base au format	
XML (2)	
Figure 15 : Le Trameur Affichage d'une partie du CADRE (annotation affichée = 1)	
Figure 16 : Le Trameur Affichage d'une partie du CADRE (annotation affichée = 2)	
Figure 17 : Le Trameur Affichage d'une section (annotation affichée = 3)	
Figure 18 : Le Trameur Onglet CADRE, recherche de chaînes	
Figure 19 : Le Trameur Onglet CADRE, recherche incrémentale de chaînes	
Figure 20 : Le Trameur Onglet CADRE, visualisation détaillée d'un item	
Figure 21 : <i>Le Trameur</i> Onglet CADRE, édition d'un item	
Figure 22 : Le Trameur Onglet CADRE, modification d'un item	
Figure 23 : Le Trameur Onglet CADRE, le vocabulaire spécifique d'une partie	
Figure 24 : Le Trameur Onglet SECTION, texte brut avec caractère « délimiteur de sections »	
Figure 25 : Le Trameur Onglet SECTION, texte structuré avec caractère « délimiteur de sections »	
Figure 26 : Le Trameur Onglet SECTION, construction de la carte des sections (3)	
Figure 27 : Le Trameur Onglet SECTION, édition de la Trame par une section	
Figure 28 : Le Trameur Onglet SECTION, paramétrage de l'affichage	
Figure 29 : Le Trameur Onglet SECTION, texte TEI sans caractère « délimiteur de sections » explicite	. 51
Figure 30 : Le Trameur Onglet SECTION, sections définies implicitement par un élement du Cadre (avec	
affichage d'une autre partie : lg)	. 52
Figure 31 : Le Trameur Onglet SECTION, sections définies implicitement par un élement du Cadre (avec	
affichage d'une autre partie : body)	
Figure 32 : Le Trameur Onglet SECTION, masquage de sections (1)	
Figure 33 : Le Trameur Onglet SECTION, masquage de sections (2)	
Figure 34 : <i>Le Trameur</i> Onglet SECTION, édition de section	
Figure 35 : Le Trameur Onglet SECTION, édition de section, sélection des annotations	
Figure 36 : Le Trameur Onglet SECTION, recherche de chaînes	
Figure 37 : Le Trameur Onglet SECTION, recherche de chaînes	
Figure 38 : Le Trameur Onglet SECTION, les mots spécifiques d'une section	
Figure 39 : Le Trameur Onglet SECTION, les mots spécifiques d'une sélection de sections	
Figure 40 : Le Trameur Onglet SECTION, items spécifiques des sections contenant un motif (cooccurrents)	
Figure 41: Le Trameur Onglet SECTION, items spécifiques des sections ne contenant pas un motif	
Figure 42 : Le Trameur Onglet CONCORDANCE, affichage édition + affichage tri	
Figure 43 : Le Trameur Onglet CONCORDANCE (forme)	
Figure 44 : Le Trameur Onglet CONCORDANCE (lemme)	
Figure 45 : Le Trameur Onglet CONCORDANCE (catégorie)	
Figure 46 : Le Trameur Onglet CONCORDANCE (forme)	. 67

CLESTHIA>SYLED>CLA²T

Figure 47: Le Trameur Onglet CONCORDANCE (rechargement)	
Figure 48 : Le Trameur Onglet CONCORDANCE, édition et modification d'un item pivot	68
Figure 49 : Le Trameur Editeur CADRE	69
Figure 50 : Le Trameur Editeur SECTION	
Figure 51 : Le Trameur Editeur CONCORDANCE	
Figure 52 : Le Trameur Sélection des annotations à colorier (1)	
Figure 53 : Le Trameur Sélection des annotations à colorier (2)	
Figure 54 : Le Trameur Marquage des annotations sur les items édités (1)	
Figure 55 : Le Trameur Marquage des annotations sur les items édités (2)	
Figure 56 : Le Trameur Marquage des annotations sur les items édités (3)	
Figure 57 : Le Trameur Sélection d'une nouvelle annotation à colorier	
Figure 58 : Le Trameur Marquage des annotations sur les items édités	
Figure 59: Le Trameur Edition d'une annotation de type « relation »	
Figure 60 : Le Trameur Sélection d'une relation à colorier	
Figure 61 : Le Trameur Marquage des annotations de type « relation » sur les items édités	
Figure 62 : Le Trameur Sélection d'annotation pour calcul de spécificités	76
Figure 63 : Le Trameur Opérations disponibles sur les items issus d'un calcul de spécificités	
Figure 64 : Le Trameur Opérations disponibles sur les items triables issus d'un calcul de spécificités	78
Figure 65: Le Trameur Onglet VENTILATION paramétrage d'une ventilation	
Figure 66: Le Trameur Onglet VENTILATION affichage d'une ventilation (mode ligne)	80
Figure 67: Le Trameur Onglet VENTILATION mode regrexp	
Figure 68 : Le Trameur Ventilation d'une sélection d'items	
Figure 69: Le Trameur Onglet VENTILATION, courbe d'accroissement du vocabulaire	
Figure 70 : Le Trameur Onglet VENTILATION, ventilation croisée d'annotations	
Figure 71: Le Trameur Onglet VENTILATION, ventilation d'une sélection d'items (1)	
Figure 72: Le Trameur Onglet VENTILATION, ventilation d'une sélection d'items (2)	
Figure 73: Le Trameur Onglet VENTILATION, ventilation d'une sélection d'items (3)	
Figure 74 : Le Trameur Onglet SEGMENT	
Figure 75 : Le Trameur Onglet SEGMENT	
Figure 76: Le Trameur Onglet COOC, importation d'une stop-liste	
Figure 77 : Le Trameur Onglet COOC, édition de la stop-liste	
Figure 78 : Le Trameur Onglet COOC, étape n°1 : module Cooccurrents (2 types d'unités contextuelle	
Figure 79 : Le Trameur Onglet COOC, étape n°2 : module Polycooccurrents	
Figure 80 : Le Trameur Onglet COOC, un pôle de type SR	93
Figure 81 : Le Trameur Onglet COOC, Coocs + SR	
Figure 82: Le Trameur Onglet COOC, Filtrage des polycoocs	
Figure 83 : Le Trameur onglet PATRON	
Figure 84 : Le Trameur zone d'enregistrement d'un patron	
Figure 85 : Le Trameur écriture d'un patron	
Figure 86 : Le Trameur liste des séquences associées au patron	
Figure 87 : Le Trameur écriture et paramétrage d'un patron	
Figure 88 : Le Trameur liste des séquences associées au patron	
Figure 89 : Le Trameur Extraction de patron (croisement d'annotation)	
Figure 90 : Le Trameur Extraction de patron (croisement d'annotation)	
Figure 91 : Le Trameur Extraction de patron (croisement d'annotation)	
Figure 92 : Le Trameur Paramétrage de l'extraction du patron visé et de l'affichage des séquences ex	
Figure 93 : Le Trameur Affichage des séquences extraites sur la base du patron visé	
Figure 94 : <i>Le Trameur</i> Ventilation-partie d'une liste de séquences sélectionnées	
Figure 95 : Le Trameur Concordance d'une liste de séquences sélectionnées	
Figure 96 : <i>Le Trameur</i> Ventilation-section d'une liste de séquences sélectionnées	
Figure 97 : Le Trameur génération du graphe de mots complet associé à un patron	
Figure 98 : Le Trameur génération du graphe de mots associé à un patron et contenant un motif	
Figure 99 : Le Trameur export au format Pajek	
Figure 100 : Le Trameur export NOM ADJ au format Pajek sur Duchn	
Figure 101: Le réseau NOM ADJ de Duchn via Pajek	
Figure 102 : Calculs textométriques complémentaires Figure 103 : Sélection du niveau d'annotation pour calculs lexicométriques complémentaires	
rigure 105. Selection du niveau à annotation pour calculs lexicometriques complementaires	106

Le *Trameur* aka Le métier textométrique

Figure 104 :	Paramétrage TFG ; choix des paramètres de calcul, de l'annotation et de la partition	107
Figure 105 :	: Tableau lexical (annotation n°1 : forme)	107
Figure 106:	: Tableau lexical (annotation n°3 : catégorie)	108
	: AFC	
	AFC (position)	
	AFC (sélection)	
	AFC (affichage des mots spécifiques de la partie)	
	Les facteurs de l'AFC	
_	PCLC d'une partition	
•	Modules R (1)	
J	: Modules R (2)	
•	: AFC via R	
•	ACP via R	
	CAH via R	
	Sélection des items pour les modules R	
	: Items sélectionnés dans la <i>Gestionnaire de Sélection</i>	
	: Paramétrage R pour travailler avec le <i>Gestionnaire de Sélection</i>	
	AFC via R sur le contenu du Gestionnaire de Sélection	
	CAH via R sur le contenu du <i>Gestionnaire de Sélection</i>	
	Barycentre temporel et coefficient de Von Neumanns salariés versus travailleurs dans corpus CFDT 1973-2006	
	Le Trameur onglet SELECTION	
	Le Trameur sélection sur la Trame (onglet CADRE)	
	Le Trameur sélection sur la Trame (onglet CADRE)	
	Le Trameur sélection sur une concordance (onglet CONCORDANCE)	
	Le Trameur sélection sur dictionnaires (onglets FORME, LEMME, CATEGORIE)	
	Le Trameur Gestionnaire de Sélection	
	Le Trameur Gestionnaire de Sélection (positions stockées dans le Gestionnaire de Sélection)	
	Le Trameur sélection directe de position dans l'éditeur	
	Le Trameur Gestionnaire de Sélection : les positions enregistrées	
	Le Trameur Gestionnaire de Sélection : sélection parmi les positions enregistrées	
	Le Trameur concordance d'une sélection de positions	
•	Le Trameur ventilation d'une sélection de positions	
_	Le Trameur modification des annotations d'une sélection de positions	
Figure 138 :	Le Trameur modification des annotations d'une sélection de positions	130
	Le Trameur modification des annotations d'une sélection de positions	
	Le Trameur insertion d'une nouvelle annotation sur une sélection de positions	
Figure 141 :	Le Trameur insertion d'une nouvelle annotation sur une sélection de positions	132
Figure 142 :	Le Trameur insertion d'une nouvelle annotation sur une sélection de positions	132
	Le Trameur insertion d'une nouvelle annotation sur une sélection de positions	
Figure 144 :	Le Trameur génération d'une liste de formes à sélectionner	133
_	Le Trameur sélection d'une liste de formes	
_	Le Trameur génération d'une liste de segments à sélectionner	
	Le Trameur génération d'une liste d'items par équation de recherche	
	Le Trameur sélection de parties	
	Le Trameur calcul de segment sur une sélection de parties	
_	Le Trameur calcul de segment sur une sélection de parties	
	Le Trameur calcul de spécificités sur une sélection de parties	
_	Le Trameur extraction de patron sur une partie sélectionnée	
	Le Trameur extraction de patron sur une partie sélectionnée	
	Le Trameur cooccurrents/polycooccurrents sur une partie sélectionnée (1)	
_	Le Trameur cooccurrents/polycooccurrents sur une partie sélectionnée (2)	
	Le Trameur cooccurrents/polycooccurrents sur une partie sélectionnée (3)	
	Cooccurrents, approche locale / globale / chronologique	
	Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (1)	
_	: Cooccurrents, approche locale	
LIRUIE TOO :	. Le Trumeur Cooccurrents sur une partie selectionnée dans un corpus chroniogique (2)	14b

CLESTHIA>SYLED>CLA²T

Figure 161 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (3)	147
Figure 162 : Cooccurrents, approche globale	
Figure 163 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (4)	148
Figure 164 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (5)	149
Figure 165 : Cooccurrents, approche chronologique	150
Figure 166 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (6)	150
Figure 167 : Le Trameur Gestion des sélections : sauvegarde et chargement	151
Figure 168 : Le Trameur sauvegarde d'une sélection (1)	
Figure 169 : Le Trameur sauvegarde d'une sélection (2)	
Figure 170 : Le Trameur sauvegarde d'une sélection (3)	
Figure 171 : Le Trameur arborescence XML d'une sélection	
Figure 172 : Le Trameur chargement d'une sélection (1)	
Figure 173 : Le Trameur chargement d'une sélection (2)	
Figure 174 : Le Trameur activation d'une sélection nommée (1)	
Figure 175 : Le Trameur activation d'une sélection nommée (2)	
Figure 176 : Le Trameur inhibation d'une sélection nommée	156
Figure 177 : Le Trameur Importation d'annotation, nommage et sélection du niveau d'annotation	
Figure 178 : Le Trameur correction d'item (annotation importée)	
Figure 179 : Le Trameur concordance d'item (annotation importée)	
Figure 180 : Le Trameur ventilation d'item (annotation importée)	
Figure 181 : Le Trameur édition de la Trame par les formes	
Figure 182 : Le Trameur ajout d'annotation sur la Trame via une forme sélectionnée	
Figure 183 : <i>Le Trameur</i> modification d'annotation sur la <i>Trame</i> via une forme sélectionnée	
Figure 184 : Le Trameur ajout / fusion d'annotation	
Figure 185 : Le Trameur annotation de sections	
Figure 186 : Le Trameur annotation de sections (0 vs 1)	
Figure 187 : Le Trameur annotation de sections sur bases alignées	
Figure 188 : Le Trameur annotation de sections sur bases alignées : recherche de l'absence	
Figure 189 : Le Trameur annotation de sections sur bases alignées : mise au jour de l'absence	
Figure 190 : Le Trameur un nuage d'annotations sur la Trame	
Figure 191 : Le Trameur un nuage d'annotations sur la Trame avec contextes	
Figure 192 : Le Trameur paramétrage du nuage d'annotations	
Figure 193 : Le Trameur recherche de relations de dépendance via l'onglet GRAPHE	
Figure 194 : Le Trameur paramétrage de la recherche de dépendance sur la Trame	
Figure 195 : Le Trameur affichage du graphe des dépendances	
Figure 196 : Le Trameur paramétrage de la recherche de dépendance avec filtrage	
Figure 197 : Le Trameur affichage du graphe des dépendances (1)	
Figure 198 : Le Trameur affichage du graphe des dépendances (2)	
Figure 199 : Le Trameur affichage du graphe des dépendances (3)	
Figure 200 : Le Trameur paramétrage de la recherche de dépendance avec filtrage	
Figure 201: Le Trameur affichage du graphe des dépendances	
Figure 202 : Le Trameur affichage de la concordance sur un des nœuds du graphe des dépendances	
Figure 203 : Le Trameur affichage de la concordance sur tous les nœuds du graphe des dépendances	
Figure 204 : Le Trameur Concordance sur tous les nœuds du graphe des dépendances (1)	
Figure 205 : Le Trameur Concordance sur tous les nœuds du graphe des dépendances (2)	
Figure 206: Le Trameur Coloration d'annotation dans le concordancier (1)	
Figure 207 : Le Trameur Coloration d'annotation dans le concordancier (2)	
Figure 208 : Le Trameur Graphe de la relation OBL	
Figure 209 : Le Trameur Recherche dans un graphe de relation (1)	
Figure 210 : Le Trameur Recherche dans un graphe de relation (2)	
Figure 211 : Le Trameur Recherche dans un graphe de relation (2)	
Figure 211 : Le Trameur Recherche dans un graphe de relation (3)	
Figure 213 : Le Trameur acces aux dependances via la Trame	
Figure 214 : Le Trameur paramétrage de l'affichage des dépendances sur un item de la Trame (2)	
Figure 215 : Le Trameur affichage de la dépendance sur la Trame (1)	
Figure 217 : Le Trameur recherche de dépendance en contexte sur un item de la Trame	
Tigure 217. Le Trumeur recherche de dépendance en contexte sur un item de la Trume	τοὄ

Le *Trameur* aka Le métier textométrique

_	218 : Le Trameur paramétrage de la recherche de dépendance sur un item de la Trame (1)	
_	219 : Le Trameur affichage des dépendances via un graphe (1)	
	220 : Le Trameur affichage des dépendances en contexte sur la Trame	
	221 : Le Trameur paramétrage de la recherche de dépendance sur un item de la Trame (2)	
_	222 : Le Trameur affichage des dépendances via un graphe (2)	
Figure	223 : Le Trameur affichage d'un graphe de dépendances en contexte (1)	. 191
Figure	224 : Le Trameur affichage d'un graphe de dépendances en contexte (2)	. 191
Figure	225 : Le Trameur affichage global des dépendances en contexte (1)	. 192
Figure	226 : Le Trameur affichage global des dépendances en contexte (2)	. 192
Figure	227 : Le Trameur affichage global des dépendances en contexte (3)	. 193
Figure	228 : Le Trameur affichage global des dépendances en contexte (4)	. 193
	229 : Le Trameur Recherche de collocation (1)	
	230 : Le Trameur Recherche de collocation (2)	
	231 : Le Trameur Recherche de collocation (3)	
	232 : Le Trameur liste de relations	
	233 : Le Trameur liste de relations (cible)	
	234 : Le Trameur liste de POS en relation	
	235 : Le Trameur les dépendants d'un item (forme, lemme ou catégorie)	
_	236 : <i>Le Trameur</i> les dépendants d'une liste d'items (forme, lemme ou catégorie)	
_	237 : Le Trameur retour en contexte (les dépendants d'un item)	
_	238 : <i>Le Trameur</i> le <i>Cadre</i> (visualisation, réorganisation)	
	239 : <i>Le Trameur</i> Chargement d'une partition à réorganiser	
	240 : <i>Le Trameur</i> Réorganisation d'une partition	
	241 : <i>Le Trameur</i> Courbe d'accroissement sur une partition initiale et la même partition tronquée et	
_	réorganisée	
	242 : Ajouter un élément au rapport	
_	243 : Rapport en cours dans l'onglet RAPPORT	
	244 : Ajouter un élément externe au rapport	
	245 : Rapport enregistré	
_	246 : Index du rapport enregistré	
	247 : Elément du rapport enregistré	
	248 : Importation d'annotation+graphe	
	249 : Chargement de la base, accès à la <i>Trame</i> par le <i>Cadre</i>	
	250 : Sélection des items constituant un nœud du graphe	
	251 : Construction du nœud	
_	252 : Insertion d'un label sur le nœud	
_	253 : Le nœud a un label	
_	254 : De nouveaux nœuds	
_	255 : La source d'un arc	
_	256 : La cible de l'arc l'arc est défini	
	257 : Ajout d'un label sur l'arc	
_	258 : L'arc a son label	
_	259 : Graphe complet construit	
_	260 : Sauvegarde du graphe	
_	261 : Alignement, partition tu (1)	
_	262 : Alignement, partition tu (2)	
_	263 : Alignement, partition volet (1)	
_	264 : Alignement, partition volet (2)	
_	265 : fichier 1, volet français (extrait)	
_	266 : fichier 1, volet anglais (extrait)	
_	267 : Edition d'un alignement dans mkAlign	
_	268 : Préparation d'une base textométrique de textes bilingues alignés	
	269 : Carte des sections d'un alignement dans le Trameur	
_	270 : Carte des sections d'un alignement dans mkAlign	
	271 : Projection d'une forme sur carte des sections dans mkAlign	
Figure	272 : Projection d'une forme sur la carte des sections dans Le Trameur	. 233
Figure	273 : Paramétrage d'un bitexte	. 234

CLESTHIA>SYLED>CLA²T

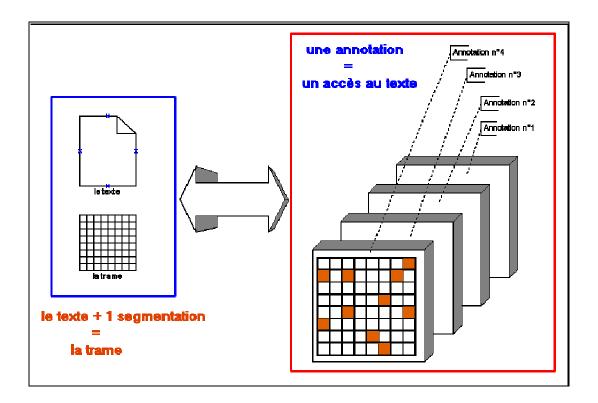
Figure 274 : Mise au jour du bitexte via la carte des sections	234
Figure 275 : Spécificités volet SOURCE de l'alignement	235
Figure 276 : Spécificités volet CIBLE de l'alignement	
Figure 277 : Spécificités des volets SOURCE et CIBLE via mkAlign	236
Figure 278 : Base textométrique de 2 traductions alignées	236
Figure 279 : Carte des sections d'un alignement de traductions dans le Trameur	
Figure 280 : Projection d'une forme sur les 2 traductions alignées	
Figure 281 : Spécificités sur volet SOURCE (TRADUCTION2) : sections contenant le motif "vieil"	238
Figure 282 : Spécificités sur volet CIBLE (TRADUCTION1) :	
sections MIROIR des sections contenant le motif "vieil"	
Figure 283 : Alignement TMX dans un navigateur (avec feuille de styles)	
Figure 284 : Alignement TMX (code source)	
Figure 285 : Alignement TMX , arbre des éléments	
Figure 286 : Carte des sections d'un alignement TMX avec le Trameur	240
Figure 287 : 1 colonne = 1 volet = 1 langue	
Figure 288 : Distribution d'une forme sur la carte de l'alignement	
Figure 289 : Contraste de volets (2 à 2)	
Figure 290 : Contraste de 2 volets (FR2, FR1)	243
Figure 291 : Spécificités volet SOURCE (FR2)	
Figure 292 : Spécificités volet CIBLE (FR1)	244
Figure 293 : Contraste de 2 volets (FR2, EN)	
Figure 294 : Spécificités volet SOURCE (FR2)	245
Figure 295 : Spécificités volet CIBLE (EN)	245
Figure 296 : Base textométrique multiannotée d'un alignement de textes bilingues	247
Figure 297 : Calculs contrastifs sur une base textométrique d'alignements	
Figure 298 : Carte de sections d'un alignement multi-volets	249
Figure 299: Edition d'un alignement multi-volets	250
Figure 300 : Opérations dans une section multi-volets	251
Figure 301 : Export de corpus alignés (HTML)	
Figure 302 : Graphe de relations de dépendances dans une section multi-volets (1)	253
Figure 303 : Graphe de relations de dépendances dans une section multi-volets (2)	254
Figure 304 : Graphe de relations de dépendances dans une section multi-volets (3)	254
Figure 305 : CFPP2000 (1)	255
Figure 306: CFPP2000 (2)	256
Figure 307 : Le petiMoteur	257
Figure 308 : Fichier de MOTIFS	259
Figure 309 : Fonctionnalités sur les MOTIFS	260
Figure 310 : Les occurrences d'un MOTIF	261
Figure 311 : Sélection d'un MOTIF	262
Figure 312 : Traitements sur MOTIF	262
Figure 313 : Le Trameur en mode console	263

2 Préambule

L'objectif principal de la textométrie est de compter des éléments (des *contenus* textuels) dans des ensembles (des *contenants* regroupant des unités élémentaires d'un texte ou des zones de texte couvrant un certain nombre ou un certain type d'unités élémentaires).

- Les *contenus* se réalisent sous la forme de ressources textuelles (une séquence de caractères organisée en phrases, en paragraphes etc.).
- Les *contenants* existent sous la forme de système de masques ou de calques que l'on peut définir sur les contenus. Il s'agit de systèmes d'annotations que l'on peut définir sur tout ou partie des zones textuelles, ces annotations constituant en retour des accès sur les parties textuelles qu'elles définissent (le marquage des phrases ou des paragraphes étant un exemple d'annotation particulier pour décrire un certain niveau de la structure du texte).

Le processus de comptage nécessite au préalable d'identifier les *contenus* et les *contenants*. Ce préalable consiste à expliciter une *segmentation* du texte conduisant à la mise au jour d'une *Trame* sur laquelle des annotations pourront se greffer ultérieurement.

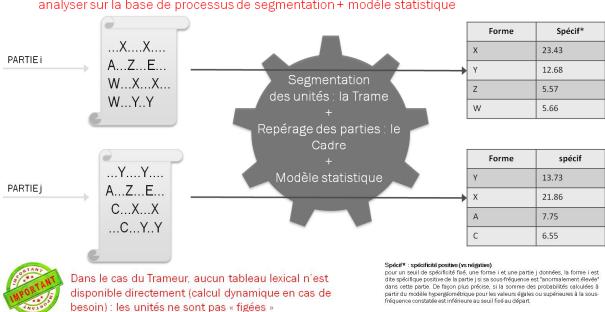


CLESTHIA>SYLED>CLA2T

L'objectif de la textométrie est de compter des éléments (des contenus textuels) dans des ensembles (des contenants) regroupant des unités élémentaires d'un texte ou des zones de texte couvrant un certain nombre ou un certain type d'unités élémentaires



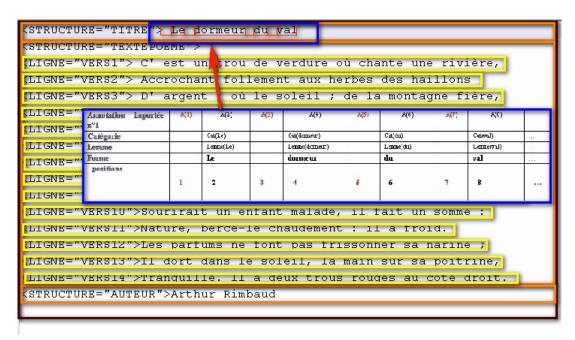
« Classiquement », la textométrie construit des tableaux lexicaux à partir des données à analyser sur la base de processus de segmentation + modèle statistique



Trame et *Cadre* textométriques

A partir d'un texte segmenté, la numérotation des items découpés dans le texte de départ permet de constituer un système de coordonnées sur le texte dans lequel chaque item est repéré par son numéro d'ordre. Nous appelons ce système de coordonnées sur la séquence textuelle : la **Trame** textométrique. Ce même système de coordonnées permet de définir et de localiser, au sein du corpus, des zones textuelles (zones formées par une suite d'items consécutifs, entre la position x_1 la position x_2 , réunion d'un certain nombre de zones de ce type, etc.).

La définition d'une **Trame** textométrique sur un corpus de textes permet de décrire les systèmes de zones qui correspondent aux contenants de l'analyse textométrique (parties, paragraphes, phrases, sections, chapitres etc.). On peut rassembler les descriptions relatives aux systèmes de contenants dans une structure de données particulière le **Cadre** textométrique.



Nous appelons **Trameur** l'outil informatique qui permet de construire une ressource textométrique **Trame/Cadre**, à partir d'un ensemble de textes rassemblés en corpus, selon les principes définis plus haut. La partie **Trame** de la ressource textuelle produite par le trameur est constituée par la suite des items isolés lors de l'opération de segmentation. La partie **Cadre** rassemble les données relatives aux différents découpages réalisés sur le corpus ainsi que les annotations déjà projetées sur les contenants ou les contenus constitués lors des étapes précédentes. La transmission d'une ressource textuelle constituée sous la forme **Trame/Cadre** constitue une solution suffisante pour servir de base à toute exploration textométrique ultérieure.

Source : [Söze-Duval, 2008], Keyser Söze-Duval. *Pour une textométrie opérationnelle*. (DOC)

3 Présentation générale

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

Le *Trameur* intègre le programme *treetagger*¹ : système d'étiquetage automatique des catégories grammaticales des mots avec lemmatisation.

Il permet aussi de généner et de gérer <u>dynamiquement</u> des annotations multiples sur les unités du texte (et de traiter les niveaux d'annotations visés).

Ce document présente tout déabord la « version Tk » du **Trameur** puis la « version console ».

TOUTES LES DISTRIBUTIONS "COMPLETES" DU TRAMEUR INTEGRENT PAR DEFAUT LES MODULES COMPLEMENTAIRES: R, TREETAGGER ET PAJEK

TOUTES LES DISTRIBUTIONS "COMPLETES" INTEGRENT UNE RESSOURCE D'ANNOTATION POUR TREETAGGER ADAPTEE AU <u>FRANÇAIS ORAL</u>, CELLE-CI A ETE REALISEE PAR CHRISTOPHE BENZITOUN (ATILF) (source: http://cnrtl.fr/corpus/perceo/)

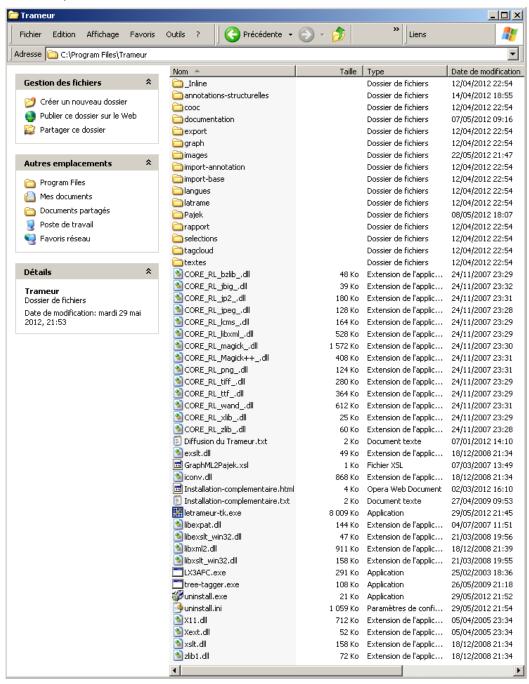
TOUTES LES DISTRIBUTIONS "ALLEGEES" N'INTEGRENT PAS LES MODULES COMPLEMENTAIRES : TREETAGGER ET PAJEK. POUR LES INSTALLER, SUIVRE LE MODE D'EMPLOI DISPONIBLE EN LIGNE.

¹ http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html

4 Installation

4.1 Installation du Trameur

Après installation du programme setup-trameur.exe, løarborescence de fichiers de løapplication a løallure suivante :



Descriptif des principales ressources du dossier deinstallation :

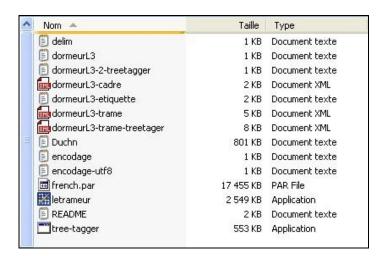
- Le dossier Inline contient des bibliothèques externes pour le *Trameur*
- Le dossier annotations-structuelles contient des ressources illustrant la construction de Graphe døannotation (cf infra)
- Le dossier cooc contient des ressources pour le calcul des cooccurrents (stop-liste)

CLESTHIA>SYLED>CLA2T

- Le dossier documentation contient ce même document en version PDF et en version HTML
- Le dossier export² est utilisé pour stocker les exports de fichier générés par le *Trameur*
- Le dossier graph est utilisé pour stocker les graphiques générés par le *Trameur*
- Le dossier images contient les différentes icônes utilisées par le *Trameur*
- Les dossiers import contiennent des modèles de fichier dømportation de bases textométriques ou de système døannotations sur les fichiers de travail utilisés dans cette documentation.
- Le dossier langues contient différents fichiers de paramétrage de langue nécessaire pour treetagger
- Le dossier Pajek contient le programme du même nom et les ressources associées
- Le dossier rapport contient les éléments utilisés pour générer les rapports produit par le *Trameur*
- Le dossier textes contient des exemples de fichier de travail
- Le programme tree-tagger.exe est lancé par le **Trameur** pour étiqueter le texte de travail
- Le fichier GraphML2Pajek.xsl est utilisé par le *Trameur* pour exporter une liste de séquences de termes associées à un patron au format *Pajek*
- Le programme le *Trameur* enfin : letrameur-tk.exe. Pour le lancer : double-clic sur son icône

4.2 Installation du Trameur (version console)

Après avoir décompressé løarchive contenant le programme, løarborescence de fichiers de løapplication a løallure suivante :



Le programme le *Trameur* se nomme ici : letrameur.exe. Pour le lancer, démarrer une fenêtre de commandes et suivre le mode dœmploi contenu dans le fichier README.txt ou dans ce document.

18

² Un dossier de même nom sera créé <u>dans le dossier contenant la base en cours de traitement</u>. Des fichiers dœxport y seront stockés (idem pour les dossiers graph, logR, rapport: ces dossiers stockeront des données générées au cours de lænalyse de la base)

5 Interface du *Trameur*

Løinterface du programme est composée døune fenêtre graphique disposant de différents onglets, chacun permettant døactiver différentes fonctionnalités.

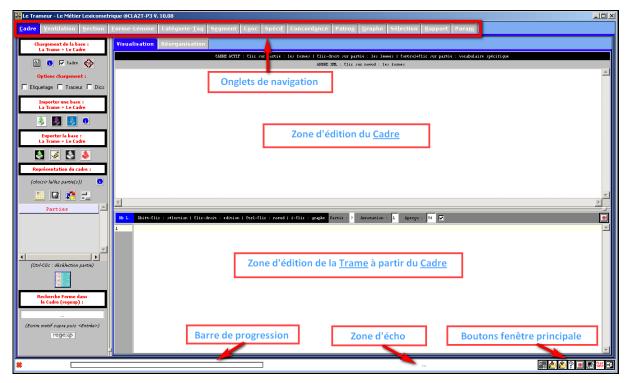


Figure 1 : *Le Trameur* Fenêtre principale

Lønglet CADRE visible au démarrage permet de charger la base de travail.

Løonglet **VENTILATION** donne accès aux fonctionnalités liées à la génération ou à løaffichage des graphes de ventilation.

Løonglet **SECTION** donne accès aux fonctionnalités liées à la *Carte des Sections*.

Løonglet **FORME-LEMME** donne accès aux dictionnaires des formes et des lemmes et aux fonctionnalités associées.

Løonglet CATEGORIE-TAG donne accès aux dictionnaires des catégories et des annotations complémentaires et aux fonctionnalités associées.

Løonglet **SEGMENT** donne accès aux fonctionnalités liées au calcul des segments répétés.

Lønglet COOC donne accès aux fonctionnalités liées aux calculs des cooccurrents et polycooccurrents.

Løonglet STAT donne accès aux fonctionnalités liées aux calculs statistiques (spécificités, AFC etc.).

Løonglet CONCORDANCE donne accès aux fonctionnalités liées aux concordances.

Løonglet PATRON donne accès aux fonctionnalités liées aux calculs dæxtraction de patron.

Løonglet **GRAPHE** donne accès aux fonctionnalités liées aux graphes de relation entre les items de la *Trame*.

Løonglet **RELATION** (non visible dans la figure 1) donne accès aux fonctionnalités liées aux relations de dépendance entre les items de la *Trame*.

Lønglet **SELECTION** donne accès aux fonctionnalités du *Gestionnaire de Sélection*.

Løonglet RAPPORT donne accès aux fonctionnalités liées à la génération du rapport døexploration døune base.

Løonglet **PARAM** donne accès aux fonctionnalités permettant de paramétrer le chargement døune base de travail.

CLESTHIA>SYLED>CLA²T

6 Descriptif des icônes³

Icône	Fonction	Localisation
	Chargement de la base	Onglet CADRE
	Chargement de la représentation graphique du Cadre	Onglet CADRE
	Chargement de la représentation graphique de la Carte des sections	Onglet SECTION
	Suppression de la représentation graphique du <i>Cadre</i>	Onglet CADRE
	Sauvegarde de la représentation graphique du <i>Cadre</i>	Onglet CADRE
€ <mark></mark>	Chargement de la représentation arborescente doun document au format XML	Onglet CADRE
±	Ajouter un élément au rapport	Fenêtre principale, tous les onglets
	Enregistrer le rapport	Fenêtre principale, onglet RAPPORT
4	Export de la base textométrique (<i>Trame+Cadre</i>)	Onglet CADRE
٠.	Export (<i>Trame+Cadre</i>): format TXT et XML	Onglet CADRE
\$	Import døune base textométrique (<i>Trame+Cadre</i>)	Onglet CADRE
<u></u>	Export du fichier <i>Cadre</i> et du fichier <i>Trame</i>	Fenêtre principale
<u></u>	Importation doune annotation sur une <i>Trame</i>	Fenêtre principale
€	Export de la concordance au format HTML	Onglet CONCORDANCE
9	Documentation du programme au format HTML	Fenêtre principale
s R	Calcul des segments répétés	Fenêtre principale, onglet SEGMENT
	Génération de liste de séquences de termes associées à un patron	Onglet PATRON
rivins.	Export de liste des séquences de termes associées à un patron au format <i>Pajek</i>	Onglet PATRON
Ac	Génération de la courbe døaccroissement du vocabulaire	Onglet VENTILATION
	Calcul des spécificités (cooc) dans les sections contenant le motif	Onglet SECTION
	Calcul des spécificités (cooc) dans les sections ne contenant pas le motif	Onglet SECTION
	Calcul des spécificités dans les sections sélectionnées	Onglet SECTION
\$	Calcul des mots spécifiques døune série sections sélectionnées	Onglet SECTION
TO CO	Edition des items	Onglets FORME-LEMME, CATEGORIE-TAG
	Ventilation des items	Onglets FORME-LEMME, CATEGORIE-TAG, SEGMENT
	Concordance des items	Onglets FORME-LEMME, CATEGORIE-TAG, SEGMENT
C	Projection sur le <i>Cadre</i> des items	Onglets FORME-LEMME, CATEGORIE-TAG, SEGMENT
	Projection sur la carte des sections des items	Onglets FORME-LEMME, CATEGORIE-TAG, SEGMENT, RELATION
	Calcul des cooccurrents et des poly-cooccurrents	Onglet COOCS
Sp. Tot	Calcul du tableau lexical dønne partition	Onglet STAT
AFC	Calcul AFC	Onglet STAT
PCLC	Calcul des Principales Caractéristiques Lexicométriques døune partition	Onglet STAT

_

 $^{^3}$ Le tableau ci-dessous intègre une grande partie des boutons disponibles dans lønterface du logiciel.

7 Chargement de la base : mise au jour du Cadre et de la Trame

A partir d'un texte segmenté, la numérotation des items découpés dans le texte de départ permet de constituer un système de coordonnées sur le texte dans lequel chaque item est repéré par son numéro d'ordre. Nous appelons ce système de coordonnées sur la séquence textuelle : la **Trame** textométrique. Ce même système de coordonnées permet de définir et de localiser, au sein du corpus, des zones textuelles (zones formées par une suite d'items consécutifs, entre la position x_1 la position x_2 , réunion d'un certain nombre de zones de ce type, etc.). La définition d'une **Trame** textométrique sur un corpus de textes permet de décrire les systèmes de zones qui correspondent aux contenants de l'analyse textométrique (parties, paragraphes, phrases, sections, chapitres etc.). On peut rassembler les descriptions relatives aux systèmes de contenants dans une structure de données particulière le **Cadre** textométrique. (Söze-Duval 2008)

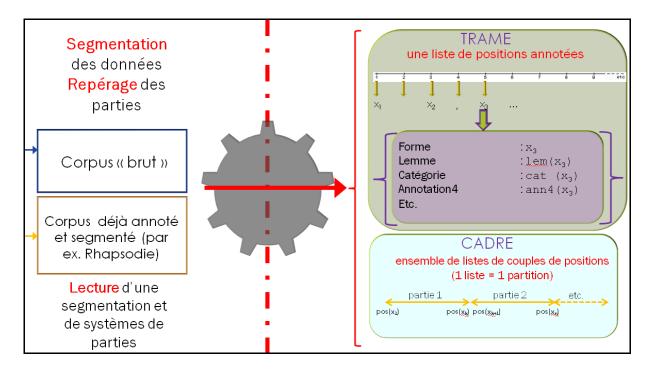
Une base textométrique peut être chargée dans Le *Trameur* :

1. **en créant une nouvelle base** (à partir de fichiers dont le format est défini *infra*)

« **Créer une nouvelle base** » déclenche la sélection du fichier visé puis son chargement (avec ou sans étiquetage via *Treetagger*) *i.e* génération døune segmentation et du repérage des parties (*cf infra* format de fichier).

2. ou en important une base déjà créée (cf export døune base textométrique)

« **Importer une base** » permet de charger une base textométrique déjà constituée (par exemple une base déjà exportée par le logiciel) *i.e* une base intégrant la description dønne *Trame* et dønn *Cadre*.



7.1 Chargement døune nouvelle base

7.1.1 Paramétrage

Løonglet **PARAM** contient différents champs permettant de paramétrer le chargement de la base :

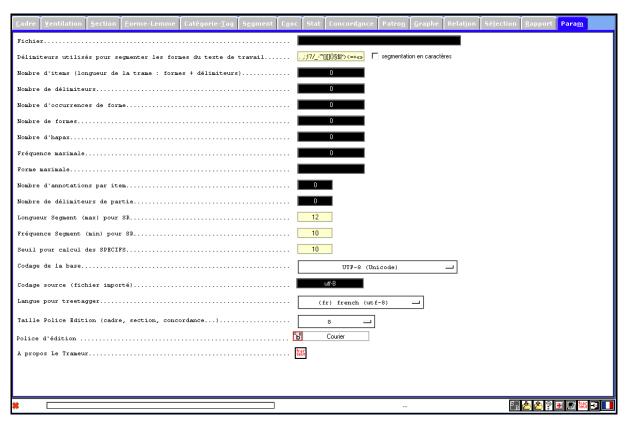


Figure 2: Le Trameur Onglet PARAM

On peut y définir les délimiteurs utilisés pour segmenter le texte, lœncodage du texte à traiter, le fichier de langue à utiliser pour lœtiquetage via *treetagger* et les paramètres pour le calcul des segments répétés. On peut aussi y modifier la police dœdition du texte chargé (nécessaire pour le traitement de certaines langues). Les autres champs sont des indicateurs renseignés par le programme à læssue du chargement de la base.

7.1.2 Chargement

Løonglet CADRE contient:

- une zone complémentaire pour paramétrer le chargement de la base
- une fenêtre dédition du *Cadre* de la base (représentation graphique de léenchâssement des parties)
- une fenêtre dédition doune partie sélectionnée dans le *Cadre*

2 cases à cocher permettent de déclencher :

- a. løétiquetage du texte traité (avec treetagger),
- b. lømpression des différents dictionnaires (*Dics*) : Løactivation de *Dics* conduit à lømpression des différents dictionnaires disponibles à løissue du chargement.

Pour charger une nouvelle base : activez le bouton

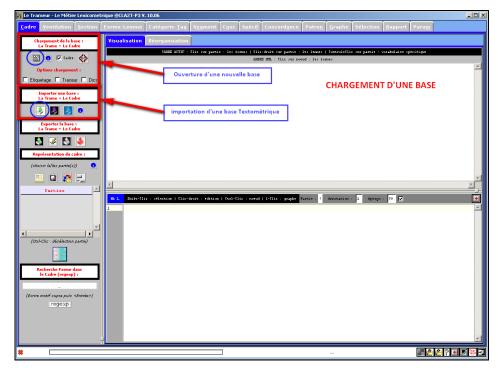


Figure 3 : Le Trameur Onglet CADRE

7.1.3 Etiquetage du texte

Le *Trameur* intégre (dans sa version complète) le programme *treetagger*⁴ : un outil qui permet d'annoter un texte avec des informations sur les parties du discours (genre de mots: noms, verbes, infinitifs et particules) et des informations de lemmatisation. Pour activer létiquetage du texte chargé au démarrage, il est nécessaire, avant de charger le texte, de cocher la case à cocher « étiquetage » (*cf* onglet **CADRE**) et de choisir le fichier paramètre de langue à utiliser pour réaliser cet étiquetage (*cf* onglet **PARAM**) :

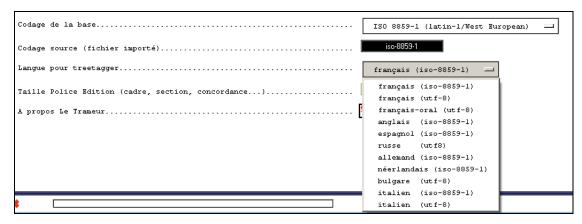


Figure 4: Le Trameur Onglet PARAM: choix de langue pour Treetagger

Les langues utilisables sont visibles dans la figure précédente. La plupart des fichiers paramètres de langue associés ont été récupérés sur le site de *treetagger*. Le fichier « français oral » est issu du projet PERCEO⁵ (ATILF (CNRS - Université de Lorraine) & INIST (CNRS) ó LIPN), responsable scientifique Christophe Benzitoun (ATILF).

⁴ http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

⁵ http://cnrtl.fr/corpus/perceo/

CLESTHIA>SYLED>CLA2T

7.1.4 Inventaire typographique

Après avoir chargé une nouvelle base, il est possible de visualiser lœnsemble des caractères composant le fichier de travail (module *Inventaire Typographique*), on peut éventuellement modifier certains caractères ou modifier le statut de certains caractères (délimiteur). Pour illustrer ces fonctionnalités, on considère que la nouvelle base chargée a été construite à partir du fichier suivant :

```
<STRUCTURE="TITRE">Le dormeur du val § £
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1">C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2">Accrochant follement aux herbes des haillons
<LIGNE="VERS3">D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4">Luit : c' est un petit val qui mousse de rayons. §
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut. §
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid. §
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit. § £
<STRUCTURE="AUTEUR">Arthur Rimbaud § £
```

Une fois la base chargée, léinventaire typographique produit le résultat suivant :

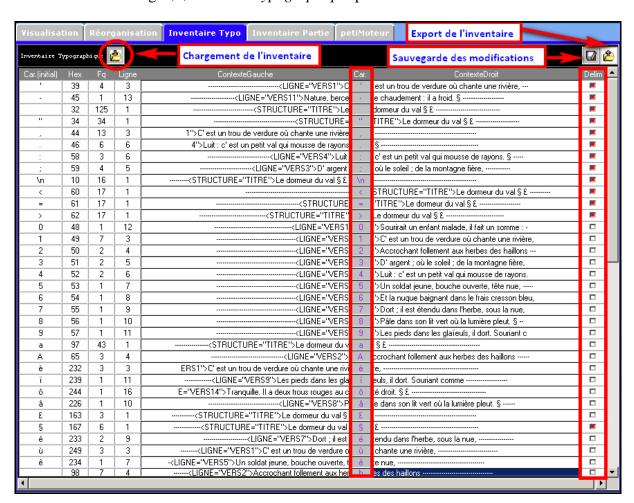


Figure 5a: Le Trameur Inventaire Typographique

La **sixième colonne** met au jour les caractères en contexte (contextes gauche et droit) dans une fenêtre døune cinquantaine de caractères : chaque cellule de cette colonne est éditable (clic sur la cellule).

La **huitième colonne** indique le statut du caractère : si la case est rouge, le caractère est un délimiteur. Un clic sur la case permet de changer le statut du caractère visé.

Les modifications étant réalisées, la sauvegarde du fichier (en tenant compte ses modifications réalisées) déclenche automatiquement le rechargement de la base dans le *Trameur* et la mise à jour éventuelle de la liste des délimiteurs (*cf* onglet **PARAM**).

Remarque : le processus de correction modifie le fichier initial, une sauvegarde de ce fichier est réalisée au préalable dans le même dossier.

Pour illustrer 2 fonctionnalités, on utilise *infra* un corpus issu du module *petiMoteur* disponible dans løonglet CADRE.

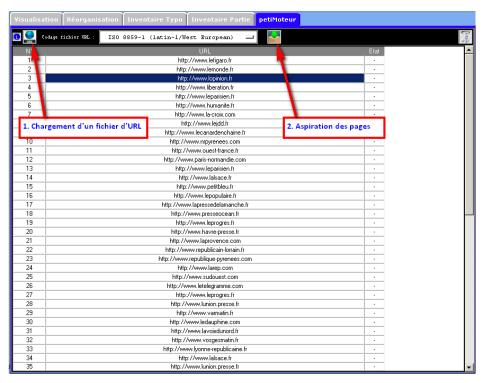


Figure 5b: Le Trameur Inventaire Typographiquesur petitMoteur (1)

Ce module permet de réaliser un corpus à partir døune liste døURLs (ici une liste de 60 urls pointant vers des pages døaccueil de journaux français en ligne). Une fois le fichier døURLs chargé (1), le module døaspiration (2) construit un fichier concaténant les contenus textuels des différentes URLs. Le fichier produit (text_URLS.txt dans le même dossier que le fichier des URLs) peut ensuite être chargé dans le *Trameur*.

CLESTHIA>SYLED>CLA2T

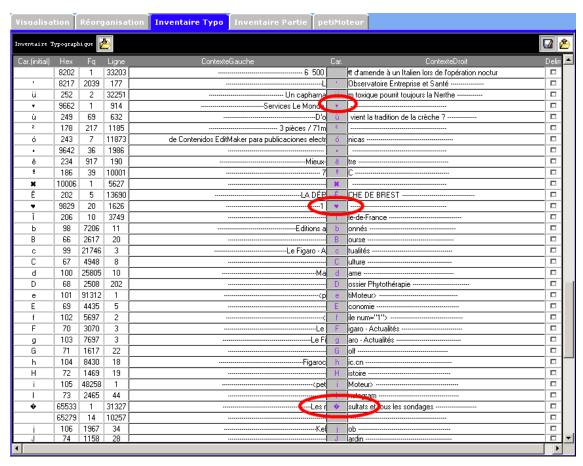


Figure 5c: Le Trameur Inventaire Typographiquesur petitMoteur (2)

Lønventaire révèle ici de nombreux caractères sur lesquels des modifications sont envisageables : erreurs de codage par exemple, attribution de statut de délimiteur etc.

7.1.5 Inventaire des parties

De la même manière, il est possible de construire un inventaire de toutes les balises utilisées pour définir le partitionnement du fichier chargé. Une fois la base chargée, léinventaire des parties a léallure suivante :

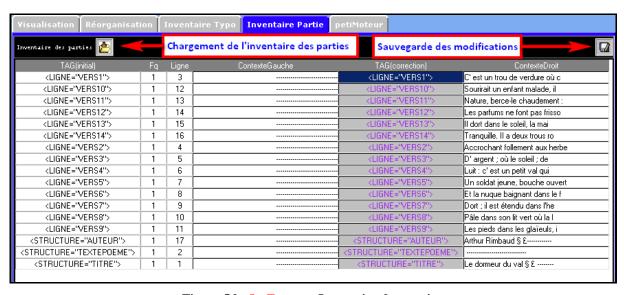


Figure 5d : Le Trameur Inventaire des parties

La **cinquième colonne** met au jour chaque balise en contexte (contextes gauche et droit) dans une fenêtre døune cinquantaine de caractères : chaque cellule de cette colonne est éditable (clic sur la cellule).

Les modifications étant réalisées, la sauvegarde du fichier (en tenant compte ses modifications réalisées) déclenche automatiquement le rechargement de la base dans le *Trameur*.

Remarque : le processus de correction modifie le fichier initial, une sauvegarde de ce fichier est réalisée au préalable dans le même dossier.

Ce processus peut être utile pour corriger le fichier chargé en supprimant notamment les caractères < et > résiduels : cf ci-dessous. Ce module permet aussi de renommer les balises ; ci-dessous, 2 opérations de renommage sont réalisées, løune døentre elles søaccompagne aussi døun ajout døune nouvelle balise.

Ci-dessous, un corpus issu du programme *Gromoteur*⁶ (même liste døURLs que précédemment) dont løinventaire des parties produit ceci :

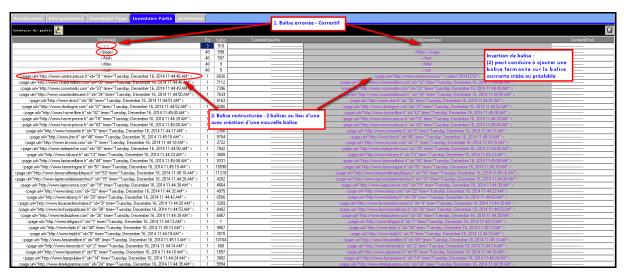


Figure 5e : Le Trameur Inventaire des parties sur Gromoteur

Løinventaire peut conduire à supprimer des balises « erronées » ou à modifier les balises existantes pour redéfinir le partitionnement de la base après sauvegarde des modifications.

_

⁶ http://gromoteur.ilpga.fr/

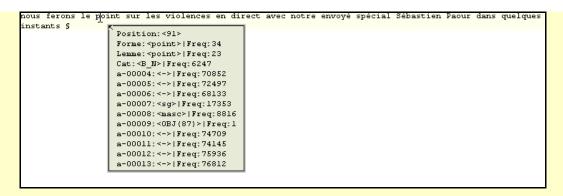
7.2 Importer une base

On peut aussi charger une base en important une base textométrique existante (une *Trame* et un *Cadre*). On active cette fonctionnalité *via* le bouton idoine

Le format du fichier dømportation døune base est présenté dans la section « <u>Exporter une base textométrique</u> ».

On trouvera des exemples de bases textométriques déjà constituées sur la page web du projet. Ci-dessous, quelques bases utilisées dans ce document :

• <u>Base "Rhapsodie2Trameur"</u> construite à partir des ressources développées dans le cadre du projet <u>Rhapsodie</u>. Chaque item de la *Trame* est associé à 13 niveaux d'annotation (micro-syntaxe).



SOURCES: projet Rhapsodie

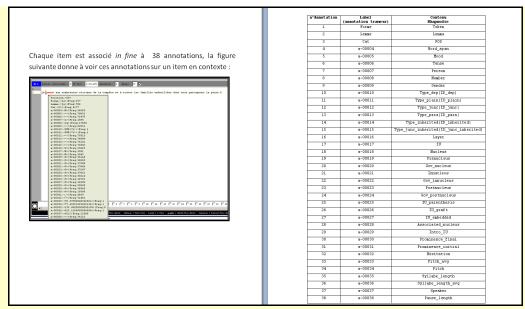
Fichiers de codage microsyntaxique version béta 10/13 (zip)

Descriptif et sources des annotations :

Annotations Rhapsodie pour le Trameur (pdf) : présentation du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique ; présentation des différents processus de traitements des annotations de dépendance.

Open publication - Free publishing Open publication - Free publishing

• <u>Base "Rhapsodie2Trameur" (v4)</u> (màj : 19.07.2014) construite à partir des ressources développées dans le cadre du projet <u>Rhapsodie</u>. Chaque item de la *Trame* est associé à 38 niveaux d'annotation (prosodie, micro et macro-syntaxe).



(clic pour vue détaillée)

SOURCES: projet Rhapsodie

Descriptif et sources des annotations :

Annotations Rhapsodie pour le Trameur (pdf) : présentation du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique ; présentation des différents processus de traitements des annotations de dépendance.

• <u>Base "Corpus de contes français"</u> (Fairy Tales Corpus, FTC) construite à partir des ressources développées par <u>Ismaïl El Maarouf</u>. Chaque item de la *Trame* est associé à 7 niveaux d'annotation.

```
petits en avant .Les <mark>mésanges</mark> s' installèrent dans les
         ?Jack a appelé les mésanges pour avoir leur version
         difficile ",ont conclu lek
                                       Position: <30823>
      compter sur l'aide des mésa
                                       Forme: <mésanges> | Freq: 13
encourageait ses assistants .Les 🕇
                                                                          ment
                                       Lemme: <mésange> | Freq: 13
        par les belettes .Les mésa
                                       Cat: < NOM > | Freq: 3801
        après la fête ,les mésange
                                       a-4:<16339>|Freq:1
                                       a-5: <animal> | Freq: 155
                                       a-6: <Addressee(16337) > | Freq: 1
                                       a-7: <0BJ (16337) > | Freq: 1
```

SOURCES: I. El Maarouf et J. Villaneau (2012). A French Fairy Tale Corpus syntactically and semantically annotated. In Proceedings of LREC 2012. Istanbul (Tk).

Pour les formes : copyright Momes.net, http://momes.net.

Lemme et POS via treetagger : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)

Pour les annotations : http://elmaarouf.legtux.org/Accueil.html

Descriptif et sources des annotations :

balise <f> (forme): copyright Momes.net, http://elmaarouf.legtux.org/data/FTC FORM.txt

balise <|> (lemme) : fichier paramètre langue (français oral, utf-8) développé par Christophe Benzitoun (ATILF) (source : http://cnrtl.fr/corpus/perceo/)

balise <c> (POS) : fichier paramètre langue (français oral, utf-8) développé par Christophe Benzitoun (ATILF) Annotations complémentaires (http://elmaarouf.legtux.org/Accueil.html) :

- Identifiants de mots : http://elmaarouf.legtux.org/data/FTC ID.txt (première balise <a>)
- Annotation référentielle : http://elmaarouf.legtux.org/data/FTC_REF.txt (seconde balise <a>)

CLESTHIA>SYLED>CLA²T

- Annotation en rôles sémantiques pour 26 verbes : http://elmaarouf.legtux.org/data/FTC SEMROLE.txt (troisième balise <a>)
- Annotation syntaxique en dépendance pour 137 verbes : http://elmaarouf.legtux.org/data/FTC_SYNROLE.txt (dernière balise <a>)

L'ensemble des annotations du corpus initial (+ lemme et POS) regroupées ici au format d'une base textométrique pour le **Trameur**.

8 Les formats dentrée (nouvelle base)

Voir aussi : Fleury Serge, *Le Trameur. Propositions de description et døimplémentation des objets textométriques*, (PDF).

Le texte précédent met au jour une description des objets textométriques et les méthodes mises en à uvre dans le *Trameur* pour travailler sur et avec ces objets dans une perspective textométrique. On y détaille aussi les opérations permises sur une *base textométrique*: format des données textuelles, modification dynamique de la *Trame*, correction ou ajout d'annotation etc.

8.1 Format texte brut

```
Le dormeur du val
C' est un trou de verdure où chante une rivière,
Accrochant follement aux herbes des haillons
D' argent ; où le soleil ; de la montagne fière,
Luit : c' est un petit val qui mousse de rayons.
```

Le texte contient une *Trame* mais pas de *Cadre*. Le *Trameur* peut travailler sur la *Trame* du texte.

8.2 Format dit « Lexico »

Exemple $n^{\circ}1$:

```
<STRUCTURE="TITRE"> Le dormeur du val

<STRUCTURE="TEXTEPOEME">

<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,

<LIGNE="VERS2"> Accrochant follement aux herbes des haillons

<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,

<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
```

Le texte contient une *Trame* et un *Cadre* (balise de début explicite pour chaque partie, balise de fin implicite). Le *Trameur* peut travailler sur la *Trame* du texte et il est possible døaccéder au texte par son *Cadre*.

8.3 Format XML

Exemple n°2:

Exemple n°3:

CLESTHIA>SYLED>CLA2T

```
<name type="place">Rome</name>
  </div>
 <div type="update">
  The text of the Convention had been
  amended according to the provisions of Protocol No. 3 (ETS
  No. 45), which entered into force on 21 September 1970, of
  Protocol No. 5 (ETS No. 55), which entered into force on 20
  December 1971 and of Protocol No. 8 (ETS No. 118), which
  entered into force on 1 January 1990, and comprised also the
  text of Protocol No. 2 (ETS No. 44) which, in accordance with
  Article 5, paragraph 3 thereof, had been an integral part of
  the Convention since its entry into force on 21 September
  1970. All provisions which had been amended or added by these
  Protocols are replaced by Protocol No. 11 (ETS No. 155), as
  from the date of its entry into force on 1 November 1998. As
  from that date, Protocol No. 9 (ETS No. 140), which entered
  into force on 1 October 1994, is repealed and Protocol No. 10
  (ETS No. 146) has lost its purpose.
 </div>
</front>
<body>
 <head id="ce.1950.en H">Convention for the Protection of Human
 Rights and Fundamental Freedoms as amended by Protocol No.
 11</head>
 <div type="preamble">
  <list type="simple">
   <item id="ce.1950.en_B.i-1">
    <seg type="membrestates">The governments signatory
    hereto, being members of the Council of Europe, </seg>
   </item>
  </list>
 </div>
</body>
</text>
```

Le texte contient une *Trame* et un *Cadre* (chaque partie est parfaitement délimitée : balises de début et de fin explicites). Le *Trameur* peut travailler sur la *Trame* du texte et il est possible døaccéder au texte par son *Cadre*.

Pour les trois exemples précédents nous donnons à voir la représentation graphique construite par le *Trameur* du *Cadre* (complet) de chacun de ces textes :

Exemple $n^{\circ}1$:

Exemple n°2:

Exemple $n^{\circ}3$:

8.4 Formats pour corpus alignés

Le *Trameur* intègre des fonctionnalités pour lœxploration de corpus alignés (2 ou plusieurs volets). On se reportera à la partie « *Bases textométriques de textes alignés* » pour une présentation de ce type de fonctionnalités.

On présente co-dessous le squelette de 3 formats de fichier permettant de structurer ce type déalignement (ici avec 2 volets). Les 3 formats correspondent quasiment aux formats vus précédemment.

Le premier marque par un balisage « à la lexico » les 2 volets du corpus, un caractère délimiteur (ici le caractère §) segmente chaque volet en unité døalignement (la première section du volet $n^{\circ}1$ est alignée avec la première section du volet $n^{\circ}2$: Source $1 \rightarrow \text{Miroir 1}$) etc.

```
<volet="1">
TEXTE Source1 $
TEXTE Source2 $
TEXTE Source3 $
TEXTE Source4 $
<volet="2">
TEXTE Miroir1 $
```

CLESTHIA>SYLED>CLA2T

```
TEXTE Miroir2 §
TEXTE Miroir3 §
TEXTE Miroir4 §
```

Dans le second format, le balisage est renforcé. Les unités døalignement sont marquées par une balise p (tout autre nom de balise est possible).

Le dernier est au format XML. Ce format peut être enrichi ou modifié :

- on peut par exemple numéroter les balises p via des attributs
- le type døattribut de la balise volet peut être différent dans les 2 balises
- etc.

```
<?xml version="1.0" encoding="utf-8"?>
<c>
<volet lang="fr">
TEXTE Source1
TEXTE Source2
TEXTE Source3
TEXTE Source4
</volet>
<volet lang="en">
TEXTE Miroir1
TEXTE Miroir3
TEXTE Miroir3
TEXTE Miroir4
TEXTE Miroir4
TEXTE Miroir4
TEXTE Miroir4
TEXTE Miroir4
TEXTE Miroir4
```

9 Les dictionnaires

9.1 Les formes

Le chargement de la base déclenche løinsertion des dictionnaires des formes, des catégories (issues de løtiquetage avec *Treetagger*) respectivement dans l'onglet **FORME-LEMME** et dans l'onglet **CATEGORIE-TAG.**

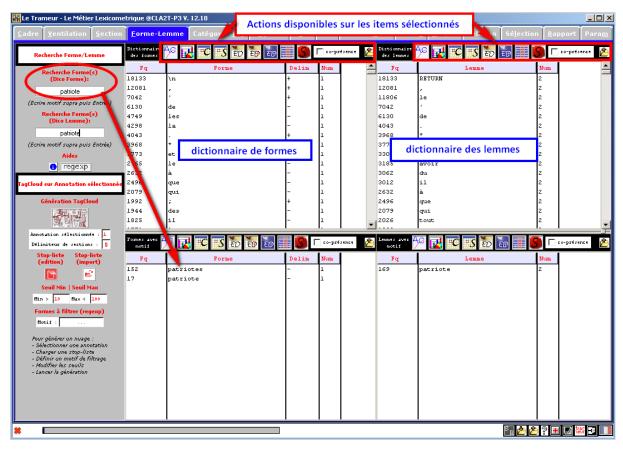


Figure 6 : Le Trameur Onglet FORME

CLESTHIA>SYLED>CLA2T

Après sélection døune forme (ou de plusieurs) dans le dictionnaire, il est possible de déclencher différentes actions :



Figure 7 : Le Trameur Actions sur Items sélectionnés à partir des dictionnaires

- la représentation de la courbe déaccroissement du vocabulaire sélectionné (les formes choisies dans le dictionnaire)
- la ventilation de cette forme (ou de plusieurs) dans la partition sélectionnée (cf onglet VENTILATION).
- la ventilation de cette forme (ou plusieurs) sur une sélection de parties selectionnées via le *Gestionnaire de Sélection* (cf onglet **SELECTION**).
- la recherche de cette forme (ou de plusieurs) dans la représentation cartographique du *Cadre infra* (*cf* onglet CADRE).
- la recherche de cette forme (ou de plusieurs) dans la représentation cartographique de la carte des sections *infra* (*cf* onglet **SECTION**). Si løption « coprésence » est activée, les sections mises au jour seront celles contenant løensemble des formes visées (ET logique)
- localition de la forme visée pour éventuellement la corriger (si la forme est mise à jour, tous les items de la *Trame* ayant cette forme seront modifiés).
- løajout døun niveau døun niveau døannotation sur la forme visée : tous les items de la *Trame* ayant cette forme seront modifiés pour cet ajout døannotation (avec la valeur døannotation fournie), les autres auront eux-aussi cette nouvelle annotation mais avec une valeur indéfinie (modifiable ultérieurement)
- la correction des différents niveaux døannotation sur la forme visée : tous les items de la *Trame* ayant cette forme seront modifiés pour ces modifications sur tous les niveaux døannotation
- **une concordance de cette forme** (ou de plusieurs) (*cf* onglet **CONCORDANCE**).
- la sélection de cette forme (ou de ces formes) *via* ses positions sur la *Trame* (*cf* onglet **SELECTION**).

Ces différentes opérations sont détaillées *infra*.

Les opérations similaires sont disponibles pour les lemmes, les catégories et les annotations importées.

On peut rechercher des formes dans le dictionnaire en définissant un motif de recherche dans la zone de saisie « Recherche Forme(s) ». Les formes trouvées seront listées au bas de l'onglet **FORME.** Les opérations décrites supra sur les formes sélectionnées sont là encore disponibles.

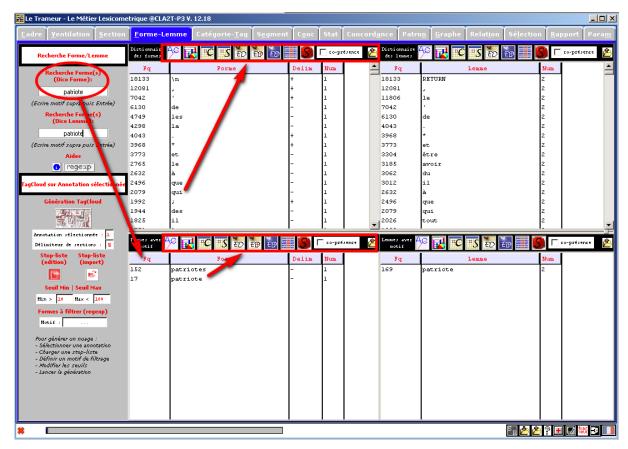


Figure 8: Le Trameur Onglet FORME

9.2 Les lemmes

(idem supra)

9.3 Les catégories

(idem supra)

9.4 Les annotations importées

(cf infra)

10 Accès à la Trame par le Cadre

10.1 Mise au jour de la Trame

Dans les figures qui suivent, le fichier de travail utilisé est le *Dormeur du Val* dont on présente ci-dessous un extrait :

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
```

La *Trame* du texte correspond à toutes les positions des **items** (délimiteurs et formes graphiques). Sur cet exemple :

En position 1, le caractère « espace » (considéré comme un délimiteur), en position 2, la forme graphique « Le », en position 3, le caractère « espace » etc.

Le *Trameur* produira une segmentation que løn peut schématiser ainsi sur le début du texte :

Forme		Le		dormeur		du		val	RC		C	í
Positions	1	2	3	4	3	9	7	8	6	10	11	í

(RC = retour à la ligne)

Le *Cadre* du texte est matérialisé ici par des balises indiquant le début døune partie (sous la forme <partie=valeur>). Ce balisage traduit une structuration en parties du texte (1 titre et 1 poème contenant des vers). Plus précisément, la partie STRUCTURE=TITRE recouvre les items de la position 1 à la position 8 et la partie STRUCTURE=POEME couvre les items à partir de la position 9 jusquøà la dernière. On verra *infra* quøil est possible de décrire le *Cadre* døun texte de manière plus formalisée *via* XML.

Après chargement de la base utilisée, le *Trameur* a l'allure suivante :

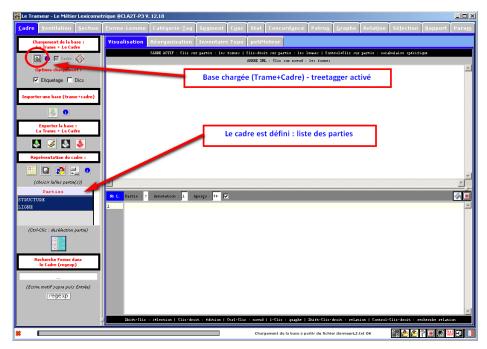


Figure 9 : Le Trameur Onglet CADRE, base chargée

Les différentes parties constituant le *Cadre* sont visibles dans la zone *Parties*. Pour visualiser tout ou partie du *Cadre*, il convient de sélectionner une ou plusieurs parties, puis doactiver le bouton . Après génération du *Cadre* (dans notre exemple, 2 parties ont été sélectionnées), le *Trameur* a l'allure suivante :

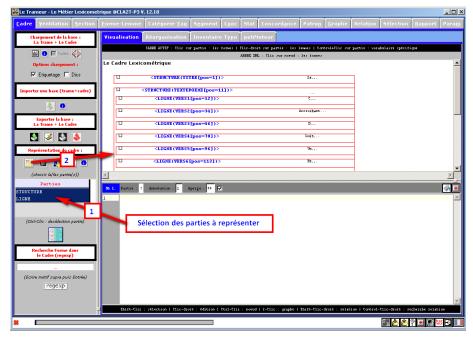


Figure 10 : Le Trameur Onglet CADRE, Cadre généré

Une représentation graphique des parties constituant le texte est construite dans la zone dédition du Cadre. Cette représentation donne à voir léemboîtement éventuel des différentes parties : dans notre exemple, les VERS sont contenus dans la partie POEME. En cliquant (via un clic-gauche) sur le nom deune partie, on active léaffichage de son contenu textuel (les formes i.e les mots) dans la zone dédition de la Trame au bas de léonglet CADRE. Dans la figure suivante, on donne à voir le contenu de la partie STRUCTURE=TEXTEPOEME :

CLESTHIA>SYLED>CLA2T

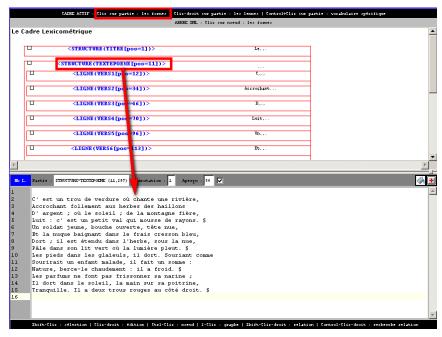


Figure 11: Le Trameur Onglet CADRE, contenu textuel døune partie (les mots)

En cliquant (*via* un *clic-droit*) sur le nom døune partie, on active løaffichage de son contenu textuel (*les lemmes*) dans la zone døédition au bas de løonglet CADRE. Dans la figure suivante, on donne à voir le contenu de la partie STRUCTURE=POEME:

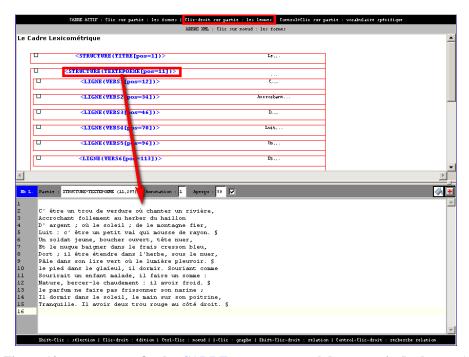


Figure 12 : Le Trameur Onglet CADRE, contenu textuel døune partie (les lemmes)

Si la base à charger est au format XML, on peut aussi construire une représentation arborescente du fichier chargé.

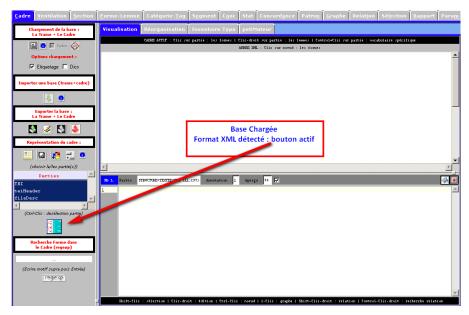


Figure 13 : *Le Trameur* Onglet CADRE, chargement døune représentation arborescente døune base au format XML (1)

Par défaut, le bouton est inactif ; dans le cas døun fichier au format XML, il devient actif et permet le déclenchement de la représentation arborescente du fichier :

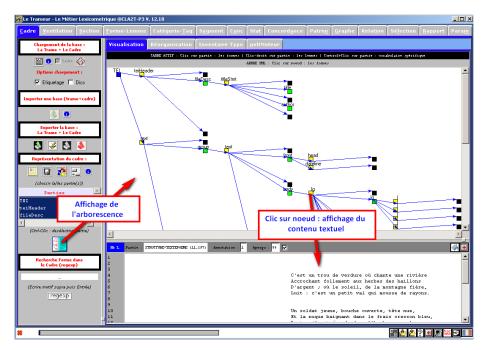


Figure 14 : *Le Trameur* Onglet CADRE, chargement døune représentation arborescente døune base au format XML (2)

10.2 Choix du niveau de løannotation à afficher

Une zone de saisie disponible dans léditeur du **CADRE** et dans léditeur de **SECTION** permet de sélectionner le numéro déannotation à utiliser pour afficher la partie ou la section visée. Rappel: par convention, les annotations associées aux items de la *Trame* sont typées de la manière suivante:

CLESTHIA>SYLED>CLA2T

- le numéro døannotation n°1 correspond à « Forme »
- le numéro døannotation n°2 correspond à « Lemme »
- le numéro døannotation n°3 correspond à « Catégorie »
- Toute autre annotation importée aura un numéro supérieur ou égal à 4.

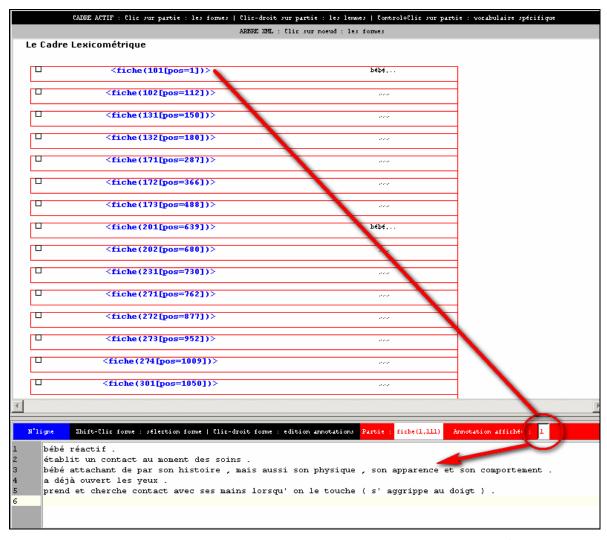


Figure 15 : Le Trameur Affichage døune partie du CADRE (annotation affichée = 1)

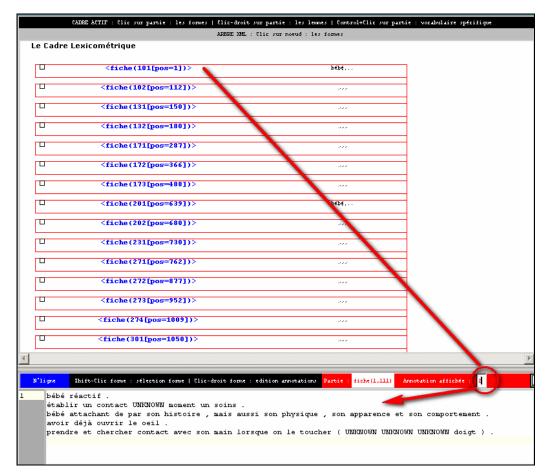


Figure 16: Le Trameur Affichage dønne partie du CADRE (annotation affichée = 2)

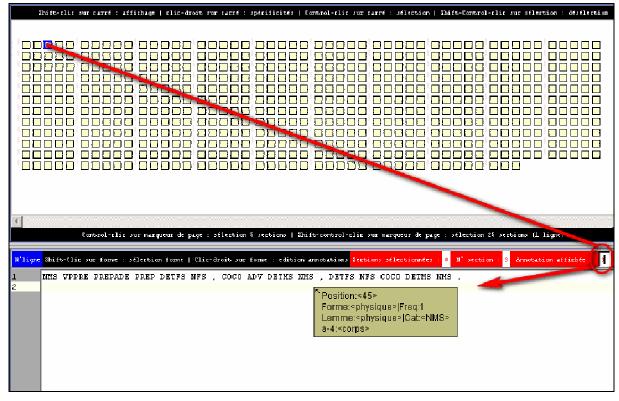


Figure 17 : Le Trameur Affichage doune section (annotation affichée = 3)

CLESTHIA>SYLED>CLA2T

10.3 Recherche de chaînes

La figure suivante illustre une des fonctionnalités classiques d'un éditeur que l'on retrouve dans le *Trameur* : la recherche de chaîne de caractères.

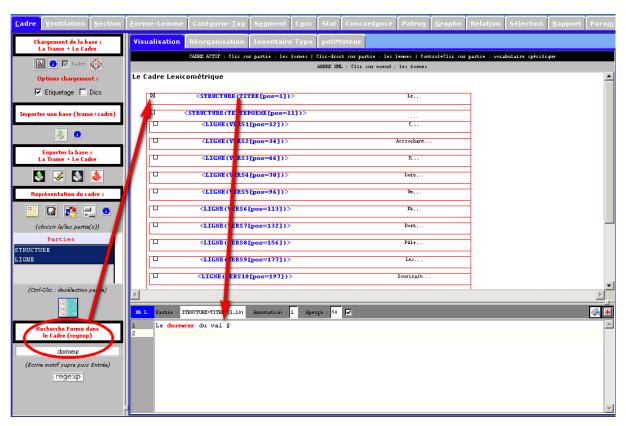


Figure 18 : Le Trameur Onglet CADRE, recherche de chaînes

On dispose dans l'onglet **CADRE** døune zone de saisie permettant de lancer des recherches de chaînes : "*Recherche*". Pour rechercher une forme, il convient de løinsérer dans la zone de saisie puis de lancer la recherche en tapant la touche « Entrée ».

Dans la figure précédente, la forme cherchée (dormeur) est présente dans la partie STRUCTURE (TITRE) : son contenu textuel est visible *via* un *clic-gauche* sur le carré associé à la partie (la croix intégrée au carré indiquant la présence de la forme visée).

On peut aussi lancer une recherche en sélectionnant une forme dans les dictionnaires disponibles (et *via* le raccourci clavier défini).

10.4 Recherche « incrémentale » de chaînes

On peut aussi activer un mode de recherche dit *incrémentale*. Le raccourci clavier *Control-s* active ce mode de recherche permettant de définir *incrémentalement* un motif de recherche.

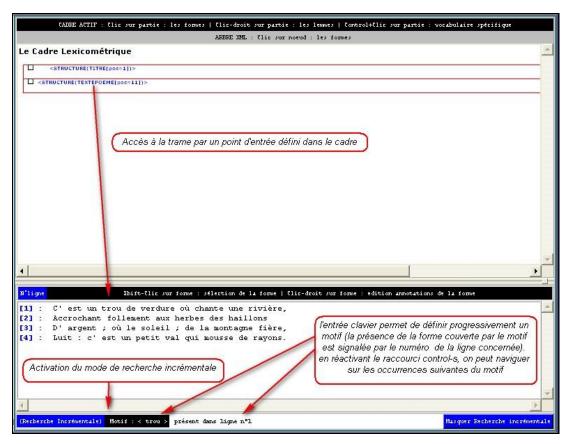


Figure 19 : Le Trameur Onglet CADRE, recherche incrémentale de chaînes

10.5 Visualisation des différentes annotations des items de la Trame

Il est possible de visualiser les différentes annotations associées aux items présents dans la zone dédition (toute ou une partie de la *Trame*). Il suffit de placer la souris sur une des formes graphiques présentes dans cette zone dédition (mot ou lemme), une boîte dédition contextuelle apparaît et donne des précisions sur la forme visée : forme graphique, fréquence, lemme, catégorie (et annotation complémentaire si disponible).

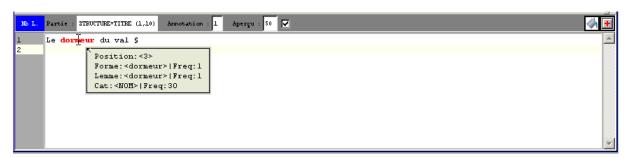


Figure 20 : Le Trameur Onglet CADRE, visualisation détaillée doun item

10.6 Edition et modification des items de la Trame

Dans la même zone dédition (forme ou lemme), il est possible déditer et de modifier les contenus de chaque item. Un *clic-droit* sur un item (forme ou lemme) déclenche léapparition de une fenêtre dédition de lettem visé dans laquelle il est possible de modifier sa forme, son lemme ou sa catégorie (et toute annotation importée) :

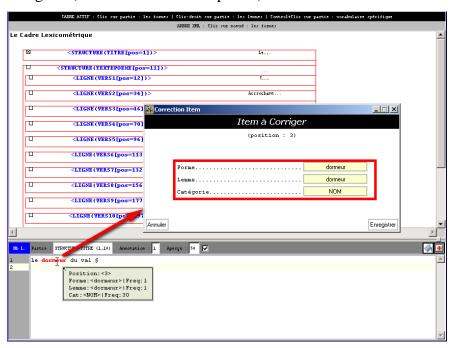


Figure 21 : Le Trameur Onglet CADRE, édition døun item

Après enregistrement des modifications réalisées, la fenêtre dédition initiale est rechargée en tenant compte des modifications apportées :

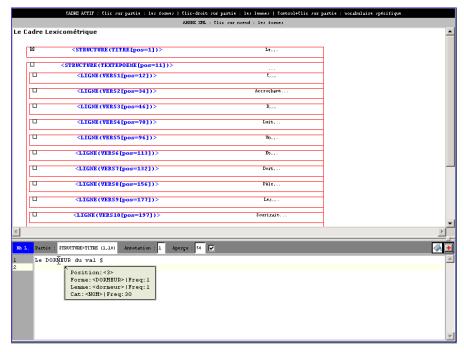


Figure 22: Le Trameur Onglet CADRE, modification døun item

Et les différents dictionnaires impactés par ces modifications sont rechargés.

10.7 Le vocabulaire spécifique døune partie

Le *Cadre* dispose døune fonctionnalité supplémentaire pour agir sur les contenus textuels des parties repérées. On a vu *supra* quøun *clic-gauche* sur une partie déclenchait løaffichage du contenu textuel associé à cette partie dans la zone døédition prévue à cet effet. Le raccourci clavier *control+clic-gauche* sur une partie déclenche le calcul du vocabulaire spécifique⁷ de la partie visée.

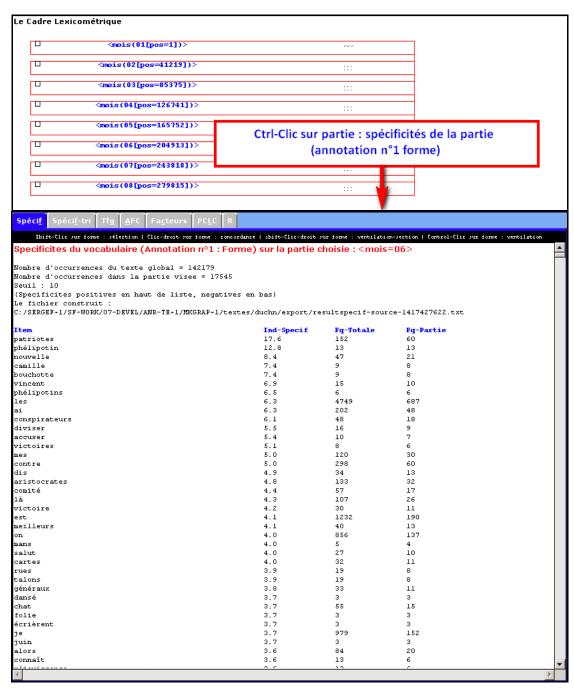


Figure 23 : Le Trameur Onglet CADRE, le vocabulaire spécifique donne partie

⁷ Cf Partie « Vocabulaire spécificique » pour le paramétrage de ce calcul

11 La carte des sections : représentation cartographique de la *Trame*

Løonglet **SECTION** permet de donner à voir une représentation graphique du texte sous la forme døune carte de sections (identique à celle produite par **Lexico3**⁸ *via* l'outil "*Carte des Sections*⁹"). Le bouton permet de construire cette représentation cartographique sur la base du caractère délimiteur choisi pour délimiter les sections.

11.1 La carte des sections : une vue globale de la Trame

La *carte des sections* permet une visualisation du corpus découpé en sections par la promotion d'un (ou de plusieurs) caractère particulier (paragraphes, point, etc.) au statut de *délimiteur de section*. La carte des sections recouvre ainsi intégralement la *Trame* textuelle (une vue globale de la *Trame*).

Si on considère les 2 exemples ci-dessous, dans chacun dœux le caractère § est intégré dans le texte pour marquer des zones.

Dans le premier exemple, le caractère § induit la présence de 6 sections.

Dans le second, on retrouve aussi six sections, mais la *Trame* textuelle dispose en outre døune *Cadre* (des jalons du type « balise » organisant le texte en parties)

```
Le dormeur du val §
2
3
    C'est un trou de verdure où chante une rivière,
    Accrochant follement aux herbes des haillons
    D'argent ; où le soleil, de la montagne fière,
   Luit : c'est un petit val qui mousse de rayons. §
7
    Un soldat jeune, bouche ouverte, tête nue,
   Et la nuque baignant dans le frais cresson bleu,
   Dort ; il est étendu dans l'herbe, sous la nue,
11
    Pâle dans son lit vert où la lumière pleut. §
12
13
    Les pieds dans les glaïeuls, il dort. Souriant comme
   Sourirait un enfant malade, il fait un somme :
15
   Nature, berce-le chaudement : il a froid. §
16
17
    Les parfums ne font pas frissonner sa narine ;
    Il dort dans le soleil, la main sur sa poitrine,
19
    Tranquille. Il a deux trous rouges au côté droit. §
20
1
               Arthur Rimbaud §
```

Figure 24 : Le Trameur Onglet SECTION, texte brut avec caractère « délimiteur de sections »

⁸ http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/

⁹ http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/manuelsL3/manuel-3.41.htm# Toc34045613

```
<STRUCTURE="TITRE">Le dormeur du val §
   <STRUCTURE="TEXTEPOEME">
   <LIGNE="VERS1">C' est un trou de verdure où chante une rivière,
4
   <LIGNE="VERS2">Accrochant follement aux herbes des haillons
   <LIGNE="VERS3">D' argent ; où le soleil ; de la montagne fière,
   <LIGNE="VERS4">Luit : c' est un petit val qui mousse de rayons. §
   <LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
   <LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
   <LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
   <LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut. §
11
   <LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
   <LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
   <LIGNE="VERS11">Nature, berce-le chaudement : il a froid. §
13
   <LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
   <LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
   <LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit. §
16
   <STRUCTURE="AUTEUR">Arthur Rimbaud §
```

Figure 25 : Le Trameur Onglet SECTION, texte structuré avec caractère « délimiteur de sections »

Pour ces deux exemples le caractère délimiteur § est utilisé dans les figures suivantes pour construire la carte des sections associée :

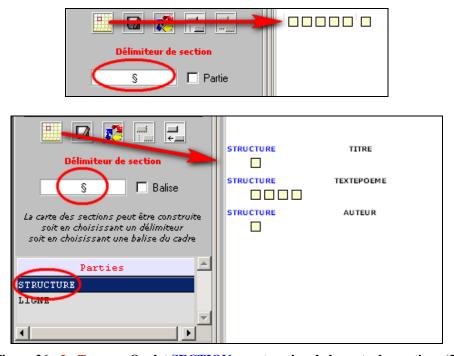


Figure 26: Le Trameur Onglet SECTION, construction de la carte des sections (3)

La seconde figure illustre aussi la possibilité døintroduire des éléments du *Cadre* dans la carte des sections.

En cliquant sur un des carrés, on active løaffichage de son contenu textuel dans la zone døédition au bas de løonglet **SECTION**.

CLESTHIA>SYLED>CLA²T

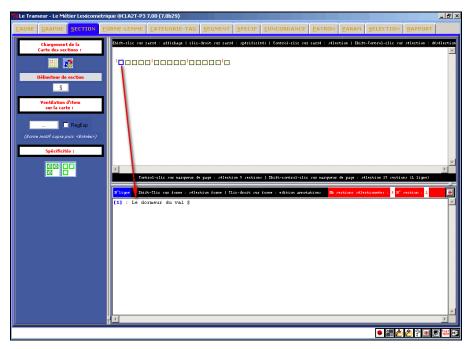


Figure 27: Le Trameur Onglet SECTION, édition de la Trame par une section

Paramétrage de løaffichage de la carte des sections



Figure 28 : Le Trameur Onglet SECTION, paramétrage de løaffichage

Le bouton active le chargement de la carte. Par défaut, cet affichage insère linéairement les sections et sans alignement à gauche après insertion døun marqueur de partie. De plus, un séparateur de blocs de sections est inséré après affichage de 5 sections : ce séparateur est utilisable ensuite pour sélectionner les sections (cf infra).

• Alignement à gauche

Ce bouton est inhibé par défaut, son activation (*clic-droit*) permet de représenter la carte en alignant les sections à gauche après chaque marquage de partie. Désactivation de cette fonction par un *clic* sur le bouton

• Insertion døun marqueur de bloc de sections

Ce bouton est actif par défaut, son inhibition permet de ne pas afficher le marqueur de blocs de sections.

Pour inhiber cette fonction: clic-droit

Pour løactiver : clic

On verra infra (cf Chantiers en cours) une illustration de løutilisation de ces 2 derniers boutons.

11.2 La carte des sections : une vue partielle de la Trame

Si on considère cette représentation du même texte (ici, format TEI) :

```
<?xml version="1.0" encoding="iso-8859-1"?>
      <?xml-stylesheet type="text/xsl" href="q2.xsl"?>
4
   自由
          <teiHeader>
5
              <fileDesc>
                  <titleStmt>
6
                      <title>Poésies</title>
8
                      <author>Arthur Rimbaud</author>
9
                  </fileDesc>
          </teiHeader>
11
              <aroun>
14
                  <text>
15
                      <front>
16
                          <head>Le Dormeur du Val</head>
17
                          <dateline>7 octobre 1870</dateline>
18
                      </frent>
    中中中
19
                      <body:
                           <1a>
21
                               <1g>
                                   <1>C'est un trou de verdure où chante une rivière</1>
                                   <1>Accrochant follement aux herbes des haillons</1>
24
                                   <1>D'argent ; où le soleil, de la montagne fière,</1>
25
                                   <1>Luit : c'est un petit val qui mousse de rayons.</1>
26
27
                               <1q>
28
                                   <1>Un soldat jeune, bouche ouverte, tête nue,</1>
29
                                   <1>Et la nuque baignant dans le frais cresson bleu,</1>
                                   <1>Dort ; il est étendu dans l'herbe, sous la nue,</1>
31
                                   <1>Pâle dans son lit vert où la lumière pleut.</1>
                               </1q>
33
                               <1g>
34
                                   <1>Les pieds dans les glaïeuls, il dort. Souriant comme</1>
35
                                   <1>Sourirait un enfant malade, il fait un somme :</1>
36
                                   <1>Nature, berce-le chaudement : il a froid.</1>
37
                               </1g>
38
                               <1g>
39
                                   <1>Les parfums ne font pas frissonner sa narine ;</1>
40
                                   <1>Il dort dans le soleil, la main sur sa poitrine</1>
41
                                   <1>Tranquille. Il a deux trous rouges au côté droit.</1>
42
43
                          </1a>
                       .
</body>
44
45
                  </text>
46
              </group>
          .
</text>
47
```

Figure 29 : Le Trameur Onglet SECTION, texte TEI sans caractère « délimiteur de sections » explicite

Dans cet exemple, aucun caractère nœst disponible explicitement pour mettre au jour de manière systématique des sections.

Cependant, certains éléments du *Cadre* permettent de délimiter des sections particulières :

- Løelément <1>...</1> permet de mettre au jour les vers (un à un) ; dans ce cas cet élément peut définir une section correspondant à un vers
- Løélément <lg>...</lg> permet de mettre au jour les strophes (une à une) : dans ce cas cet élément peut définir une section correspondant à une strophe.

CLESTHIA>SYLED>CLA²T

La mise en à uvre dans le *Trameur* de cette prise en charge des éléments du *Cadre* pour construire la carte des sections se déroule de la manière suivante :



Figure 30 : Le Trameur Onglet SECTION, sections définies implicitement par un élement du Cadre (avec affichage døune autre partie : 1g)

Dans la figure précédente, la case à cocher « Partie » permet de spécifier que la construction de la carte utilise un élément du *Cadre* (une balise) ; le nom de la balise « 1 » à utiliser est inséré dans la zone de saisie. Comme précédemment, la partie lg est utilisée pour marquer certaines parties dans la carte construite.

Dans lœxemple qui suit, la balise utilisée pour la représentation de la carte est la balise lg:



Figure 31 : *Le Trameur* Onglet SECTION, sections définies implicitement par un élement du *Cadre* (avec affichage døune autre partie : body)

A la différence de la carte des sections définie précédemment, la carte des sections construite via un élément du *Cadre* offre ainsi **une vue partielle** de la *Trame* : celle induite par les différentes localisations des parties associées à lælément sélectionné dans le *Cadre*.

11.3 La carte des sections : masquage de sections

Quelque soit le mode sélectionné pour afficher la carte des sections (caractère délimiteur ou élement du cadre), il est possible de masquer temporairement løaffichage de sections pour une ou plusieurs parties données. Dans løexemple qui suit, la carte des sections est initialement dessinée en affichant toutes les sections :

```
(+/-): marguage/affichage des sections en cliquant sur la partie

- mois 01

- mois 02

- mois 03

- mois 04

- mois 05

- mois 05

- mois 06

- mois 06

- mois 07
```

Figure 32: Le Trameur Onglet SECTION, masquage de sections (1)

Løentête de la figure précédente mentionne la possibilité de masquer les sections des parties dans le nom est précédé du caractère + (réciproquement, døafficher les sections des parties dont le nom est précédé du caractère -). La figure qui suit illustre cette fonctionnalité : toutes les parties sont masquées sauf deux.

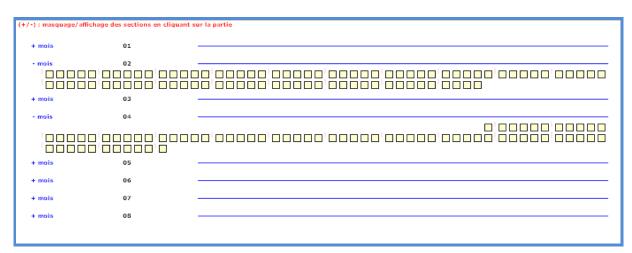


Figure 33 : Le Trameur Onglet SECTION, masquage de sections (2)

Pour inhiber tous les masquages, il faut réinitialiser la carte des sections via le bouton :



12 Accès à la *Trame* par la carte des sections : principales fonctionnalités

12.1 Edition de section

Une fois la carte construite, les sections sont éditables dans léditeur de l'onglet **SECTION**

- Soit en cliquant sur la section visée
- Soit en activant la combinaison *shift+clic* : cette solution permet de maintenir la sélection de la section visée pour des calculs ultérieurs (*cf infra*)

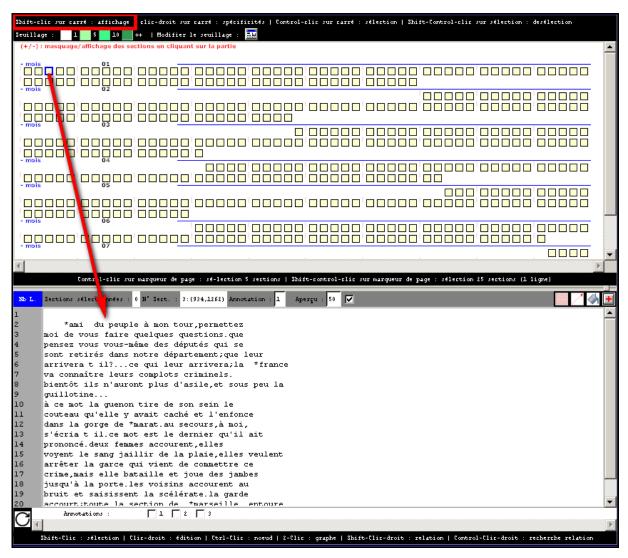
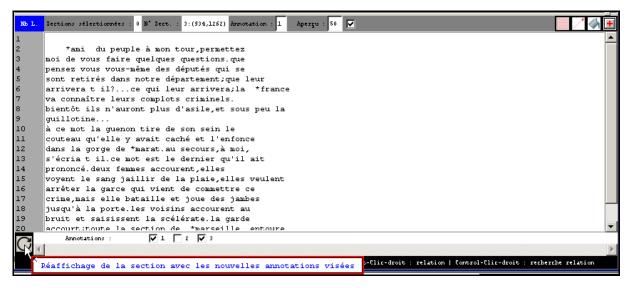


Figure 34 : Le Trameur Onglet SECTION, édition de section

On peut ensuite recharger l\(\phi\) affichage du contenu de la section en s\(\text{electionnant les annotations} \) à visualiser :



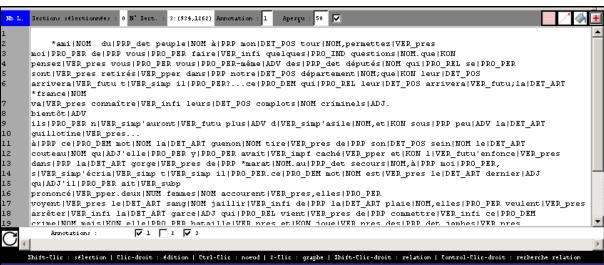


Figure 35 : Le Trameur Onglet SECTION, édition de section, sélection des annotations

12.2 Recherche de chaînes

On dispose dans l'onglet **SECTION** døune zone de saisie permettant de lancer des recherches de chaînes : "*Ventilation de*". Pour rechercher une forme, il convient de løinsérer dans la zone de saisie puis de lancer la recherche en tapant la touche « Entrée ».

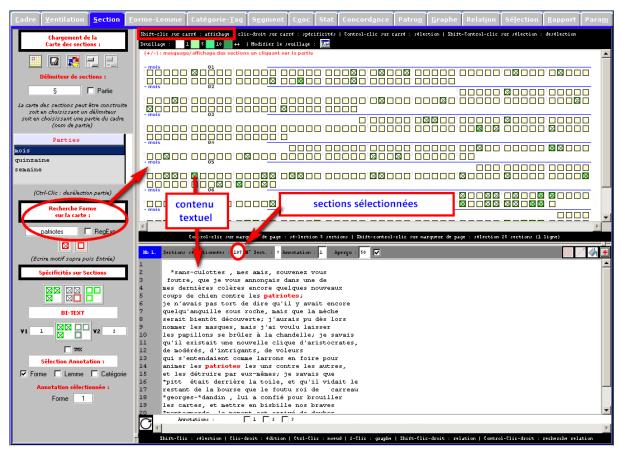


Figure 36 : Le Trameur Onglet SECTION, recherche de chaînes

Dans la figure précédente, la forme cherchée (dormeur) est présente dans la section représentée par le carré marqué døune croix matérialisant la présence de la forme cherchée. La coloration verte de cette section signifie que celle-ci est désormais sélectionnée pour des opérations spécifiques sur cette section (calcul de spécificité, recherche de cooccurrents, *cf infra*).

Le contenu textuel est visible *via* le raccourci clavier *shift-clic* sur le carré associé à la section : ce raccourci permet de maintenir cette sélection sélectionnée contrairement au simple *clic* sur la section qui permet son affichage mais aussi sa désélection (*cf supra*).

On peut aussi lancer une recherche en sélectionnant une forme dans les dictionnaires disponibles (et *via* le raccourci clavier défini).

12.3 Recherche « incrémentale » de chaînes

Un mécanisme similaire à celui présenté ci-dessus sur løéditeur de la *Trame* dans løonglet **CADRE** est disponible sur løonglet **SECTION**. Le raccourci clavier *Control-r* active ce mode de recherche permettant de définir *incrémentalement* un motif de recherche.

12.4 Visualisation des différentes annotations des items de la Trame

Un mécanisme similaire à celui présenté ci-dessus sur lœditeur de la *Trame* dans lønglet CADRE est disponible sur lønglet SECTION.

Løéditeur de la *Trame* de løonglet **SECTION** donne à voir les items de la *Trame* et les annotations associées.

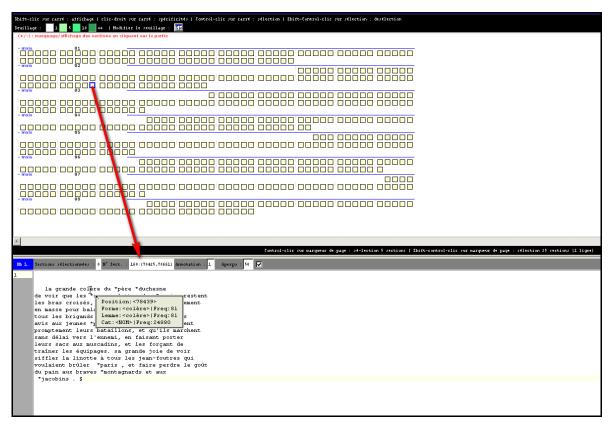


Figure 37 : Le Trameur Onglet SECTION, recherche de chaînes

12.5 Edition et modification des items de la Trame

De la même manière, on peut modifier les items de la *Trame* à travers lœditeur de la *Trame* de lønglet **SECTION**. Un *clic-droit* sur un item déclenche løncès aux différentes annotations de løntem visé, toute modification sera enregistrée dans les dictionnaires associés.

12.6 Le vocabulaire spécifique døune section

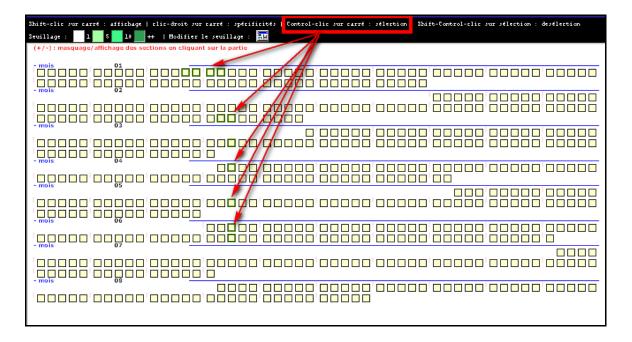
Un *clic droit* sur un carré déclenche le calcul du vocabulaire spécifique¹⁰ de la section considérée.

```
.
Specificites du vocabulaire sur la partie choisie : <Section n°457 (227227,227507) | delimiteur de section : §
Nombre d'occurrences du texte global = 142179
Nombre d'occurrences dans la partie visee = 128
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit
C:/SERGEF~1/SF-WORK/O7-DEVEL/ANR-TE~1/MKGRAP~1/textes/duchn/export/resultspecif-source-1417428081.txt
                                                           Ind-Specif
                                                                            Fg-Totale
on
                                                           4.9
                                                                            856
plus
                                                                            798
-
entraîner
                                                           4.0
                                                                            1
sali
                                                           3.7
persécute
                                                           3.6
sachant
                                                           3.4
naufrage
                                                           3.4
égarer
```

Figure 38 : Le Trameur Onglet SECTION, les mots spécifiques donne section

12.7 Le vocabulaire spécifique døune sélection de sections de la carte

Il est possible de sélectionner une série de sections dans la carte des sections et de calculer le vocabulaire spécifique¹¹ de cette sélection. Pour sélectionner une section, il convient doactiver la combinaison de touches : Control+clic. Après avoir été sélectionné, le contour du carré associé à la section devient vert. Après avoir sélectionné une série de sections, on peut activer le calcul du vocabulaire spécifique de cette sélection via le bouton disponible dans longlet **SECTION.**



¹⁰ Cf Partie « Vocabulaire spécificique » pour le paramétrage de ce calcul

¹¹ idem

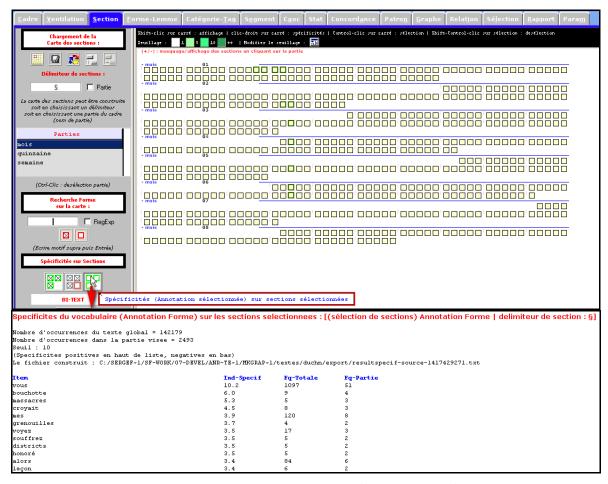
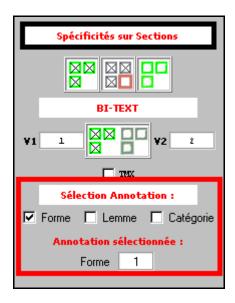


Figure 39 : Le Trameur Onglet SECTION, les mots spécifiques døune sélection de sections

Ce calcul peut être fait sur tous les niveaux donnotation de la base chargée :

Avant de lancer ce calcul, il est possible de sélectionner løannotation utilisée pour réaliser le calcul. Pour cela il suffit de sélectionner løannotation visée :

• soit en cochant løannotation visée forme, lemme ou catégorie



• soit en donnant le numéro de løannotation visée (puis entrée) dans la la zone de saisie

12.8 Le vocabulaire spécifique des sections de la carte contenant un motif : recherche de cooccurrents

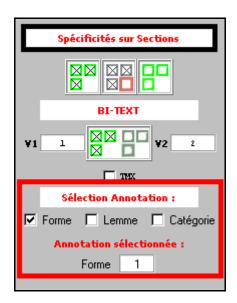
Lønglet **SECTION** dispose døun bouton permettant de calculer les cooccurrents døun item ou døun motif donné.

recherche des cooccurrents døun item						
(calcul des spécificités sur les sections contenant un motif donné)						

Ce calcul peut être fait sur tous les niveaux donnotation de la base chargée :

Avant de lancer ce calcul, il est possible de sélectionner løannotation utilisée pour réaliser le calcul. Pour cela il suffit de sélectionner løannotation visée :

• soit en cochant lænnotation visée forme, lemme ou catégorie



• soit en donnant le numéro de loannotation visée (puis entrée) dans la la zone de saisie

Recherche døun motif sur la carte puis recherche de cooccurrence :

A partir de la carte des sections, on peut rechercher un item ou un motif. On obtient une série de carré marqué par une croix indiquant la présence de løtem ou du motif. On peut ensuite lancer un calcul mettant au jour le vocabulaire spécifique de lønsemble de ces sections. On obtient le vocabulaire spécifique de l'ensemble des sections contenant løtem ou le motif initial : liste døtems qui portent soit un indice de spécificité positif soit un indice de spécificité négatif. Dans le premier cas, on aboutit en gros à une liste des cooccurrents de løtem ou du motif cherché, dans le second cas on obtient des items qui n'apparaissent pas ou peu avec løtem ou le motif cherché.

-

¹² Cf Partie « Vocabulaire spécificique » pour le paramétrage de ce calcul

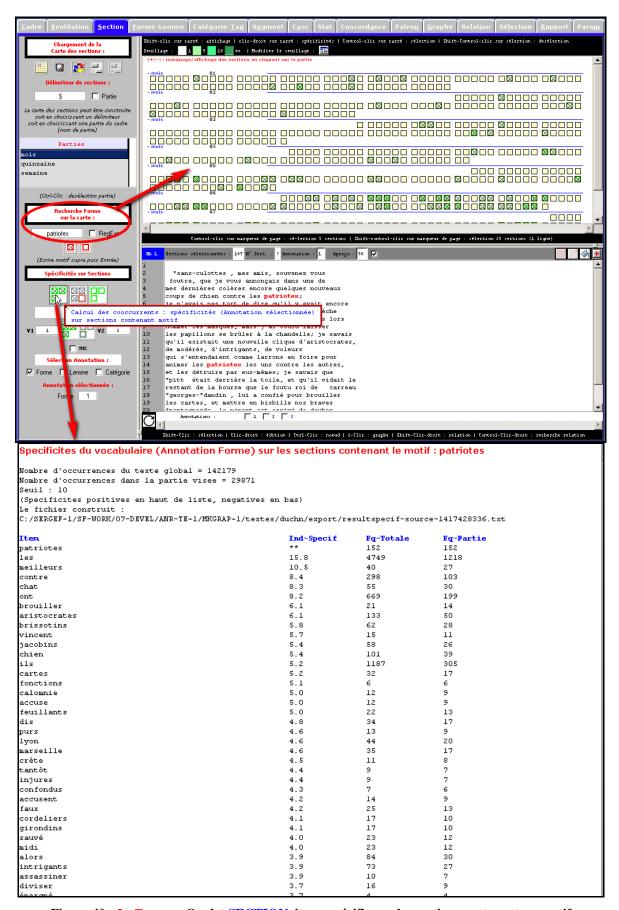


Figure 40 : *Le Trameur* Onglet SECTION, items spécifiques des sections contenant un motif (cooccurrents)

12.9 Le vocabulaire spécifique des sections de la carte ne contenant pas un motif

Un calcul de spécificité peut aussi être lancé sur les sections ne contenant pas le motif cherché dans la carte des sections :

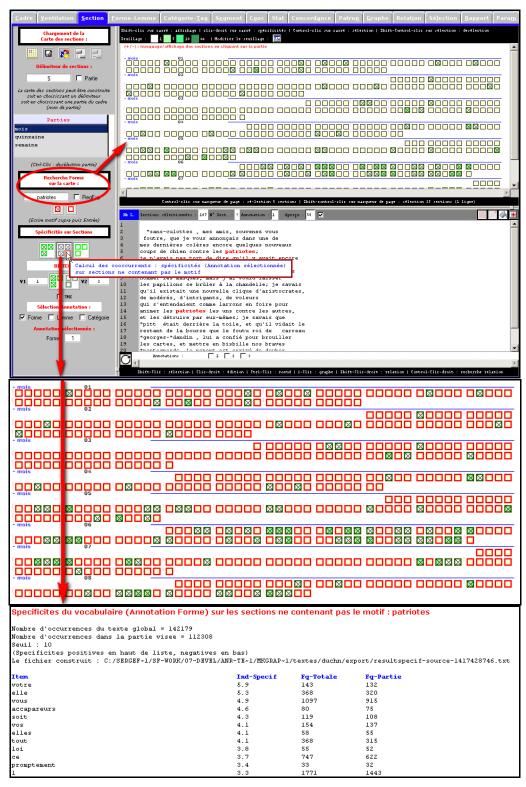


Figure 41 : Le Trameur Onglet SECTION, items spécifiques des sections ne contenant pas un motif

Là encore, ce calcul peut être fait sur tous les niveaux døannotation de la base chargée : cf supra.

12.10 Fonctionnalités complémentaires sur la carte des sections

Control-Clic sur le trait vertical entre 2 séries de 5 sections : sélection des 5 sections à sa droite



ShiftóControl-Clic sur ce même trait vertical : sélection des 25 sections à sa droite (une ligne dans la carte des sections)



Shift-Control-Clic sur un carré dans la carte des sections : désélection de la section visée.

13 Concordances

Lønglet CONCORDANCE permet døafficher les concordances des différentes annotations disponibles sur chacun des items de la *Trame*: forme, lemme, catégorie, annotations complémentaires. Les concordances produites sont disponibles dans 2 sous-onglets distincts: un sous-onglet avec édition des items de la concordance, un sous-onglet permettant de trier les contextes de la concordance.



	te-partie Tr <u>i</u> -Co	ncordance		
Ĵt	D			
n°	Partie	ContexteGauche	Pôle	ContexteDroit
1	mois=01	dernier que nos ennemis doivent porter aux	patriotes	les mêmes jean-foutres qui ont tant
2	mois=01	constitution va être partout acceptée;bientôt les	patriotes	de *lyon et de *
3	mois=01	tous les échappés de *coblentz .les	patriotes	, en se rencontrant,pleurent et gémissent;
4	mois=01	grossir sont d'un côté et les	patriotes	de l'autre;le combat est commencé
5	mois=01	et il serait à désirer que les	patriotes	fussent d'un aussi bon accord que
6	mois=01	l'eau froide.aussi,foutre,tous les	patriotes	sont ils à présent sur leurs gardes
7	mois=01	vive *louis-XVII ;les meilleurs	patriotes	seront massacrés;les scélérats qui mitonnent cette
8	mois=01	toutes nos villes de guerre; de faux	patriotes	qui s'entendent avec nous, brouilleront
9	mois=01	;personne ne s'entendra que nous;les	patriotes	s'armeront les uns contre les autres
10	mois=01	royal ,qui tirent la langue sur les	patriotes	et qui ne cessent de faire de
11	mois=02	calme et dissipera toutes les inquiétudes, nos	patriotes	iront rendre de pareilles visites aux accapareurs
12	mois=02	je lancerais la foudre sur les faux,	patriotes	;*montagnards, il est des traîtres parmi
13	mois=02	été opprimés par cette foutue canaille;les	patriotes	ont été jetés dans des cachots,et
14	mois=02	les uns contre les autres,les faux	patriotes	les royalistes,les feuillants,les fédéralistes,n
15	mois=02	ils ont accusé les magistrats les plus,	patriotes	de tous les maux qu'ils faisaient
16	mois=02	ils bravent ,ils insultent,ils outragent les	patriotes	depuis longtemps ils se disposent à un
17	mois=02	l'ombre les poignards pour égorger les	patriotes	si nous avions différé de quelques
18	mois=03	, du soir au matin, gouaillaient les	patriotes	, qui se vantaient dans les cafés,
19	mois=03	'est toi qui as fait massacrer les	patriotes	de *marseille et de *
20	mois=03	tous les traîtres,de tous les faux	patriotes	de tous les royalistes traitons les comme,
21	mois=03	avaient juré la mort de tous les	patriotes	je ne m'attends pas que l
22	mois=03	*vendée , qui ont fait massacrer les	patriotes	à *lyon ,et à *
23	mois=03	en défendant ce traître, quand tous les,	patriotes	quand toutes les sociétés populaires demandaient sa
24	mois=03	seul jour de triomphe pour que les	patriotes	perdent le goût du pain dans toute
25	mois=04	maudite assemblée nationale,et à tous les	patriotes	qui lui ont rogné les ongles,
26	mois=04	tous les scélérats qui ont égorgé les,	patriotes	le 10 août;c'est vous,
27	mois=04	jeté le chat aux jambes des meilleurs	patriotes	, quand toutes vos manigances ont été
28	mois=04	manoeuvres,lâches et méprisables coquins,que les	patriotes	de *marseille , de *bordeaux ,
29	mois=04	ont été vite en besogne avec les	patriotes	de *marseille et de *lyon
30	mois=04	murs que des traîtres à exterminer. les	patriotes	y ont tous été égorgés, c'
31	mois=04	vient d'être jugé; voilà comme les	patriotes	répondent à la calomnie. à ton
32	mois=04	brissotins n'avait pas été quillotinés, les	patriotes	l'auraient été, vous faites bien

Figure 42 : Le Trameur Onglet CONCORDANCE, affichage édition + affichage tri

13.1 Modes de déclenchement døune concordance

Ces concordances peuvent être réalisées :

- à partir des dictionnaires, døune liste de segments répétés, døune liste de séquences de termes etc. en sélectionnant les items visés dans la liste visée et en activant la requête de concordance à partir du bouton disponible dans løonglet associé,
- ou en renseignant directement les zones de saisie idoines décrites dans les figures qui suivent.

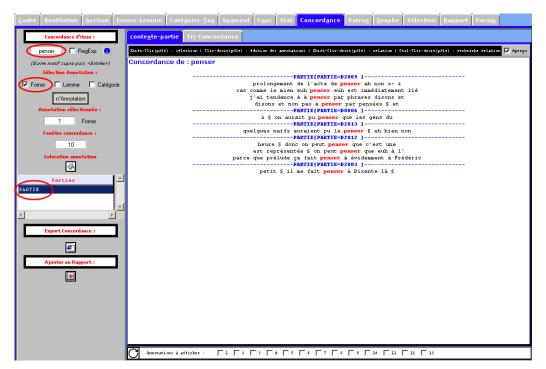


Figure 43: Le Trameur Onglet CONCORDANCE (forme)

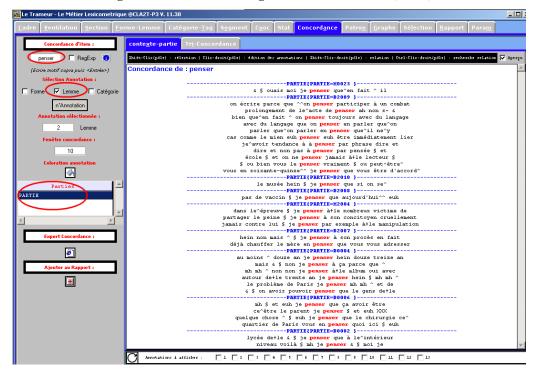


Figure 44 : Le Trameur Onglet CONCORDANCE (lemme)

CLESTHIA>SYLED>CLA2T

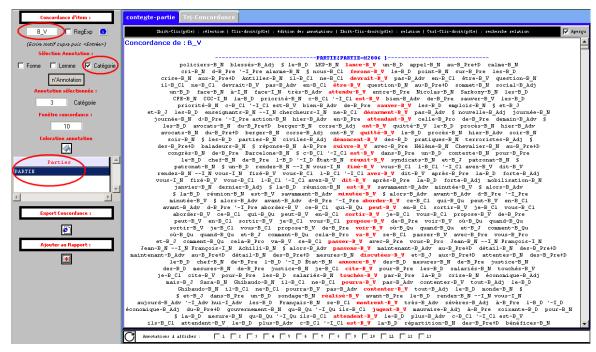


Figure 45 : Le Trameur Onglet CONCORDANCE (catégorie)

13.2 Paramétrage døune concordance

Løaffichage de la concordance peut être paramétré de la manière suivante :

- choisir une ou plusieurs parties
- déterminer la longueur de la fenêtre graphique de la concordance (par défaut 10 formes à droite et 10 à gauche)
- choisir le type de forme à rechercher en cochant løune des cases : Forme, Lemme, Catégorie (ou annotation importée *cf infra*)
- choisir løitem visé (une forme, un lemme, une valeur de catégorie ou une annotation importée)

Dans les exemples précédents, løaffichage demandé est celui døune forme, døun lemme puis døune catégorie en tenant compte de la seule partition disponible.

On peut aussi activer la concordance en utilisant le mode RegExp: la recherche de la forme visée se fait en écrivant un motif exprimé sous la forme døune expression régulière.

Une fois la concordance chargée, il est possible de la recharger en modifiant les annotations à afficher pour cette concordance. On trouve au bas de la fenêtre du concordancier une liste de cases à cocher, chacune dœlle est associée à une des annotations disponibles : en activant tel ou tel bouton, le rechargement de la concordance intègre løaffichage des annotations sélectionnées.

Si on considère cette première concordance (seule løannotation n°1 (forme) est affichée):

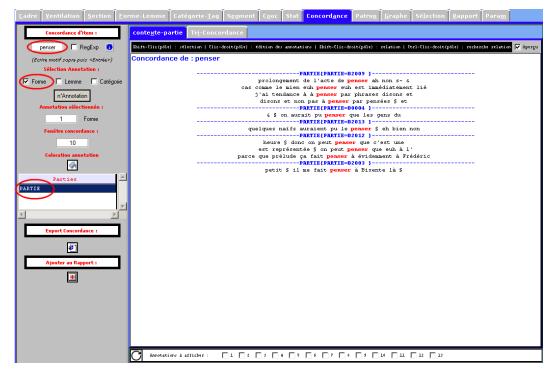


Figure 46 : Le Trameur Onglet CONCORDANCE (forme)

La sélection de nouvelles annotations et le rechargement de la concordance modifie løaffichage de cette concordance :

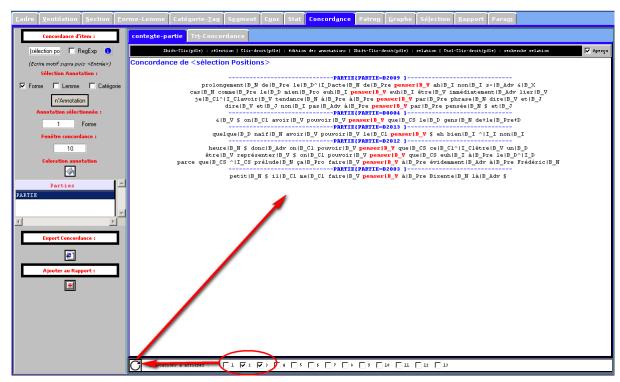


Figure 47 : Le Trameur Onglet CONCORDANCE (rechargement)

En activant le bouton 🗐, il est possible dœxporter une concordance au format HTML.

13.3 Editeur de concordance

Comme dans lønglet CADRE et dans lønglet SECTION, il est possible dans lønglet CONCORDANCE de modifier les annotations associées à tous les items de la concordance affichée.

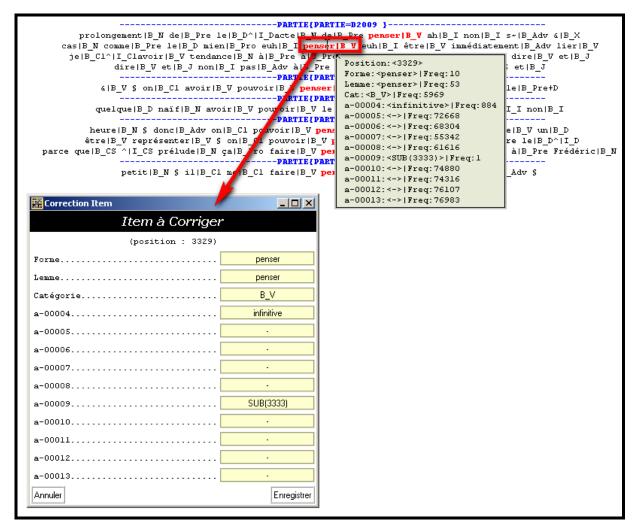


Figure 48: Le Trameur Onglet CONCORDANCE, édition et modification døun item pivot

Toute modification entraîne le rechargement des différentes ressources impactées par ces modifications.

14 Marquage des annotations de la *Trame*

14.1 Annotation « simple »

Le *Trameur* dispose de 3 éditeurs vus *supra* :

- Editeur du *Cadre*
- Editeur de la Carte des sections
- Editeur du concordancier

Ces 3 éditeurs disposent døune fonctionnalité permettant de colorer dans la zone dødition les items possédant une annotation donnée. Cette fonctionnalité est activée par le bouton présent dans chacun des onglets contenant ces éditeurs.

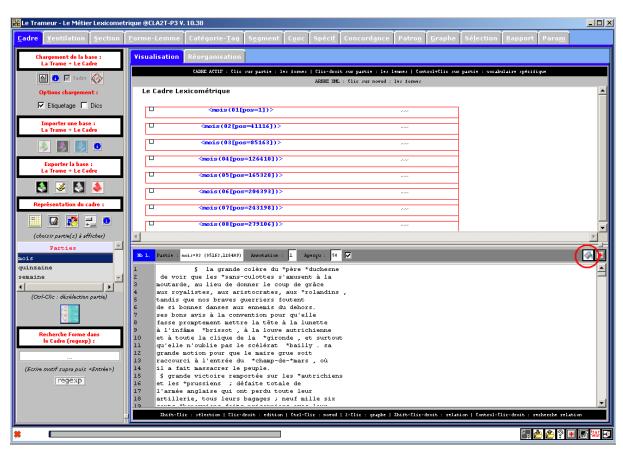


Figure 49: Le Trameur Editeur CADRE

CLESTHIA>SYLED>CLA2T

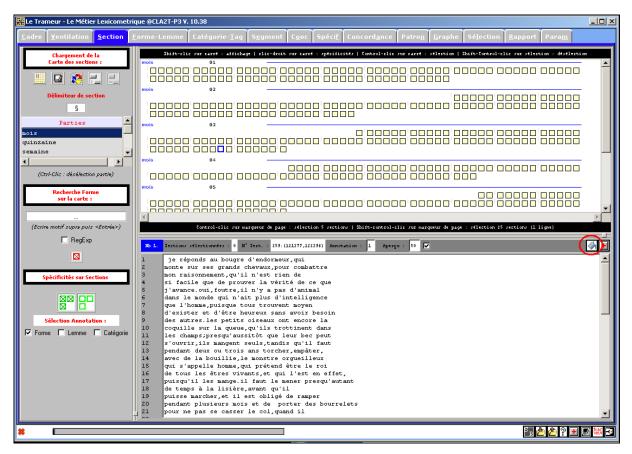


Figure 50: Le Trameur Editeur SECTION

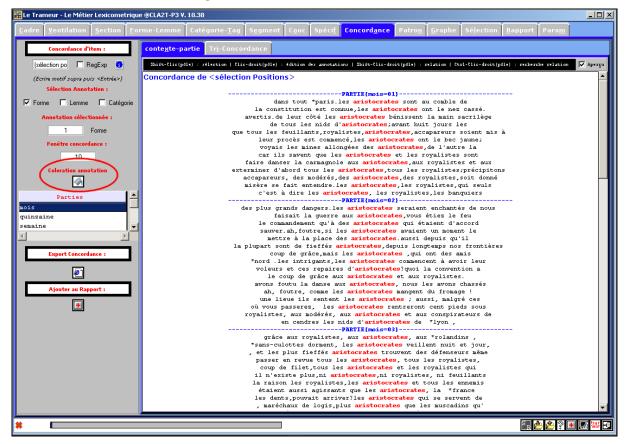


Figure 51: Le Trameur Editeur CONCORDANCE

Løactivation du bouton déclenche le menu suivant :



Figure 52 : Le Trameur Sélection des annotations à colorier (1)

Ce menu permet de choisir la valeur de løannotation à colorier, le numéro de løannotation correspondante et la couleur à utiliser : tous les niveaux døannotation sont accessibles pour cette fonctionnalité.

Par exemple, dans la figure suivante, on choisit løétiquette « NOM » associée à løannotation n°3 et la couleur visée par la palette de choix :

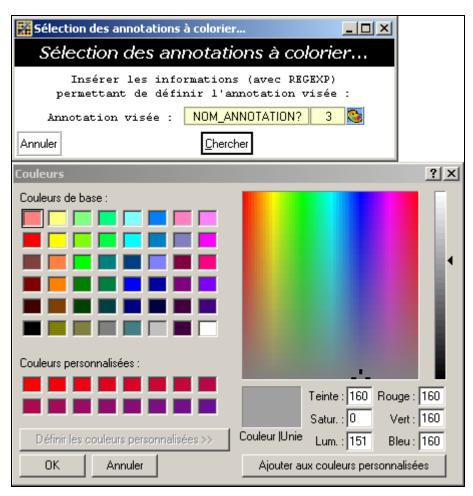


Figure 53 : Le Trameur Sélection des annotations à colorier (2)

CLESTHIA>SYLED>CLA²T

Le résultat dans les différents éditeurs est le suivant :

Editeur de la Carte des sections

```
Sections sélectionnées : 0 N° Sect. : 259:(121277,122296) Annotation : 1 Aperçu : 50 🔽
je réponds au bougrs d'endormeur, qui monte sur ses grands chevaux, pour combattre mon raisonnement, qu'il n'est rien de si facile que de prouver la vérice de ce que l'evance, oui foutre, il n'ys pas d'aminal dans le monde qui n'ait plus d'intelligence que l'honse, puisque tous trouvent moyen d'exister et d'être heureux sans avoir besoin des autres, les petits piseaux ont encore la coquille sur la queue, qu'ils trottinent dans les champs presqu'aussitôt que leur bed peut s'ouvrir, ils mangent seuls, tandis qu'il faut pendant deux ou trois ans torcher, empâter, avec de la bouillie, le monstre orqueilleux qui s'appelle homme, qui prétend être le roi de tous les stres vivants, et qui l'est en effet, puisqu'il les mange.il faut le mener presqu'autant de temps à la lisiere, avant qu'il puisse marcher, et il est obligé de ramper pendant plusieurs mois et de porter des bourrelets pour ne pas se casser le col, quand il
```

Figure 54 : Le Trameur Marquage des annotations sur les items édités (1)

Editeur du *Cadre*

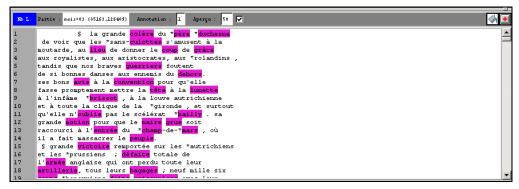


Figure 55 : Le Trameur Marquage des annotations sur les items édités (2)

Editeur du concordancier

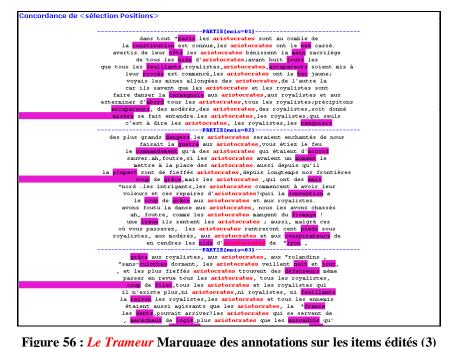


Figure 56 : Le Trameur Marquage des annotations sur les items édités (3)

On peut « empiler » les couleurs et marquer plusieurs annotations par un système de couleurs variées dans la zone dédition : on choisit ci-dessous une autre valeur déannotation (ADJ) et on applique la coloration dans lééditeur de la carte des sections :

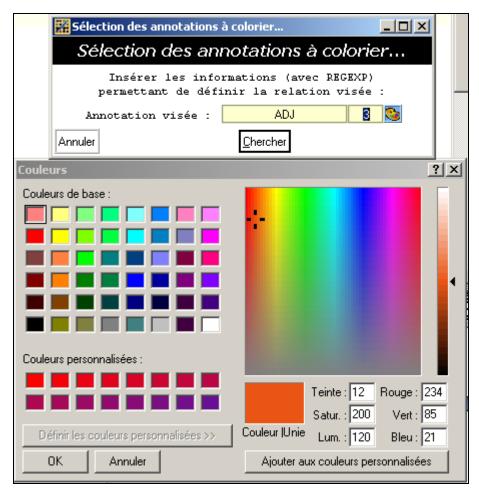


Figure 57 : Le Trameur Sélection døune nouvelle annotation à colorier

Le résultat dans les éditeurs est le suivant :

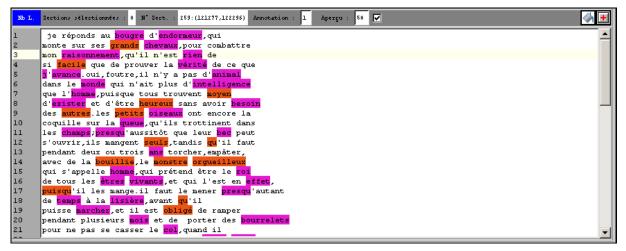


Figure 58 : Le Trameur Marquage des annotations sur les items édités

14.2 Annotation de type « relation »

Si løannotation recherchée est de type « relation » entre 2 items (cf Graphe de relations infra), la source et la cible de la relation sont marquées dans løéditeur.

Pour illustrer cette fonctionnalité, on utilise dans læxemple qui suit la base textométrique intitulée <u>Base "Corpus de contes français"</u> (Fairy Tales Corpus, FTC) construite à partir des ressources développées par <u>Ismaïl El Maarouf</u> (base décrite en détail dans la partie « Graphes de relations *infra*).

On considère par exemple ici la section suivante de cette base :

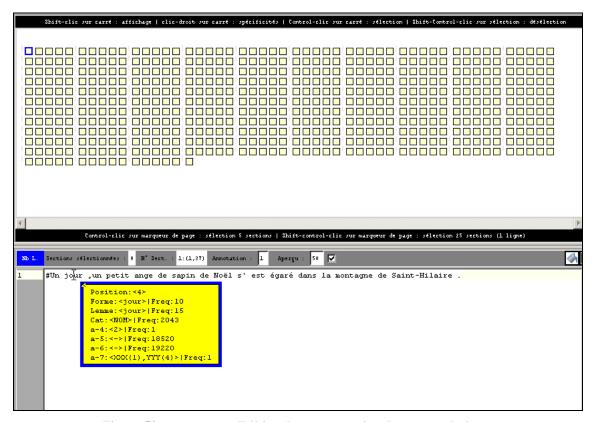


Figure 59: Le Trameur Edition døune annotation de type « relation »

Dans celle-ci, løtem en position n°4 (la forme « jour ») possède des annotations suivantes :

- Annotation n°1 (forme): jour
 Annotation n°2 (lemme): jour
- Annotation $n^{\circ}3$ (POS): NOM
- Annotation n°4 (identifiant): 2 (cette annotation permet dødentifier chaque item par un identifiant unique. Remarque: la position de chaque item est une autre manière dødentifier les items de la *Trame*)
- Annotation n°5 : annotation référentielle non définie
- Annotation n°6: annotation sémantique non définie
- Annotation $n^{\circ}7$: XXX(1), YYY(4)

Cette dernière annotation induit des relations sur loitem courant : par exemple celle notée YYY (4). Cette valeur døannotation signifie que løtem courant est en relation de type YYY avec un autre item de la *Trame* (celui dont lødentifiant est 4)

Si on souhaite afficher par exemple les relations de type YYY, løactivation du bouton déclenche le menu suivant :



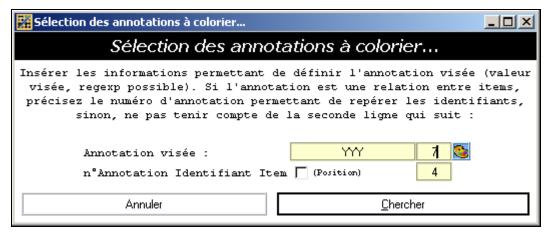


Figure 60 : Le Trameur Sélection døune relation à colorier

La première ligne permet comme précédemment de définir la valeur de lonnotation visée (YYY stockée dans løannotation n°7)

La seconde permet de sélectionner løannotation dans laquelle sont enregistrés les identifiants des items:

- soit lødentifiant est stocké dans un niveau døannotation particulier : cøest le cas ici, løannotation n°4 a été construite pour enregistrer un identifiant unique sur chaque item de type non délimiteur
- soit il correspond à la position de løtem (cette identification par la position peut être considérée comme løidentification par défaut et par construction de chacun des items de la *Trame*)

Le résultat produit est cette fois-ci le suivant :

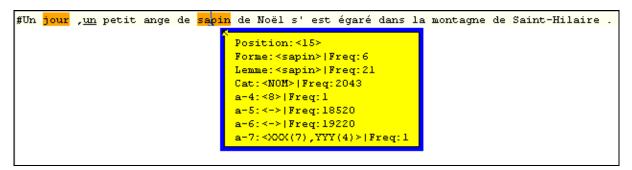


Figure 61 : Le Trameur Marquage des annotations de type « relation » sur les items édités

Løtem en position n°4 est colorié (source de la relation YYY (4)) Løtem associé à identifiant n°4 (cible de la relation) est souligné (la forme « un ») Dans cet exemple, lettem en position n°15 a une relation similaire vers la même cible.

15 Vocabulaire spécifique

Løanalyse des spécificités permet de porter un jugement sur la fréquence de chacune des unités textuelles dans chacune des parties du corpus¹³.

Tous les résultats donn calcul de spécificités sont insérés dans longlet **STAT** (ou **SPECIF**). Ils sont associés à des processus de calcul disponibles dans différents onglets.

Tous les modules intégrant un calcul de spécificité peuvent déclencher ce calcul sur tous les niveaux d'annotation disponibles dans la base en cours d'analyse : <u>tous les niveaux</u> d'annotation ajoutés, importés etc. sont accessibles pour le calcul des spécificités.

Localisation du paramétrage de l@annotation pour le calcul des spécificités et modules associés :

Onglet SECTION

- calcul des spécificités des sections sélectionnées
- calcul des spécificités des sections contenant un motif donné (i.e cooccurrent du motif visé)

Onglet STAT (ou SPECIF)

- calcul des spécificités sur une partie donnée (sélectionnée via PCLC)
- spécificités totales sur une partition (TFG)
- AFC

Onglet SELECTION - sélection sur *Cadre* (i.e sélection de parties)

• calcul des spécifs sur une partie donnée

Le paramétrage de løannotation visée est similaire dans ses trois onglets :

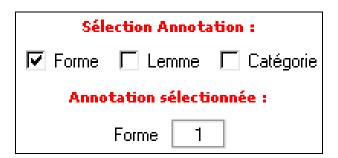


Figure 62 : Le Trameur Sélection déannotation pour calcul de spécificités

3 cases à cocher permettent de sélectionner les annotations : forme (annotation $n^\circ 1$), lemme (annotation $n^\circ 2$) ou catégorie (annotation $n^\circ 3$). Pour les autres niveaux døannotation, il suffit de donner son numéro dans la zone de saisie (puis entrée), le nom de løannotation choisie sera visible en regard de son numéro : dans la figure précédente, løannotation $n^\circ 1$ (forme) est sélectionnée.

 $^{^{13}}$ Sur la méthode des spécificités on consultera par exemple : (Lafon, 1984) ou en ligne : $\underline{\text{http://weblex.ens-lsh.fr/doc/weblex/contraste.html\#specif}}$

Toutes les opérations (concordance, ventilation, etc) activables sur les résultats de spécificités sont disponibles sur tous les niveaux døannotation (cf infra).

Les résultats døun calcul de spécificités sont insérés dans 2 sous-onglets de løonglet **STAT** (ou **SPECIF**), chacun døux permet døactiver des opérations similaires :

a. sous-onglet **SPECIF**:

Les résultats sont affichés ici en format texte. Les actions disponibles sur les items affichés sont activables par des raccourcis clavier (encadrés en rouge) :

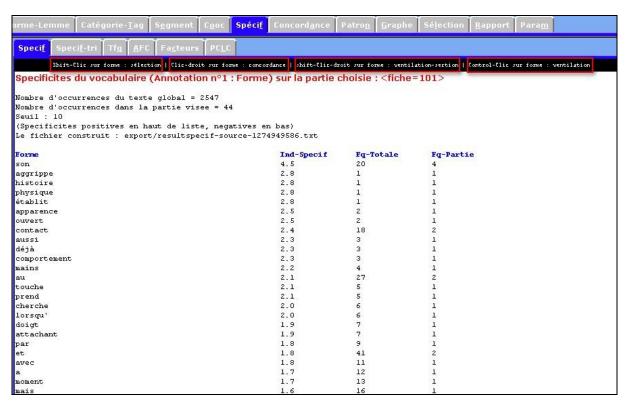


Figure 63 : Le Trameur Opérations disponibles sur les items issus døun calcul de spécificités

- Sélection : tous les items ayant la forme visée sont sélectionnés (cf Gestionnaire de Sélection)
- Concordance: calcul døune concordance sur toutes les positions des items ayant la forme visée
- *Ventilation-section* : ventilation sur la carte des sections de toutes les positions des items ayant la forme visée
- *Ventilation*: ventilation sur le **CADRE** de toutes les positions des items ayant la forme visée

b. sous-onglet **SPECIF-TRI**

Dans cet onglet, les résultats affichés sous forme de colonnes donnent à voir pour chaque item : løitem visé, son indice de spécificité, sa fréquence totale, sa fréquence dans la partie et le numéro døannotation de løitem (1 : forme, 2 : lemme, 3 : catégorie, etc.).

CLESTHIA>SYLED>CLA2T

orme-Lemme Catégorie- <u>T</u> ag S <u>e</u> gment I	C <u>o</u> oc Spéci <u>f</u> Conc	ord <u>a</u> nce	Patro <u>n</u> <u>G</u> r	aphe Sé <u>l</u> ec							
Speci <u>f</u> Speci <u>f</u> -tri Tf <u>q</u> <u>A</u> FC Fa <u>c</u> teurs	PC <u>L</u> C										
Actions sur Résultats spécifs 🔃 🖽 🖟 🗀 Со-présence											
Forme	Ind-Sp	FQ	fq	Num							
son	4.5	20	4	1							
aggrippe	2.8	1	1	1							
histoire	2.8	1	1	1							
physique	2.8	1	1	1							
établit	2.8	1	1	1							
apparence	2.5	2	1	1							
ouvert	2.5	2	1	1							
contact	2.4	18	2	1							
aussi	2.3	3	1	1							
déjà	2.3	3	1	1							
comportement	2.3	3	1	1							
mains	2.2	4	1	1							
au	2.1	27	2	1							
touche	2.1	5	1	1							
prend	2.1	5	1	1							
cherche	2.0	6	1	1							
lorsqu'	2.0	6	1	1							
doigt	1.9	7	1	1							
attachant	1.9	7	1	1							
par	1.8	9	1	1							
et	1.8	41	2	1							
avec	1.8	11	1	1							
a	1.7	12	1	1							
moment	1.7	13	1	1							
mais	1.6	16	1	1							
un	1.6	17	1	1							
ses	1.6	17	1	1							
le	1.6	19	1	1							
on	1.5	20	1	1							

Figure 64 : Le Trameur Opérations disponibles sur les items triables issus døun calcul de spécificités

Dans cet onglet, les processus de calcul sont activables via les boutons disponibles au dessus des listes produites. Il suffit de sélectionner un ou plusieurs items dans la liste puis déactiver le bouton idoine (ventilation, ventilation sur Cadre, ventilation sur carte des sections, concordance).

Dans le cas døune ventilation sur la carte des sections, si løption « co-présence » est activée, les sections mises au jour seront celles contenant lønsemble des formes sélectionnées dans la liste (ET logique).

Enfin, dans cet onglet, il est possible de trier chacune des colonnes en cliquant sur le titre de la colonne visée.

16 Graphiques de ventilation

16.1 Ventilation par partie døun item

Løonglet **VENTILATION** permet døafficher des graphiques de ventilation de formes, de lemmes ou de catégories dans une partition sélectionnée :

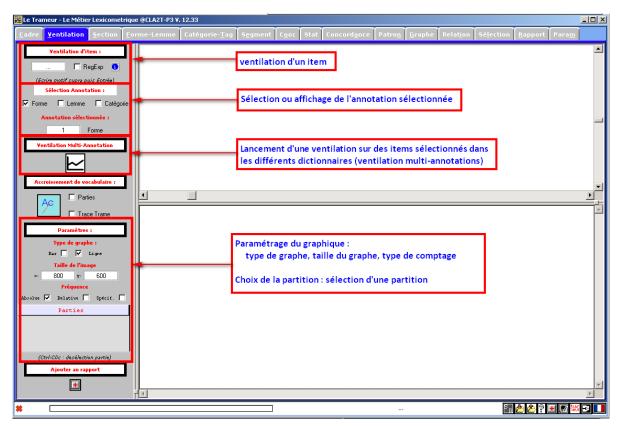


Figure 65 : Le Trameur Onglet VENTILATION paramétrage døune ventilation

Dans læxemple suivant, løtem recherché est celui correspondant à la forme « patriote ».

On peut aussi activer la construction du graphe de ventilation en utilisant le mode RegExp: la recherche de la forme visée se fait en écrivant un motif exprimé sous la forme døune expression régulière.

CLESTHIA>SYLED>CLA²T

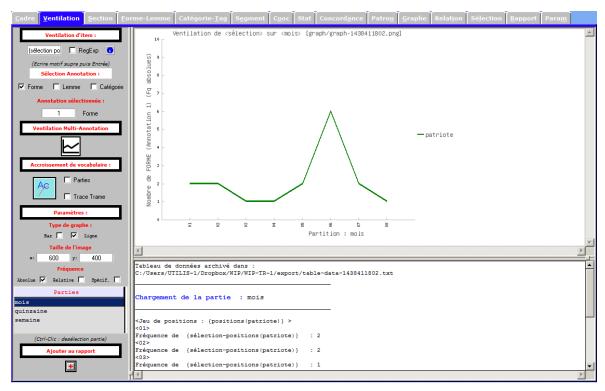


Figure 66 : Le Trameur Onglet VENTILATION affichage døune ventilation (mode ligne)

\bpatriot.*\b représente toutes les formes commençant par patriotø et qui se termine par nømporte quelle chaîne de caractères (dans notre texte de travail, les chaînes couvertes par cette expressions sont par exemple patrioteø patriotesø etc). Le résultat de la ventilation de ce motif est le suivant :

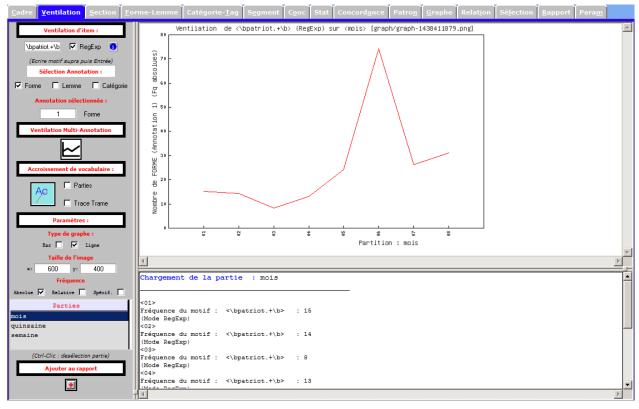


Figure 67: Le Trameur Onglet VENTILATION mode regrexp

16.2 Ventilation par partie døune sélection døitems

Il est aussi possible de construire des graphiques de ventilation døune sélection døtem (forme, lemme, catégorie, segment) à partir des listes disponibles. Dans les figures qui suivent, on sélectionne des formes dans le dictionnaire des formes, puis on active la ventilation de ces formes :

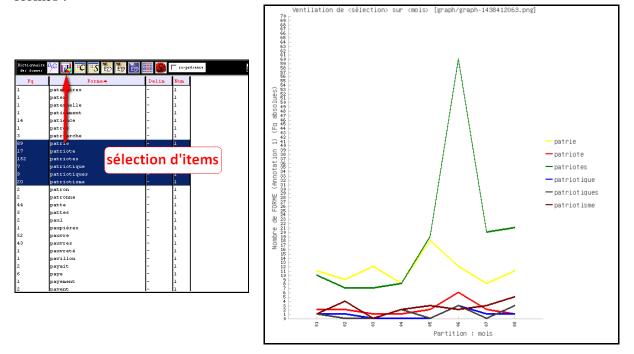
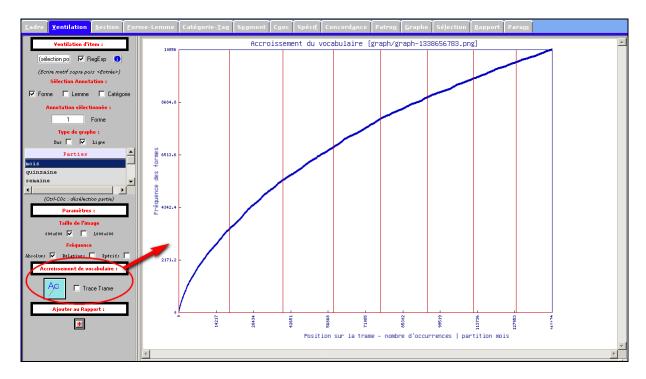


Figure 68 : Le Trameur Ventilation døune sélection døtems

16.3 Courbe døaccroissement du vocabulaire

Løonglet **VENTILATION** permet aussi døafficher la courbe døaccroissement du vocabulaire sur la base de travail :

- soit sur le corpus complet, en choisissant éventuellement une partition
- soit sur chaque partie de la partition sélectionnée



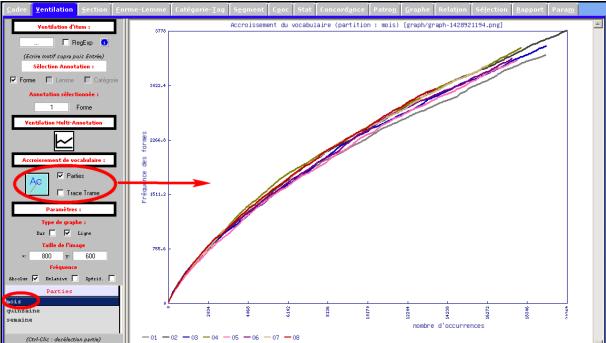


Figure 69 : Le Trameur Onglet VENTILATION, courbe deaccroissement du vocabulaire

16.5 Ventilation døune sélection døitems associés à différents niveaux døannotation

Il est possible de construire un graphique de ventilation sur une sélection døtems relevant de différents dictionnaires (croisement døannotation). Après les avoir sélectionnés, løactivation du bonton visible sur la figure suivante déclenche la construction de cette distribution. Cidessous, un graphique de ventilation donnant à voir la distribution de la forme « patriotes » et du lemme « patriote » :



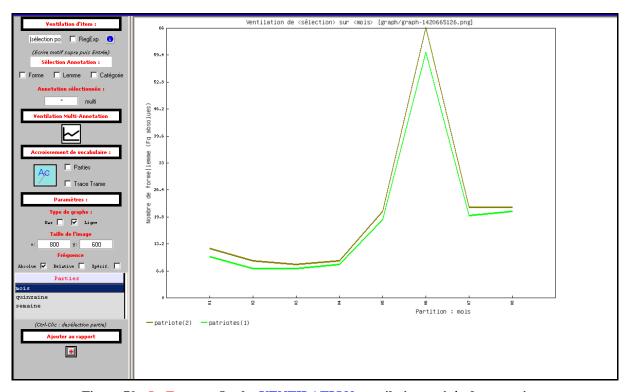


Figure 70 : Le Trameur Onglet VENTILATION, ventilation croisée d@annotations

16.4 Ventilation døune sélection døitems sur une sélection de parties

Le *Gestionnaire de Sélection* permet de sélectionner un certain nombre de parties (*cf infra*). Une fois ces parties sélectionnées, il est possible de construire un graphique de ventilation døune sélection døtems sur cette sous-partition. Les figures suivantes illustrent le processus en 2 temps :

1. Sélection de parties :

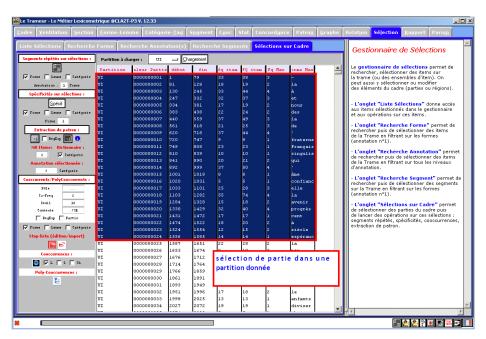


Figure 71 : Le Trameur Onglet VENTILATION, ventilation døune sélection døitems (1)

2. Ventilation døune sélection døéléments sur les parties visées :

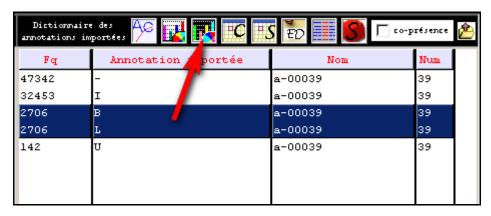


Figure 72 : Le Trameur Onglet VENTILATION, ventilation døune sélection døitems (2)

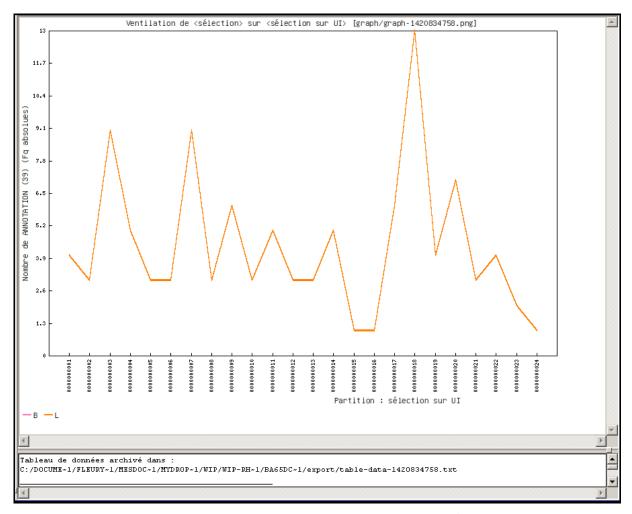


Figure 73 : Le Trameur Onglet VENTILATION, ventilation døune sélection døitems (3)

17 Segments répétés

La fenêtre principale dispose doun bouton permettant de calculer les segments répétés de la base puis domporter la liste obtenue dans longlet SEGMENT. Le calcul des segments est paramétrable (cf onglet PARAM): on peut choisir la longueur maximale des segments à rechercher ou leur fréquence minimale. Si ces 2 derniers paramètres ne sont pas renseignés tous les segments de fréquence au moins égale à 2 et de longueur au plus égale à 12 items sont recherchés. Le calcul des segments répétés peut se faire en selectionnant le niveau d'annotation à prendre en compte pour construire ces segments: Annotation n°1 (forme), Annotation n°2 (lemme) ou Annotation n°3 (catégorie) etc.

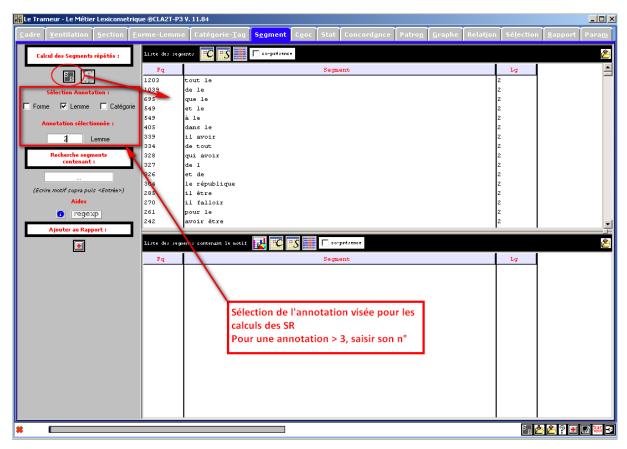


Figure 74: Le Trameur Onglet SEGMENT

Remarque : Les segments apparaissent dans le volet supérieur de longlet SEGMENT.

Après sélection døun segment (ou de plusieurs), il est possible de déclencher :

- la recherche de cette forme (ou de plusieurs) dans la représentation cartographique du *Cadre infra* (*cf* onglet CADRE).
- la recherche de cette forme (ou de plusieurs) dans la représentation cartographique de la carte des sections *infra* (*cf* onglet **SECTION**).
- une concordance des segments sélectionnés.

On peut aussi rechercher des segments contenant une forme donnée en utilisant les zones de saisie disponibles sur la gauche de lønglet **SEGMENT**.

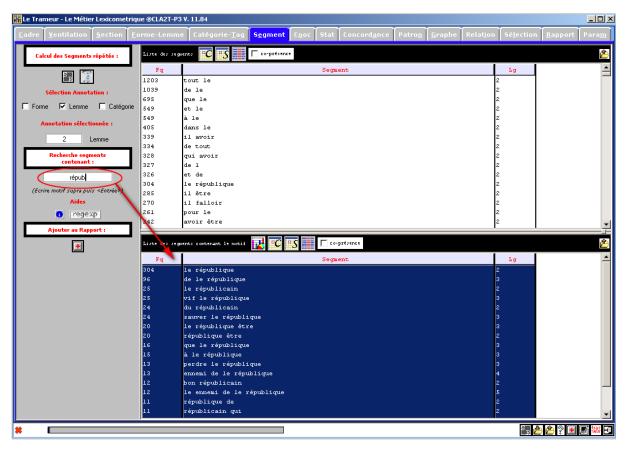


Figure 75 : *Le Trameur* Onglet **SEGMENT**

Après sélection døun segment (ou de plusieurs), il est possible de déclencher :

- la recherche de ce segment (ou de plusieurs) dans la représentation cartographique du *Cadre infra* (*cf* onglet CADRE).
- la recherche de ce segment (ou de plusieurs) dans la représentation cartographique de la carte des sections *infra* (*cf* onglet **SECTION**).
- la concordance de ce segment (ou de plusieurs) (cf onglet **CONCORDANCE**).
- la ventilation de ce segment (ou de plusieurs) (cf onglet **VENTILATION**).

18 Cooccurrences ó Poly-Cooccurrences

L'onglet COOC permet de lancer des modules de calculs de cooccurrences et de polycooccurrences sur la base chargée. Les modules disponibles søinscrivent dans la démarche mise en ò uvre dans le travail de William Martinez (2002, 2003, 2006).

- Une *cooccurrence* désigne løapparition de deux mots en même temps et dans le même contexte.
- Le terme *poly-cooccurrence* désigne les attractions lexicales au-delà de la cooccurrence binaire.

Le module de poly-cooccurrences intégré reprend løalgorithme décrit dans [Martinez, 2006] :

- On calcule pour le pôle A les cooccurrents spécifiques A₁, A₂ et A₃
 - \circ Dans leurs contextes communs, on calcule pour les pôles $A+A_1$ les cooccurrents spécifiques A_{11} et A_{12}
 - Les pôles A+A₁+A₁₁ ont pour cooccurrent spécifique A₁₁₁
 - On examine les cooccurrents de A+A₁+A₁₁+A₁₁₁.
 - Les pôles A+A₁+A₁₁+A₁₁₁ n'ont pas de cooccurrent spécifique : l'exploration s'interrompt pour ce chemin
 - o Les pôles A+A₁+A₁₂ ont pour cooccurrent spécifique A₁₂₁
 - On examine les cooccurrents de : $A+A_1+A_{11}+A_{121}$:
 - Etc.
 - O Dans leurs contextes communs, on calcule pour les pôles A+A2 les cooccurrents spécifiques A21 et A22
 - Etc.
- Durant lœxploration, différents filtrages conditionnent l'épuisement des explorations contextuelles et réduisent le bruit dans les résultats pour privilégier løinformation la plus spécifique : seuils maximaux de fréquence et de spécificité du cooccurrent.

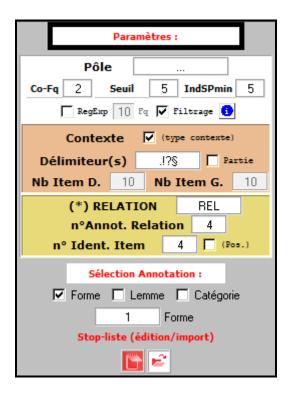
18.1 Paramétrage et lancement des modules

Pour le calcul des cooccurrents, la zone de paramétrage permet de :

- fournir la forme pôle (simple ou complexe via une regexp)
- définir une co-fréquence minimale
- définir le seuil utilisé par le calcul de spécificité mis en ò uvre par ce calcul
- définir løindice de spécificité minimale pour filtrer les candidats cooccurrents
- définir le type døunité contextuelle à utiliser pour le calcul (2 types de contextes sont possibles)
- préciser løannotation à utiliser pour le calcul (forme, lemme, catégorie, etc.)
- prendre en compte une stop-liste (ou ajouter des éléments dans une stop-liste)

Les valeurs fournies par défaut pour certains paramètres sont en général adéquates pour ce type de calcul.

La figure suivante donne à voir cette zone de paramétrage :



Dans la zone de paramétrage du calcul des cooccurrents, une case à cocher permet de définir le pôle sous la forme døune expression régulière.

Løunité contextuelle retenue pour la calcul peut être définie de 2 manières différentes

- Løunité contextuelle est définie en fournissant
 - o soit une liste de caractères délimiteurs permettant de définir des contextes délimités par ces caractères,
 - o soit en spécifiant une zone du *Cadre* (une balise de partie) à utiliser pour délimiter les zones visées (une partie donc) ; dans ce cas il faut cliquer sur la case à cocher « Partie » après avoir fourni la balise à prendre en compte pour définir les contextes
- Løunité contextuelle est définie par un nombre døtem à prendre en compte à droite et à gauche du pôle (par défaut 10 items de chaque côté)

La première solution est le choix par défaut. Pour la modifier, il suffit de cliquer sur la case à cocher « type de contexte ».

Enfin, le calcul des cooccurrents/polycooccurrents peut se faire en sélectionnant le niveau d'annotation à prendre en compte pour construire ces séquences : Annotation $n^{\circ}1$ (forme), Annotation $n^{\circ}2$ (lemme) ou Annotation $n^{\circ}3$ (catégorie) etc.

18.2 Mise en à uvre du calcul

Les figures qui suivent illustrent la démarche mise en à uvre pour construire des listes de cooccurrents puis des graphes donnant à voir les poly-cooccurrents.

Utilisation døune stop-liste

CLESTHIA>SYLED>CLA2T

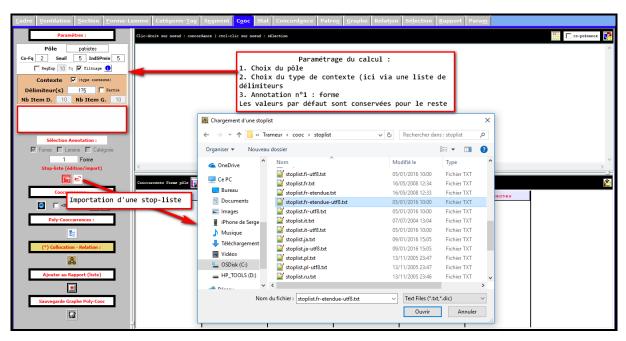


Figure 76: Le Trameur Onglet COOC, importation døune stop-liste

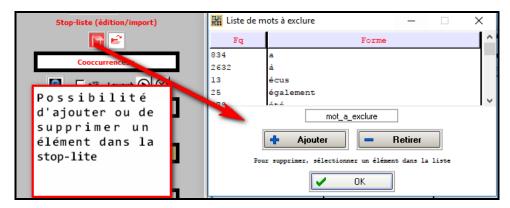
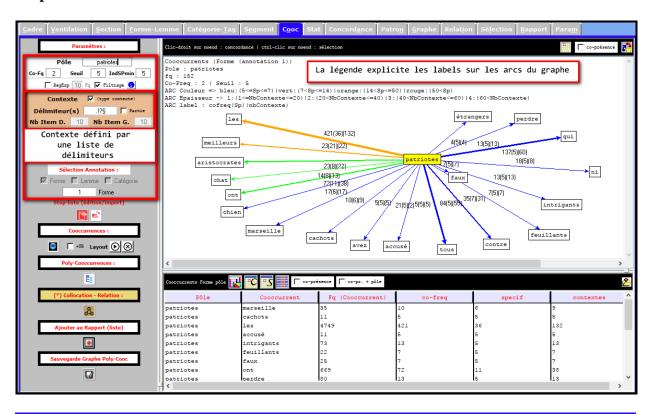


Figure 77 : Le Trameur Onglet COOC, édition de la stop-liste

Etape n°1: Calcul des cooccurrents



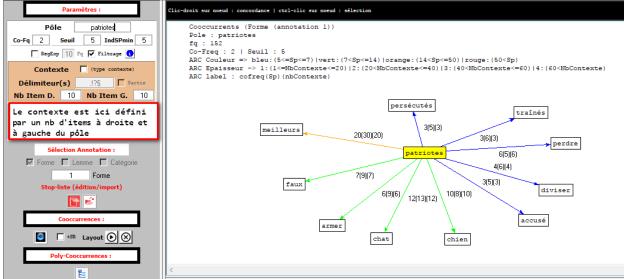


Figure 78 : Le Trameur Onglet COOC, étape n°1 : module Cooccurrents (2 types døunités contextuelles)

Etape n°2: Calcul des polycooccurrents

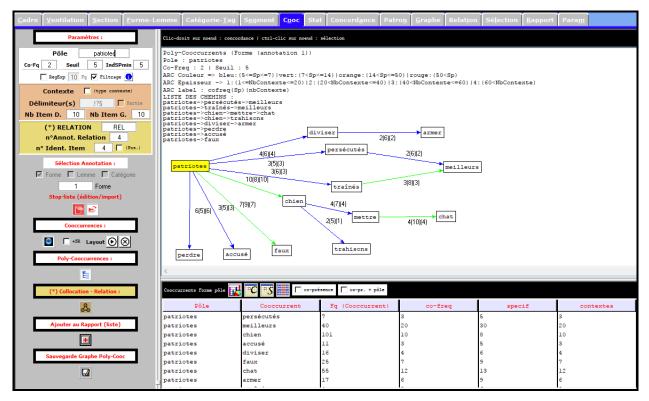


Figure 79: Le Trameur Onglet COOC, étape n°2: module Polycooccurrents

Dans les 2 graphes produits, les arcs connectant des nò uds portent un label regroupant 3 indications (disponibles aussi dans les colonnes de la liste des cooccurrents) : co-fréquence (spécificité) (contextes). Par exemple, pour la forme chien à partir du pôle patriotes on trouve la notation 10 (8) 10 qui indique que le cooccurrent apparaît 10 fois à côté du pôle avec une spécificité de 8 dans un total de 10 contextes. La représentation de ces arcs permet aussi de mettre en avant de manière graphique 2 de ces indications statistiques : la couleur du trait dépend de løindice de spécificité et løépaisseur du trait dépend du nombre de contextes (la légende précise les valeurs dans les 2 cas).

18.3 Cooccurrents et segments répétés

Le calcul des cooccurrents / polycooccurrents peut aussi être activé en intégrant les segments répétés si ce module a déjà été lancé.

Choix døun pôle de type SR

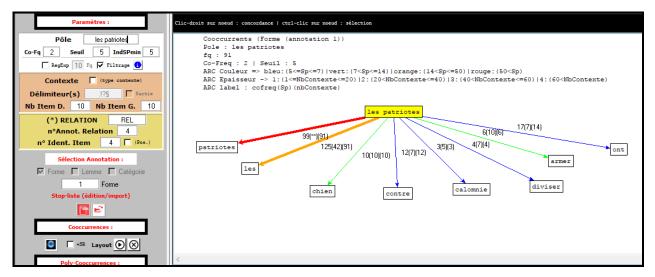


Figure 80 : Le Trameur Onglet COOC, un pôle de type SR

Si le calcul des SR a déjà été activé, il est possible de calculer les cooccurrents / polycooccurrents døun pôle exprimé sous la forme døun segment répété.

Intégration des SR dans les calculs de cooccurrence

De même, il est possible døntégrer les segments disponibles dans le calcul des cooccurrents / polycooccurrents : la case à cocher idoine rend cette option disponible.

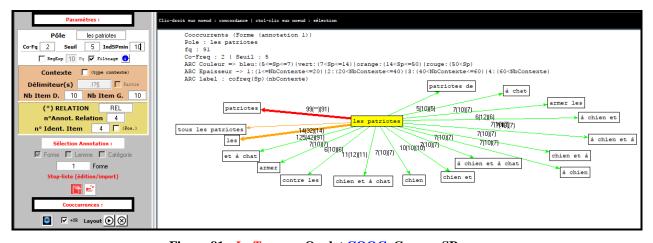


Figure 81 : Le Trameur Onglet COOC, Coocs + SR

18.4 Sélection des chemins de cooccurrences à examiner

CLESTHIA>SYLED>CLA2T

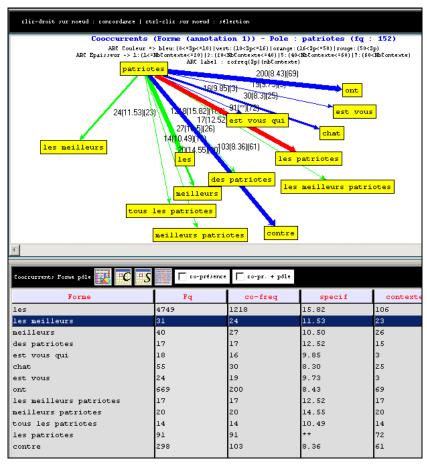


Figure 82: Le Trameur Onglet COOC, Filtrage des polycoocs

Le calcul des chemins de polycoocurrence se fait sur la base de la liste des cooccurrents du pôle choisi.

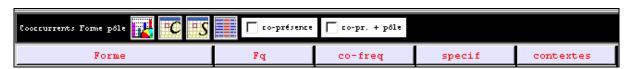
- Par défaut le calcul des polycooccurrents est réalisé en traitant tous les cooccurrents du pôle.
- Il est aussi possible de lancer le calcul des polycooccurrents sur une sélection de cooccurrents du pôle.

Dans la figure ci-dessus, la sélection døun seul cooccurrent du pôle choisi permettra de ne calculer que la polycooccurrence initiée par cette sélection (coprésence du pôle et du cooccurrent sélectionnée).

18.5 Fonctionnalités complémentaires sur les résultats produits

Les résultats produits par ces 2 modules sont insérés dans l\u00c3onglet COOC et sont associ\u00e9s \u00e0 des processus de calcul d\u00e9j\u00e0 disponibles dans d\u00e9autres onglets :

a. Actions sur la liste des cooccurrents :



Il suffit de sélectionner un ou plusieurs éléments dans la liste des cooccurrents puis donctiver lopération visée (ventilation, ventilation sur Cadre, ventilation sur carte des sections, concordance).

- b. Actions sur le graphe des cooccurrent ou des polycoccurrents
- le raccourci *clic-droit* sur un nò ud du graphe déclenche le calcul de la concordance de la forme associée au nò ud visé
- le raccourci control-clic sur un nò ud permet de le sélectionner

Des nò uds du graphe étant sélectionnés, on peut ensuite les projeter sur la carte des sections :



Si løoption « co-présence » est activée, les sections mises au jour seront celles contenant lønsemble des formes visées (ET logique) : on peut ainsi visualiser en contexte les chemins de polycooccurrences calculés précédemment.

19 Extraction de patrons et graphes de mots

19.1 Extraction de patrons (recherche sur un niveau døannotation)

Une base de travail étant chargée dans le *Trameur*, il est possible de réaliser une extraction de patron sur la base des annotations associées à chaque item de la *Trame*. Les séquences de termes à construire pour un patron donné peuvent être définies sur les catégories morphosyntaxiques déterminées par *treetagger* ou sur les annotations importées. Pour lancer cette recherche de patrons, il faut se placer dans løonglet **PATRON**:

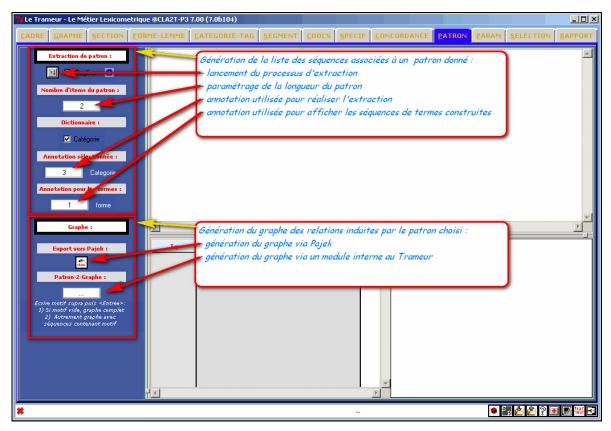


Figure 83: Le Trameur onglet PATRON

Dans les figures qui suivent, nous allons illustrer cette extraction de patron en utilisant notre fichier de travail et en recherchant tout døabord le patron « NOM ADJ» sur løannotation n°3 (les catégories construites par treetagger) puis le patron « orange rouge bleu » sur løannotation n°4 importée, les couleurs associées à chaque item de la Trame (cf partie suivante pour une présentation de lømportation døannotation). Avant de lancer la génération de la liste de séquences de termes, il convient de choisir le niveau døannotation choisi pour définir le patron ; par défaut, ce niveau (le numéro 3) est celui qui est associé aux catégories construites par treetagger. Pour choisir un numéro correspondant à un système døannotation importé, il suffit de donner son numéro (numéro 4¹⁴ ou supérieur). Pour enregistrer le patron visé, il faut activer le bouton le saisie permettant de définir les constituants du patron. Le nombre de zones de saisie correspond à la valeur (modifiable) associée à løétiquette :

96

¹⁴ Rappel: par défaut, les numéros døannotation 1, 2 et 3 sont associés respectivement aux formes, aux lemmes et aux catégories (que *treetagger* soit activé ou non).

« Nombre døtems du patron » dans lønglet PATRON. Løcriture des constituants døun patron peut se faire en utilisant des expressions régulières si ce paramétrage a été au préalable défini dans lønglet PATRON.

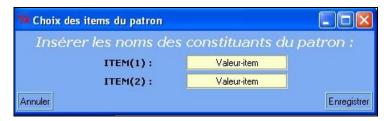


Figure 84 : Le Trameur zone døenregistrement døun patron

Après avoir renseigné les valeurs des éléments constitutifs du patron visé, le bouton « *Enregistrer* » déclenche la génération des séquences de termes (*via* leurs formes) associées au patron visé. Cette liste sera intégrée dans la zone dédition gauche de løonglet **PATRON**.

Exemple 1:

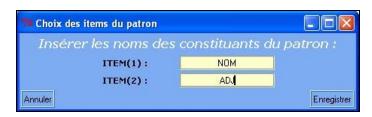


Figure 85 : Le Trameur écriture doun patron

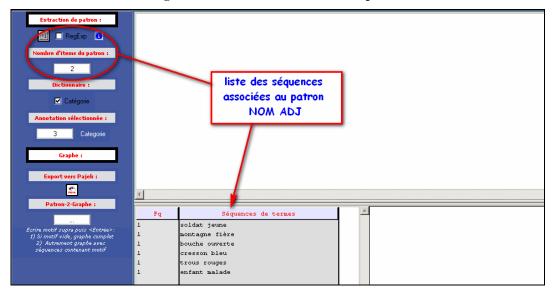


Figure 86 : Le Trameur liste des séquences associées au patron

Exemple 2:



Figure 87 : Le Trameur écriture et paramétrage døun patron

CLESTHIA>SYLED>CLA²T

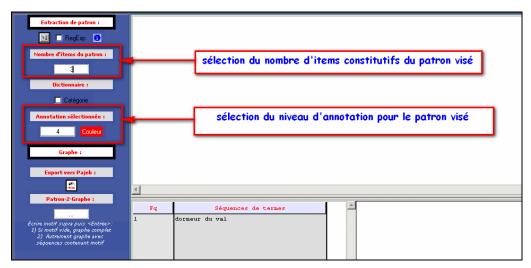


Figure 88 : Le Trameur liste des séquences associées au patron

dans notre texte de travail, les couleurs associés aux formes sont, pour le début du texte :

Le-vert dormeur-orange du-rouge val-bleu C-vert '-jaune

19.2 Extraction de patrons (avec croisement døannotations)

Cette fonctionnalité permet dœxtraire un patron en décrivant initialement ces composants sur des niveaux dœnnotations différents. Pour illustrer cette fonctionnalité, nous allons utiliser la base Prématurés 96¹⁵ : dans cette base, chaque item de la *Trame* est associé à 4 annotations : forme (niveau 1), lemme (niveau 2), catégorie (niveau 3), annotation sémantique (niveau 4).

1. Paramétrage de læxtraction :

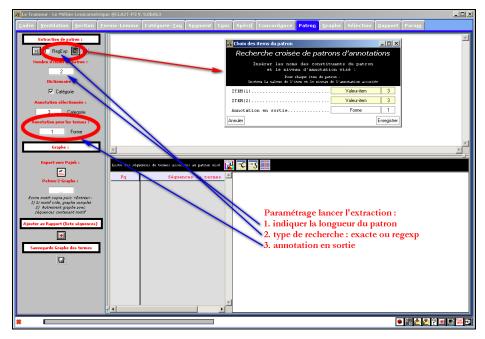


Figure 89 : Le Trameur Extraction de patron (croisement deannotation)

_

¹⁵ http://tal.univ-paris3.fr/trameur/bases/base-p96-trameur.zip

2. Paramétrage du patron

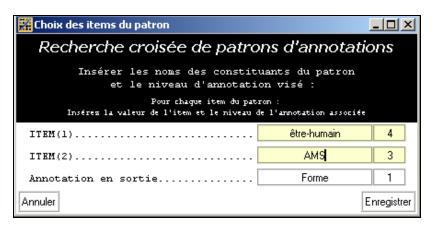


Figure 90 : Le Trameur Extraction de patron (croisement deannotation)

On cherche donc ici un patron de longueur 2, dont le premier composant porte la valeur « être-humain » (annotation de niveau 4 : sémantique) et dont le second porte la valeur AMS (Adjectif Masculin Singulier) (annotation de niveau 3 : POS). Les termes extraits seront affichés suivant leur forme graphique.

Résultat produit : patron « être-humain AMS »

Liste des séguences de termes associées au patron visé								
Fq	Séquences de termes							
10	bébé calme							
5	bébé tonique							
2	enfant calme							
2	bébé réactif							
2	bébé agréable							
1	bébé excitable							
1	bébé inconfortable							
1	bébé sage							
1	enfant mignon							
1	bébé douloureux							
1	bébé sédaté							
1	bébé chétif							
1	bébé bien							
1	enfant sédaté							
1	bébé gigoteur							
1	enfant attendrissante							
1	bébé adorable							

Figure 91 : Le Trameur Extraction de patron (croisement døannotation)

Si la case à cocher *RegExp* est activée, lécriture des composants du patron peut se faire en décrivant ces composants par une expression régulière.

19.3 Actions sur les séquences extraites à partir døun patron donné

Les séquences de termes associées à un patron donné sont réutilisables pour des opérations déjà disponibles dans les autres onglets :

- Concordance: calcul døune concordance sur toutes les positions des séquences sélectionnées
- Ventilation-section : ventilation sur la carte des sections de toutes les séquences sélectionnées

CLESTHIA>SYLED>CLA2T

- Ventilation-Cadre : ventilation sur le Cadre de toutes les séquences sélectionnées
- *Ventilation-partie* : ventilation de toutes les séquences sélectionnées sur une partie donnée

Les figures qui suivent illustrent de nouveau la démarche à suivre pour lancer une recherche de patron puis pour déclencher les opérations sur des éléments de la liste de termes produite :

a. Paramétrage de l\(extraction de patron

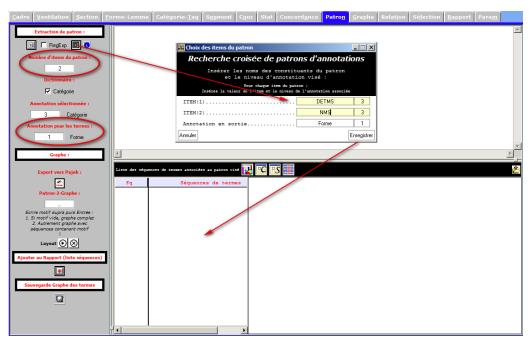


Figure 92 : *Le Trameur* Paramétrage de lœxtraction du patron visé et de lœffichage des séquences extraites

b. Sélection des termes puis activation des calculs visés :

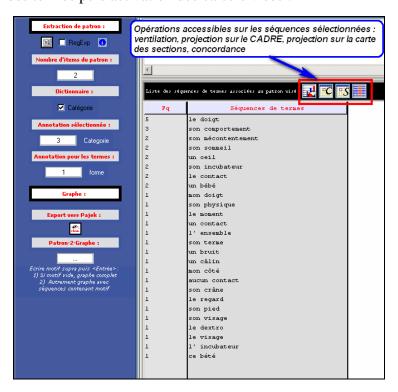


Figure 93 : Le Trameur Affichage des séquences extraites sur la base du patron visé

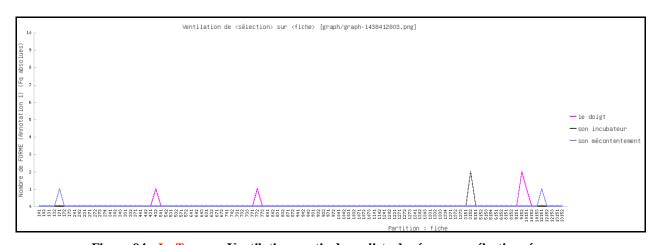


Figure 94 : Le Trameur Ventilation-partie donne liste de séquences sélectionnées

Figure 95 : Le Trameur Concordance doune liste de séquences sélectionnées

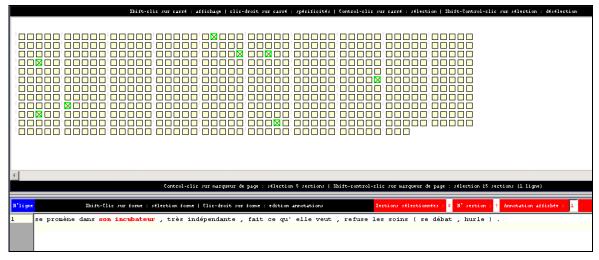


Figure 96 : Le Trameur Ventilation-section døune liste de séquences sélectionnées

19.4 Graphes de mots avec Le Trameur

On peut ensuite représenter la liste de séquences de termes associés à un patron sous la forme doun graphe de mots donnant à voir la proximité entre ces mots induite par le patron visé. Dans loexemple présenté infra, le texte de travail est un extrait doun débat télévisé.

Pour générer ce graphe, il convient de renseigner la zone de saisie « Patron-2-graphe » :

1. si le motif saisi est vide, le programme construit le graphe complet donnant à voir lœnsemble des relations entre les mots des séquences construites

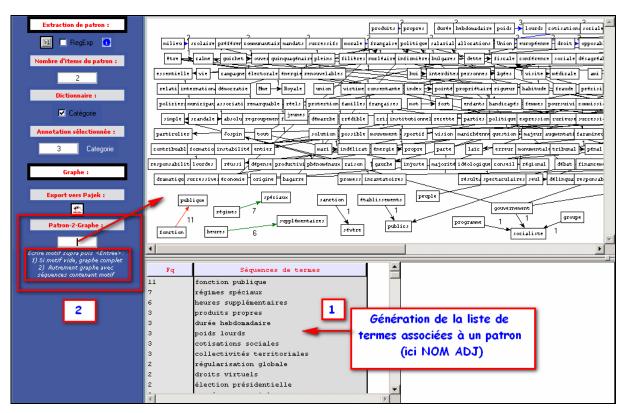


Figure 97 : Le Trameur génération du graphe de mots complet associé à un patron

2. si le motif saisi nœst pas vide, le programme construit le graphe donnant à voir lænsemble des relations entre les mots des séquences construites et contenant le motif.

Dans læxemple ci-dessous, le motif « public » permet de donner à voir toutes les relations entre les éléments des séquences construites contenant ce motif.

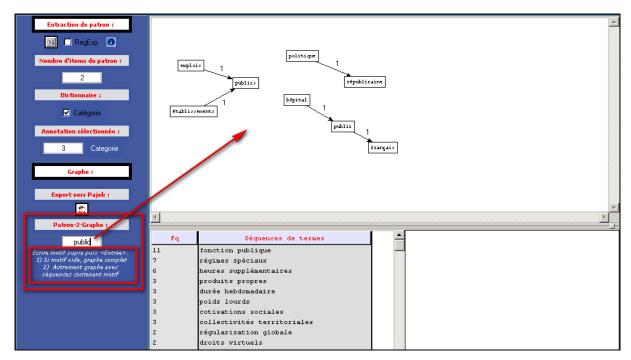


Figure 98 : Le Trameur génération du graphe de mots associé à un patron et contenant un motif

19.5 Graphes de mots avec Pajek

On peut aussi exporter la liste de séquences de termes associées à un patron dans un format compatible avec l ϕ outil de représentation de réseaux $Pajek^{16}$. On se reportera à la page en ligne « $\underline{Des\ Textes\ aux\ graphes}$ » pour avoir des informations complémentaires sur la démarche suivie.

Cet export se déroule en 2 temps :

- 1. Transformation du fichier contenant la liste des séquences associées à un patron au format GRAPHML¹⁷.
- 2. Génération de cette même liste au format *Pajek* en utilisant une feuille de styles fournie avec le *Trameur* (GraphML2Pajek.xsl disponible dans le répertoire de travail du *Trameur*).

¹⁶ Hypertoire: http://vlado.fmf.uni-lj.si/pub/networks/pajek/. Mode d'emploi: http://vlado.fmf.uni-lj.si/pub/networks/pajek/. Mode d'emploi: http://vlado.fmf.uni-lj.si/pub/networks/pajek/. Mode d'emploi: http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

¹⁷ http://graphml.graphdrawing.org/

CLESTHIA>SYLED>CLA2T

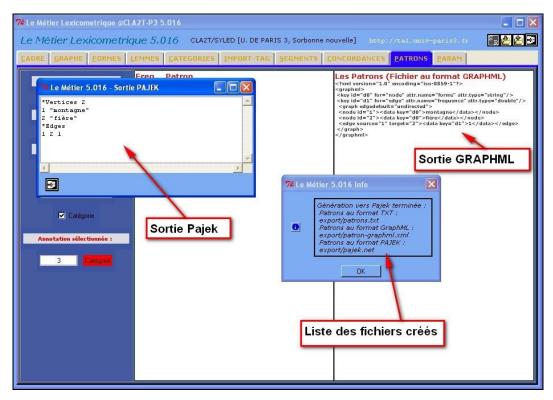


Figure 99 : Le Trameur export au format Pajek

La figure qui suit est une illustration døun graphe construit avec *Pajek*. Le texte de travail est le texte Duchn.txt (*Le Père Duchesne*) disponible dans le dossier textes. Le patron visé est là encore NOM ADJ.

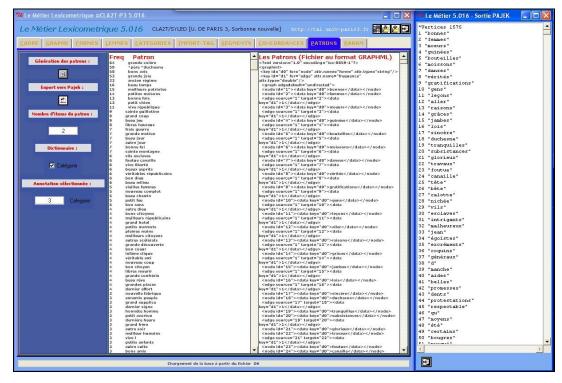


Figure 100 : Le Trameur export NOM ADJ au format Pajek sur Duchn

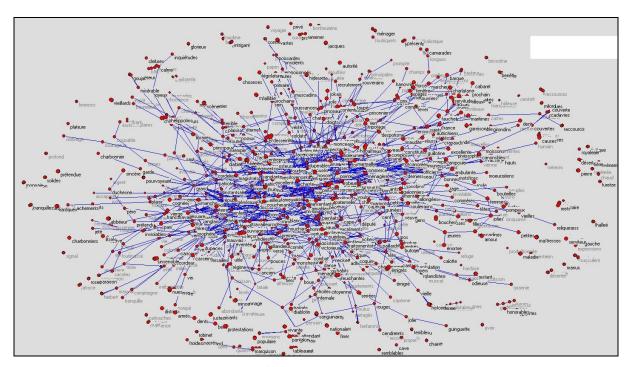


Figure 101 : Le réseau NOM ADJ de Duchn $via\ Pajek$

Dans Pajek, cliquer sur « Draw », puis sur « Layout » et choisir løoption, puis sur « Export » pour exporter le résultat.

20 Calculs textométriques complémentaires

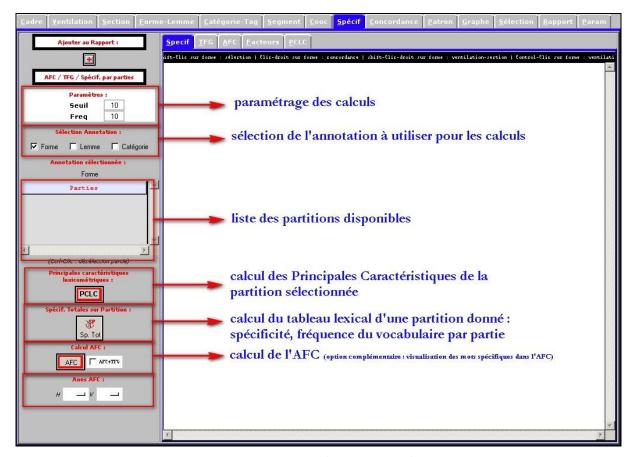


Figure 102 : Calculs textométriques complémentaires

Dans cette partie, nous décrivons les fonctionnalités disponibles dans lønglet **STAT** (ou **SPECIF**). Ces différents calculs se font en sélectionnant au préalable le niveau d'annotation à prendre en compte dans les calculs ; pour le moment, on peut sélectionner : **Annotation n°1** (forme), **Annotation n°2** (lemme) ou **Annotation n°3** (catégorie).

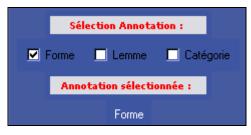


Figure 103 : Sélection du niveau d'annotation pour calculs lexicométriques complémentaires

20.1 Tableau lexical døune partition

Le bouton Sp. Tot permet de calculer le *tableau lexical* des items (forme ou lemme ou catégorie) døune partition sélectionnée dont la fréquence est supérieure à un seuil donné et dont løindice de spécificité est lui aussi supérieur à un seuil donné. Les lignes de ce tableau à double entrée sont constituées par les ventilations et les indices de spécificité des différents items dans les différentes parties de la partition choisie. Le terme générique k(i,j) du tableau

est égal au nombre de fois que løtem i est attestée dans la partie j ou à son indice de spécifité dans la même partie.

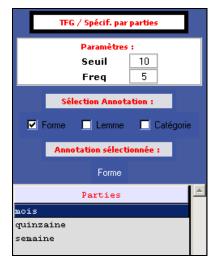


Figure 104 : Paramétrage TFG ; choix des paramètres de calcul, de løannotation et de la partition

Pour lancer ce calcul, il faut :

- sélectionner le niveau døannotation visé ;
- sélectionner une des partitions disponibles dans la liste fournie sur la partie gauche de lønglet STAT (ou SPECIF);
- fixer les valeurs des paramètres (fréquence, seuil) ;

Løactivation du bouton déclenche ensuite le calcul dont les résultats søaffichent dans løonglet $\overline{\text{TFG}}$

Specif TFG	<u>A</u> FC	<u>F</u> acto	eurs	PCLC												
Forme (Fq > 5)	Fq.	Sp. 01	Fq.	Sp. 02	Fq.	Sp.	Fq. 04	Sp. 04	Fq. 05	Sp. 05	Fq. 06	Sp. 06	Fq.	Sp.	Fq. 08	Sp. 08
accaparent	1	1.29	2	1.82	0	-1.32	1	1.31	0	-1.29	0	-1.29	1	1.34	0	-1.22
accapareur	0	-1.38	1	-1.10	1	1.23	0	-1.35	1	1.26	0	-1.34	0	-1.32	3	2.85
ccapareurs	18	2.71	20	3.10	16	2.16	9	-1.36	2	-3.76	2	-3.70	10	1.36	3	-2.35
ccaparé	0	-1.51	5	3.64	2	1.53	1	-1.14	0	-1.47	0	-1.46	0	-1.42	0	-1.35
accepter	1	1.29	3	2.63	0	-1.32	0	-1.29	0	-1.29	1	1.32	0	-1.26	0	-1.22
cceptée	4	3.40	1	-1.10	0	-1.38	1	1.25	0	-1.35	0	-1.34	0	-1.32	0	-1.26
ccord	3	1.39	1	-1.56	2	-1.23	2	-1.20	1	-1.45	5	2.31	1	-1.39	2	1.31
ccusant	1	1.29	1	1.27	0	-1.32	1	1.31	1	1.31	1	1.32	0	-1.26	0	-1.22
ccusation	2	1.53	0	-1.54	3	2.08	0	-1.47	1	-1.13	1	-1.13	0	-1.42	1	1.26
	0		0		3	1.67	2	1.34	1	-1.26		2.29	0	-1.63	2	1.49
	2	-1.15	0		0		0		2	1.27	6	3.36	1	-1.29	3	1.84
	0	-1.63	0		0		0		2	1.44	7	5.43	0	-1.53	1	-1.12
		1.35	2		0	-1.70	1		1	-1.23		1.84	2	1.44	0	-1.48
	1		0		0		0		0	-1.52		2.05	0	-1.48	5	4.13
	1	-1.22	ő		2		1		1	-1.20	ő		4	2.69	Ĭ	-1.12
		1.96	ő	-1.94					0	-1.81	_	1.28	1	-1.29	ļ,	-1.22
			0	-1.34			0		1	1.31		1.32	1	1.34	0	-1.22
		1.29	ő		2		0		0	-1.29	ō	-1.29	1	1.34	ľ	1.40
		2.24	0		0	-1.44	ľ		3	2.33	o	-1.40	0	-1.37	l,	1.30
		1.35	0	-1.74	ľ		0		3	1.82		1.84	1	-1.20	i,	-1.15
	1		0		0		0		2	1.58	0		4	3.08	l,	1.26
	0		0		0		-	1.31	0	-1.29	_	1.32	0	-1.26	3	3.12
	0		0			2.70		1.31	0	-1.29	0		1	1.34	0	-1.22
	-	1.61	2		3 2		0	-1.41	0	-1.29	0	-1.29 -1.40	0	-1.34	,	1.30
	0		0			1.61			0				0		ļ,	
	_	1.53			2			1.66 2.04	0	-1.41 -1.34		1.68 -1.80		-1.37	[,	1.30
	_		2		0				1		0		_	1.68	1	-1.22
	1	-1.42	3		5		0		0		0	-1.86		2.08	2	1.37
	0	-1.44	1		2		0		0	-1.41	0	-1.40		2.43	1	1.30
		1.72			2		0		0	-1.35	0	-1.34	1	1.29	0	-1.26
		1.30	1		1		0		2	1.34		1.75	1	-1.23	2	1.49
		3.00	4		7	1.44	3		5	-1.26	4	-1.43	3	-1.59	5	1.40
-		1.61	1		0		0		0	-1.41		1.68		1.73	0	-1.31
	1	-3.66	10	1.50	5				12	2.27		1.37	5	-1.48	8	1.72
	26	-1.36	21	-2.19				2.67	30	1.73		6.26	14	-2.65	9	-3.28
	0	-1.38	1	-1.10	1	1.23		1.77	0	-1.35		1.26	1	1.29	0	-1.26
ient	3	1.60	11	I -1.37	12	1.27	11	-1.30	11	- 1.30	11	l-1.29	12	1.35	12	1.45
l .																

Figure 105 : Tableau lexical (annotation n°1 : forme)

CLESTHIA>SYLED>CLA2T

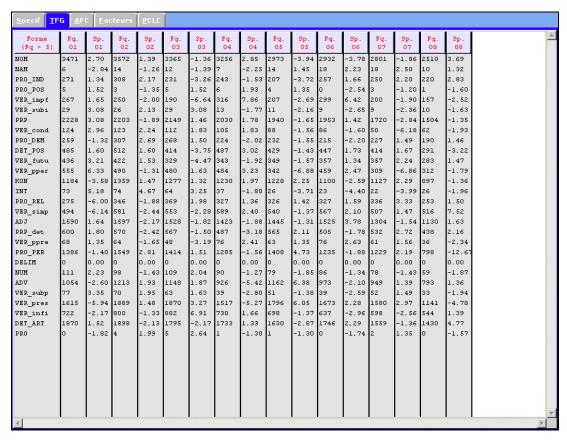


Figure 106 : Tableau lexical (annotation n°3 : catégorie)

Pour chaque ligne, associée à un item donné, on trouve successivement, la fréquence et løindice de spécificité de cet item dans toutes les parties de la partition sélectionnée.

20.2 AFC: Analyse Factorielle des Correspondances

Le bouton permet de réaliser une analyse factorielle des correspondances sur l'ensemble des parties du corpus. Le paramétrage permet de fixer entre autres le nombre des unités textuelles prises en compte dans l'analyse. Par défaut, l'analyse prend en compte les unités dont la fréquence est supérieure à 10.

Pour lancer ce calcul, il faut :

- sélectionner une des partitions disponibles dans la liste fournie sur la partie gauche de lønglet **STAT** (ou **SPECIF**);
- fixer les valeurs des paramètres (fréquence, seuil) ;

Løactivation du bouton déclenche ensuite le calcul dont les résultats søaffichent dans løonglet **AFC**. Les parties du corpus apparaissent sur le plan des deux premiers axes factoriels extraits par l'analyse.

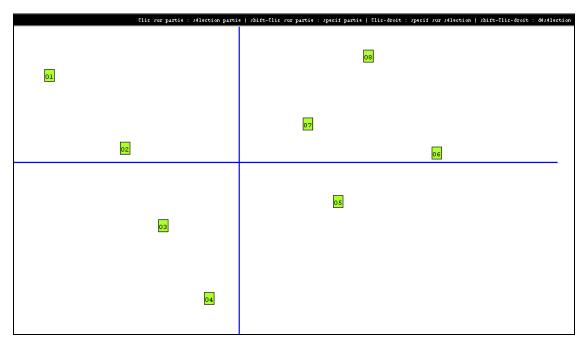


Figure 107 : AFC

En positionnant la souris sur une des parties, ses coordonnées dans le graphique sont affichées :

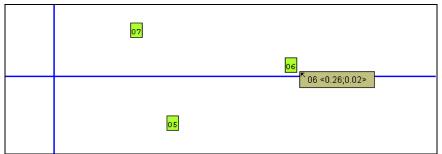


Figure 108 : AFC (position)

On peut ensuite sélectionner (clic gauche), directement sur la carte produite, une partie ou un groupe de parties. Les contours des parties sélectionnées apparaissent alors en surbrillance rouge. Cette sélection permet de calculer des spécificités sur le groupe de parties sélectionnées *via* un clic-droit sur le graphique.

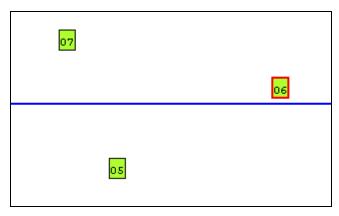


Figure 109 : AFC (sélection)

On peut aussi directement activer le calcul des spécificités døune partie donnée *via* la combinaison de touche *shift+clic* sur la partie visée.

Le calcul de løAFC peut aussi être lancé avec une option complémentaire permettant døafficher dans le graphique final les dix items les plus spécifiques sur la partie visée. Pour activer cette option, il faut cocher la case AFC+TFG:



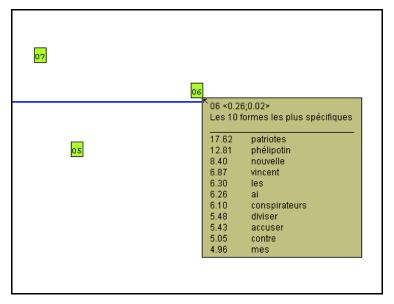


Figure 110 : AFC (affichage des mots spécifiques de la partie)

Le calcul de løAFC sur une partition donnée déclenche aussi løaffichage :

• dans løonglet **FACTEURS**: des facteurs construits lors du calcul de løAFC permettant de résumer (de décrire brièvement) les variables actives initiales;

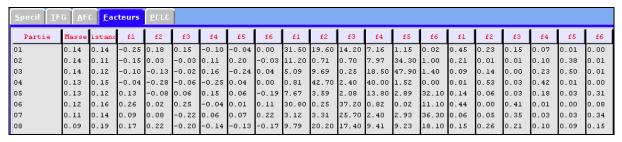


Figure 111 : Les facteurs de løAFC

• dans lønglet PCLC : des principales caractéristiques lexicométriques des différentes parties de la partition sélectionnée.

Specif TFG	Specif IFG AFC Facteurs PCLC							
Partition	Valeur Partie	début	fin	fq item (forme)	FQ item (occur)			
mois	01	1	41214	3439	19256			
mois	02	41215	85366	3778	20343			
mois	03	85367	126728	3563	19278			
mois	04	126729	165735	3542	18004			
mois	05	165736	204892	3373	17814			
mois	06	204893	243793	3392	17545			
mois	07	243794	279786	3297	16295			
mois	08	279787	310144	2970	13644			

Figure 112: PCLC døune partition

Un premier calcul de løAFC ayant été lancé, il est possible de relancer le calcul en sélectionnant des parties de la partition visee à ne pas prendre en compte pour le calcul : <u>les parties selectionnées dans la liste des PCLC ne seront pas integrées</u> au calcul.

20.3 Intégration de modules R (phase 1)

Cette partie décrit la mise à jour mise en ò uvre à partir de la version 11.00 du *Trameur* permettant de lancer des modules R (AFC, ACP, CAH) via la librairie *factoMineR* (http://factominer.free.fr/index_fr.html)

Pour illustrer ces fonctionnalités, on utilise ici la base textométrique intitulée « Base des états finaux du corpus BROUILLONS (ANR ECRITURES¹⁸) » :

Cette base rassemble 19 « états finaux » issus de 19 dossiers ¹⁹ du corpus BROUILLONS (ANR ECRITURES). Chaque item de la *Trame* est associé à 3 niveaux døannotation : forme, lemme, catégorie. On présente ci-dessous un extrait de la *Trame* de cette base via une concordance :

```
--PARTIE{dossier=ravet}-
 un lycée professionnel standard mais nous craignons que adèle ne
      compte tenu de ce que nous connaisson
                                            Position:<482>
relation avec les garçons que nous trouvic
compte tenu de ce que nous observ:
                                           Lemme: <Nous_Sujet_Sentiment> | Freq:13
         un autre garçon accueilli) nous
           . § a ce jour nous n'avons auct Cat:<Nous_Sujet_Sentiment>|Freq:13
           vacances et elle a pu nous dire que sa grand-
  madame richard, assistante familiale, nous a informé que adèle
            de sa famille. § nous avons reçu adèle pour
              de la dap). § nous lui avons rappelé qu'
          famille non préparé. § nous allons proposer à la
         fragilités de adèle, il nous semble qu'un accueil
  l'institut camille lamartine, nous avons organisé des rencontres
        adèle a été demandeuse .nous allons en organiser une
      foyer des oiseaux. § nous sommes actuellement en contact
             elle.
                    § toutefois, nous pensons qu'il v
```

Les items associés à la forme (annotation n°1) « nous » ont été annotées via le *Trameur* : leur catégorie (annotation n°3) a été modifiée pour prendre les valeurs suivantes :

¹⁸ http://www.univ-paris3.fr/anr-ecritures-96530.kjsp?RH=1295620557102

¹⁹ Chaque dossier est initialement constitué doun ensemble chronologique de textes allant du brouillon initial à la version finale, on ne retient ici que ce dernier état.

```
Nous_Faire
Nous_OBJ
Nous_Sujet
Nous_Sujet_Constat
Nous_Sujet_Dire
Nous_Sujet_Dire_Faire
Nous_Sujet_Dire_Preco
Nous_Sujet_Faire
Nous_Sujet_Faire
Nous_Sujet_Pensee
Nous_Sujet_Pensee
Nous_Sujet_Pensee_Faire
Nous_Sujet_Preco
Nous_Sujet_Savoir
Nous_Sujet_Savoir
```

Mise en ò uvre des modules R dans le Trameur

Les modules R sont disponibles dans l\u00f3onglet STAT qui a d\u00e9sormais l\u00e7allure suivante :

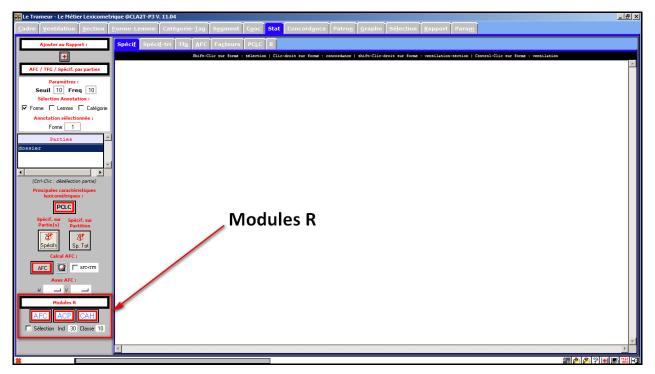


Figure 113: Modules R (1)

Le fichier de données utilisés par R est généré dynamiquement dans le dossier contenant la base de travail. Les 3 modules produisent des sorties PDF (rassemblant les graphiques produits par R) qui sont stockées dans le sous-dossier *export* du dossier contenant la base de travail.

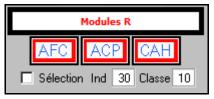


Figure 114: Modules R (2)

Par défaut, les 3 modules utilisent løannotation sélectionnée dans le paramétrage de løonglet **STAT** (en haut à gauche) et les items de fréquence supérieure à celle du paramètre *freq* (même zone de paramétrage, en haut à gauche).

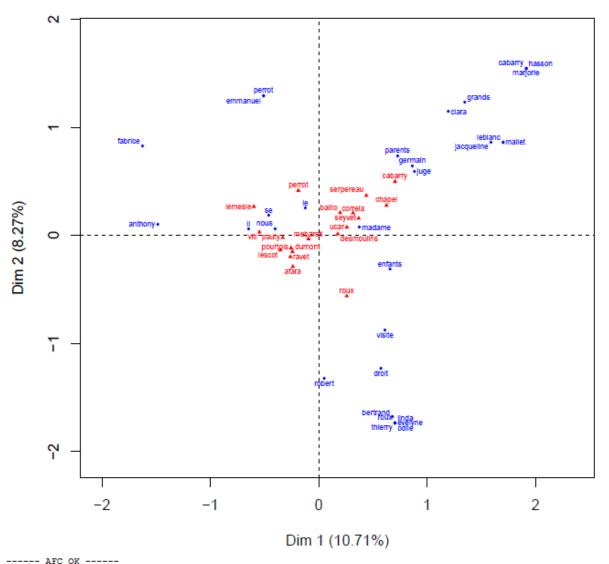
- Le paramètre *Ind* permet de spécifier le nombre døtem (forme, lemme, catégorieí) qui seront projetés sur les graphiques de løAFC ou de løACP.
- Le paramètre *Classe* permet de spécifier le nombre de classes à mettre au jour pour le module CAH.
- Le paramétre *Sélection* permet de modifier le tableau de données en entrée de R en ne gardant que les items préalablement sélectionnés via le *Gestionnaire de Sélection*.

Illustration sur la base « Etats finaux BROUILLONS »

Calcul sur base complète

La base étant chargée, on active le calcul de løAFC via R (paramétrage par défaut) via le bouton idoine, on donne à voir ci-dessous un extrait de la sortie PDF produite :

COLs + ROWs (contrib 30)

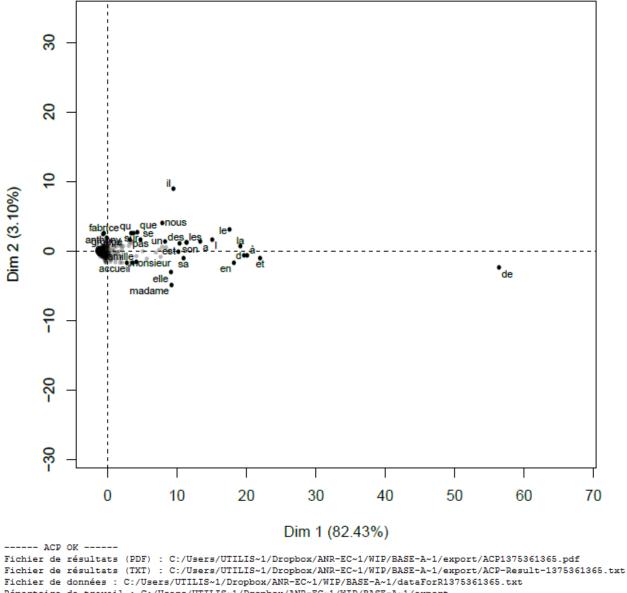


Fichier de résultats (PDF): C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/AFC1375361223.pdf
Fichier de résultats (TXT): C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/AFC-Result-1375361223.txt
Fichier de données: C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/dataForR1375361223.txt
Répertoire de travail: C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export
Répertoire de logs: C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/logR
Répertoire de R: C:/SergeFleury/SF-WORK/07-DEVEL/ANR-textometrie/mkGraphTrameCadre/R/bin/x64/R.exe

Figure 115 : AFC via R

De même pour løACP:

Axes=1,2; Choix=ind; Contrib 30

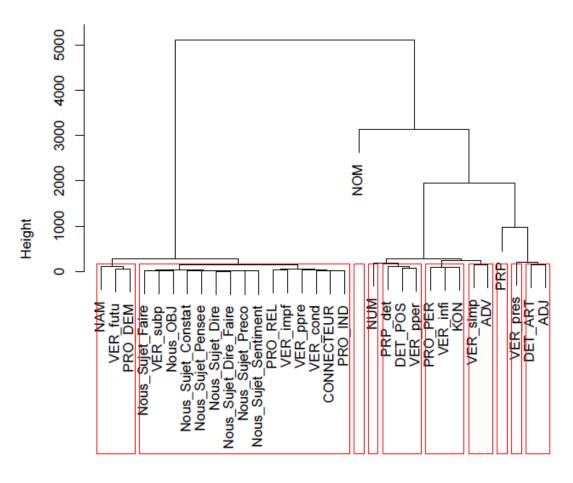


Répertoire de travail : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export Répertoire de logs : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/logR Répertoire de R : C:/SergeFleury/SF-WORK/07-DEVEL/ANR-textometrie/mkGraphTrameCadre/R/bin/x64/R.exe

Figure 116: ACP via R

De même pour le module CAH en modifiant ici le calcul pour travailler sur l\u00e1annotation n°3 (cat\u00e9gorie) :

Ward Hierarchical Clustering / euclidean



d hclust (*, "ward")

Figure 117: CAH via R

Calcul via le Gestionnaire de Sélection

On souhaite désormais que les calculs soient faits uniquement sur les items de type VERBE et ceux correspondant aux différents types de « nous ».

Pour repérer ces items, on commence par les sélectionner via le *Gestionnaire de Sélection* (cf *infra*) : recherche en 2 temps, les « nous » puis les verbes (cf figures suivantes) :

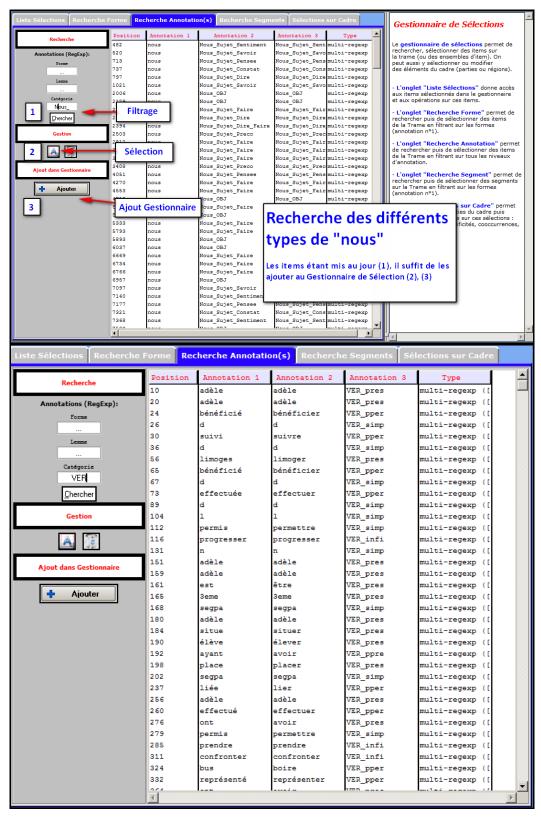


Figure 118 : Sélection des items pour les modules R

Au final, le *Gestionnaire de Sélection* contient tous les items recherchés, on prend soin de tous les sélectionner :

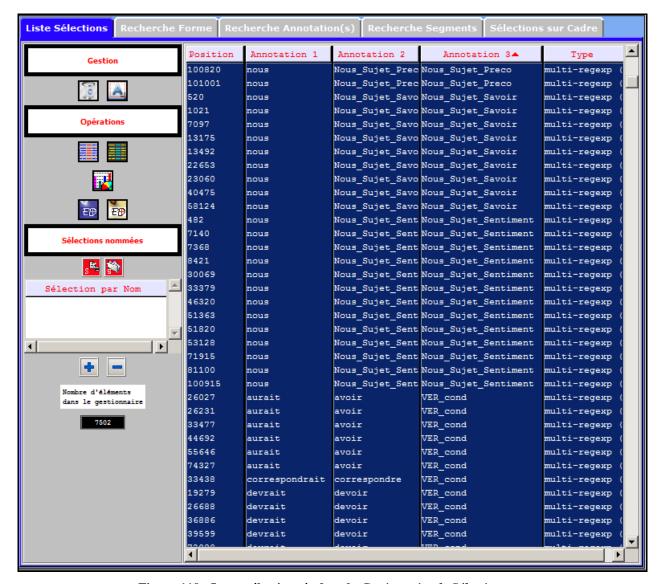


Figure 119 : Items sélectionnés dans la Gestionnaire de Sélection

Les items pertinents étant sélectionnés, on peut désormais lancer les modules R sur ces items uniquement :

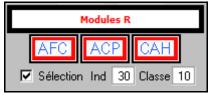


Figure 120 : Paramétrage R pour travailler avec le Gestionnaire de Sélection

Il peut être nécessaire/utile de modifier aussi les paramétres *Ind* et/ou *Classe*.

Calcul de løAFC

COLs + ROWs (contrib 30)

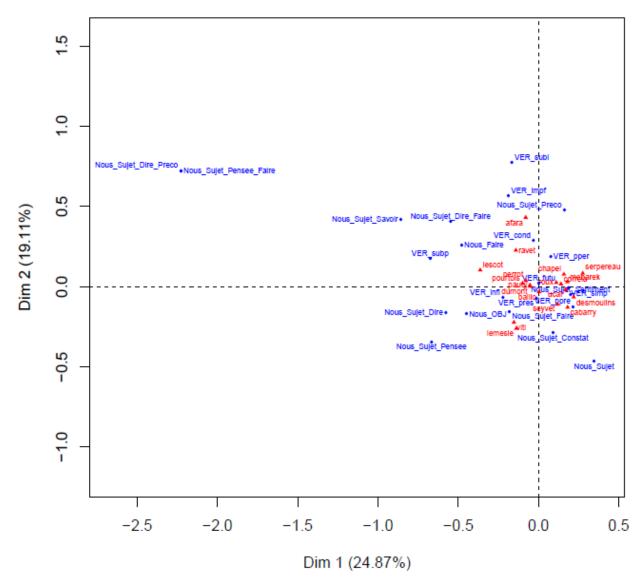


Figure 121 : AFC via R sur le contenu du Gestionnaire de Sélection

Calcul de la CAH

Ward Hierarchical Clustering / euclidean

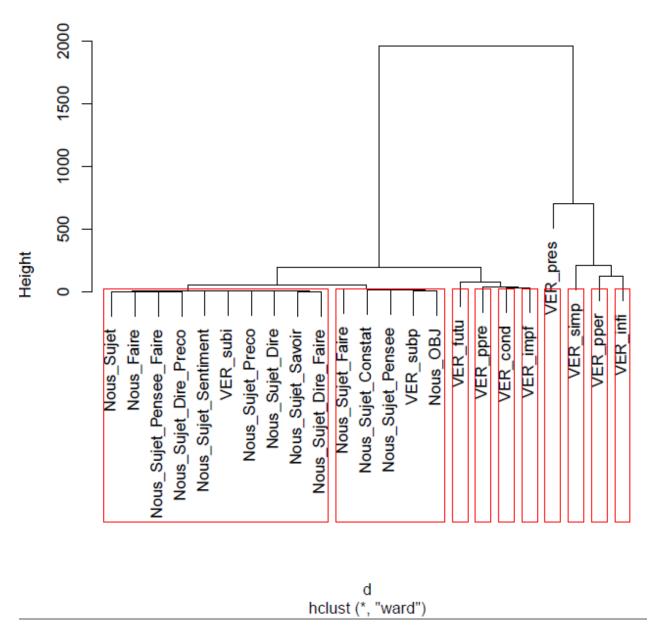


Figure 122 : CAH via R sur le contenu du Gestionnaire de Sélection

20.4 Corpus chronologiques : barycentre temporel, coefficient de Von Neumann

Référence : "Les séries textuelles chronologiques", André Salem, <u>Histoire & Mesure</u>, 1991 <u>6-1-2</u> pp. 149-175

Le calcul du barycentre temporel et de løindice de Von Neuman, présenté dans løarticle

BT+VN

døAndré Salem, est disponible dans løonglet STAT via le bouton

Le résultat permet la mise au jour de « termes évolutifs » ; il se présente sous la forme døun tableau qui a løallure suivante (illustration ici sur le corpus CFDT présenté dans løarticle cité précédemment) :

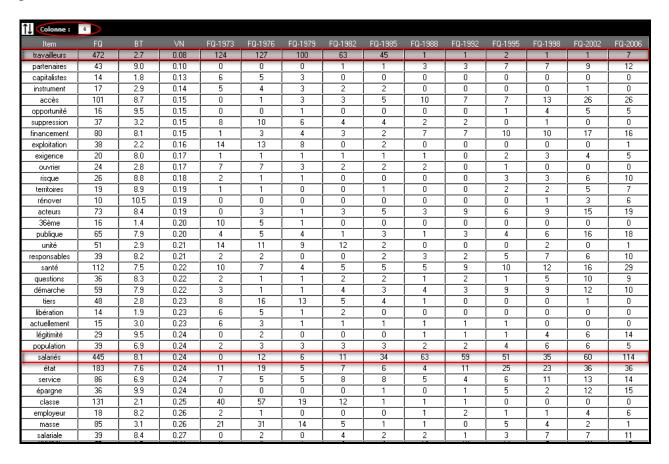


Figure 123 : Barycentre temporel et coefficient de Von Neumann

Pour tous les items de fréquence supérieure à la fréquence minimale fixée, le tableau donne à voir la ventilation en fréquence de chaque item et la valeur de leur barycentre temporel et de leur indice de Van Neuman. Chaque colonne est triable (il suffit de spécifier la colonne à trier et døactiver le tri, ci-dessus tri sur la colonne n°4).

Le tableau précédent met en avant un des résultats présentés dans lorticle de référence cité supra, ci-dessous la courbe de ventilation illustrant le comportement des 2 items mis en avant sur ce corpus : salariés vs travailleurs.

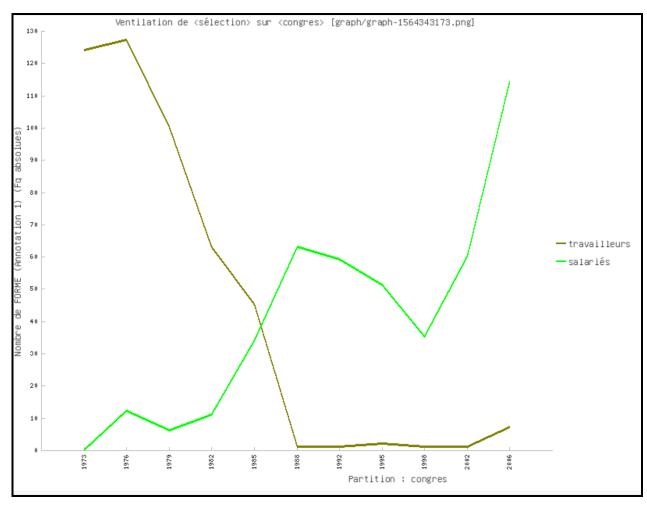


Figure 124 : salariés versus travailleurs dans corpus CFDT 1973-2006

21 Sélection. Gestionnaire de sélection

La sélection est un objet textométrique qui permet de décrire de manière homogène la plupart des objets sur lesquels portent les opérations textométriques. Une sélection est un sous-ensemble détems appartenant à la **Trame**.

Pour construire des unités complexes et étudier leur répartition dans différentes zones d'un corpus de textes, un logiciel de textométrie doit disposer dœun gestionnaire de sélection permettant de constituer, de modifier, de gérer, les sélections à partir de toutes les vues possibles sur le texte et les sélections déjà construites. Le gestionnaire de sélections prend appui sur toutes les présentations possibles du texte ou des comptages dœunités réalisés à partir de ce dernier pour définir dœutres unités qui seront elles-mêmes soumises à des évaluations textométriques.

(Söze-Duval 2008)

Le *Trameur* dispose de fonctionnalités permettant de sélectionner des positions de la *Trame*, de les mémoriser et de constituer des unités réutilisables dans des opérations de ventilation ou de concordance. Løonglet **SELECTION** enregistre au fur et à mesure toutes les sélections réalisées ; il donne aussi accès à løoutil *Gestionnaire de Sélection* qui permet de travailler à partir de ces sélections.



Figure 125: Le Trameur onglet SELECTION

Le *Gestionnaire de Sélection* permet de « sélectionner» des items de la *Trame* (via leur position) ou de « sélectionner » des zones de la *Trame* (via les parties disponibles). Dans les deux cas, on peut ensuite lancer des opérations sur les sélections réalisées.

21.1 Sélection de positions sur la Trame

On distingue deux modes pour sélectionner des positions de la *Trame* :

- Le mode sélection directe à partir de la *Trame* : ce mode est disponible sur les objets éditeur (onglet CADRE et VENTILATION), løbjet concordance (onglet CONCORDANCE) et les objets dictionnaire (onglets FORME, LEMME, CATEGORIE).
- o Le mode **génération døune sélection**: ce mode est accessible dans løoutil *Gestionnaire de Sélection*. Dans ce mode, il est possible de choisir dans un ensemble døunités (formes, segments) vérifiant une propriété visée.

21.1.1 Sélection directe à partir de la *Trame*

21.1.1.1 Sélectionner dans loéditeur du *Cadre*

On dispose dans lønglet **CADRE** døun raccourci clavier (*shift-clic*) permettant de sélectionner les positions visées. Après avoir édité la *Trame* døune partie du *Cadre* (accès à la *Trame* par le *Cadre*), les items de la *Trame* imprimés sont sélectionnables : løactivation du raccourci clavier sur un item déclenche (1) sa sélection (une trace de cette opération est visible dans lønglet **SELECTION**) et (2) une coloration jaune autour de løtem visé :

```
Thift-tlis sur forme: selection de la forme | Clic-droit sur forme: edition annocations de la forme

[11]: C' est un trou de verdure où chante une rivière,
[22]: Accrochant follement aux herbes des haillons
[33]: D' argent; où le soleil; de la montagne fière,
[41]: Luit: c' est un petit val qui mousse de rayons.
```

Figure 126 : Le Trameur sélection sur la Trame (onglet CADRE)

Dans la figure précédente, les positions associées aux items surlignés en jaune sont désormais sélectionnées dans le *Gestionnaire de Sélection*.

21.1.1.2 Sélectionner dans løéditeur des sections

Le même mécanisme est disponible dans lééditeur de la *Trame* par les sections.

21.1.1.3 Sélectionner dans løéditeur des graphiques

Un mécanisme similaire est disponible sur lœditeur de lønglet **VENTILATION** quand celuici est utilisé pour construire la courbe døaccroissement du vocabulaire. Cet éditeur affiche une trace du calcul de la courbe døaccroissement donnant à voir pour chaque position de la *Trame* les valeurs retenues à cette position. Løactivation du raccourci clavier sur une position déclenche sa sélection.

```
Position: 1
Position : 3
              Nombre d'occurrence
                                         Nombre de formes
 osition: 5
              Nombre d'occurrence : 3
                                        Nombre de formes : 3
Position : 7
                                        Nombre de formes :4
              Nombre d'occurrence :
Position: 11 | Nombre d'occurrence :
                                      5 | Nombre de formes :5
               Nombre d'occurrence
                                          Nombre de formes : 6
Position: 15 | Nombre d'occurrence :
                                          Nombre de formes :
Position: 17
Position: 19
               Nombre d'occurrence : 8 |
Nombre d'occurrence : 9 |
                                         Nombre de formes :8
Nombre de formes :9
Position: 21
               Nombre d'occurrence : 10
                                          | Nombre de formes :10
               Nombre d'occurrence : 11
Position: 23
                                          Nombre de formes : 11
Position: 25
               Nombre d'occurrence : 12
                                          Nombre de formes : 12
Position: 27
                Nombre d'occurrence :
Position: 29
               Nombre d'occurrence : 14
                                          Nombre de formes : 14
                Nombre d'occurrence : 15
Position: 34
Position: 36
               Nombre d'occurrence : 16
                                          Nombre de formes : 16
Position : 38
               Nombre d'occurrence : 18 |
                                          Nombre de formes : 18
Position: 42
                Nombre d'occurrence : 19
Position: 44
               Nombre d'occurrence : 20 | Nombre de formes : 20
Position: 48
                Nombre d'occurrence
                                      21
                                          Nombre de formes
Position: 50
               Nombre d'occurrence : 22 | Nombre de formes : 22
                Nombre d'occurrence : 23
                                          Nombre de formes : 22
Position: 56 | Nombre d'occurrence: 24 | Nombre de formes: 23
```

Figure 127 : Le Trameur sélection sur la Trame (onglet VENTILATION)

21.1.1.4 Sélectionner dans les concordances

Le même mécanisme de sélection est disponible dans lønglet **CONCORDANCE** sur les pôles des concordances construites. Le même raccourci clavier (*shift-clic*) permet de sélectionner la position de la forme pôle visée :



Figure 128 : Le Trameur sélection sur une concordance (onglet CONCORDANCE)

21.1.2 Sélectionner dans les dictionnaires

On peut aussi sélectionner les ensembles de positions associées aux entrées des différents dictionnaires (onglets **FORME**, **LEMME**, **CATEGORIE**).

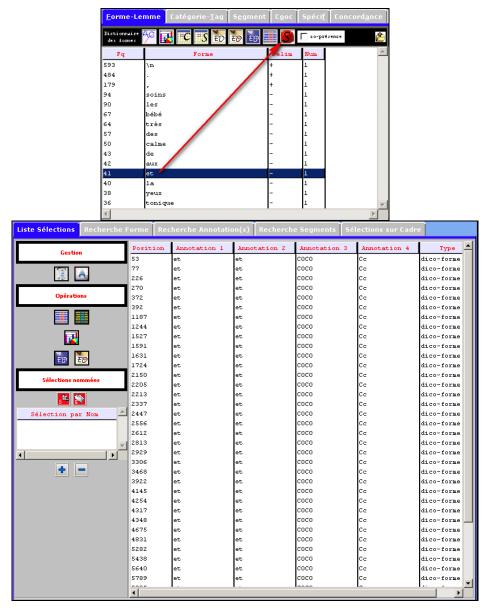


Figure 129 : Le Trameur sélection sur dictionnaires (onglets FORME, LEMME, CATEGORIE)

Dans la première figure qui précède, les items surlignés vont être sélectionnés : lænsemble de leurs positions respectives sont mémorisées (visibles dans le seconde figure) dans le *Gestionnaire de Sélection*.

21.2 Traitements doune sélection dans le Gestionnaire de sélection

Løoutil *Gestionnaire de Sélection* est visible dans løonglet **SELECTION**. Le *Gestionnaire de Sélection* permet :

- De visualiser les items sélectionnés (onglet Liste Sélections)
- De lancer des opérations sur une sélection de positions disponibles dans le Gestionnaire de Sélection
- De rechercher des items en filtrant sur le niveau døannotation n°1 (forme) (onglet **Recherche Forme**) : les items résultants peuvent ensuite être ajoutés au *Gestionnaire* de Sélection
- De rechercher des items en filtrant sur les différents niveaux døannotation disponibles (onglet **Recherche Annotation(s)**) : les items résultants peuvent ensuite être ajoutés au *Gestionnaire de Sélection*
- De rechercher des segments parmi ceux disponibles (quel que soit le niveau déannotation choisi pour le calcul des segments) (onglet Recherche Segments): les segments résultants peuvent ensuite être ajoutés au Gestionnaire de Sélection
- De sélectionner des « zones » du *Cadre* (une ou plusieurs parties) pour y lancer des opérations textométriques (onglet **Sélections sur** *Cadre*)

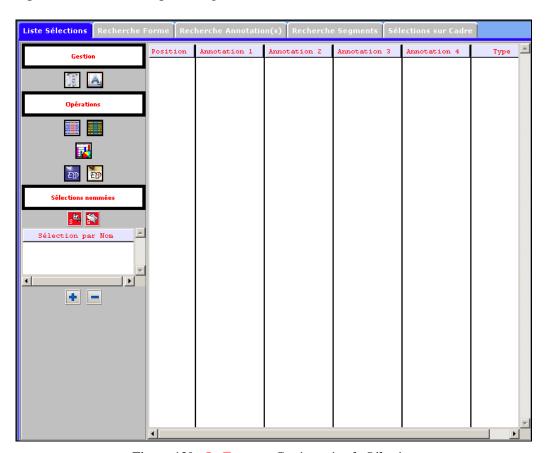


Figure 130 : Le Trameur Gestionnaire de Sélection

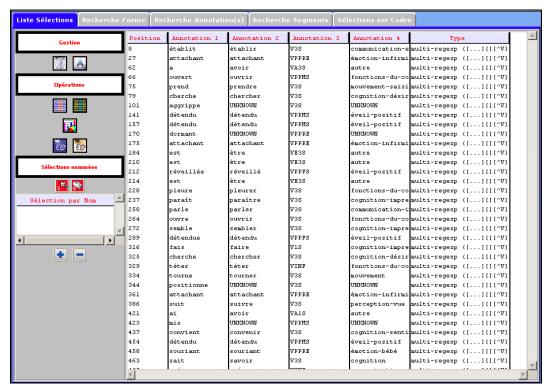


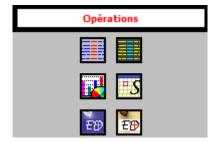
Figure 131 : Le Trameur Gestionnaire de Sélection (positions stockées dans le Gestionnaire de Sélection)

On peut ensuite agir sur les positions sélectionnées *via* les boutons disponibles dans la partie supérieure de la fenêtre du *Gestionnaire de Sélection*.

On peut supprimer toutes les sélections réalisées ou les sélectionner toutes :



On peut, après sélection des positions mémorisées, lancer des opérations sur ces positions : concordance, ventilation, carte des sections, modification des annotations, insertion døune nouvelle annotation



21.3 Opérations sur sélections : concordance, ventilation, modification annotation, nouvelle annotation

21.3.1 Concordance

Løexemple ci-dessous montre le texte de travail avant chargement :

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut.
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid.
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit.
<STRUCTURE="AUTEUR">Arthur Rimbaud
```

La première figure qui suit montre lœnsemble des positions sélectionnées (contour jaune) dans lœditeur de la *Trame* :

```
C' est un trou de verdure où chante une rivière,
23
       Accrochant follement aux herbes des haillons
      D' argent ; où le soleil ; de la montagne fière,
      Luit : c' est un petit val qui mousse de rayons.
      Un soldat jeune, bouche ouverte, tête nue,
      Et la nuque baignant dans le frais cresson bleu,
      Dort ; il est étendu dans l'herbe, sous la nue,
      Pâle dans son lit vert où la lumière pleut.
     Les pieds dans les glaïeuls, il dort. Souriant comme
10
     Sourirait un enfant <mark>malade</mark>, il fait un somme :
11
     Nature, berce-le chaudement : il a froid.
12
      Les parfums ne font pas frissonner sa narine ;
13
      Il dort dans le soleil, la main sur sa poitrine,
      Tranquille. Il a deux trous rouges au côté droit.
```

Figure 132 : Le Trameur sélection directe de position dans lééditeur

La figure suivante montre lœnsemble des positions mémorisées dans le *Gestionnaire de Sélection* :

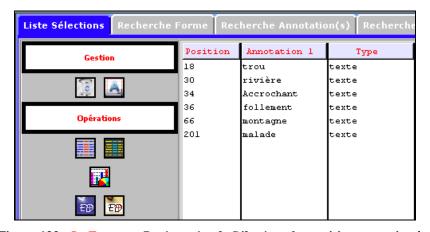


Figure 133 : Le Trameur Gestionnaire de Sélection : les positions enregistrées

La figure suivante montre la sélection réalisée parmi lœnsemble des positions mémorisées (surlignées en noir) dans le *Gestionnaire de Sélection* :

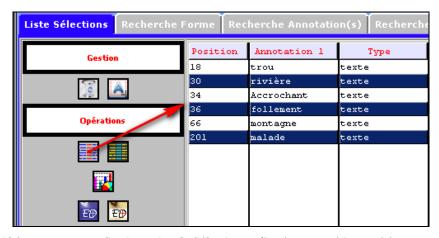


Figure 134 : Le Trameur Gestionnaire de Sélection : sélection parmi les positions enregistrées

Enfin la figure suivante montre le résultat de la concordance réalisée sur les positions sélectionnées :

```
de verdure où chante une <mark>rivière</mark>, Accrochant follement aux
chante une rivière, Accrochant <mark>follement</mark> aux herbes des haillons
Souriant comme Sourirait un enfant <mark>malade</mark>, il fait un somme
```

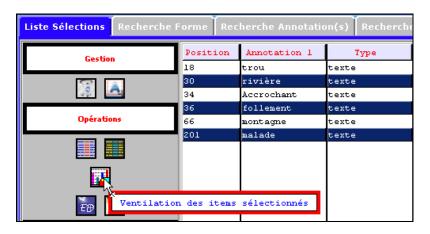
Figure 135 : Le Trameur concordance døune sélection de positions

La concordance réalisée est la concordance des positions sélectionnées :

o position 30 (rivière)
o position 36 (follement)
o position 201 (malade)

21.3.2 Ventilation

En reprenant la même sélection, la figure suivante montre le résultat de la ventilation réalisée sur les positions sélectionnées sur la partition VERS.



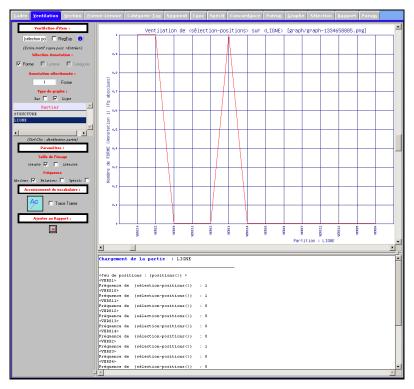


Figure 136 : Le Trameur ventilation døune sélection de positions

La ventilation réalisée est la ventilation des formes associées aux positions sélectionnées :

- o position 30 (rivière) : 1 occurrence de cette forme dans la partie VERS1
- o position 36 (follement) : 1 occurrence de cette forme dans la partie VERS2
- o position 201 (malade) : 2 occurrences de cette forme dans la partie VERS10

21.3.3 Modification des annotations sur les items sélectionnés

Toujours en reprenant la même sélection, la figure suivante montre comment modifier les annotations de tous les items sélectionnés :



CLESTHIA>SYLED>CLA²T

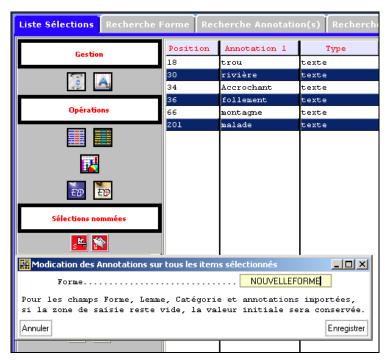


Figure 137 : Le Trameur modification des annotations døune sélection de positions

Une zone de saisie permet de mettre à jour les annotations de tous les items correspondant aux positions sélectionnées. Pour modifier une annotation, il suffit de saisir la nouvelle annotation dans la zone de saisie idoine : les zones de saisie « vides » permettent de conserver les annotations initiales.

Après modification, le Gestionnaire de Sélection est mis à jour :

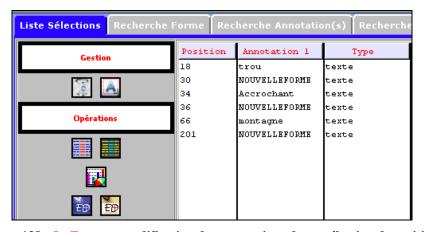


Figure 138 : Le Trameur modification des annotations døune sélection de positions

Ainsi que le dictionnaire des formes :

Dictionnaire des formes	🏸 🔁 🔟 🔣 📆 📆 📆	S	Co-pr	résence 产
Fq	Forme	Delim	Num	1
16	\n	+	1	1
13	,	+	1	
6		+	1	
5	dans	-	1	
5	la	 -	1	
4	le	-	1	
4	•	+	1	
4	il	 -	1	
4	un	 -	1	
4	;	+	1	
3	NOUVELLEFORME	-	1	
3	est	-	1	
3	où	-	1	
3	de	 -	1	
3	:	+	1	▼
4				Þ

Figure 139 : Le Trameur modification des annotations døune sélection de positions

21.3.4 Ajout døune nouvelle annotation sur les items sélectionnés

Toujours en reprenant la même sélection, la figure suivante montre comment modifier les annotations de tous les items sélectionnés :

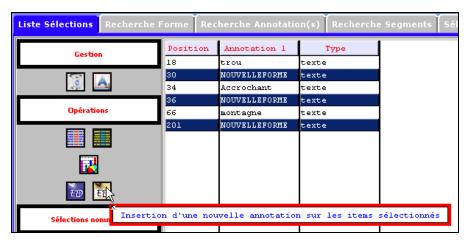


Figure 140 : Le Trameur insertion donne nouvelle annotation sur une sélection de positions

Comme dans létape précédente, après avoir sélectionné des items dans le *Gestionnaire de Sélection*, une zone de saisie permet de définir une nouvelle couche déannotation sur tous les items correspondant aux positions sélectionnées.

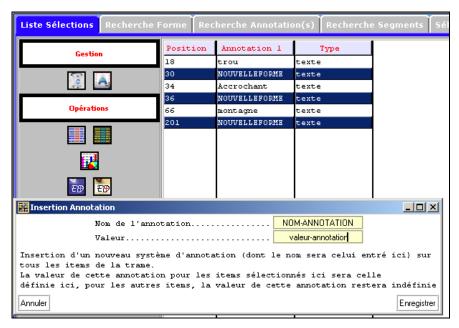


Figure 141: Le Trameur insertion donne nouvelle annotation sur une sélection de positions

Pour insérer une annotation, il suffit de saisir le nom de la nouvelle annotation et sa valeur dans la zone de saisie idoine. Les autres items de la *Trame* auront aussi cette nouvelle couche døannotation mais avec une valeur indéfinie (« unknown »).

Après modification, le Gestionnaire de Sélection est mis à jour :

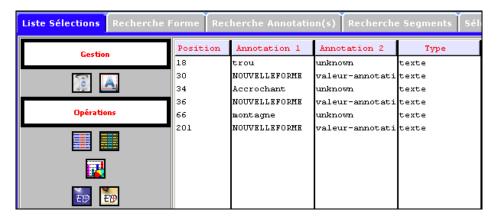


Figure 142 : *Le Trameur* insertion døune nouvelle annotation sur une sélection de positions Ainsi que le dictionnaire des annotations :



Figure 143: Le Trameur insertion dønne nouvelle annotation sur une sélection de positions

21.4 Méthodes de génération de sélection via le Gestionnaire

21.4.1 Génération et sélection døune liste de formes

Løonglet Recherche Forme du *Gestionnaire de Sélection* permet de lancer la recherche døune liste de formes possédant une propriété décrite par un motif exprimé sous la forme døune expression régulière. Dans la figure suivante, on recherche une liste de formes possédant la propriété : \bd (forme commençant par d).

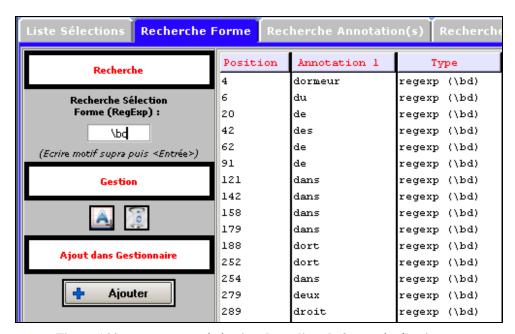


Figure 144 : Le Trameur génération døune liste de formes à sélectionner

On peut ensuite choisir les positions à sélectionner *i.e.* choisir celles qui seront ajoutées laux sélections en cours dans løonglet **Liste Sélections**.

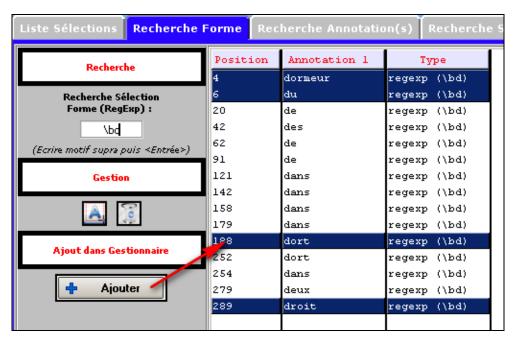


Figure 145 : Le Trameur sélection doune liste de formes

21.4.2 Génération et sélection døune liste de segments

Løonglet Recherche Segments du Gestionnaire de Sélection permet de lancer la recherche døune liste de segments contenant un motif exprimé sous la forme døune expression régulière. Dans la figure suivante, on recherche une liste de segments contenant la forme « le ».

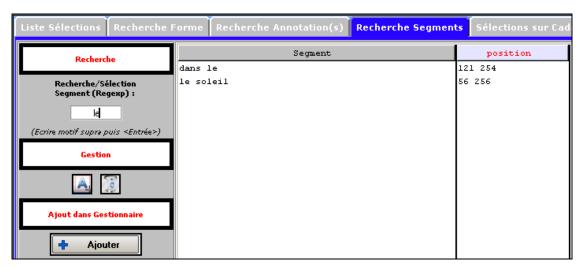


Figure 146 : Le Trameur génération døune liste de segments à sélectionner

Comme précédemment, on peut ensuite choisir les segments à sélectionner et les ajouter aux sélections en cours dans lønglet Liste Sélections. Cet ajout consiste à mémoriser la position initiale du segment visé : les positions ajoutées sont celles associées à la première forme du segment visé (dans læxemple ci-dessus : le segment « dans le » est présent à la position 121 et 254 de la *Trame* (positions respectives des 2 formes « dans » amorçant le segment), ces positions mémorisées seront aussi associées à la longueur du segment ajouté (dans læxemple ci-dessus, on mémorise donc les positions 121 et 254 et la longueur 2).

21.4.3 Génération et sélection multicritère

Après chargement døune base textométrique, chaque item de la *Trame* est associé à un liste døannotation, cette liste peut être enrichie pour ajout dynamique døune nouvelle couche døannotation (cf *Ajout/modification døannonation sur la Trame*). Løonglet Recherche døannotation(s) permet de sélectionner des items de la *Trame* en paramétrant une recherche multicritère sur les différentes annotations disponibles pour chaque item.

Recherche							
Annotations (RegExp):							
	Forme						
	Lemme						
	Catégorie						

Lønterface de cet onglet est reconstruite dynamiquement dès quøune annotation supplémentaire est intégrée dans la base en cours de traitement.

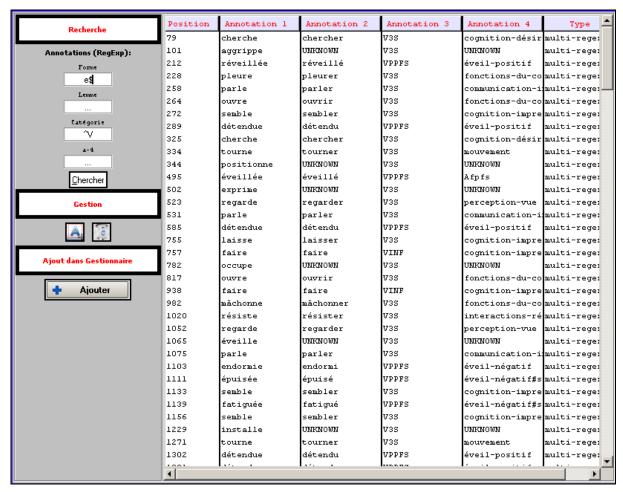


Figure 147 : Le Trameur génération doune liste doitems par équation de recherche

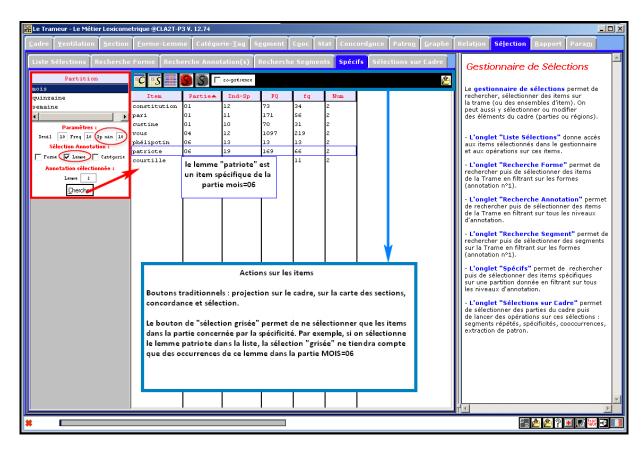
Dans læxemple précédent, la recherche définie :



permet de sélectionner les items de la *Trame* ayant un niveau døannotation n°1 (Forme) se terminant par la lettre e et un niveau døannotation n°3 (Catégorie) commençant par la lettre V, les 2 autres niveaux døannotation étant indifférents. La liste produite illustre les items de la *Trame* possédant ces propriétés.

21.4.4 Génération døune liste døtems spécifiques

Lønglet Spécifs du Gestionnaire de Sélection permet de lancer la recherche døune liste døtems spécifiques sur une partition donnée. Dans la figure suivante, on recherche des lemmes spécifiques (ici un indice de spécificité supérieur à 10) sur la partition MOIS de la base chargée.



Le résultat donne à voir les items (lemmes) spécifiques sur les parties de la partition visée. Ces éléments peuvent ensuite être utilisés pour des opérations déjà présentées : ventilation, concordance, carte des sections, sélection.

Le bouton de sélection « grisé » permet døajouter au *Gestionnaire de Sélection* uniquement les occurrences de la partie visée.

21.5 Sélection de parties (comme ensembles de sélection de positions sur la *Trame*)

Le *Cadre* døune base textométrique est défini comme un système permettant des accès aux textes, il regroupe des systèmes de parties (partition recouvrante ou non de la surface textuelle). On peut de fait assimiler ces accès aux textes à des sélections de positions sur la *Trame*.

Si on considère le texte de travail suivant présenté dans 2 formats possibles en entrée du *Trameur* :

```
<STRUCTURE="TITRE">Le dormeur du val

<STRUCTURE="TEXTEPOEME">

<LIGNE="VERS1"> C'est un trou de verdure où chante une rivière,

<LIGNE="VERS2"> Accrochant follement aux herbes des haillons

<LIGNE="VERS3"> D'argent ; où le soleil ; de la montagne fière,

<LIGNE="VERS4"> Luit : c'est un petit val qui mousse de rayons.
```

```
<?xml version="1.0" encoding="iso-8859-1"?>
<poeme>
<structure niveau="TITRE">Le dormeur du val</structure>
<structure niveau="TEXTEPOEME">
cligne type="VERS1"> C'est un trou de verdure où chante une rivière,</ligne>
cligne type="VERS2"> Accrochant follement aux herbes des haillons</ligne>
cligne type="VERS3"> D'argent ; où le soleil ; de la montagne fière,</ligne>
cligne type="VERS4"> Luit : c'est un petit val qui mousse de rayons.
</rr>
</ra>
```

Une segmentation *choisie* de la partie textuelle de ce texte (pour ces 2 formats) peut conduire à une *Trame* décrite ainsi :

Forme	Le		dormeur		du		val	\n	í	rayons	
Su											
sitio										7	s
Pos	1	2	3	4	S.	9	7	∞	`=	sod	sod

(pos désignant la dernière position de la *Trame*)

Pour le texte dans son premier format, les partitions disponibles peuvent être décrites ainsi :

Partition STRUCTURE

La partie STRUCTURE_TITRE couvre la *Trame* de la position 1 à la position 8

La partie STRUCTURE_TEXTEPOEME couvre la *Trame* de la position 9 à la position pos

Partition LIGNE

La partie LIGNE_VERS1 couvre la *Trame* de la position 9 à la position í Etc.

Pour le texte dans son second format, les partitions disponibles peuvent être décrites ainsi :

Partition poeme

La partie poeme couvre la *Trame* de la position 1 à la position pos

Partition structure

La partie structure TITRE couvre la *Trame* de la position 1 à la position 8

La partie structure_TEXTEPOEME couvre la *Trame* de la position 9 à la position pos

Partition ligne

La partie ligne_VERS1 couvre la *Trame* de la position 9 à la position í Etc.

Chacun des parties décrites est ainsi une sélection de positions sur la *Trame* couvrant les items compris entre la position initiale et la position finale sur la *Trame* décrites pour chaque partie.

Løoutil *Gestionnaire de Sélection* dispose døun onglet permettant de sélectionner des zones sur la *Trame* en søappuyant sur les parties c'est-à-dire sur les sélections de positions définies par le *Cadre* de la base textométrique.

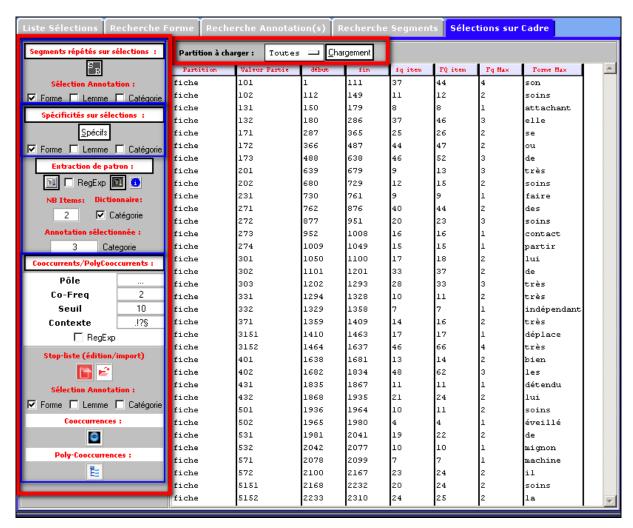


Figure 148 : Le Trameur sélection de parties

Une ou plusieurs parties étant sélectionnées, 4 types døpérations sont disponibles sur les parties sélectionnées :

- calcul des segments
- calcul de spécicifité
- extraction de patron
- calcul de cooccurrents / polycooccurrents

Les figures suivantes illustrent ces différents modules. Dans tous les cas, après avoir sélectionné une ou plusieurs parties, les calculs invoqués se font sur la sélection visée.

21.5.1 Calcul de segments

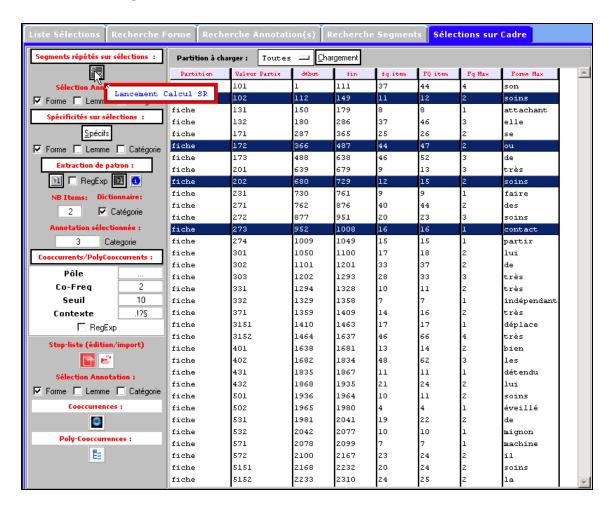




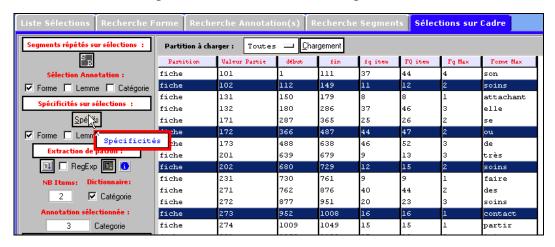
Figure 149 : Le Trameur calcul de segment sur une sélection de parties

On commence par sélectionner des parties (ici 4 parties) puis on déclenche le calcul des SR avec le bouton idoine. Les segments résultants ne sont en fait disponibles que dans 2 des parties sélectionnées : une concordance des segments construits donne à voir leurs réalisations sur la *Trame* :

Figure 150 : Le Trameur calcul de segment sur une sélection de parties

21.5.2 Calcul de spécificité

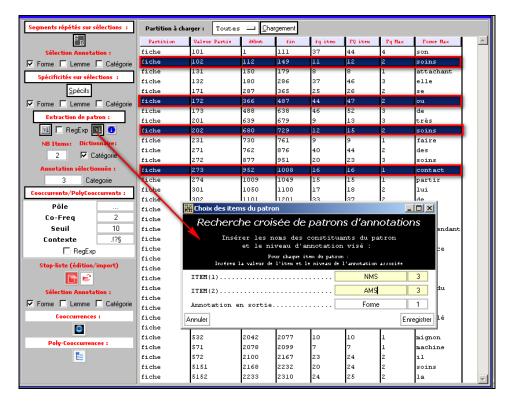
Après avoir sélectionné des parties, on lance le calcul des spécificités :



	1 - 0540		
Nombre d'occurrences du texte globa			
Nombre d'occurrences dans la partie Seuil : 10	Visee = 90		
Seull : 10 (Specificites positives en haut de	Notes and the second second		
(specificates positives en haut de Le fichier construit :	liste, negatives en bas;		
Le fichier construit : C:/SERGEF~1/SF-WORK/07-DEVEL/ANR-TE	1 (MEZCEAR 1 (DACEC 1 (DACE R 1 ((1	1004663030 +
C:/SERGEF~1/SF-WORK/U/-DEVEL/ANK-IE	~1/MKGRAP~1/BASES-~1/BASE-P~1/exp	orc/resulcspecii	-source-1334662323.t
Forme	Ind-Specif	Fq-Totale	Fq-Partie
j'	2.5	1	1
ai	2.5	1	1
teint	2.5	1	1
pâle	2.5	1	1
situation	2.5	1	1
souriant	2.5	1	1
forte	2.5	1	1
jusqu'	2.5	1	1
lache	2.5	1	1
impossibilité	2.5	1	1
mis	2.5	1	1
mauvais	2.5	1	1
interaction	2.5	1	1
couverture	2.5	1	1
bain	2.5	1	1
mâchonne	2.5	1	1
sensible	2.5	1	1
accrocheur	2.5	1	1
créer	2.5	1	1
timide	2.5	1	1
exigeante	2.5	1	1
hypotonique	2.5	1	1
raide	2.5	1	1
douleur	2.5	1	1
convient	2.5	1	1
infecté	2.5	1	1
ou	2.4	9	2
suit	2.2	2	1
détendu	2.2	12	2
infirmière	2.2	2	1
oedématié	2.2	2	1

Figure 151 : Le Trameur calcul de spécificités sur une sélection de parties

21.5.3 Extraction de patron



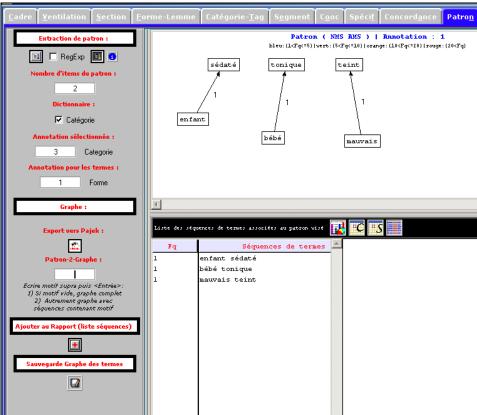


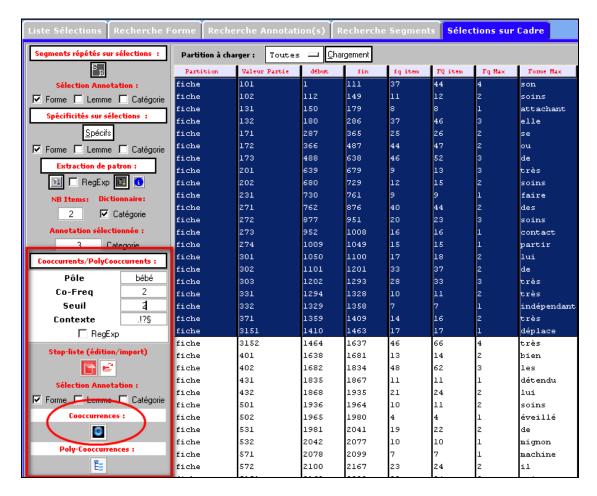
Figure 152 : Le Trameur extraction de patron sur une partie sélectionnée

Comme précédemment, une concordance des termes produits illustrent la localisation de ces derniers sur la *Trame* (dans løune des parties sélectionnées) :

Figure 153 : Le Trameur extraction de patron sur une partie sélectionnée

21.5.4 Calcul de cooccurrents

Cas n°1 : unité contextuelle déterminée par des délimiteurs



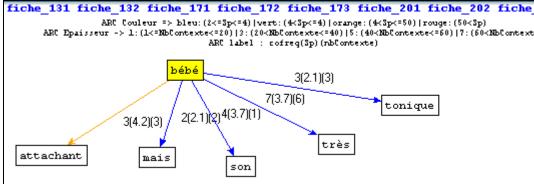


Figure 154 : Le Trameur cooccurrents/polycooccurrents sur une partie sélectionnée (1)

Le calcul des polycooccurrents peut ensuite être lancé.

Cas n°2 : unité contextuelle déterminée par le *Cadre*

On reprend ici lexemple de texte de travail vu plus haut :

```
<?xml version="1.0" encoding="iso-8859-1"?>
       <mark><?</mark>xml-stylesheet type="text/xsl" href="q2.xsl"<mark>?></mark>
3
    =<TEI>
4
          <teiHeader>
5
              <fileDesc>
6
                   <titleStmt>
7
                       <title>Poésies</title>
8
                       <author>Arthur Rimbaud</author>
9
                   </titleStmt>
10
               </fileDesc>
11
          </teiHeader>
12
          <text>
13
               <group>
14
                   <text>
15
                       <front>
16
                           <head>Le Dormeur du Val</head>
17
                           <dateline>7 octobre 1870</dateline>
18
                       </front>
19
                       <body>
20
                           <10>
21
                                    <1>C'est un trou de verdure où chante une rivière</1>
23
                                    <1>Accrochant follement aux herbes des haillons</1>
24
                                    <1>D'argent ; où le soleil, de la montagne fière,</1>
25
                                    <1>Luit : c'est un petit val qui mousse de rayons.</1>
                                </1g>
26
27
                               <1g>
28
                                    <1>Un soldat jeune, bouche ouverte, tête nue,</1>
29
                                    <1>Et la nuque baignant dans le frais cresson bleu,</1>
                                    <1>Dort ; il est étendu dans l'herbe, sous la nue,</1>
31
                                    <1>Pâle dans son lit vert où la lumière pleut.</1>
                                </1g>
33
                               <1q>
34
                                    <1>Les pieds dans les glaïeuls, il dort. Souriant comme</1>
3.5
                                    <1>Sourirait un enfant malade, il fait un somme :</1>
36
                                    <1>Nature, berce-le chaudement : il a froid.</1>
                                </1g>
38
                                <1g>
39
                                    <1>Les parfums ne font pas frissonner sa narine ;</1>
40
                                    <1>Il dort dans le soleil, la main sur sa poitrine</1>
41
                                    <1>Tranquille. Il a deux trous rouges au côté droit.</1>
42
                               </1q>
43
                           </1g>
44
                       </body>
45
                   </text>
46
               </aroun>
47
           </text>
48
      </TEI>
```

Figure 155 : Le Trameur cooccurrents/polycooccurrents sur une partie sélectionnée (2)

Dans la figure qui suit, la partition 1g a été chargée dans le *Gestionnaire de Sélection* et certaines instances de strophes sélectionnées; seules les strophes sélectionnées seront utilisées pour définir les contextes utilisés par le calcul.

CLESTHIA>SYLED>CLA²T

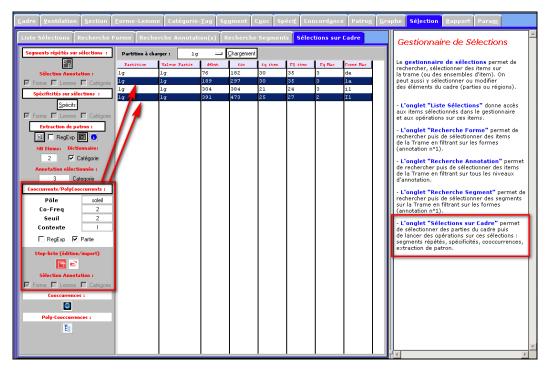


Figure 156 : Le Trameur cooccurrents/polycooccurrents sur une partie sélectionnée (3)

Comme précédemment, le calcul des cooccurrents utilisent des paramètres similaires. On peut soit définir des unités contextuelles (au sein des parties sélectionnées) en mettant en avant des caractères délimiteurs (cf cas n°1 supra), soit choisir døutiliser une zone du Cadre pour déterminer les contextes.

Si cette seconde option est choisie, <u>il est important de søassurer</u> que la partie utilisée pour définir le contexte est une sous-partie des parties sélectionnées dans le *Cadre*.

Dans la figure précédente, les parties du *Cadre* sélectionnées correspondent à la première et à la troisième strophe ; la partie utilisée pour définir les contextes est la balise 1, les contextes retenus correspondent donc aux vers constitutifs de ces 2 strophes :

```
Un soldat jeune, bouche ouverte, tête nue, Et la nuque baignant dans le frais cresson bleu, Dort; il est étendu dans l'herbe, sous la nue, Pâle dans son lit vert où la lumière pleut. Les pieds dans les glaïeuls, il dort. Souriant comme Sourirait un enfant malade, il fait un somme: Nature, berce-le chaudement: il a froid.
```

Et au final, les contextes retenus pour le calcul rassemblent les vers contenant le pôle : dans le cas présent, aucun !

21.5.5 Calcul de cooccurrents sur corpus chronologique

On considère dans cette partie que lon travaille sur le corpus chronologique « Le Père Duchesne » (cf Documentation en ligne Lexico3²⁰), noté ici DUCHN.

-

²⁰ http://www.tal.univ-paris3.fr/lexico/manuels.htm

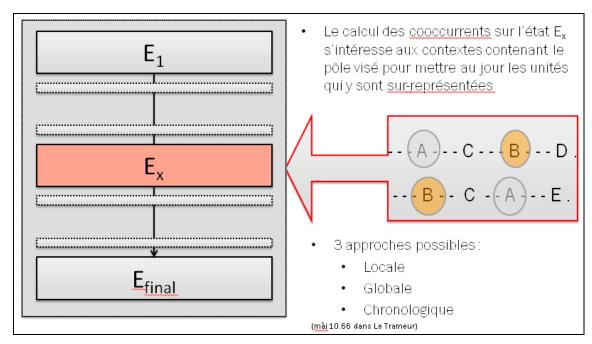


Figure 157: Cooccurrents, approche locale / globale / chronologique

Approche « locale » : on travaille sur une sélection de partie pour une partition donnée

Partition sélectionnée : MOIS

Sélection de parties déclinées ci-dessous.

Calcul des cooccurrents de « patriotes »

Cas n°1: 1 partie sélectionnée (MOIS=6), cooccurrence « locale » sur cette partie

Liste Sélections Recherche F	Forme Reche	rche Annotati	ion(s) R	echerche	Segment	Sélec	tions sur	Cadre	
Segments répétés sur sélections :	Partition à charger: mois — Chargement								
S _R	Partition	Valeur Partie	début	fin	fq item	FQ item	Fq Max	Forme Max	
Sélection Annotation :	mois	01	1	41135	3439	19256	886	de	
Forme ☐ Lemme ☐ Catégorie	mois	02	41136	85199	3778	20343	875	de	
Spécificités sur sélections :	mois	03	85200	126470	3563	19278	853	de	
	mois 04		126471	165407	3542	18004	753	de	
<u>S</u> pécifs	mois 05		165408	204487	3373	17814	746	de	
Forme ☐ Lemme ☐ Catégorie	mois	06	204488	243308	3392	17545	757	de	
Annotation sélectionnée :	mois	07	243309	279232	3297	16295	669	de	
Forme 1	mois	08	279233	309530	2970	13644	591	de	
Extraction de patron :									
🔟 🗆 RegExp 🔃 🚺									
NB Items: Dictionnaire:									
2 Catégorie									
Annotation sélectionnée :									
3 Categorie									
Cooccurrents/PolyCooccurrents:									
Pôle patriotes									
Co-Freq 4									
Seuil 6									
Contexte §									
☐ RegExp ☐ Partie									
Stop-liste (édition/import)									
Sélection Annotation :									
▼ Forme Lemme Catégorie									
Cooccurrences :									
□ □ L □ G □ Ch									

Figure 158 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (1)

- Option L (local) (activée en bas à gauche)
- On « oublie » le reste du corpus (pas de fond) : les paramètres lexicométriques globaux choisis sont ceux de cette partie.
- T : Nb occurrence de la partie
- t : nb occcurrence des contextes contenant le pôle
- F : fréquence døune forme sur la partie
- f : fréquence døune forme dans les contextes contenant le pôle

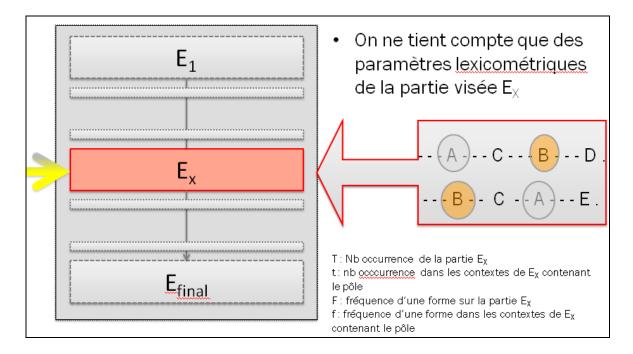


Figure 159: Cooccurrents, approche locale

Le réseau produit a løallure suivante :

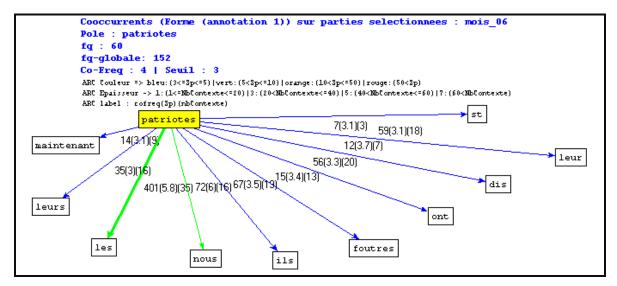


Figure 160 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (2)

Réitérer le même calcul sur dœutres parties permet døbtenir un « réseau chronologique de cooccurrents » du même pôle.

Cas n°2: 1 partie sélectionnée (MOIS=6), cooccurrence « globale » sur cette partie

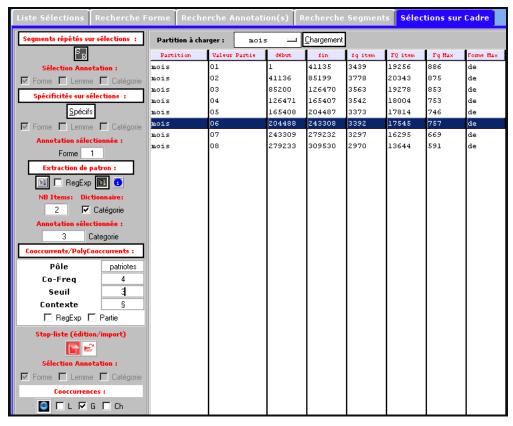


Figure 161 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (3)

- Option G (global) (activée en bas à gauche)
- On conserve le reste du corpus (prise en compte du fond) : les paramètres lexicométriques globaux choisis sont ceux du corpus complet.
- T: Nb occurrence du corpus complet
- t : nb occcurrence des contextes contenant le pôle
- F: fréquence døune forme dans le corpus complet
- f : fréquence døune forme dans les contextes contenant le pôle

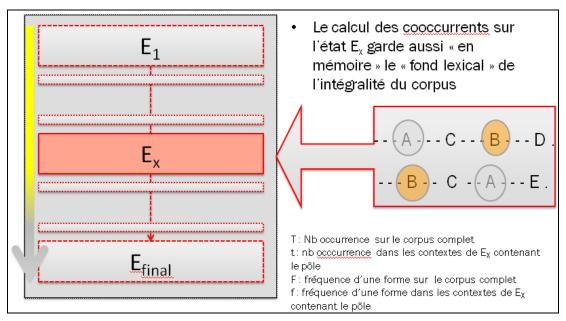


Figure 162: Cooccurrents, approche globale

Le réseau produit a løallure suivante :

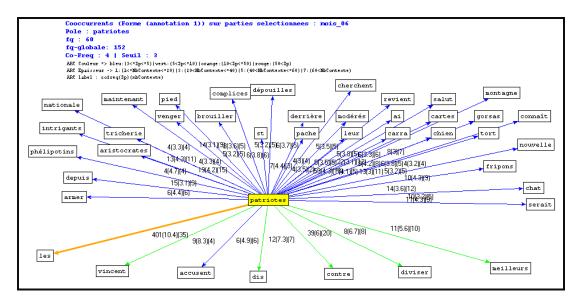
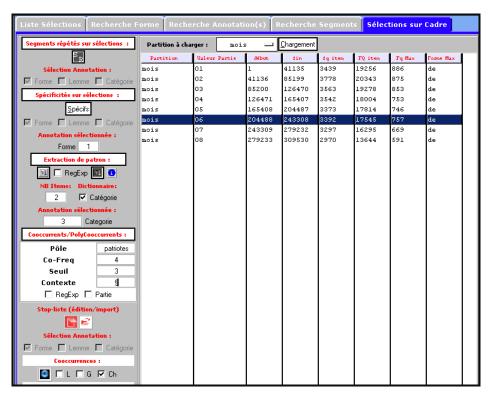


Figure 163 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (4)



Cas n°3: 1 partie sélectionnée (MOIS=6), cooccurrence « chronologique » sur cette partie

Figure 164 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (5)

- Option Ch (chrono) (activée en bas à gauche)
- On conserve les parties du corpus antérieures chronologiquement à celle choisie (prise en compte partielle du fond) : les paramètres lexicométriques globaux choisis sont ceux du corpus constitué par les parties antérieures et la partie sélectionnée.
- T : Nb occurrence dans les parties antérieures et la partie sélectionnée
- t : nb occcurrence des contextes contenant le pôle
- F : fréquence døune forme dans les parties antérieures et la partie sélectionnée
- f : fréquence døune forme dans les contextes contenant le pôle

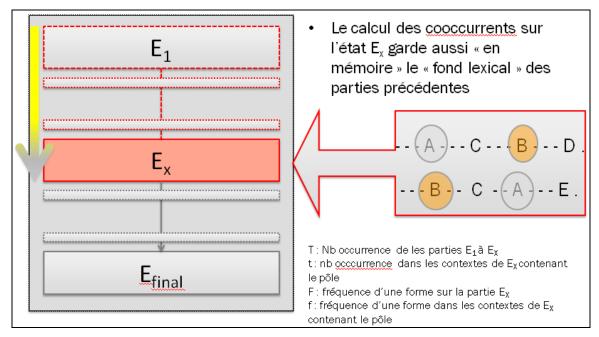


Figure 165: Cooccurrents, approche chronologique

Le réseau produit a løallure suivante :

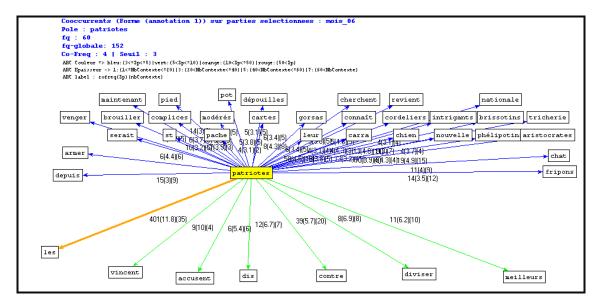


Figure 166 : Le Trameur cooccurrents sur une partie sélectionnée dans un corpus chronlogique (6)

La forme « accusent » serait ici un cooccurrent chronologiquement « très spécifique » de «patriotes » sur cette partie.

21.6 Gestion des sélections : sauvegarde et chargement

Un processus permet de **sauvegarder** des sélections : la sauvegarde døune sélection permet de **nommer** cette sélection (ce nom étant ensuite utilisé pour activer ou inhiber les positions concernées par la sélection), de la **décrire** (édition døune zone de texte pour qualifier de manière précise le contenu de la sélection).

Un autre processus permet aussi de **charger** une sélection préalablement construite. Le sélection chargée est ensuite *activable* ou *inhibable*.



Figure 167 : Le Trameur Gestion des sélections : sauvegarde et chargement

21.6.1 Sauvegarde døune Sélection

Les figures qui suivent présentent le processus de sauvegarde.

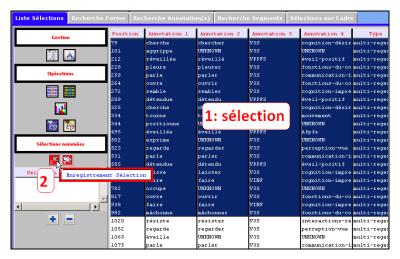


Figure 168 : Le Trameur sauvegarde døune sélection (1)



Figure 169 : Le Trameur sauvegarde døune sélection (2)

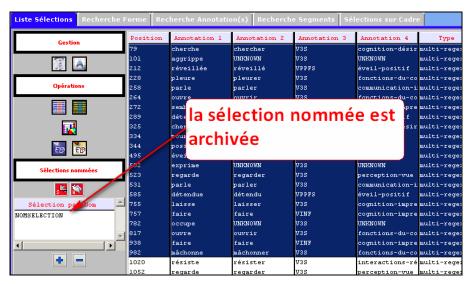


Figure 170 : Le Trameur sauvegarde døune sélection (3)

Le processus décrit ci-dessus génère un fichier (dans le dossier selections du répertoire courant du *Trameur*) regroupant les différentes informations identifiant la sélection.

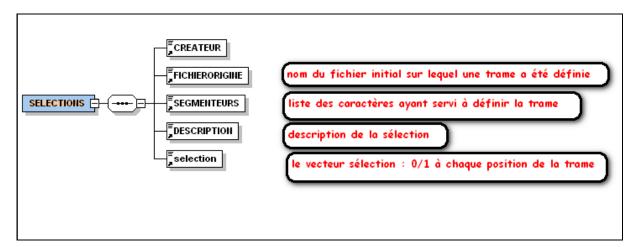
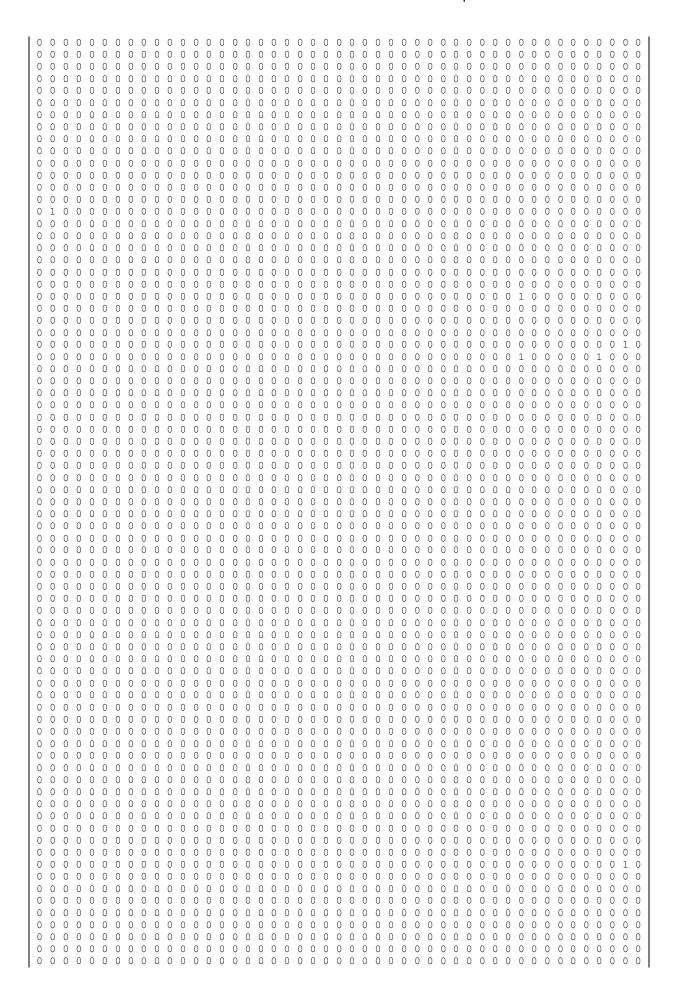
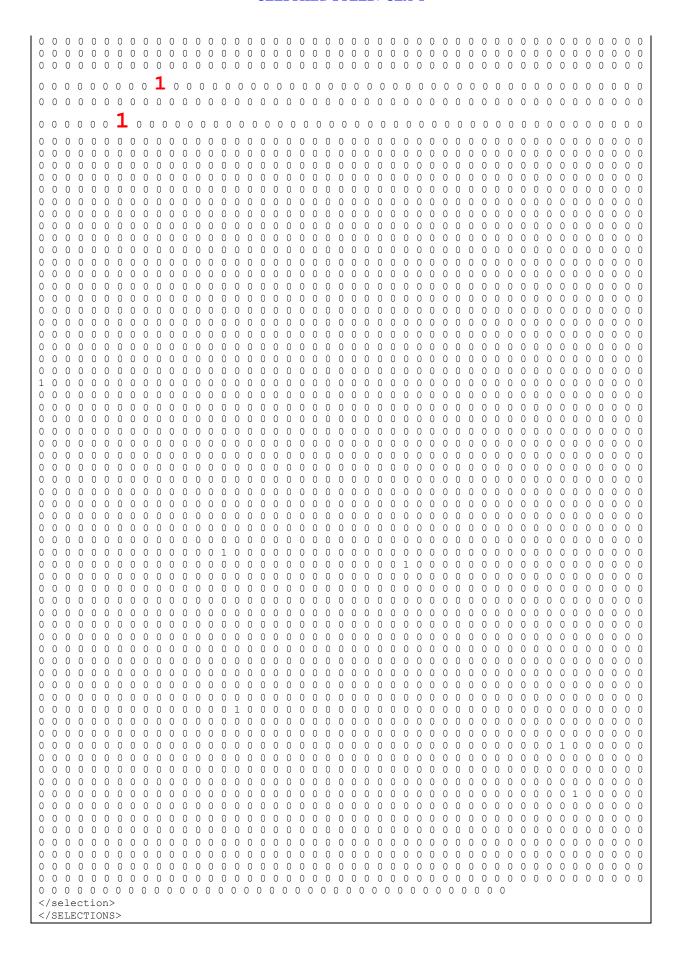


Figure 171 : Le Trameur arborescence XML døune sélection

On donne ci-dessous le contenu døun fichier produit (selection_NOMSELECTION_1334663471.xml), les positions sélectionnées ont ici été marquées « lisiblement » (1 rouge):

```
<?xml version="1.0" encoding="utf-8"?>
<SELECTIONS>
<CREATEUR>Le Trameur 8.00 (8.0b010)</CREATEUR>
<FICHIERORIGINE>base-trameur-p96-annot.txt</fichierorigine>
<SEGMENTEURS><![CDATA[ :,</pre>
?).(!]]></SEGMENTEURS>
<DESCRIPTION><![CDATA[la liste sélectionnée contient le résultat de la recherche de tous les</pre>
adjectifs associés \tilde{\mathbf{A}} l'annotation "émotion-bébé".
]]></DESCRIPTION>
<selection nom="emotion-bebe"</pre>
              dimension="7181">
0 0
                                       0 0 0 0 0 0
0
                        0
                         0 0 0 0 0 0 0
                                0
                                 0
                                  0
                                    0
                                     0
                                      0 0
                                   0
```





21.6.2 Chargement døune Sélection

Dans les figures qui suivent, on charge une sélection (correspondant à celle enregistrée cidessus dans le fichier « selection NOMSELECTION 1334663471.xml »).

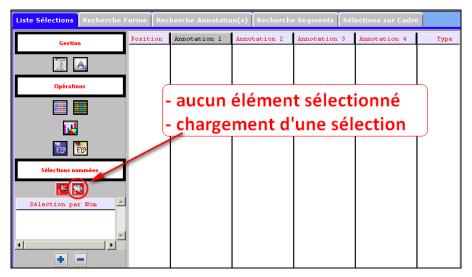


Figure 172 : Le Trameur chargement døune sélection (1)

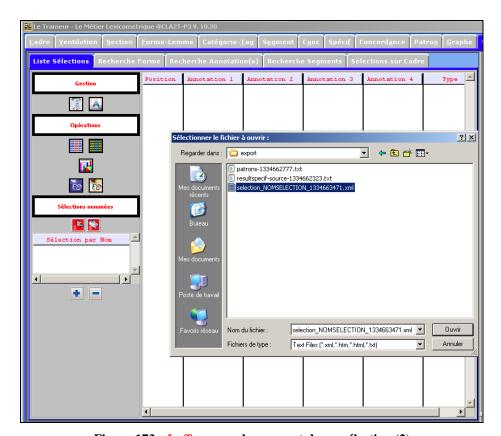


Figure 173 : Le Trameur chargement døune sélection (2)

La sélection chargée peut ensuite être activée ou inhibée :

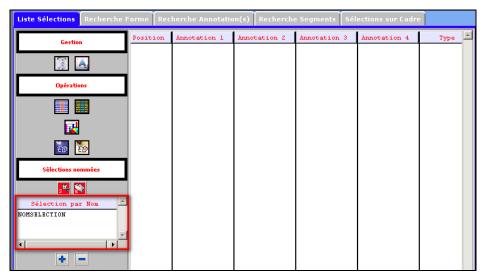


Figure 174 : Le Trameur activation doune sélection nommée (1)

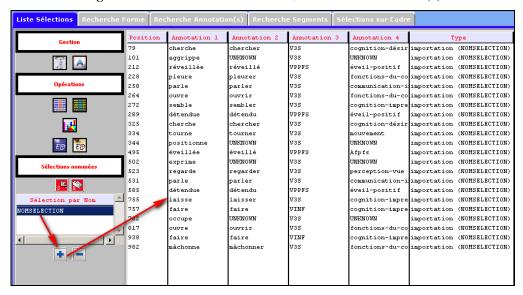


Figure 175 : Le Trameur activation døune sélection nommée (2)

Løactivation døun *Clic-droit* sur la sélection nommée permet døaccéder à sa description (que løon peut modifier ou enrichir).

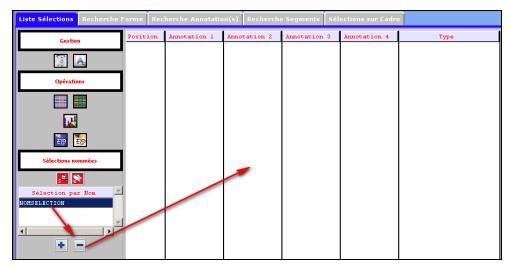


Figure 176 : Le Trameur inhibation doune sélection nommée

22 Importation døannotation sur une *Trame*

Il est possible déimporter un ou plusieurs systèmes déannotations sur une *Trame* et un *Cadre* déjà chargés dans le *Trameur*. Ce processus consiste à annoter chaque position de la *Trame* en lui associant une nouvelle « étiquette ».

Pour réaliser cette importation, <u>il est nécessaire de disposer døune description précise de la Trame</u> sur laquelle on doit réaliser la projection døune nouvelle annotation : en particulier, le nombre døtems et le statut de chacun des items (forme ou délimiteur). <u>Pour disposer døune telle description de la Trame</u>, il est nécessaire de prendre appui sur le fichier construit à <u>løissue døun export de la base en cours de traitement</u> (*cf* plus loin dans ce manuel, présentation du module døxport). Le fichier døannotation à importer peut être construit automatiquement dans un éditeur à partir du fichier døxport de la base.

Chaque item (défini par une position unique) est associé à sa forme graphique initiale (løannotation n°1 dite « Forme »), éventuellement son lemme (løannotation n°2 dite « Lemme ») et sa catégorie (løannotation n°3 dite « Catégorie ») si *treetagger* a été paramétré, et une nouvelle « étiquette » issue du système døannotations importé.

Annotation importée n°1	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	A(7)	A(8)	
Catégorie		Cat(Le)		Cat(dormeur)		Cat(du)	} !	Cat(val)	í
Lemme		Lemme(Le)		Lemme(dormeur)		Lemme(du)		Lemme(val)	í
Forme		Le		dormeur		du	[val	í
Positions	1	2	3	4	5	9	7	8	í

Le format du fichier døannotations à importer søappuie sur le fichier de tramage issu de løexport présenté *infra*.

Dans ce qui suit, nous considérons que nous allons associer une *couleur* à chaque item de la *Trame* :

couleur	bleu	vert	rouge	orange	bleu	rouge	jaune	bleu	
Forme		Le		dormeur		du		val	í
Positions	1	2	3	4	5	9	<i>L</i>	8	Í

CLESTHIA>SYLED>CLA²T

```
entête
Le fichier døannotation complet à importer a løallure suivante :
<?xml version="1.0" encoding="iso-8859-1"?>
<importannotationsurtrame>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), importa<mark>b</mark>le et
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
Samedi 26 Mai 2007
12:02:30
  Ce document n'est pas encore publie.
</publicationStmt>
<sourceDesc>
Fichier d'import d'annotations sur une Trame deja generee sur le fichier initial
                                                                      C:/SFleury/DEVEL/WIP-ANR-
Fichier
                  initialement
                                         traite
                                                          :
textometrrie/mkGraphTrameCadre/textes/dormeurL3.txt
Encodage : iso-8859-1
</sourceDesc>
</fileDesc>
                                                                                   Trame des
</teiheader>
<items>
<item type="delim" pos="1"><a>bleu</a></item>
                                                                                   annotatio
<item type="forme" pos="2"><a>vert</a></item>
<item type="delim" pos="3"><a>rouge</a></item>
<item type="forme" pos="4"><a>orange</a></item>
<item type="delim" pos="5"><a>bleu</a></item>
<item type="forme" pos="6"><a>rouge</a></item>
<item type="delim" pos="7"><a>jaune</a></item>
<item type="forme" pos="8"><a>bleu</a></item>
<item type="delim" pos="9"><a>noir</a></item>
<item type="delim" pos="10"><a>bleu</a></item>
<item type="delim" pos="11"><a>vert</a></item>
<item type="forme" pos="12"><a>vert</a></item>
<item type="delim" pos="13"><a>jaune</a></item>
<item type="delim" pos="14"><a>vert</a></item>
<item type="forme" pos="15"><a>rouge</a></item>
<item type="delim" pos="16"><a>bleu</a></item>
<item type="forme" pos="17"><a>vert</a></item>
<item type="delim" pos="18"><a>rouge</a></item>
<item type="forme" pos="19"><a>bleu</a></item>
<item type="delim" pos="20"><a>rouge</a></item>
<item type="forme" pos="21"><a>noir</a></item>
<item type="delim" pos="22"><a>jaune</a></item>
<item type="forme" pos="23"><a>noir</a></item>
<item type="delim" pos="24"><a>vert</a></item>
<item type="forme" pos="25"><a>bleu</a></item>
<item type="delim" pos="26"><a>rouge</a></item>
<item type="forme" pos="27"><a>noir</a></item>
<item type="delim" pos="28"><a>bleu</a></item>
<item type="forme" pos="29"><a>noir</a></item>
<item type="delim" pos="30"><a>noir</a></item>
<item type="forme" pos="31"><a>bleu</a></item>
<item type="delim" pos="32"><a>jaune</a></item>
<item type="delim" pos="33"><a>vert</a></item>
<item type="delim" pos="34"><a>rouge</a></item>
<item type="delim" pos="35"><a>jaune</a></item>
<item type="forme" pos="36"><a>bleu</a></item>
<item type="delim" pos="37"><a>bleu</a></item>
<item type="forme" pos="38"><a>vert</a></item>
<item type="delim" pos="39"><a>noir</a></item>
<item type="forme" pos="40"><a>rouge</a></item>
<item type="delim" pos="41"><a>bleu</a></item>
<item type="forme" pos="42"><a>noir</a></item>
<item type="delim" pos="43"><a>bleu</a></item>
<item type="forme" pos="44"><a>jaune</a></item>
<item type="delim" pos="45"><a>noir</a></item>
<item type="forme" pos="46"><a>bleu</a></item>
<item type="delim" pos="47"><a>bleu</a></item>
<item type="delim" pos="48"><a>vert</a></item>
```

```
<item type="delim" pos="49"><a>bleu</a></item>
<item type="forme" pos="50"><a>rouge</a></item>
<item type="delim" pos="51"><a>noir</a></item>
<item type="delim" pos="52"><a>bleu</a></item>
<item type="forme" pos="53"><a>bleu</a></item>
<item type="delim" pos="54"><a>rouge</a></item>
<item type="delim" pos="55"><a>jaune</a></item>
<item type="delim" pos="56"><a>vert</a></item>
<item type="forme" pos="57"><a>bleu</a></item>
<item type="delim" pos="58"><a>rouge</a></item>
<item type="forme" pos="59"><a>bleu</a></item>
<item type="delim" pos="60"><a>noir</a></item>
<item type="forme" pos="61"><a>noir</a></item>
<item type="delim" pos="62"><a>bleu</a></item>
<item type="delim" pos="63"><a>noir</a></item>
<item type="delim" pos="64"><a>rouge</a></item>
<item type="forme" pos="65"><a>bleu</a></item>
<item type="delim" pos="66"><a>jaune</a></item>
<item type="forme" pos="67"><a>vert</a></item>
<item type="delim" pos="68"><a>rouge</a></item>
<item type="forme" pos="69"><a>jaune</a></item>
<item type="delim" pos="70"><a>vert</a></item>
<item type="forme" pos="71"><a>rouge</a></item>
<item type="delim" pos="72"><a>bleu</a></item>
<item type="delim" pos="73"><a>rouge</a></item>
<item type="delim" pos="74"><a>jaune</a></item>
<item type="delim" pos="75"><a>bleu</a></item>
<item type="forme" pos="76"><a>rouge</a></item>
<item type="delim" pos="77"><a>vert</a></item>
<item type="delim" pos="78"><a>jaune</a></item>
<item type="delim" pos="79"><a>noir</a></item>
<item type="forme" pos="80"><a>bleu</a></item>
<item type="delim" pos="81"><a>rouge</a></item>
<item type="delim" pos="82"><a>jaune</a></item>
<item type="forme" pos="83"><a>rouge</a></item>
<item type="delim" pos="84"><a>rouge</a></item>
<item type="forme" pos="85"><a>vert</a></item>
<item type="delim" pos="86"><a>bleu</a></item>
<item type="forme" pos="87"><a>jaune</a></item>
<item type="delim" pos="88"><a>rouge</a></item>
<item type="forme" pos="89"><a>bleu</a></item>
<item type="delim" pos="90"><a>vert</a></item>
citem type="forme" pos="91"><a>jaune</a></item>
citem type="delim" pos="92"><a>bleu</a></item>
<item type="forme" pos="93"><a>rouge</a></item>
<item type="delim" pos="94"><a>vert</a></item>
<item type="forme" pos="95"><a>bleu</a></item>
<item type="delim" pos="96"><a>jaune</a></item>
<item type="forme" pos="97"><a>rouge</a></item>
<item type="delim" pos="98"><a>noir</a></item>
<item type="delim" pos="99"><a>bleu</a></item>
<item type="forme" pos="100"><a>vert</a></item>
  items>
</importannotationsurtrame>
```

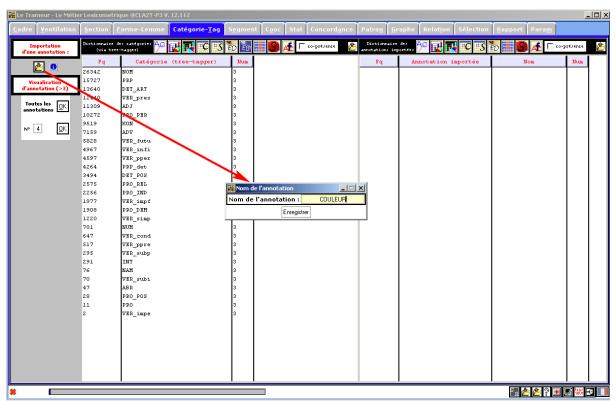
Pour charger ce fichier døannotations, il suffit døactiver le bouton (onglet CATEGORIE-TAG)

A la fin de lømportation il convient de nommer le système importé, ce nom sera affiché dans la zone « *Annotation sélectionnée* » dans les onglets **VENTILATION**, **CONCORDANCE** et **PATRON** quand le numéro de cette annotation sera sélectionné dans la zone de choix døun numéro døannotation.

Par convention:

- le numéro døannotation n°1 correspond à « Forme »
- le numéro døannotation n°3 correspond à « Catégorie »

Toute annotation importée aura un numéro supérieur ou égal à 4.



(enregistrement du nom de løannotation importée)

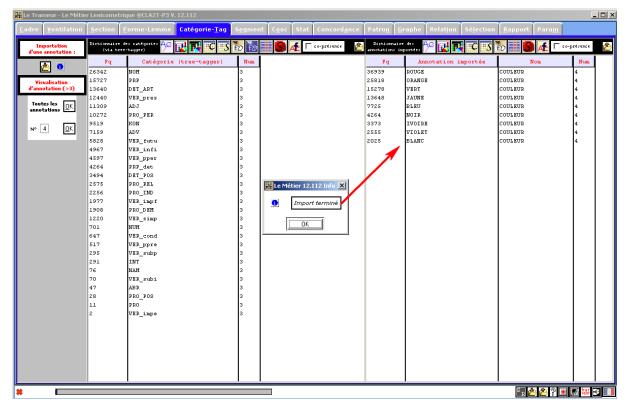


Figure 177: Le Trameur Importation dønnotation, nommage et sélection du niveau dønnotation

Une fois chargé, ce nouveau système døannotations est disponible pour les différentes opérations présentées précédemment : concordance, ventilation, recherche de patron, édition/modification des items.

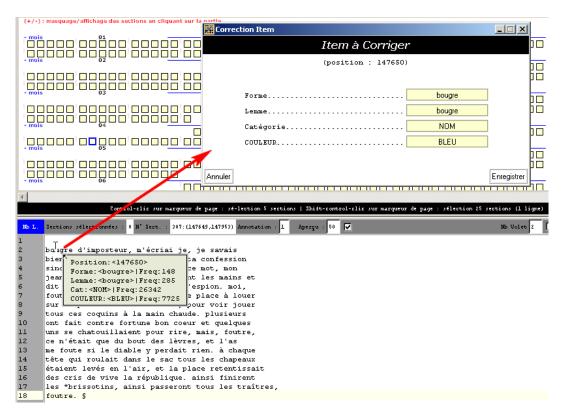


Figure 178 : Le Trameur correction døitem (annotation importée)

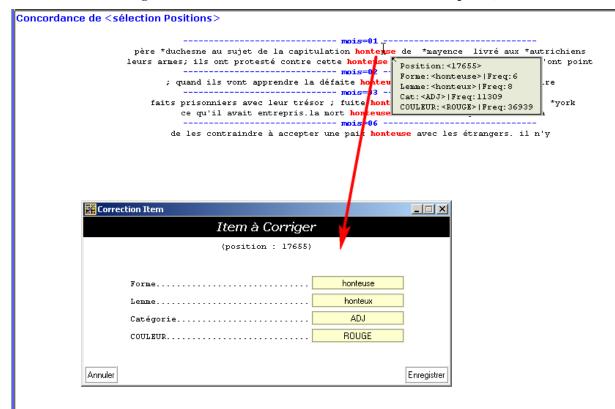


Figure 179 : Le Trameur concordance døitem (annotation importée)

CLESTHIA>SYLED>CLA²T

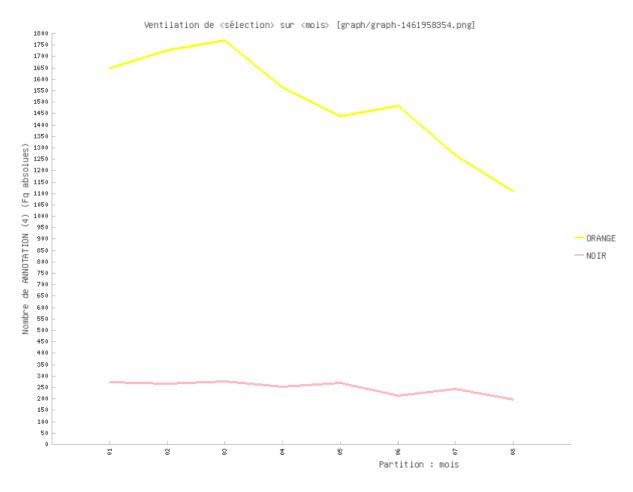


Figure 180 : Le Trameur ventilation døitem (annotation importée)

23 Ajout/modification d@annotation sur une Trame

Pour modifier une annotation, il est possible døagir sur chaque item de la *Trame* via les différents éditeurs : *Cadre*, *Section*, *Concordance* (*cf* parties précédentes). Ce type de modification ne modifie quøun seul item. Dans un processus døannotation plus large, il faut pouvoir réaliser des modifications à plus grande échelle.

Ajouter une annotation est une opération réalisable en important une nouvelle annotation décrite dans un fichier préconstruit à partir de la *Trame* (cf partie précédente).

Ces deux opérations peuvent être aussi être réalisées dynamiquement sur une base chargée dans le *Trameur* (en sélectionnant des items dans les dictionnaires disponibles) ou en modifiant la base préalablement exportée.

23.1 Ajout/modification døannotation par sélection døitems dans un dictionnaire

La modification et løinsertion døannotations ont déjà été présentées via le *Gestionnaire de Sélection*. Ces 2 opérations sont aussi disponibles sur les contenus des dictionnaires.

Dans le *Trameur*, chaque position de la *Trame* peut être associée à plusieurs systèmes døannotation. On peut importer un système døannotation sur løintégralité des positions de la *Trame* (cf partie précédente). On peut aussi insérer ou modifier des annotations en sélectionnant des items de la *Trame* dans le dictionnaire des formes par exemple (ou via les autres dictionnaires). Dans løexemple qui suit, on travaille sur le texte « Le Père Duchesne » ; le chargement du fichier de travail a été paramétré pour construire un étiquetage à løissue de la segmentation. On dispose donc en fin de chargement du fichier de 3 couches døannotation sur chaque item de la *Trame* : le numéro døannotation n°1 correspondant à « Forme », le numéro døannotation n°2 correspondant à « Lemme », le numéro døannotation n°3 correspondant à « Catégorie ». Toute annotation importée/insérée aura un numéro supérieur ou égal à 4. Dans chaque onglet donnant à voir un dictionnaire, on peut agir sur les items disponibles :

lédition de la forme visée pour éventuellement la corriger (si la forme est mise à jour, tous les items de la *Trame* ayant cette forme seront modifiés).

løajout døun niveau døun niveau døannotation sur la forme visée : tous les items de la *Trame* ayant cette forme seront modifiés pour cet ajout døannotation (avec la valeur døannotation fournie), les autres auront eux-aussi cette nouvelle annotation mais avec une valeur indéfinie (modifiable ultérieurement)

la correction des différents niveaux døannotation sur la forme visée : tous les items de la *Trame* ayant cette forme seront modifiés pour ces modifications sur tous les niveaux døannotation

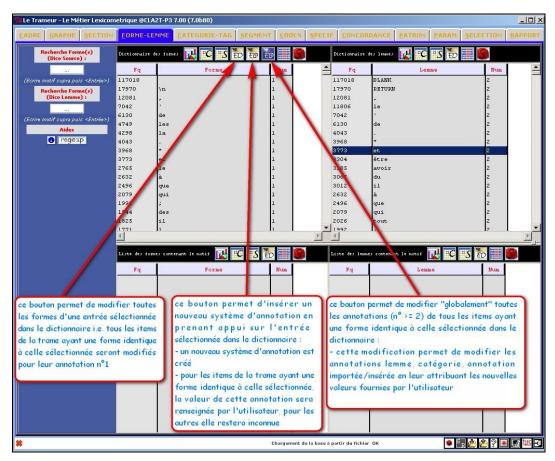
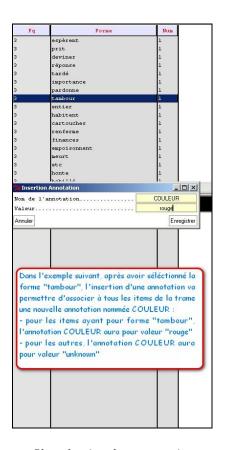


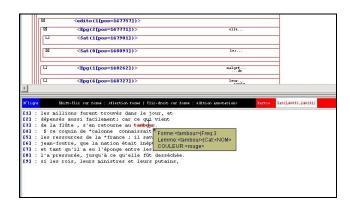
Figure 181 : Le Trameur édition de la Trame par les formes

23.1.1 Ajout døannotation

Les figures suivantes illustrent le mécanisme déajout de annotation sur tous les items de la *Trame* en sélectionnant au préalable une forme donnée dans le dictionnaire des formes (ou via les autres dictionnaires).

løajout døun niveau døun niveau døannotation sur la forme visée : tous les items de la *Trame* ayant cette forme seront modifiés pour cet ajout døannotation (avec la valeur døannotation fournie), les autres auront eux-aussi cette nouvelle annotation mais avec une valeur indéfinie (modifiable ultérieurement)





Phase 1: ajout døune annotation

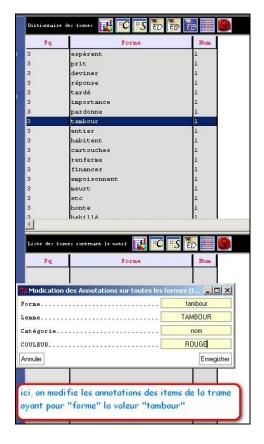
Phase 2 : visualisation sur la **Trame** de la présence de la nouvelle annotation

Figure 182 : Le Trameur ajout doannotation sur la Trame via une forme sélectionnée

23.1.2 Modification deannotation

Les figures suivantes illustrent le mécanisme de modification døune annotation sur tous les items de la *Trame* en sélectionnant au préalable une forme donnée dans le dictionnaire des formes (ou via les autres dictionnaires).

la correction des différents niveaux døannotation sur la forme visée : tous les items de la *Trame* ayant cette forme seront modifiés pour ces modifications sur tous les niveaux døannotation





Phase 1: modification doune annotation

Phase 2 : visualisation sur la **Trame du résultat**de la mise à jour de løannotation
tambour → TAMBOUR

Figure 183 : Le Trameur modification døannotation sur la Trame via une forme sélectionnée

Au cours de la modification des annotations sur un item sélectionné, si les valeurs des champs Lemme, Catégorie et annotations complémentaires restent vides, les valeurs initiales seront respectivement conservées pour chacune des annotations concernées.

IMPORTANT: toute insertion ou modification réalisée dynamiquement via les fonctionnalités précédentes ne modifie pas le fichier chargé ou la base importée initialement.

<u>Il convient dœxporter la base</u> en cours de traitement à løissue døune session de travail pour conserver les traces des modifications réalisées (insertion ou modification døannotation) : la base exportée est ensuite ré-importable ultérieurement et le travail døannotation peut être poursuivi et enrichi.

23.2 Ajout/fusion de couche døannotation

Une base textométrique étant chargée, il est possible døy insérer une nouvelle couche døannotation, de dupliquer une couche døannotation existante ou de fusionner deux ou plusieurs couches døannotations existantes.

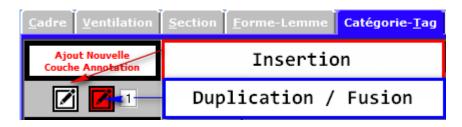


Figure 184: Le Trameur ajout / fusion døannotation

Dans la figure précédente, le bouton de gauche permet de générer automatiquement une nouvelle couche døannotation : à løissue de cette insertion, tous les items de la *Trame* auront une annotation supplémentaire « vide ».

Le bouton de droite permet soit de dupliquer la couche døannotation spécifiée dans la zone de saisie par son numéro (1 : forme ; 2 : lemme ; 3 : catégorie ; etc.) soit de fusionner les couches døannotions existantes spécifiées dans la même zone de saisie sous la forme suivante : i&j où i et j sont les numéros des annotations visées (pour fusionner les annotations 1, 2 et 3, on écrira : 1&2&3).

23.3 Ajout/modification døannotation dans une base exportée

La modification ou lønsertion dønnotation peut aussi être mise en ò uvre sur une base textométrique préalablement exportée (cf plus loin dans ce manuel, présentation du module døxport).

Une fois une base exportée, il est possible de lœditer (via un éditeur de texte type *notepad++* par exemple). Une base textométrique est un fichier XML décrivant la *Trame* et le *Cadre* du fichier initialement chargé. Løajout ou la modification døune annotation ne modifie que la représentation de la *Trame*.

La figure suivante donne à voir un extrait doune *Trame*: on y voit la représentation doun extrait de la segmentation (les items résultants) et sur chaque item, les valeurs de tous les niveaux doannotation définis sur la *Trame*. Dans cet exemple, elle comporte six niveaux doannotation. Chaque item a un type (forme ou délimiteur), une position sur la *Trame* et une description de ses différentes annotations via un système de balises:

- <f> sa forme graphique </f>
- <c> sa catégorie morpho-syntaxique </c>
- <1> son lemme </1>
- $\langle a \rangle$ annotation $n^{\circ}4 \langle a \rangle$
- <a> annotation $n^{\circ}5 $
- $\langle a \rangle$ løindication annotation n°6 $\langle a \rangle$

CLESTHIA>SYLED>CLA²T

```
<item type="forme" pos="6"><i>would</i><<>\ti>Would</i>><\ti>would</i></a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a
<item type="delim" pos="7"><f> </f><<p>DELIM</c><1>BLANK</1><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-<
</a></item>
<item type="forme" pos="8"><f>have</f><<1>have</l><a>-</a><a>-</a><a>-</a>
</a></item>
</a></item>
<item type="forme" pos="10"><f>only</f><c>JJ</c><l>only</l></a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-<
</a></item>
<item type="delim" pos="11"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-<
</a></item>
<item type="forme" pos="12"><f>a</f><c>DT</c><l>a</l><a>-</a><a>-</a></item>
<item type="delim" pos="13"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-<
</a></item>
  <item type="forme" pos="14"><f>small</f><c>JJ</c><l>small</l></a>-</a><a>-</a><a>-</a>
</a></item>
<item type="delim" pos="15"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><
</a></item>
<item type="forme" pos="16"><f>effect</f><c>NN</c><l>effect</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a
  </a></item>
```

Les 3 premières balises sont toujours utilisées pour décrire la forme graphique, la catégorie et le lemme de løtem visée (les balises c et 1 ne sont visibles que si une annotation morphosyntaxique a été intégrée (via *tree-tagger* ou tout autre programme)).

Les annotations supplémentaires sont toutes associées à une balise a. Løordre des balises a dans chaque item doit maintenir løordonnancement voulu des annotations : le premier couple de balises <a>... correspond à løannotation n°4, le second couple de balises <a>... correspond à løannotation n°5 etc. Le type des annotations (ici de 4 à 6) nøest pas explicitement nommé dans la base, cøest à løutilisateur de les connaître pour les identifier et y réaliser les modifications souhaitées. Dans løexemple donné, toutes les valeurs des annotations 4, 5 et 6 ont une valeur identique : un tiret.

Pour modifier une valeur døannotation donnée, il suffit de modifier la valeur située dans la balise correspondante.

Pour ajouter une nouvelle couche dønnotation, il convient dønjouter sur tous les items un couple de balise <a>... et døy insérer la valeur dønnotation idoine.

Ces opérations peuvent être réalisées par des méthodes classiques disponibles dans un « bon éditeur » : « rechercher/remplacer » par exemple.

```
Pour ajouter une annotation:
rechercher: </a></item>
remplacer par: </a></a></item>
```

Reste ensuite à modifier le contenu de la nouvelle annotation <a> sur chaque item.

23.4 Annotation de sections

A løssue døune ventilation sur la carte des sections il est possible de døannoter les sections concernées par cette recherche.

Dans læxemple ci-dessous, un motif a été projeté sur la carte :

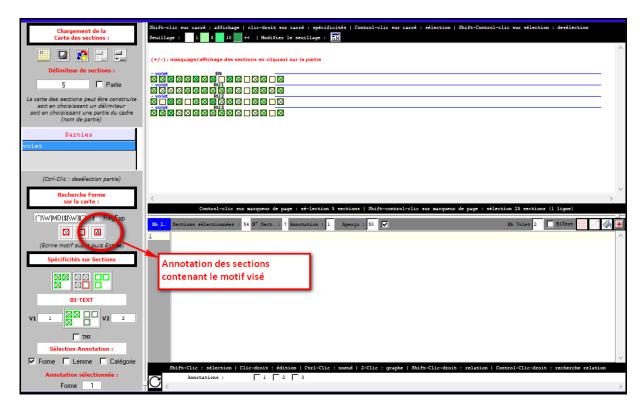


Figure 185: Le Trameur annotation de sections

Løactivation du bouton dø « Annotation des sections » va réaliser les opérations suivantes :

- 1. Ajout døune couche døannotation n°X sur les items de la *Trame* avec pour valeur -1
- 2. Le premier item døune section contenant le motif visé aura comme valeur 1 pour løannotation $n^{\circ}X$
- 3. Le premier item døune section ne contenant pas le motif visé aura comme valeur 0 pour løannotation $n^{\circ}X$

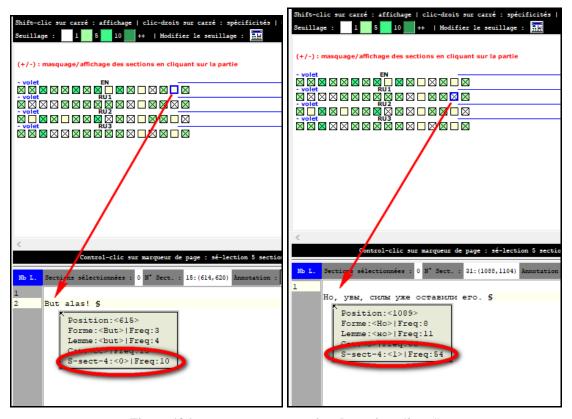


Figure 186 : *Le Trameur* annotation de sections (0 vs 1)

<u>Dans le cas de bases alignées</u>, si une des sections ne contient pas le motif visé, toutes les sections alignées sont marquées avec la valeur 0 sur le premier item de ces sections :

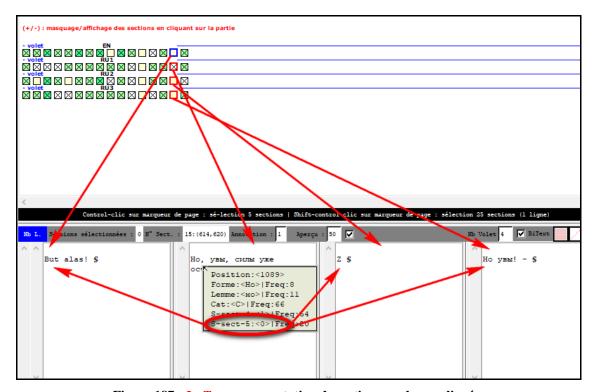


Figure 187 : Le Trameur annotation de sections sur bases alignées

Ce marquage des sections permet par exemple de mettre au jour les sections contenant ou ne contenant pas le motif visé.

Dans lœxemple ci-dessous, on søintéresse aux sections alignées ne contenant pas le motif visé et on les projette dans la carte des sections :

Fq	Annotation importée	Nom
10	0	S-sect-4
855	-1	S-sect-4
54	1	S-sect-4
20	0	S-sect-5
855	-1	S-sect-5
44	1	S-sect-5
		l

Figure 188 : Le Trameur annotation de sections sur bases alignées : recherche de løabsence

On peut ainsi isoler, à løissue døune recherche de motif dans la carte des sections (ici alignées), des sections particulières ; dans cet exemple, les sections alignées mises au jour se caractérisent par løabsence du motif initialement cherché dans løune des sections alignées :



Figure 189 : Le Trameur annotation de sections sur bases alignées : mise au jour de løabsence

24 Des nuages de døannotations sur la *Trame*

Ce module permet de générer un fichier (au format XML) donnant à voir une liste døtems de la *Trame* (*via* un niveau døannotation choisi), chacun de ces items étant associé à løensemble des sections (zones de texte paramétrable) associées à cette annotation.

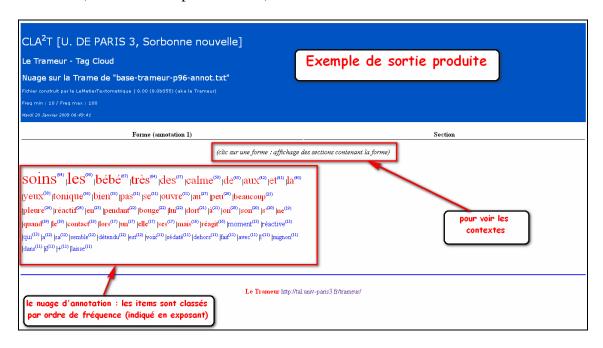


Figure 190 : Le Trameur un nuage déannotations sur la Trame

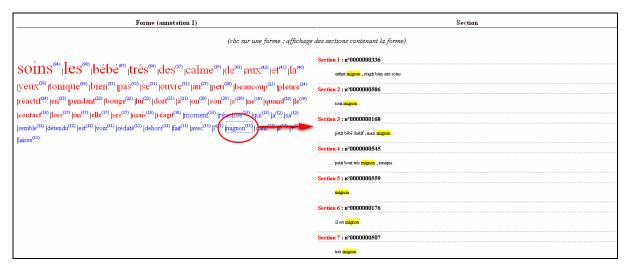


Figure 191 : Le Trameur un nuage déannotations sur la Trame avec contextes

Le paramétrage du calcul de ce nuage se fait dans løonglet DIC.

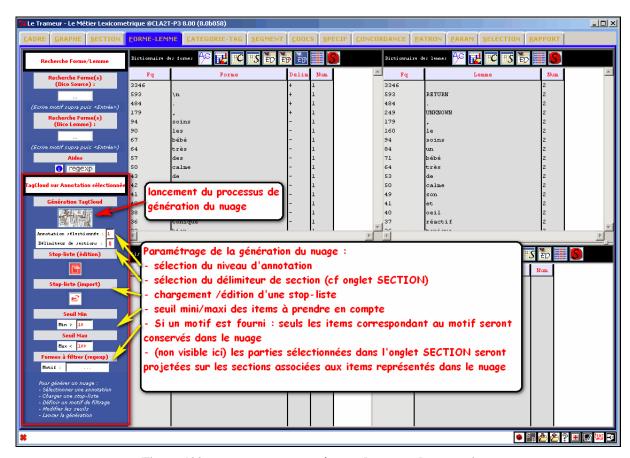


Figure 192 : Le Trameur paramétrage du nuage d@annotations

<u>Des exemples de nuages døannotations</u> sur la *Trame* sont visibles sur la page web du *Trameur*.

25 Relations de dépendance entre les items de *Trame* (via leurs annotations)

Cette partie introduit le traitement des relations de dépendances disponibles entre les items de la *Trame*. Pour illustrer ces fonctionnalités, nous utilisons les bases textométriques suivantes (disponibles sur la page du logiciel) :

1. <u>Base "Corpus de contes français"</u> (Fairy Tales Corpus, FTC) construite à partir des ressources développées par <u>Ismaïl El Maarouf</u>. Chaque item de la *Trame* est associé à 7 niveaux d'annotation.

SOURCES: I. El Maarouf et J. Villaneau (2012). *A French Fairy Tale Corpus syntactically and semantically annotated.* In Proceedings of LREC 2012. Istanbul (Tk).

Pour les formes : copyright Momes.net, http://momes.net.

Lemme et POS via treetagger : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)

Pour les annotations : http://elmaarouf.legtux.org/Accueil.html

Descriptif et sources des annotations :

 $balise <\!\!f\!\!> (forme): copyright \underline{Momes.net}, \underline{http://elmaarouf.legtux.org/data/FTC\ FORM.txt}$

balise <l> (lemme) : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)

balise <c> (POS) : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)

Annotations complémentaires (http://elmaarouf.legtux.org/Accueil.html):

Identifiants de mots : http://elmaarouf.legtux.org/data/FTC_ID.txt (première balise <a>)

- Annotation référentielle : http://elmaarouf.legtux.org/data/FTC_REF.txt (seconde balise <a>)
- Annotation en rôles sémantiques pour 26 verbes : http://elmaarouf.legtux.org/data/FTC_SEMROLE.txt (troisième balise <a>)
- Annotation syntaxique en dépendance pour 137 verbes :

http://elmaarouf.legtux.org/data/FTC_SYNROLE.txt (dernière balise <a>)

L'ensemble des annotations du corpus initial (+ lemme et POS) regroupées ici au format d'une base textométrique pour le *Trameur*.

2. <u>Base "Rhapsodie2Trameur"</u> construite à partir des ressources développées dans le cadre du projet <u>Rhapsodie</u>. Chaque item de la *Trame* est associé à 13 niveaux d'annotation.

SOURCES: projet Rhapsodie

Fichiers de codage microsyntaxique version béta 10/13 (zip)

Descriptif et sources des annotations :

Présentation du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique (pdf)

3. <u>Base "Rhapsodie2Trameur" (v4)</u> construite à partir des ressources développées dans le cadre du projet <u>Rhapsodie</u>. Chaque item de la *Trame* est associé à 38 niveaux d'annotation (prosodie, micro et macro-syntaxe).

25.1 Extrait de la Trame : « un jour, un petit ange de sapin de í »

Dans les exemples présentés *infra*, nous allons travailler sur un extrait de la base des contes dont lømage suivante donne à voir le début de la *Trame* telle quœlle est représentée dans la base textométrique disponible :

```
<items>
                                                                                                                                                              pos = "1" > <f> \# </f> < c> DELIM </c> <1> \# </l> < contesdemontagne </a> <a> -</a> <a> -</a> </in> <interpolation | |
 <item type="delim"
                                                                                                                                                                pos="2"><f>Un</f><c>NAM</c><1>Un</1><a>1</a><a>-</a><a>-</a><a>-</a></item>
 <item type="forme"
 <item type="delim" pos="3"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/>
 <item type="delim" pos="5"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="6"><f>,</f><c>NOM</c><1>,</f><c>NOM</c><1>,</f><a>-</a><a>-</a><a>-</a><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><a>-</a><br/><
 <item type="delim" pos="8"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a></a></a></a></a></a></a></a></a></item>
<item type="forme" pos="9"><f>petit</f><c>ADJ</c><1>petit</1><a>5</a><a>-</a></a><a>-</a></a></a></b></br></pr>
 <item type="delim" pos="10"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a><<a>-</a></a></a><-/a></a><-/a></a>
<item type="forme" pos="13"><f>de</f><c>PRP</c><1>de</1><a>7</a><a>-</a><a>-</a><a>-</a><a>XXX(11)</a></item
 <item type="delim" pos="14"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a>-</a>-<a>-</a>-<a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</
<item type="forme"
                                                                                                                                                                pos="15"><f>sapin</f><c>NOM</c><l>sapin</l><a>8</a><a>-</a><a>-</a><a>XXX(7),YYY(4)</a></item>
 <item type="delim" pos="16"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a
 <item tvpe="forme" pos="17"><f>de</f><c>PRP</c><1>de</l><a>9</a><a>-</a><a>-</a><a>-</a><a>-</a><br/>forme" pos="17"><f>de</f><a>-</a><a>-</a><a>-</a><br/>forme" pos="17"><f>de</f><a>-</a><a>-</a><a>-</a><br/>forme" pos="10"</a><br/>forme" pos="10"<br/>forme" p
   <item tvpe="delim" pos="18"><f> </f><c>DELIM</c><1>BLANK</1><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-<
```

Cet extrait met en avant des relations de dépendances sur plusieurs items de la *Trame* :

• Løttem en position 4 (dont la forme est « jour ») porte dans son annotation n°7 la valeur :

$$XXX(1)$$
, $YYY(4)$

Cette information traduit la présence de 2 relations de dépendance pour cet item :

- une nommée XXX, løautre YYY; la première associe løitem courant avec un item dont løidentifiant est 1,
- la seconde associe løtem courant avec un item dont løtdentifiant est 4.

Dans cette base, lødentification des éléments de la *Trame* ne coïncide pas avec leur position sur la *Trame*, une annotation particulière (la première balise <a> correspondant à løannotation n°4) porte cette information.

On a donc:

```
« jour » (position 4) → « Un » (position 2, identifiant 1) par la relation XXX « jour » (position 4) → « un » (position 7, identifiant 4) par la relation YYY
```

25.2 Recherche de relations

Lønglet **GRAPHE** permet de lancer une requête de recherche de relations entre items de la *Trame*.

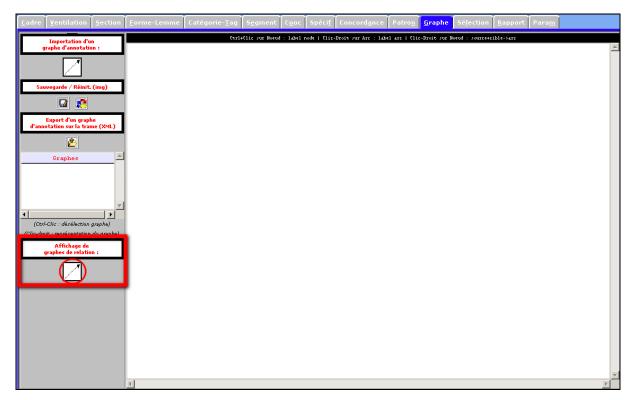


Figure 193 : Le Trameur recherche de relations de dépendance via løonglet GRAPHE

25.2.1 Sélection des relations à afficher



Figure 194 : Le Trameur paramétrage de la recherche de dépendance sur la Trame

Pour afficher les relations disponibles dans la base, un menu permet de définir ces relations.

Dans læxemple ci-dessus, on définit le modèle « XXX | YYY » permettant de représenter soit la relation XXX soit la relation YYY en spécifiant le numéro de lænnotation portant les relations de dépendance à scruter.

On précise aussi ici où repérer les identifications des items concernés (annotation n°4).

Dans la base traitée ici, chaque item de la *Trame* est associé à un niveau døannotation (numéro 4) portant une valeur permettant døidentifier ces items. Si ce niveau døannotation nøest pas disponible, il est possible døutiliser une « niveau døannotation virtuel » pour identifier les items : celui fourni par la position des items sur la *Trame*. Pour activer cette identification des items par leur position, il suffit de cocher la case disponible « (Position) ».

25.2.2 Graphe des relations

Løactivation de la recherche produit le graphe des relations disponibles sur le modèle visé :

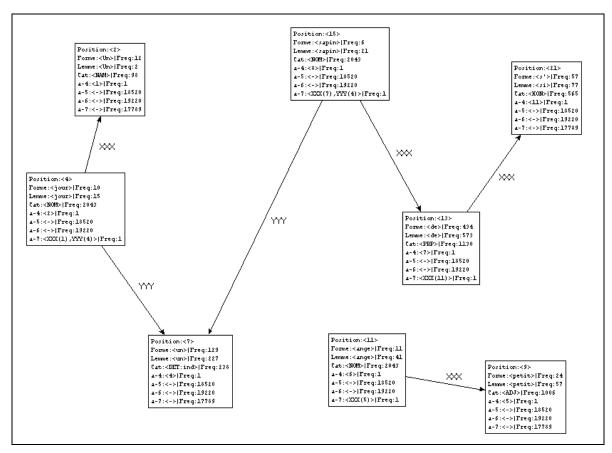


Figure 195 : Le Trameur affichage du graphe des dépendances

25.3 Sélection des relations avec filtrage sur le nò ud source de la relation

On peut affiner le modèle de recherche de relations en filtrant sur les annotations associées aux nò uds des relations visées. Dans læxemple suivant, on søintéresse aux objets (relation OBJ) des items contenant la chaîne de caractère « connu » au niveau de løannotation $n^{\circ}1$ (forme) :

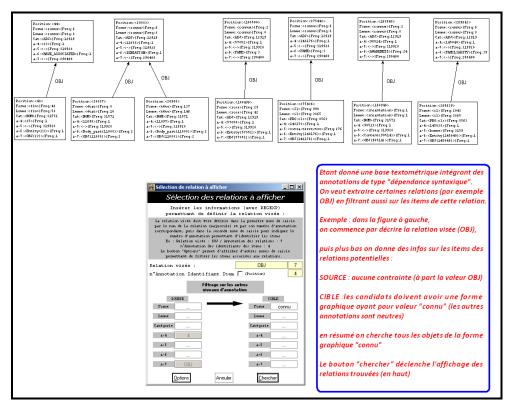


Figure 196 : Le Trameur paramétrage de la recherche de dépendance avec filtrage

Autre illustration sur la base « Corpus de conte » sur lequel on a cherché la relation « SUJ | OBJ », le graphique donne à voir un extrait du graphe produit :

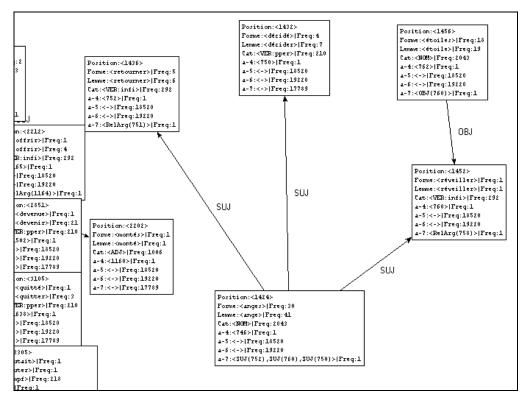


Figure 197 : Le Trameur affichage du graphe des dépendances (1)

De même, sur la base *Rhapsodie2Trameur*, la figure suivante illustre la recherche des « objets » du lemme « affirmer » :

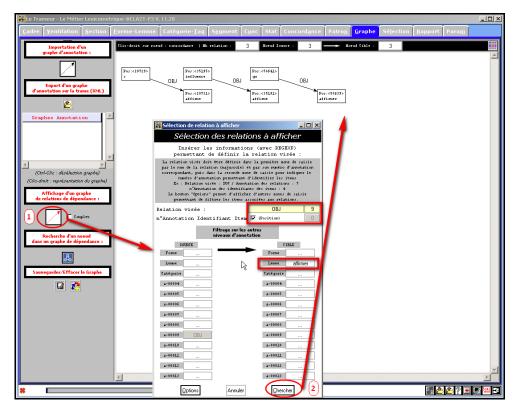


Figure 198 : Le Trameur affichage du graphe des dépendances (2)

Dans la figure précédente, on cherche des relations de dépendance de type \mbox{OBJ} en contraignant un des composants de la relation : il doit porter le lemme "affirmer" (i.e. on cherche les objets d'affirmer). Dans la figure suivante, la relation cherchée est double (SUB ou \mbox{OBJ}), la cible de la relation impose une valeur pour le lemme (« penser ») et pour løannotation $n^{\circ}9$ (ROOT)

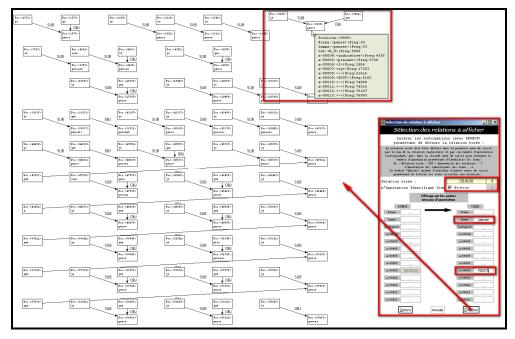
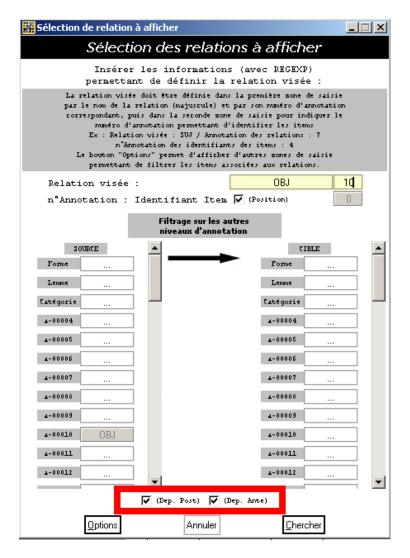


Figure 199 : Le Trameur affichage du graphe des dépendances (3)

Dans les 2 graphes précédents, løaffichage des nò uds est celui produit par défaut : affichage de la position de løitem sur la *Trame* et de sa forme graphique. On peut visualiser løensemble des annotations døun nò ud donné via le mode Aperçu disponible dans tous les éditeurs (en passant la souris au-dessus de løbjet visé) : ci-dessus, løitem en position 9899 (lemme : penser) est mis au jour. Dans les précédentes figures, løaffichage døun nò ud est directement rendu « complet » via løoption du même nom disponible et activable avant de charger un graphe.

25.4 Sélection des relations avec filtrage sur la position du dépendant

Une option supplémentaire permet de filtrer la recherche de relation en contraignant la position du dépendant : postposition ou antéposition.



Par défaut, la recherche met au jour les dépendants antéposés et postposés, si on décoche une des options, la recherche ne søappliquera que sur les dépendants correspondant à løoption cochée.

25.5 Actions sur les nò uds du graphe : retour en contexte

Dans læxemple qui suit, on søintéresse aux relations SUJ ou OBJ concernant les items ayant pour forme « anges » dans la base des contes :

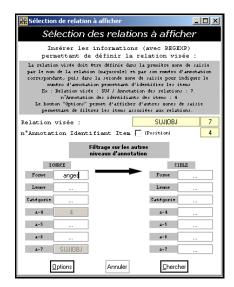


Figure 200 : Le Trameur paramétrage de la recherche de dépendance avec filtrage

Le graphe produit a løallure suivante :

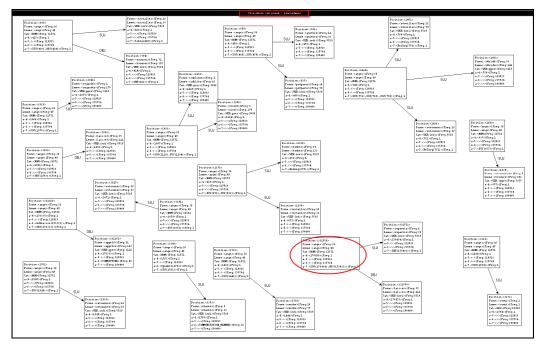


Figure 201 : Le Trameur affichage du graphe des dépendances

La combinaison de touche « clic-droit » sur un nò ud du graphe permet døafficher une concordance de løtem visé :

```
une nouvelle vie à un animal .Mais Dieu ne voulait pas laisser repartir ses anges comme ça !Alors ,soudain ,il eut une idée .-Qu' aimeriez -vous ,
```

Figure 202 : Le Trameur affichage de la concordance sur un des nò uds du graphe des dépendances

Løensemble des nò uds peut aussi être projeté globalement sur le concordancier. Dans la figure suivante, on reprend le graphe produit sur la base *Rhapsodie2Trameur* montrant le résultat de la recherche des « objets » du lemme « affirmer » :

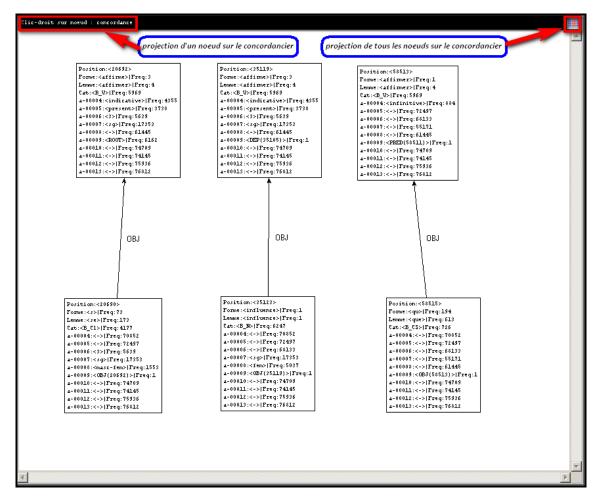


Figure 203 : *Le Trameur* affichage de la concordance sur tous les nò uds du graphe des dépendances La concordance produite après projection de tous les noeuds a løallure suivante :

Figure 204 : Le Trameur Concordance sur tous les nò uds du graphe des dépendances (1)

Les sources de la relation sont coloriées en rouge (pôle de la concordance), les cibles en vert (i.e ici le verbe est en vert, son objet en rouge). On peut aussi varier løaffichage des annotations dans la concordance en matérialisant par exemple la catégorie de chaque item :

Figure 205 : Le Trameur Concordance sur tous les nò uds du graphe des dépendances (2)

Toutes les zones dédition du *Trameur* permettent de mettre au jour les annotations de la *Trame* (cf « Marquage des annotations de la *Trame* »). Si on considère la figure suivante, elle présente un extrait de la concordance construite à partir du graphe ayant permis dœxtraire la relation OBJ:

```
S de la crise aux Antilles il ne devrait pas en etre question au sommet social de l'Élysée cet savamment minutée § alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va se passer passer avec vous Jean-François Achilli § alors passons maintenant au détail des mesures discutées et aux attentes des syndicats des mesures de justice je cite pour les salariés touches par la crise économique § mais Sara Ghibaudo il ne vous d'aujourd'hui les Français se montrent très sévères à l'égard de la politique économique du gouvernement qu' des voitures § bonjour sébastien Paour § vous vous trouvez au Gosier § où en est la situation § merci § tout le monde est maintenant en tout cas suspendu à ce que Nicolas Sarkozy annoncera demain lors de sa a promulgué son plan de relance adopté ce week-end par le congres § sept cent quatre-vingt-sept milliards se dollars destinés à sauver ou créer plus de trois millions et demi à ralentir les saisies immobilières devrait être annoncé aujourd'hui par le président américain § cela concerne plusieurs milliards toujours en garde à vue § les enquêteurs cherchent maintenant à asvoir si elle a été complice dans cette évasion a été complice dans cette évasion selle se trouvait dans le parloir de la prison au moment où les la foire d'empoigne § la tension est encore montée d'un cran hier § les avocats du berger corse procès en attendant que la cour d'assisse se prononce sur leur denande de supplément d'information § lundi le d- ce sont des procédés terroristes § il le dit à Laurent Doulsant § direction maintenant Barcelone pour un congrès Caurent Doulsant § direction maintenant Barcelone pour un congrès consacré au téléphone mobile en crise lui aussi § pour la nul zéro partout pour Lille au Mans § valenciennes sort
```

Figure 206: Le Trameur Coloration d@annotation dans le concordancier (1)

Le marquage de certains items est possible en sélectionnant les annotations à mettre au jour :

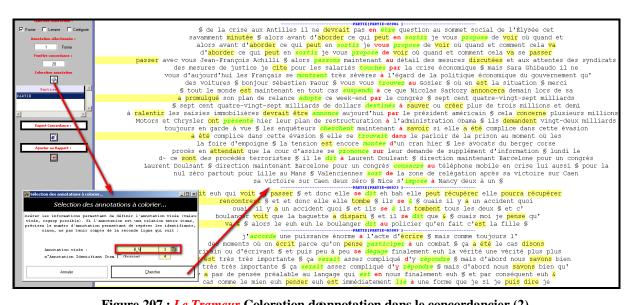


Figure 207 : Le Trameur Coloration d@annotation dans le concordancier (2)

25.6 Rechercher dans un graphe de dépendance

On présente tout døabord le résultat produit par la requête suivante « recherche de la relation OBL » :



Le résultat produit a løallure suivante :

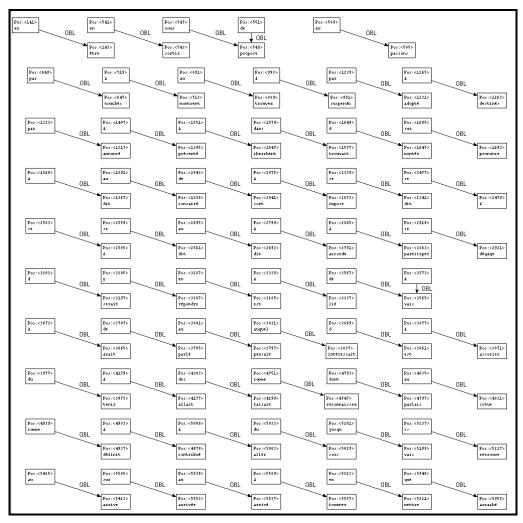


Figure 208: Le Trameur Graphe de la relation OBL

On peut mettre au jour dans ce graphe certains no uds en filtrant leurs annotations. Le processus de filtrage des no uds du graphe permet de sélectionner certains no uds sur la base

des différentes annotations disponibles. Dans léexemple suivant, on veut mettre au jour les nò uds de catégorie B_V :

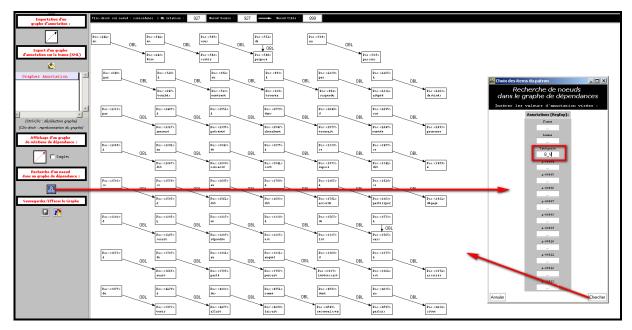


Figure 209 : Le Trameur Recherche dans un graphe de relation (1)

Le résultat produit a løallure suivante :

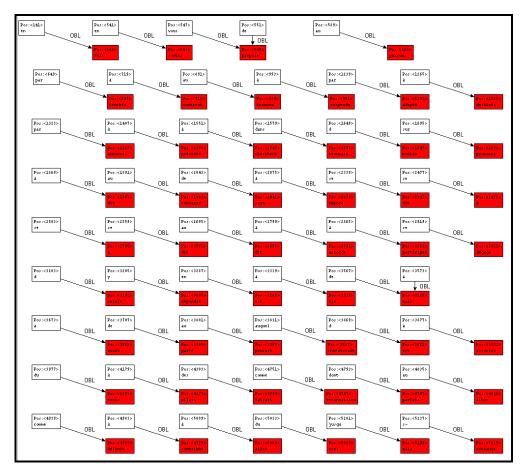


Figure 210 : Le Trameur Recherche dans un graphe de relation (2)

Dans la figure qui suit, 3 filtrages successifs sont réalisés pour marquer successivement les nò uds ayant pour catégorie B_V, puis B_N et enfin B_Pre; chaque requête est précédée par la modification de la couleur à utiliser pour le marquage des nò uds visés (verbe en bleu, nom en vert et préposition en orange):



Figure 211 : Le Trameur Recherche dans un graphe de relation (3)

25.7 Affichage de relation en contexte

Le *Trameur* permet des accès à la *Trame* via différents éditeurs (éditeur du *CADRE*, de la Carte des sections, du concordancier). Ces éditeurs permettent de donner à voir les relations définies sur les items présents dans løéditeur.

Dans læxemple qui suit, la figure présente læditeur de la **Carte des sections** contenant løtem (en position 15) dont læannotation n°1 (forme) est « sapin » (le mode aperçu signale des relations définies sur cet item via le niveau dæannotation n°7)

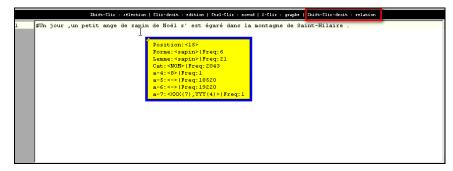


Figure 212 : Le Trameur accès aux dépendances via la Trame

La combinaison de touches « shift-clic droit » active une fenêtre permettant de sélectionner les relations à rechercher sur cet item :

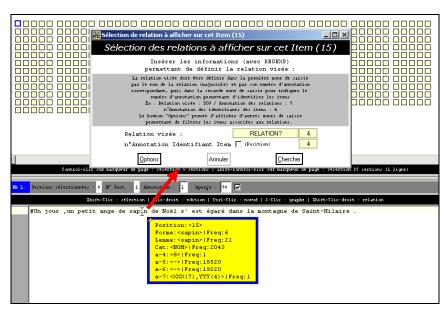


Figure 213 : Le Trameur paramétrage de løaffichage des dépendances sur un item de la Trame (1)

Comme précédemment, le paramétrage de cette recherche permet de spécifier le nom de la relation visée et le numéro døannotation dans lequel elle doit être cherchée ; dans la figure qui suit, on cherche les relations XXX ou YYY intégrées dans løannotation n°7 :

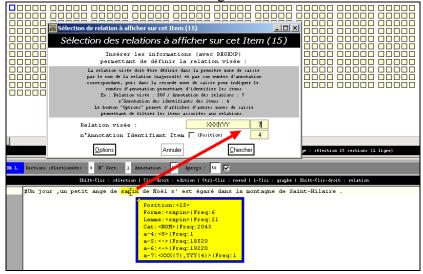


Figure 214 : Le Trameur paramétrage de løaffichage des dépendances sur un item de la Trame (2)

Le résultat de la recherche produit deux types døaffichage :

• La matérialisation dans léditeur des items concernés par relations visées



Figure 215 : Le Trameur affichage de la dépendance sur la Trame (1)

• Løaffichage du graphe des relations

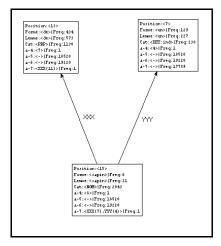


Figure 216 : Le Trameur affichage de la dépendance via un graphe (2)

25.8 Recherche de relation en contexte

Les éditeurs de la *Trame* permettent aussi de rechercher déventuelles relations sur les items édités.

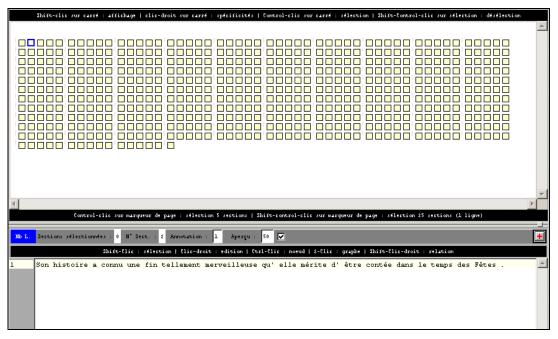


Figure 217 : Le Trameur recherche de dépendance en contexte sur un item de la Trame

Dans læxemple qui suit, læditeur de la **Carte des sections** affiche une section contenant lætem n°44 dans læannotation n°1 (forme) est « connu ». La combinaison de touches « controlclic droit » active une fenêtre permettant de paramétrer la recherche de relations concernant cet item. Le paramétrage de cette recherche de relations consiste à spécifier le numéro de læannotation dans laquelle la recherche de relations doit sæffectuer (ici le numéro 7).

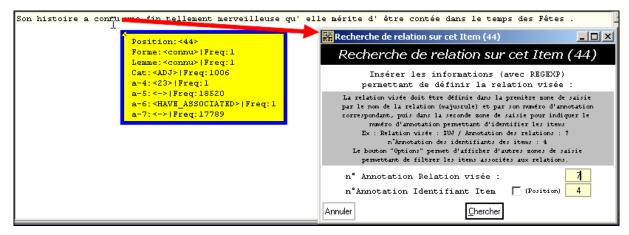


Figure 218 : Le Trameur paramétrage de la recherche de dépendance sur un item de la Trame (1)

Le résultat de la recherche produit là aussi 2 types de résultats.

1. Affichage du graphe des relations concernant le nò ud initialement visé

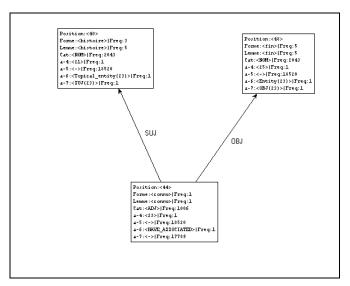


Figure 219 : Le Trameur affichage des dépendances via un graphe (1)

2. La matérialisation dans léditeur des items concernés



Figure 220 : Le Trameur affichage des dépendances en contexte sur la Trame

De même la recherche de dépendance peut aussi être réalisée sur le concordancier. Si on considère la concordance suivante (en haut de la figure) :



Figure 221 : Le Trameur paramétrage de la recherche de dépendance sur un item de la Trame (2)

La combinaison de touches « control-clic droit » sur un item de la concordance (ici løtem visé est la première occurrence de la forme graphique « affirme ») permet de rechercher toutes les relations de dépendance pointant sur cet item. On commence par indiquer où chercher les relations de dépendance (numéro døannotation portant ce type døinformation : ici le $n^\circ 9$) et comment sont indexer les items sur la trame (ici par leur position). La recherche conduit à la production døin graphe mettant au jour toutes les relations sur løtem visé :

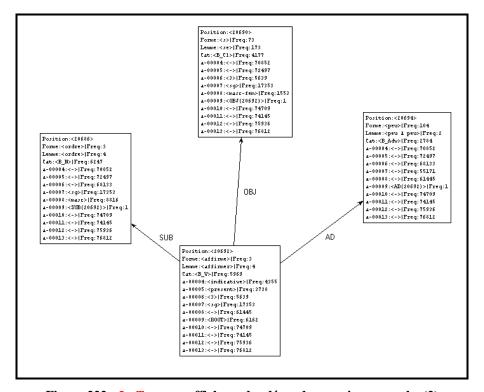


Figure 222 : Le Trameur affichage des dépendances via un graphe (2)

Le graphe donne à voir les 3 relations pointant sur loitem.

25.9 Graphe de relations en contexte

Une fonctionnalité permet de donner à voir (sous la forme donne graphe) lognemble des relations disponibles sur une zone de la *Trame* éditée dans la Carte des sections. La base étant chargée, la figure suivante donne à voir la Carte des sections et logdition donne section :

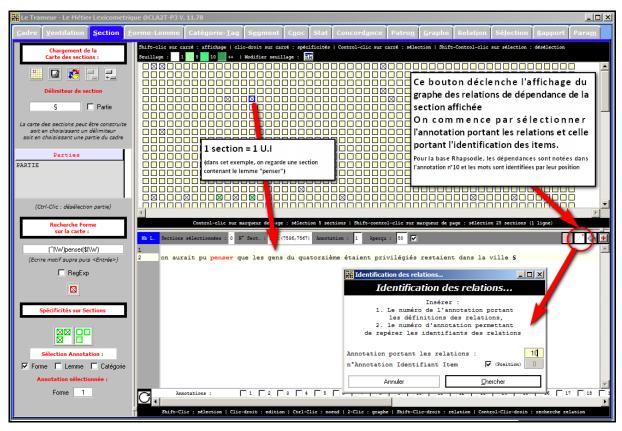


Figure 223 : Le Trameur affichage doun graphe de dépendances en contexte (1)

Dès quøine section est éditée, une icône est activée et permet de déclencher løaffichage du graphe des relations de dépendance entre les items de la zone textuelle éditée.

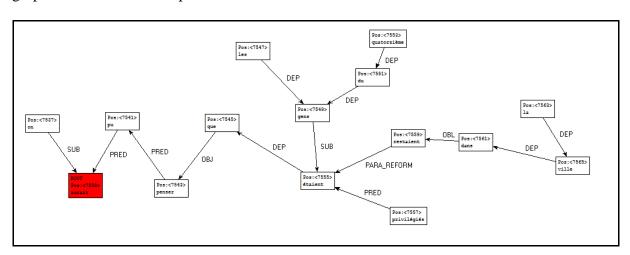


Figure 224 : Le Trameur affichage donn graphe de dépendances en contexte (2)

25.10 Visualisation de toutes les relations dans un contexte donné

Une fonctionnalité complémentaire permet de donner à voir (sous la forme donne matrice) lonsemble des relations disponibles sur une zone de la *Trame* éditée dans la **Carte des sections**. La base étant chargée, la figure suivante donne à voir la **Carte des sections** (le caractère point est utilisé comme délimiteur de sections) et loédition de la première section :

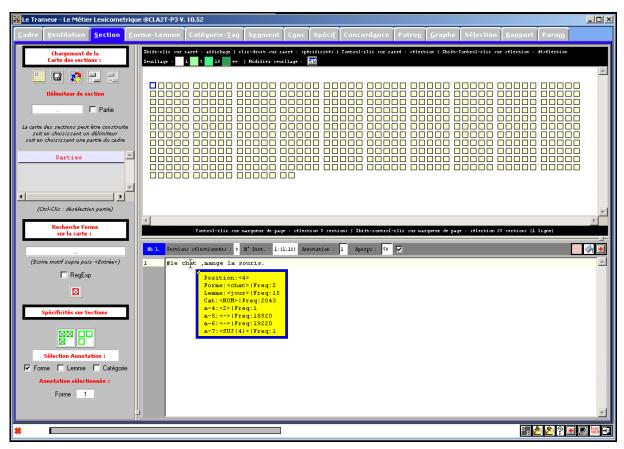


Figure 225 : Le Trameur affichage global des dépendances en contexte (1)

Dès quøine section est éditée, une icône est activée et permet de déclencher løaffichage des relations de dépendance entre les items de la zone textuelle éditée.



Figure 226 : Le Trameur affichage global des dépendances en contexte (2)

Un menu permet de sélectionner le numéro døannotation portant les relations visées (ici løannotation n°7) et la manière dans les items sont identifiables dans ces relations (ici via løannotation n°4)



Figure 227 : Le Trameur affichage global des dépendances en contexte (3)

Løactivation de la recherche produit une matrice donnant à voir les relations disponibles sur la section visée :

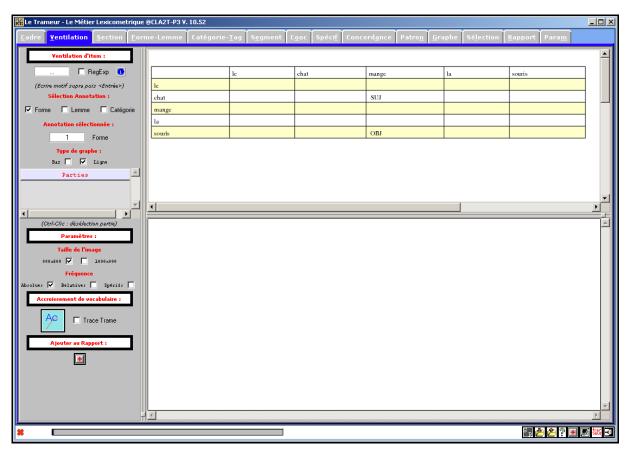


Figure 228 : Le Trameur affichage global des dépendances en contexte (4)

25.11 Recherche de collocation : spécificités sur relation

Une version particulière du module de calcul des cooccurrences permet de prendre en compte les relations entre les items de la *Trame* :

Etant donné une forme pôle, la recherche de ses cooccurrents peut être contrainte par la prise en compte døune relation entre ce pôle est ses candidats cooccurrents. Pour un item X donné (le pôle), on søintéresse aux items Y cooccurrents de X et en relation REL avec X (les collocatifs de X):

La figure suivante présente les cooccurrents du lemme penser en contraignant ses cooccurrents à être en position døbjet :

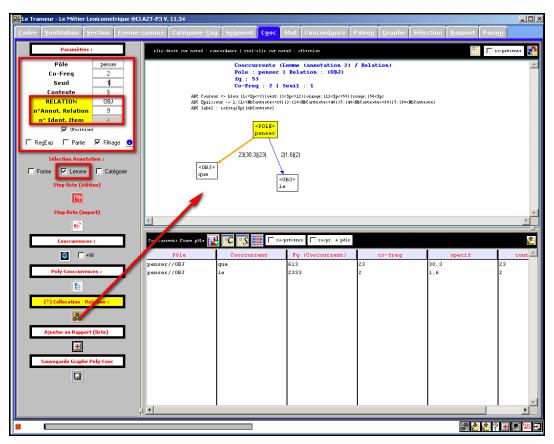


Figure 229 : Le Trameur Recherche de collocation (1)

La relation entre le pôle et ses cooccurrents peut être « non déterminée » par avance, dans la figure suivante, la relation est exprimée sous la forme : .* (regexp) ; elle vise donc toutes les relations du type : penser-OBJ-y, penser-SUB-y etc.

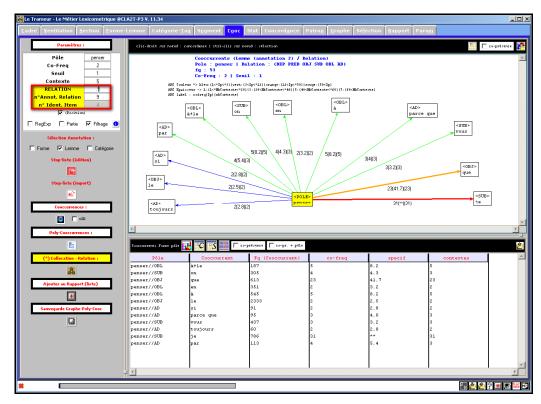


Figure 230 : Le Trameur Recherche de collocation (2)

Une fois le graphe construit, un clic-droit sur un des noeuds montrent les contextes dans lesquels la relation se réalise (i.e les contextes utilisés par le calcul).

Dans cet exemple, 3 contextes dans lesquels *vous* est sujet de *penser* (*cf* arc penser -> vous) ont été mis au jour, un clic-droit sur le nò ud *vous* les montrent dans le concordancier (les 2 items de la relation y sont colorés automatiquement : *vous* en rouge (pôle de la concordance) et *penser* en vert).

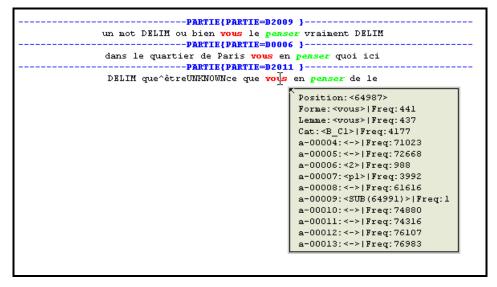


Figure 231 : *Le Trameur* Recherche de collocation (3)

25.12 Recherche source-cible døune relation

Løonglet **Relation** contient døautres fonctionnalités permettant de traiter les relations de dépendance. Une première permet de lister les termes *source* et/ou *cible* døune relation visée. On peut soit chercher la liste des termes *source* de cette relation, soit la liste des termes *cible*.

Dans la figure suivante, on søntéresse à la relation OBJ (les annotations de dépendance étant portées ici par løannotation n°10 de la base traitée et les items mis en relation sont indexés par leur position sur la *Trame*), la requête « chercher » produit la liste des items de la *Trame* portant cette annotation (cf colonne a-00010):

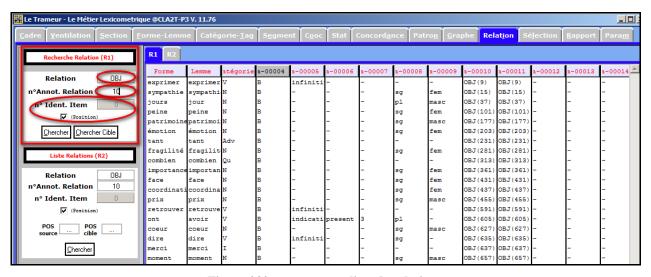


Figure 232 : *Le Trameur* liste de relations

Dans la figure suivante, on søintéresse à la même relation, la requête « chercher cible » produit cette fois-ci les cibles des relations portées par les items précédents :

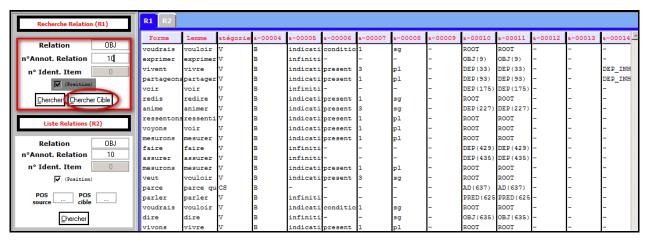


Figure 233 : *Le Trameur* liste de relations (cible)

On peut enfin construire la liste des termes de POS contraints par un relation de dépendance. Dans la figure qui suit, la relation visée est la relation OBJ; on paramètre ensuite les catégories *source* et *cible* visées (annotation n°3) visées (ici .* permet de les récupérer toutes).

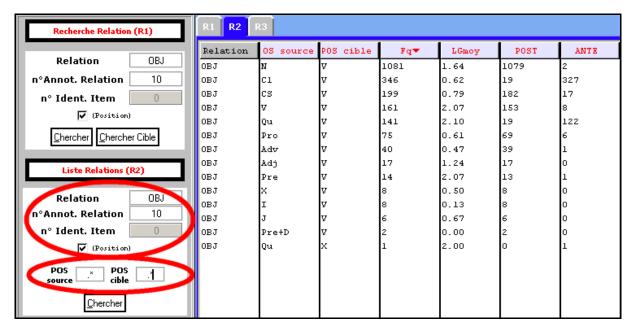


Figure 234 : Le Trameur liste de POS en relation

Le résultat produit donne à voir les fréquences des patrons de dépendance (via leur POS) : par exemple ici, la relation de dépendance OBJ met 40 fois en relation un item de type Adv et un item de type V avec : Adv-OBJ->V ; dans la base traitée ici, la relation OBJ met majoritairement en relation les items de type V et les items de type N : N-OBJ->V (fq : 1081).

Les trois dernières colonnes donnent des informations complémentaires sur chaque type de relation :

- **LGmoy** (longueur moyenne) : le nombre de mots entre løtem source (dépendant) et løtem cible (gouverneur).
- **POST** : le nombre de relation avec un dépendant postposé.
- ANTE : le nombre de relation avec un dépendant antéposé.

Enfin, dans la figure suivante, la requête exprimée permet døafficher, pour un item donné (ici le lemme « penser »), les dépendants de cet item dans la relation de dépendance visée (ici SUB ou OBJ) :

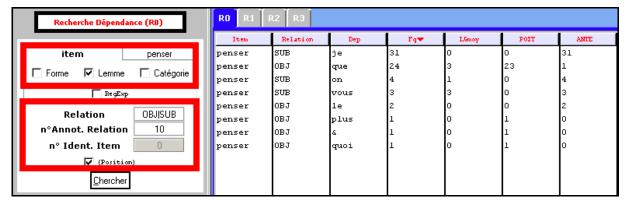


Figure 235 : Le Trameur les dépendants døun item (forme, lemme ou catégorie)

Pour produire la liste des dépendants de plusieurs items, il est possible de les décrire *via* une expression régulière (ici tous les mots commençant par « pens ») :

CLESTHIA>SYLED>CLA²T

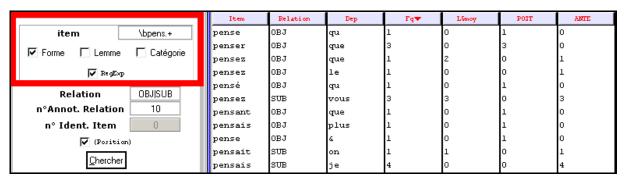


Figure 236 : Le Trameur les dépendants døune liste døtems (forme, lemme ou catégorie)

Les items de la liste peuvent être projetés sur le concordancier via le raccourci clavier « shiftclic » sur le ou les items visé(s) :





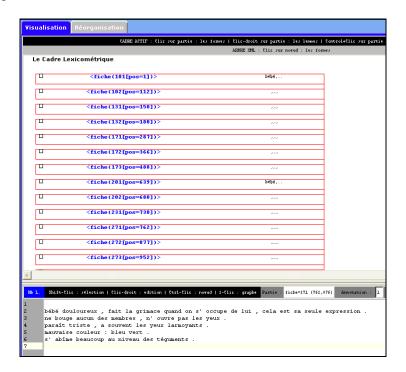
Figure 237 : Le Trameur retour en contexte (les dépendants døun item)

26 Le Cadre: visualisation, réorganisation

Le *Cadre* est présenté supra comme une interface permettant de définir des accès à la *Trame*. Nous présentons ici un chantier qui étend les outils et fonctionnalités associés à la *Trame*. Il søagit essentiellement de permettre à løutilisateur de réorganiser le *Cadre* afin de conduire des opérations supplémentaires à partir de la réorganisation réalisée en agissant dynamiquement sur la structuration du *Cadre*. Nous illustrerons ces modifications sur 2 fonctionnalités déjà disponibles :

- Réorganisation dynamique du *Cadre* pour construire un nouveau corpus de travail
- Réorganisation dynamique (et provisoire) du *Cadre* pour un calcul lexicométrique donné (accroissement du vocabulaire)

Lønglet CADRE est composé de 2 volets (2 sous-onglets) : løn dédié à sa visualisation, lønutre à sa réorganisation.



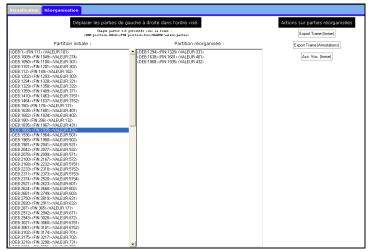


Figure 238 : Le Trameur le Cadre (visualisation, réorganisation)

Ce second volet permet de gérer la réorganisation dynamique du *Cadre*. Une fois la base textométrique chargée, un bouton permet de charger une partition donnée en vue de sa réorganisation dynamique :



Figure 239 : Le Trameur Chargement døune partition à réorganiser

Løutilisateur peut donc définir dynamiquement une nouvelle organisation døune partition initialement définie :

- En modifiant løordre des parties
- En ne sélectionnant quœune sélection restreinte de parties (ici la réorganisation peut être vue comme une opération de *sélection*)

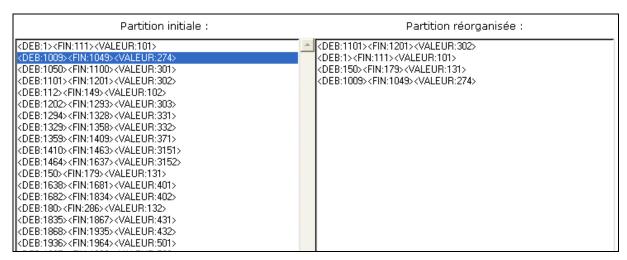


Figure 240 : Le Trameur Réorganisation døune partition

Une fois la réorganisation réalisée, il est possible døactiver différentes opérations sur le **CADRE** réorganisé dynamiquement. Il est important de souligner quo ce stage aucune modification no sest faite sur la base initiale : la *Trame* et le *Cadre* de départ restent dans leur état initial. La réorganisation visée est simplement décrite par la position des parties pour la réorganisation escomptée.

A ce stade du développement, les actions disponibles sont :

- 1. Export de la base textométrique en tenant compte de la réorganisation visée : cet export permet de regénérer une *Trame* et un *Cadre* sur la base de løordre décrit pour la partition choisie.
 - **Application**: réorganisation døun corpus pour rétablir par exemple une chronologie textuelle non réalisée dans la phase initiale. Cette réorganisation est vue ici comme un outil de préparation des données (ou de reformatage) avant analyse. Une fois la base exportée, il suffit de recharger la nouvelle base et de poursuivre son exploration.
- 2. Accroissement du vocabulaire sur la base en tenant compte de la réorganisation réalisée.

Ici, la réorganisation reste virtuelle, elle ne concerne que le calcul opéré. La base initiale reste inchangée. On peut par exemple comparer 2 courbes døaccroissement du vocabulaire sur une même *Trame* døtems organisés différemment sur la base de 2 partitions équivalentes mais structurées séquentiellement de manière différente. Le résultat produit donne døailleurs à voir le calcul de løaccroissement du vocabulaire sur la base initiale et sur la base réorganisée.

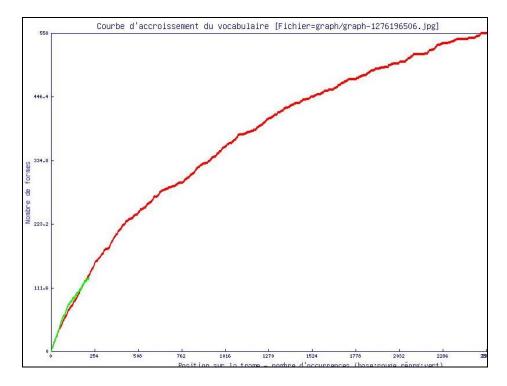


Figure 241 : *Le Trameur* Courbe déaccroissement sur une partition initiale et la même partition tronquée et réorganisée

Dans les 2 cas, løpération réalisée peut se réaliser sur lænsemble des parties réorganisées ou sur un sous-ensemble ne couvrant pas totalement lænsemble de la *Trame* initiale, dans ce dernier cas, on extrait un sous corpus ou on calcule løaccroissement du vocabulaire sur une partir de la *Trame* initiale réorganisée virtuellement.

27 Exporter une base textométrique

Le *Trameur* permet dømporter une *base textométrique i.e.* une ressource textuelle décrivant une *Trame* et un *Cadre* textométriques. Cette ressource peut être construite par le *Trameur* pour construire, à un moment donné, un état textométrique donnant à voir simultanément la *Trame* et le *Cadre* du texte traité. Elle peut ensuite être <u>réimportée</u> par le processus mis en ò uvre pour charger ce type de ressource.

Le processus dexportation de dune base textométrique est disponible dans le donglet CADRE. La ressource construite après activation de cette exportation est disponible dans le dossier export du répertoire contenant le fichier chargé.

On donne à voir ci-dessous le texte de travail « Le dormeur du Val » (chargé avec løoption étiquetage) et le fichier døexportation obtenu :

Texte initial

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons. 
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut.
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid.
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit.
<STRUCTURE="AUTEUR">Arthur Rimbaud
```

Fichier exporté (et ré-importable comme base textométrique



```
<?xml version="1.0" encoding="utf8"?>
<baselexicometrique>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base textometrique (le metier = le Cadre et la Trame),
importable et echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
Samedi 02 Juin 2012
22:15:57
       Ce document n'est pas encore publie.
</publicationStmt>
<sourceDesc>
<title>
<br>Le Trameur 10.38.2. Samedi 02 Juin 2012
22:15:57 </br></title>
<content><l><br>Fichier traite </br><br> dormeurL3.txt</br></l>
<l><br>Encodage </br><br> iso-8859-1</br></l>
<l><br>Nombre d'items </br><br> 295</br></l>
```

```
<l><br>Nombre de délimiteurs </br></pr>
<1><br>Nombre d'occurrences de forme </br>< 127</br></1>
<1><br>Nombre de formes </br><br> 97</br></1>
<1><br>Nombre d'hapax </br><br> 80</br></1>
<1><br>Fréquence maximale </br><br> 5</br></1>
<l><br>Forme maximale </br><br> dans</br></l>
<l><br>Délimiteurs </br><![CDATA[. ,:;!?/_-"'()[]{}$$|*><=+
       11></br></1>
<l><br>Etiquetage Treetagger </br><br>OUI</br></l>
<l><br>Langue pour Treetagger </br><br>français</br></l>
</content></sourceDesc>
</fileDesc>
</teiheader>
<Trame>
<codage>utf8</codage>
<delimiteur><![CDATA[. ,:;!?/_-"'()[]{}$$|*><=+
       ]]></delimiteur>
<items>
<item type="delim" pos="1"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="2"><f>Le</f><c>DET ART</c><1>le</l></item>
<item type="delim" pos="3"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="4"><f>dormeur</f><c>NOM</c><l>dormeur</l></item>
<item type="delim" pos="5"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="6"><f>du</f><c>PRP det</c><1>du</1></item>
<item type="delim" pos="7"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="8"><f>val</f><c>NOM</c><l>val</l></item>
<item type="delim" pos="9"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="10"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="11"><f>C</f><c>NUM</c><1>C</l></item>
<item type="delim" pos="12"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="delim" pos="13"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="14"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="15"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="16"><f>un</f><c>DET ART</c><1>un</1></item>
<item type="delim" pos="17"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="18"><f>trou</f><c>NOM</c><1>trou</l></item>
<item type="delim" pos="19"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="20"><f>de</f><c>PRP</c><1>de</l></item>
<item type="delim" pos="21"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="22"><f>verdure</f><c>NOM</c><1>verdure</1></item>
<item type="delim" pos="23"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="24"><f>où</f><c>PRO REL</c><l>où</l></item>
<item type="delim" pos="25"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="26"><f>chante</f><c>VER pres</c><1>chanter</1></item>
<item type="delim" pos="27"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="28"><f>une</f><c>DET_ART</c><1>un</1></item>
<item type="delim" pos="29"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="30"><f>rivière</f><c>NOM</c><1>rivière</1></item>
<item type="delim" pos="31"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="32"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="33"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="34"><f>Accrochant</f><c>NAM</c><1>Accrochant</1></item>
<item type="delim" pos="35"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="36"><f>follement</f><c>ADV</c><1>follement</l></item>
<item type="delim" pos="37"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="38"><f>aux</f><c>PRP_det</c><1>au</1></item>
<item type="delim" pos="39"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="40"><f>herbes</f><c>VER_pres</c><1>herber</1></item>
<item type="delim" pos="41"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="42"><f>des</f><c>PRP det</c><l>du</l></item>
<item type="delim" pos="43"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="44"><f>haillons</f><c>NOM</c><l>haillon</l></item>
<item type="delim" pos="45"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="46"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="47"><f>D</f><c>NAM</c><l>D</l></item>
<item type="delim" pos="48"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="delim" pos="49"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="50"><f>argent</f><c>NOM</c><1>argent</l></item>
<item type="delim" pos="51"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="delim" pos="52"><f>;</f><c>DELIM</c><1>;</1></item>
<item type="delim" pos="53"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="54"><f>où</f><c>PRO REL</c><l>où</l></item>
<item type="delim" pos="55"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="56"><f>le</f><c>DET ART</c><1>le</l></item>
<item type="delim" pos="57"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="58"><f>soleil</f><c>NOM</c><l>soleil</l></item>
```

CLESTHIA>SYLED>CLA²T

```
<item type="delim" pos="59"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="delim" pos="60"><f>;</f><c>DELIM</c><1>;</l></item>
<item type="delim" pos="61"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="62"><f>de</f><c>PRP</c><l>de</l></item>
<item type="delim" pos="63"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="64"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="65"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="66"><f>montagne</f><c>NOM</c><1>montagne</l></item>
<item type="delim" pos="67"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="68"><f>fière</f><c>ADJ</c><1>fier</1></item>
<item type="delim" pos="69"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="70"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="71"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="72"><f>Luit</f><c>NAM</c><1>Luit</l></item>
<item type="delim" pos="73"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="delim" pos="74"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="delim" pos="75"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="76"><f>c</f><c>VER_simp</c><1>c</1></item>
<item type="delim" pos="77"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="delim" pos="78"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="79"><f>est</f><c>VER pres</c><l>être</l></item>
<item type="delim" pos="80"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="81"><f>un</f><c>DET_ART</c><1>un</1></item>
<item type="delim" pos="82"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="83"><f>petit</f><c>ADJ</c><1>petit</1></item>
<item type="delim" pos="84"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="85"><f>val</f><c>NOM</c><l>val</l></item>
<item type="delim" pos="86"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="87"><f>qui</f><c>PRO_REL</c><l>qui</l></item>
<item type="delim" pos="88"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="89"><f>mousse</f><c>NOM</c><1>mousse</1></item>
<item type="delim" pos="90"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="91"><f>de</f><c>PRP</c><1>de</l></item>
<item type="delim" pos="92"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="93"><f>rayons</f><c>NOM</c><1>rayon</1></item>
<item type="delim" pos="94"><f>.</f><c>DELIM</c><1>.</1></item>
<item type="delim" pos="95"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="96"><f>Un</f><c>NAM</c><1>Un</1></item>
<item type="delim" pos="97"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="98"><f>soldat</f><c>NOM</c><l>soldat</l></item>
<item type="delim" pos="99"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="100"><f>jeune</f><c>ADJ</c><1>jeune</l></item>
<item type="delim" pos="101"><f>, </f><c>DELIM</c><1>, </l></item>
<item type="delim" pos="102"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="103"><f>bouche</f><c>VER_pres</c><1>boucher</1></item>
<item type="delim" pos="104"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="105"><f>ouverte</f><c>ADJ</c><1>ouvert</1></item>
<item type="delim" pos="106"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="107"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="108"><f>tête</f><c>NOM</c><1>tête</l></item>
<item type="delim" pos="109"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="110"><f>nue</f><c>VER pres</c><l>nue</l></item>
<item type="delim" pos="111"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="112"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="113"><f>Et</f><c>NAM</c><1>Et</l></item>
<item type="delim" pos="114"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="115"><f>la</f><c>DET ART</c><1>le</l></item>
<item type="delim" pos="116"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="117"><f>nuque</f><c>NOM</c><l>nuque</l></item>
<item type="delim" pos="118"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="119"><f>baignant</f><c>VER ppre</c><l>baigner</l></item>
<item type="delim" pos="120"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="121"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="delim" pos="122"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="123"><f>le</f><c>DET ART</c><1>le</1></item>
<item type="delim" pos="124"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="125"><f>frais</f><c>ADJ</c><l>frais</l></item>
<item type="delim" pos="126"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="127"><f>cresson</f><c>NOM</c><l>cresson</l></item>
<item type="delim" pos="128"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="129"><f>bleu</f><c>ADJ</c><1>bleu</1></item>
<item type="delim" pos="130"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="131"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="132"><f>Dort</f><c>NAM</c><1>Dort</1></item>
<item type="delim" pos="133"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="delim" pos="134"><f>;</f><c>DELIM</c><1>;</l></item>
```

```
<item type="delim" pos="135"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="136"><f>il</f><c>PRO PER</c><l>il</l></item>
<item type="delim" pos="137"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="138"><f>est</f><c>VER pres</c><l>être</l></item>
<item type="delim" pos="139"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="140"><f>étendu</f><c>VER_pper</c><l>étendre</l></item>
<item type="delim" pos="141"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="142"><f>dans</f><c>PRP</c><1>dans</1></item>
<item type="delim" pos="143"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="144"><f>l</f><c>VER_simp</c><l>l</l></item>
<item type="delim" pos="145"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="forme" pos="146"><f>herbe</f><c>NOM</c><l>herbe</l></item>
<item type="delim" pos="147"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="148"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="149"><f>sous</f><c>PRP</c><l>sous</l></item>
<item type="delim" pos="150"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="151"><f>la</f><c>DET ART</c><1>le</l></item>
<item type="delim" pos="152"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="153"><f>nue</f><c>VER_pres</c><l>nue</l></item>
<item type="delim" pos="154"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="155"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="156"><f>Pâle</f><c>NAM</c><1>Pâle</1></item>
<item type="delim" pos="157"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="158"><f>dans</f><c>PRP</c><1>dans</l></item>
<item type="delim" pos="159"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="160"><f>son</f><c>DET POS</c><l>son</l></item>
<item type="delim" pos="161"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="162"><f>lit</f><c>VER_pres</c><l>lire</l></item>
<item type="delim" pos="163"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="164"><f>vert</f><c>ADJ</c><1>vert</1></item>
<item type="delim" pos="165"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="166"><f>où</f><c>PRO REL</c><l>où</l></item>
<item type="delim" pos="167"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="168"><f>la</f><c>DET ART</c><1>le</l></item>
<item type="delim" pos="169"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="170"><f>lumière</f><c>NOM</c><1>lumière</l></item>
<item type="delim" pos="171"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="172"><f>pleut</f><c>VER pres</c><l>pleuvoir</l></item>
<item type="delim" pos="173"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="174"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="175"><f>Les</f><c>DET_ART</c><1>le</l></irtem>
<item type="delim" pos="176"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="177"><f>pieds</f><c>NOM</c><1>pied</1></item>
<item type="delim" pos="178"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="179"><f>dans</f><c>PRP</c><1>dans</l></item>
<item type="delim" pos="180"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="181"><f>les</f><c>DET ART</c><1>le</1></item>
<item type="delim" pos="182"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="183"><f>glaïeuls</f><c>NOM</c><1>glaïeul</l></item>
<item type="delim" pos="184"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="185"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="186"><f>il</f><c>PRO PER</c><l>il</l></item>
<item type="delim" pos="187"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="188"><f>dort</f><c>VER_pres</c><1>dormir</1></item>
<item type="delim" pos="189"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="190"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="191"><f>Souriant</f><c>NAM</c><1>Souriant</1></item>
<item type="delim" pos="192"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="193"><f>comme</f><c>KON</c><1>comme</l></item>
<item type="delim" pos="194"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="195"><f>Sourirait</f><c>NAM</c><1>Sourirait</1></item>
<item type="delim" pos="196"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="197"><f>un</f><c>DET_ART</c><1>un</1></item>
<item type="delim" pos="198"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="199"><f>enfant</f><c>NOM</c><1>enfant</1></item>
<item type="delim" pos="200"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="201"><f>malade</f><c>ADJ</c><1>malade</1></item>
<item type="delim" pos="202"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="203"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="204"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="delim" pos="205"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="206"><f>fait</f><c>VER_pres</c><1>faire</1></item>
<item type="delim" pos="207"><f> </f><c>DELIM<\overline{/}c><1>BLANK</1></item>
<item type="forme" pos="208"><f>un</f><c>DET ART</c><1>un</1></item>
<item type="delim" pos="209"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="210"><f>somme</f><c>NOM</c><1>somme</l></item>
```

CLESTHIA>SYLED>CLA²T

```
<item type="delim" pos="211"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="delim" pos="212"><f>:</f><c>DELIM</c><1>:</1></item>
<item type="delim" pos="213"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="214"><f>Nature</f><c>NAM</c><1>Nature</l></item>
<item type="delim" pos="215"><f>,</f><c>DELIM</c><1>,</1></item>
<item type="delim" pos="216"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="217"><f>berce</f><c>VER_pres</c><l>bercer</l></item>
<item type="delim" pos="218"><f>-</f><c>DELIM</c><1>-</1></item>
<item type="forme" pos="219"><f>le</f><c>DET ART</c><l>le</l></item>
<item type="delim" pos="220"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="221"><f>chaudement</f><c>ADV</c><l>chaudement</l></item>
<item type="delim" pos="222"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="delim" pos="223"><f>:</f><c>DELIM</c><1>:</l></item>
<item type="delim" pos="224"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="225"><f>il</f><c>PRO PER</c><l>il</l></item>
<item type="delim" pos="226"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="227"><f>a</f><c>VER pres</c><l>avoir</l></item>
<item type="delim" pos="228"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="229"><f>froid</f><c>ADJ</c><l>froid</l></item>
<item type="delim" pos="230"><f>.</f><c>DELIM</c><1>.</1></item>
<item type="delim" pos="231"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="232"><f>Les</f><c>DET ART</c><1>le</1></item>
<item type="delim" pos="233"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="234"><f>parfums</f><c>NOM</c><1>parfum</1></item>
<item type="delim" pos="235"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="236"><f>ne</f><c>ADV</c><1>ne</1></item>
<item type="delim" pos="237"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="238"><f>font</f><c>VER_pres</c><l>faire</l></item>
<item type="delim" pos="239"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="240"><f>pas</f><c>ADV</c><1>pas</l></item>
<item type="delim" pos="241"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="242"><f>frissonner</f><c>VER infi</c><1>frissonner</1></item>
<item type="delim" pos="243"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="244"><f>sa</f><c>DET POS</c><l>son</l></item>
<item type="delim" pos="245"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="246"><f>narine</f><c>NOM</c><1>narine</1></item>
<item type="delim" pos="247"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="delim" pos="248"><f>;</f><c>DELIM</c><1>;</1></item>
<item type="delim" pos="249"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="250"><f>Il</f><c>NAM</c><l>Il</l></item>
<item type="delim" pos="251"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="252"><f>dort</f><c>VER pres</c><l>dormir</l></item>
<item type="delim" pos="253"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="254"><f>dans</f><c>PRP</c><1>dans</l></item>
<item type="delim" pos="255"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="256"><f>le</f><c>DET_ART</c><1>le</l></irem>
<item type="delim" pos="257"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="258"><f>soleil</f><c>NOM</c><1>soleil<//></item>
<item type="delim" pos="259"><f>,</f><c>DELIM</c><1>,</1></item>
<item type="delim" pos="260"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="261"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="262"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="263"><f>main</f><c>NOM</c><1>main</1></item>
<item type="delim" pos="264"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="265"><f>sur</f><c>PRP</c><1>sur</l></item>
<item type="delim" pos="266"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="267"><f>sa</f><c>DET POS</c><1>son</1></item>
<item type="delim" pos="268"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="269"><f>poitrine</f><c>NOM</c><1>poitrine</1></item>
<item type="delim" pos="270"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="271"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="272"><f>Tranquille</f><c>NAM</c><1>Tranquille</1></item>
<item type="delim" pos="273"><f>.</f><c>DELIM</c><1>.</1></item>
<item type="delim" pos="274"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="275"><f>Il</f><c>NAM</c><l>Il</l></item>
<item type="delim" pos="276"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="277"><f>a</f><c>VER pres</c><l>avoir</l></item>
<item type="delim" pos="278"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="279"><f>deux</f><c>NUM</c><1>deux</l></item>
<item type="delim" pos="280"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="281"><f>trous</f><c>NOM</c><1>trou</l></item>
<item type="delim" pos="282"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="283"><f>rouges</f><c>ADJ</c><1>rouge</1></item>
<item type="delim" pos="284"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="285"><f>au</f><c>PRP_det</c><1>au</1></item>
<item type="delim" pos="286"><f> </f><c>DELIM</c><1>BLANK</1></item>
```

```
<item type="forme" pos="287"><f>côté</f><c>NOM</c><1>côté</1></item>
<item type="delim" pos="288"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="289"><f>droit</f><c>NOM</c><1>droit</1></item>
<item type="delim" pos="290"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="291"><f>RETURN</f><c>DELIM</c><1>RETURN</l></item>
<item type="forme" pos="292"><f>Arthur</f><c>NAM</c><1>Arthur</l></item>
<item type="delim" pos="293"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="294"><f>Rimbaud</f><c>NAM</c><1>Rimbaud</1></item>
<item type="delim" pos="295"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="296"><f> </f><c>DELIM</c><1>BLANK</1></item>
</items>
</Trame>
<Cadre>
<acces>
<partition nom="ligne">
</partition>
<partition nom="structure">
</partition>
</acces>
</Cadre>
</baselexicometrique>
```

Le fichier dœxport contient tout dœbord la *Trame* (liste des positions des items et pour chacune dœlles les annotations disponibles: forme, lemme²¹, catégorie²², annotations importées) puis le *Cadre* (les intervalles de positions regroupant certaines unités de la *Trame* en parties structurées).

²¹ Si treetagger a préalablement été activé

²² Si treetagger a préalablement été activé

28 Exportation du *Cadre* et de la *Trame*

Il est possible déexporter le *Cadre* et la *Trame* en tenant compte des éventuelles corrections apportées sur les mots, les lemmes ou les catégories ; de même les systèmes déannotation importés (*cf infra*) seront intégrés dans les fichiers déexport.

En activant le bouton $\stackrel{\triangle}{\leftarrow}$, le *Trameur* enregistre dans le dossier export 3 fichiers : la *Trame*, le *Cadre* et la *Trame* annotée.

Nous allons illustrer sur un exemple les formats des fichiers déexport.

28.1 Fichier de travail

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut.
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid.
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit.
<STRUCTURE="AUTEUR">Arthur Rimbaud
```

28.2 Fichiers døexport

1. La *Trame* reconstituée :

```
Le dormeur du val
 C' est un trou de verdure où chante une rivière,
 Accrochant follement aux herbes des haillons
 D' argent ; où le soleil ; de la montagne fière,
 Luit : c' est un petit val qui mousse de rayons.
Un soldat jeune, bouche ouverte, tête nue,
Et la nuque baignant dans le frais cresson bleu,
Dort ; il est étendu dans l'herbe, sous la nue,
Pâle dans son lit vert où la lumière pleut.
Les pieds dans les glaïeuls, il dort. Souriant comme
Sourirait un enfant malade, il fait un somme :
Nature, berce-le chaudement : il a froid.
Les parfums ne font pas frissonner sa narine ;
Il dort dans le soleil, la main sur sa poitrine,
Tranquille. Il a deux trous rouges au côté droit.
Arthur Rimbaud
```

2. Le Cadre:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<Cadre>
                                  entête
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame),
importable et echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
Samedi 26 Mai 2007
10:50:52
 Ce document n'est pas encore publie.
</publicationStmt>
<sourceDesc>
Fichier d'export genere par LeMetierLexicometrique 5.009 (aka le Trameur)
(http://www.tal.univ-paris3.fr/trameur/).
Fichier initilement traite : C:/SFleury/DEVEL/WIP-ANR-
textometrrie/mkGraphTrameCadre/textes/dormeurL3.txt
Encodage : iso-8859-1
</sourceDesc>
</fileDesc>
</teiheader>
<acces>
<partition nom="ligne">
- 
</partition>
<partition nom="structure">
</partition>
</acces>
</Cadre>
```

3. La *Trame* annotée :

```
<?xml version="1.0" encoding="iso-8859-1"?>
<trametreetagger>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), importable e
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
Samedi 26 Mai 2007
10:41:33
  Ce document n'est pas encore publie.
</publicationStmt>
<sourceDesc>
Fichier d'export genere par LeMetierLexicometrique 5.009 (aka le Trameur)
(http://www.tal.univ-paris3.fr/trameur/).
Fichier initilement traite : C:/SFleury/DEVEL/WIP-ANR-
textometrrie/mkGraphTrameCadre/textes/dormeurL3.txt
Encodage : iso-8859-1
</sourceDesc>
</fileDesc>
</technology
<items>
<item type="forme" pos="1"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="2"><f>Le</f><c>DET ART</c><1>le</l></item>
<item type="forme" pos="3"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="4"><f>dormeur</f><c>NOM</c><l>dormeur</l></item>
<item type="forme" pos="5"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="6"><f>du</f><c>PRP_det</c><l>du</l></ir>
<item type="forme" pos="7"><f></f><c>BLANK</c><l>BLANK</l></ir>
<item type="forme" pos="8"><f>val</f><c>NOM</c><1>val</l></item>
<item type="forme" pos="9"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="10"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="11"><f>C</f><c>NUM</c><1>C</1></item>
<item type="forme" pos="12"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="forme" pos="13"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="14"><f><c>BLANK</c><1>BLANK</l></i>
<item type="forme" pos="15"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="16"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="forme" pos="17"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="18"><f>trou</f><c>NOM</c><1>trou</1></item>
<item type="forme" pos="19"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="20"><f>de</f><c>PRP</c><1>de</1></item>
<item type="forme" pos="21"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="22"><f>verdure</f><c>NOM</c><1>verdure</1></item>
<item type="forme" pos="23"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="24"><f>où</f><c>PRO REL</c><l>où</l></item>
<item type="forme" pos="25"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="26"><f>chante</f><c>VER pres</c><l>chanter</l></item>
<item type="forme" pos="27"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="28"><f>une</f><c>DET_ART</c><1>un</1></item>
<item type="forme" pos="29"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="30"><f>rivière</f><c>NOM</c><1>rivière</1></item>
<item type="forme" pos="31"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="32"><f></f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="33"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="34"><f>Accrochant</f><c>NAM</c><l>Accrochant</l></item>
<item type="forme" pos="35"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="36"><f>follement</f><c>ADV</c><1>follement</l></item>
<item type="forme" pos="37"><f> </f><c>BLANK</c><1>BLANK<//1></item>
<item type="forme" pos="38"><f>aux</f><c>PRP det</c><1>au</1></item>
<item type="forme" pos="39"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="40"><f>herbes</f><c>VER pres</c><l>herber</l></item>
<item type="forme" pos="41"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="42"><f>des</f><c>PRP det</c><1>du</1></item>
<item type="forme" pos="43"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="44"><f>haillons</f><c>NOM</c><1>haillon</l></item>
<item type="forme" pos="45"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="46"><f> </f><c>BLANK</c><1>BLANK</l></item>
```

entête

```
<item type="forme" pos="47"><f>D</f><c>NAM</c><1>D</1></item>
<item type="forme" pos="48"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="forme" pos="49"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="50"><f>argent</f><c>NOM</c><1>argent</1></item>
<item type="forme" pos="51"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="52"><f>;</f><c>DELIM</c><1>;</1></item>
<item type="forme" pos="53"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="54"><f>où</f><c>PRO REL</c><l>où</l></item>
<item type="forme" pos="55"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="56"><f>le</f><c>DET_ART</c><1>le</l></item>
<item type="forme" pos="57"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="58"><f>soleil</f><c>NOM</c><l>soleil</l></item>
<item type="forme" pos="59"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="60"><f>;</f><c>DELIM</c><1>;</l></item>
<item type="forme" pos="61"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="62"><f>de</f><c>PRP</c><1>de</l></item>
<item type="forme" pos="63"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="64"><f>la</f><c>DET_ART</c><1>le</l></item>
<item type="forme" pos="65"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="66"><f>montagne</f><c>NOM</c><1>montagne</1></item>
<item type="forme" pos="67"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="68"><f>fière</f><c>ADJ</c><1>fier</1></item>
<item type="forme" pos="69"><f>,</f><c>DELIM</c><1>,</1></item>
<item type="forme" pos="70"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="71"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="72"><f>Luit</f><c>NAM</c><1>Luit</1></item>
<item type="forme" pos="73"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="74"><f>:</f><c>DELIM</c><1>:</l></item>
<item type="forme" pos="75"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="76"><f>c</f><c>VER simp</c><1>c</1></item>
<item type="forme" pos="77"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="forme" pos="78"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="79"><f>est</f><c>VER pres</c><l>être</l></item>
<item type="forme" pos="80"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="81"><f>un</f><c>DET_ART</c><1>un</1></item>
<item type="forme" pos="82"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="83"><f>petit</f><c>ADJ</c><1>petit</1></item>
<item type="forme" pos="84"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="85"><f>val</f><c>NOM</c><1>val</1></item>
<item type="forme" pos="86"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="87"><f>qui</f><c>PRO REL</c><l>qui</l></item>
<item type="forme" pos="88"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="89"><f>mousse</f><c>NOM</c><1>mousse</1></item>
<item type="forme" pos="90"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="91"><f>de</f><c>PRP</c><l>de</l></l>
<item type="forme" pos="92"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="93"><f>rayons</f><c>NOM</c><l>rayon</l></item>
<item type="forme" pos="94"><f>.</f><c>DELIM</c><1>.</1></item>
<item type="forme" pos="95"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="96"><f>Un</f><c>NAM</c><1>Un</1></item>
<item type="forme" pos="97"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="98"><f>soldat</f><c>NOM</c><l>soldat</l></item>
<item type="forme" pos="99"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="100"><f>jeune</f><c>ADJ</c><l>jeune</l></item>
<item type="forme" pos="101"><f>,</f><c>DELIM</c><1>,</1></item>
<item type="forme" pos="102"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="103"><f>bouche</f><c>VER_pres</c><1>boucher</1></item>
<item type="forme" pos="104"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="105"><f>ouverte</f><c>ADJ</c><1>ouvert</1></item>
<item type="forme" pos="106"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="107"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="108"><f>tête</f><c>NOM</c><l>tête</l></item>
<item type="forme" pos="109"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="110"><f>nue</f><c>VER pres</c><1>nuer</1></item>
<item type="forme" pos="111"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="112"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="113"><f>Et</f><c>NAM</c><1>Et</l></item>
<item type="forme" pos="114"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="115"><f>la</f><c>DET ART</c><1>le</l></item>
<item type="forme" pos="116"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="117"><f>nuque</f><c>NOM</c><1>nuque</1></item>
<item type="forme" pos="118"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="119"><f>baignant</f><c>VER_ppre</c><1>baigner</l></item>
```

```
<item type="forme" pos="120"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="121"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="forme" pos="122"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="123"><f>le</f><c>DET ART</c><1>le</l></item>
<item type="forme" pos="124"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="125"><f>frais</f><c>ADJ</c><l>frais</l></item>
<item type="forme" pos="126"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="127"><f>cresson</f><c>NOM</c><1>cresson</1></item>
<item type="forme" pos="128"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="129"><f>bleu</f><c>ADJ</c><1>bleu</l></item>
<item type="forme" pos="130"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="131"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="132"><f>Dort</f><c>NAM</c><1>Dort</1></item>
<item type="forme" pos="133"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="134"><f>;</f><c>DELIM</c><1>;</1></item>
<item type="forme" pos="135"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="136"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="forme" pos="137"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="138"><f>est</f><c>VER pres</c><l>être</l></item>
<item type="forme" pos="139"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="140"><f>étendu</f><c>VER_pper</c><l>étendre</l></item>
<item type="forme" pos="141"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="142"><f>dans</f><c>PRP</c><1>dans</l></item>
<item type="forme" pos="143"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="144"><f>l</f><c>VER simp</c><l>l</l></item>
<item type="forme" pos="145"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="forme" pos="146"><f>herbe</f><c>NOM</c><l>herbe</l></item>
<item type="forme" pos="147"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="148"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="149"><f>sous</f><c>PRP</c><1>sous</1></item>
<item type="forme" pos="150"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="151"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="152"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="153"><f>nue</f><c>VER pres</c><1>nuer</1></item>
<item type="forme" pos="154"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="155"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="156"><f>Pâle</f><c>NAM</c><l>Pâle</l></item>
<item type="forme" pos="157"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="158"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="forme" pos="159"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="160"><f>son</f><c>DET_POS</c><l>son</l></item>
<item type="forme" pos="161"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="162"><f>lit</f><c>VER_pres</c><l>lire</l></item>
<item type="forme" pos="163"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="164"><f>vert</f><c>ADJ</c><1>vert</1></item>
<item type="forme" pos="165"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="166"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="forme" pos="167"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="168"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="169"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="170"><f>lumière</f><c>NOM</c><1>lumière</1></item>
<item type="forme" pos="171"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="172"><f>pleut</f><c>VER pres</c><1>pleuvoir</1></item>
<item type="forme" pos="173"><f>.</f><c>DELIM</c><1>.</l></item>
<item type="forme" pos="174"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="175"><f>Les</f><c>DET_ART</c><1>le</l></item>
<item type="forme" pos="176"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="177"><f>pieds</f><c>NOM</c><1>pied</1></item>
<item type="forme" pos="178"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="179"><f>dans</f><c>PRP</c><1>dans</l></item>
<item type="forme" pos="180"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="181"><f>les</f><c>DET ART</c><1>le</1></item>
<item type="forme" pos="182"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="183"><f>glaïeuls</f><c>NOM</c><l>glaïeul</l></item>
<item type="forme" pos="184"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="185"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="186"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="forme" pos="187"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="188"><f>dort</f><c>VER_pres</c><l>dormir</l></item>
<item type="forme" pos="189"><f>.</f><c>DELIM</c><1>.</1></item>
<item type="forme" pos="190"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="191"><f>Souriant</f><c>NAM</c><1>Souriant</l></item>
<item type="forme" pos="192"><f> </f><c>BLANK</c><1>BLANK</1></item>
```

```
<item type="forme" pos="193"><f>comme</f><c>KON</c><1>comme</l></item>
<item type="forme" pos="194"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="195"><f>Sourirait</f><c>NAM</c><1>Sourirait</1></item>
<item type="forme" pos="196"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="197"><f>un</f><c>DET_ART</c><1>un</1></item>
<item type="forme" pos="198"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="199"><f>enfant</f><c>NOM</c><1>enfant</1></item>
<item type="forme" pos="200"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="201"><f>malade</f><c>ADJ</c><1>malade</1></item>
<item type="forme" pos="202"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="203"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="204"><f>il</f><c>PRO PER</c><l>il</l></item>
<item type="forme" pos="205"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="206"><f>fait</f><c>VER pres</c><l>faire</l></item>
<item type="forme" pos="207"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="208"><f>un</f><c>DET ART</c><1>un</1></item>
<item type="forme" pos="209"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="210"><f>somme</f><c>NOM</c><1>somme</f></item>
<item type="forme" pos="211"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="212"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="forme" pos="213"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="214"><f>Nature</f><c>NAM</c><1>Nature</l></item>
<item type="forme" pos="215"><f>,</f><c>DELIM</c><1>,</1></item>
<item type="forme" pos="216"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="217"><f>berce</f><c>VER_pres</c><1>bercer</1></item>
<item type="forme" pos="218"><f>-</f><c>DELIM</c><1>-</1></item>
<item type="forme" pos="219"><f>le</f><c>DET_ART</c><1>le</l></item>
<item type="forme" pos="220"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="221"><f>chaudement</f><c>ADV</c><1>chaudement</1></item>
<item type="forme" pos="222"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="223"><f>:</f><c>DELIM</c><1>:</1></item>
<item type="forme" pos="224"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="225"><f>il</f><c>PRO PER</c><l>il</l></item>
<item type="forme" pos="226"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="227"><f>a</f><c>VER pres</c><l>avoir</l></item>
<item type="forme" pos="228"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="229"><f>froid</f><c>ADJ</c><l>froid</l></item>
<item type="forme" pos="230"><f>.</f><c>DELIM</c><1>.</l></item>
<item type="forme" pos="231"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="232"><f>Les</f><c>DET ART</c><1>le</1></item>
<item type="forme" pos="233"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="234"><f>parfums</f><c>NOM</c><l>parfum</l></item>
<item type="forme" pos="235"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="236"><f>ne</f><c>ADV</c><1>ne</1></item>
<item type="forme" pos="237"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="238"><f>font</f><c>VER_pres</c><1>faire</1></item>
<item type="forme" pos="239"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="240"><f>pas</f><c>ADV</c><1>pas</1></item>
<item type="forme" pos="241"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="242"><f>frissonner</f><c>VER infi</c><1>frissonner</1></item>
<item type="forme" pos="243"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="244"><f>sa</f><c>DET POS</c><1>son</1></item>
<item type="forme" pos="245"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="246"><f>narine</f><c>NOM</c><1>narine</1></item>
<item type="forme" pos="247"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="248"><f>;</f><c>DELIM</c><1>;</l></item>
<item type="forme" pos="249"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="250"><f>Il</f><c>NAM</c><l>Il</l></item>
<item type="forme" pos="251"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="252"><f>dort</f><c>VER_pres</c><l>dormir</l></item>
<item type="forme" pos="253"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="254"><f>dans</f><c>PRP</c><1>dans</1></item>
<item type="forme" pos="255"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="256"><f>le</f><c>DET ART</c><l>le</l></item>
<item type="forme" pos="257"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="258"><f>soleil</f><c>NOM</c><1>soleil</1></item>
<item type="forme" pos="259"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="260"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="261"><f>la</f><c>DET ART</c><1>le</l></item>
<item type="forme" pos="262"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="263"><f>main</f><c>NOM</c><1>main</1></item>
<item type="forme" pos="264"><f> </f><c>BLANK</c><1>BLANK</l></item>
```

```
<item type="forme" pos="265"><f>sur</f><c>PRP</c><1>sur</1></item>
<item type="forme" pos="266"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="267"><f>sa</f><c>DET_POS</c><1>son</1></item>
<item type="forme" pos="268"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="269"><f>poitrine</f><c>NOM</c><1>poitrine</l></item>
<item type="forme" pos="270"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="forme" pos="271"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="272"><f>Tranquille</f><c>NAM</c><1>Tranquille</1></item>
<item type="forme" pos="273"><f>.</f><c>DELIM</c><1>.</1></item>
<item type="forme" pos="274"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="275"><f>II</f><c>NAM</c><1>II</1></item>
<item type="forme" pos="276"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="277"><f>a</f><c>VER_pres</c><1>avoir</1></item>
<item type="forme" pos="278"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="279"><f>deux</f><c>NUM</c><1>deux</l></item>
<item type="forme" pos="280"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="281"><f>trous</f><c>NOM</c><1>trou</1></item>
<item type="forme" pos="282"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="283"><f>rouges</f><c>ADJ</c><1>rouge</l></item>
<item type="forme" pos="284"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="285"><f>au</f><c>PRP det</c><l>au</l></item>
<item type="forme" pos="286"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="287"><f>côté</f><c>NOM</c><1>côté</l></item>
<item type="forme" pos="288"><f> </f><c>BLANK</c><1>BLANK</l></item>
<item type="forme" pos="289"><f>droit</f><c>NOM</c><l>droit</l></item>
<item type="forme" pos="290"><f>.</f><c>DELIM</c><1>.</l></item>
<item type="forme" pos="291"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="292"><f>Arthur</f><c>NAM</c><1>Arthur</1></item>
<item type="forme" pos="293"><f> </f><c>BLANK</c><1>BLANK</1></item>
<item type="forme" pos="294"><f>Rimbaud</f><c>NAM</c><1>Rimbaud</1></item>
<item type="forme" pos="295"><f>
</f><c>RETURN</c><1>RETURN</1></item>
<item type="forme" pos="296"><f> </f><c>BLANK</c><1>BLANK</l></item>
</items>
</trametreetagger>
```

29 Le rapport : enregistrement des résultats produits

Les résultats qui intéressent l'utilisateur pour une exploitation ultérieure peuvent être rassemblés dans un dossier construit dans le répertoire rapport. Ce dossier aisément manipulable à l'aide d'un navigateur web (*Firefox, Internet Explorer, Safari*, etc.) contient un fichier døindex qui permet la navigation parmi les résultats sélectionnés. Le rapport peut être consulté dès que l'utilisateur l'a enregistré.

29.1 Ajouter un élément au rapport

Pour ajouter un document au rapport, il suffit de se positionner sur un onglet (ou une fenêtre) contenant un résultat produit par le programme puis de cliquer sur l'icône *Ajouter au rapport* présente dans la fenêtre principale ou dans la fenêtre de résultats (cœst le cas pour les spécificités).

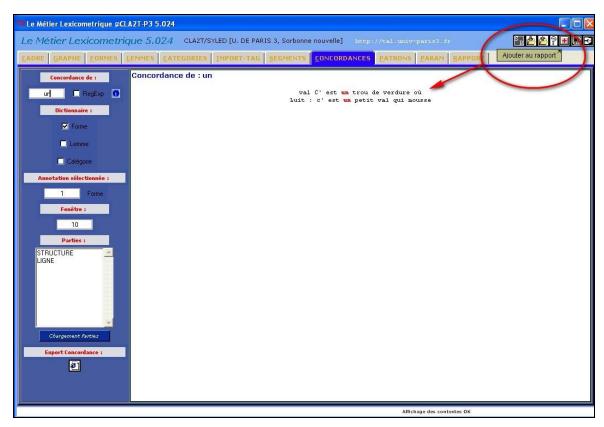


Figure 242 : Ajouter un élément au rapport

Une trace des éléments ajoutés au rapport est visible de manière synthétique dans lønglet **RAPPORT**.

Le rapport est constitué par un cartouche et la liste des éléments ajoutés par lœutilisateur. Chacun des éléments du rapport dispose dœune zone de texte « libre » dans laquelle lœutilisateur peut y noter ses observations. Pour éditer ou modifier cette note, il suffit dœuctiver lœucône de lœutilisateur présent près de chaque élément du rapport dans læuglet RAPPORT.

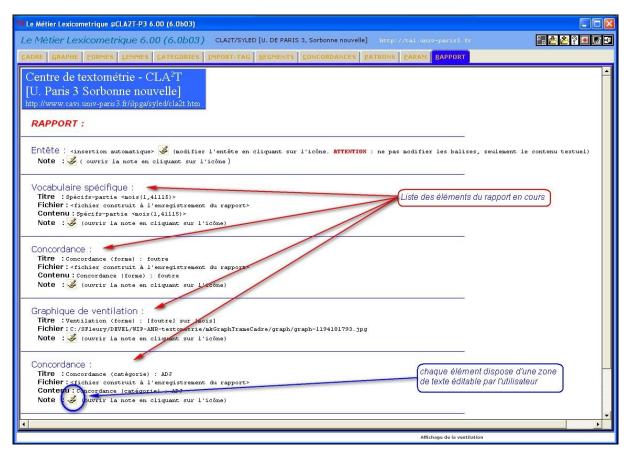


Figure 243: Rapport en cours dans løonglet RAPPORT

29.2 Ajouter un élément externe au rapport

Cette fonctionnalité permet à løutilisateur døinsérer dans le rapport des images externes (non produites directement par le *Trameur*). Par exemple, si løon souhaite insérer une image donnant à voir la carte des sections, on peut le faire en procédant de la manière suivante :

- créer une copie décran de la carte des sections (par exemple avec un outil comme *FastStone Capture* http://www.faststone.org/) et sauvegarder la copie décran au format jpeg ou gif
- insérer lømage dans le rapport en utilisant le menu idoine dans lønglet RAPPORT

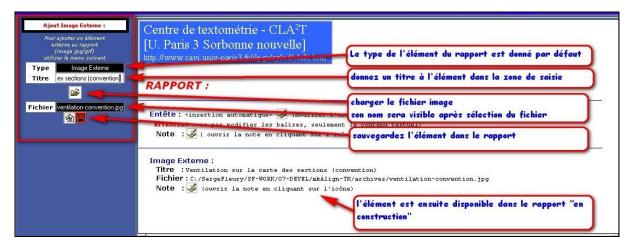


Figure 244 : Ajouter un élément externe au rapport

29.2 Enregistrer le rapport

On peut à tout moment visualiser le rapport en cliquant sur le bouton *Enregistrer le rapport*Une fois activé, ce bouton déclenche la génération dœun « dossier rapport ». Ce dossier contient, dans la distribution initiale du *Trameur*, des feuilles de styles XSL (pour chaque type de résultat à afficher) et une feuille de styles CSS (pour gérer læffichage général des toutes les pages du rapport). Les fichiers construits pour un rapport sont au format XML et ils sont rassemblés dans un sous-dossier horodaté du dossier rapport.

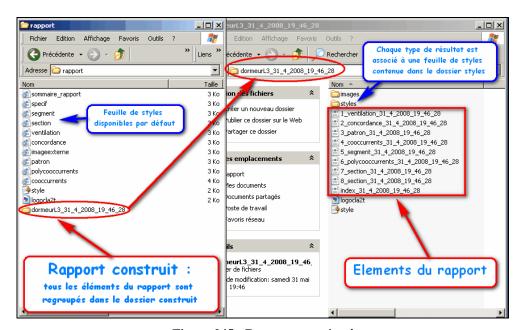


Figure 245 : Rapport enregistré

Après enregistrement, le rapport apparaît dans le navigateur web paramétré par défaut sur la machine utilisée (ici *Firefox*) :

Centre de textométrie - CLA2T [U. Paris 3 Sorbonne nouvelle] http://www.cavi.univ-paris3.fr/ilpga/syled/cla2t.htm Le Trameur 7.00 (7.0b30). Lundi 28 Avril 2008 23:04:16 Fichier traite Encodage iso-8859-1 Nombre d'items 295 Nombre de délimiteurs 168 Nombre d'occurrences de forme 127 Nombre de formes 96 Nombre d'hapax Fréquence maximale Forme maximale dans Délimiteurs . ,::!?/_-""()[]()\$\$1*×=+ Etiquetage Treetagger OUI Langue pour Treetagger La Trame dormeurL3-trame.xml Le Cadre dormeurL3-radre.xml Fichier des formes construit a partir de la trame Dictionnaire des formes dormeurL3-forme.dic Fichier a treetagger dormeurL3-2-treetagger.txt Fichier issu de treetagger dormeurL3-etiquette.txt Fichier issu de treetagger reconstruisant la trame initiale ${\tt dormeurL3-trame-treetagger.xml}$ Dictionnaire des lemmes dormeurL3-lemme.dic Dictionnaire des categories dormeurL3-categorie.dic Eléments du rapport : e1. Contenu de la partie STRUCTURE(10,291) e2. Contenu de section 1 e3. Ventilation (sélections/forme) : [sélection] sur [LIGNE] e4. Concordance (forme) : Sélection/Positions

Figure 246 : Index du rapport enregistré

Chaque élément du rapport contient un menu de navigation (à gauche) pour accéder à tous les résultats construits (les différents éléments du rapport) et (à droite) de la description de léélément visé.

Cette description doun élément du rapport (sur la droite) est constituée :

- o døun *corps* (le contenu de lø*élément du rapport* pour les fichiers résultats ou une liste des *éléments du rapport* pour le fichier døindex)
- o døune *note* : zone de texte libre renseignée par løutilisateur

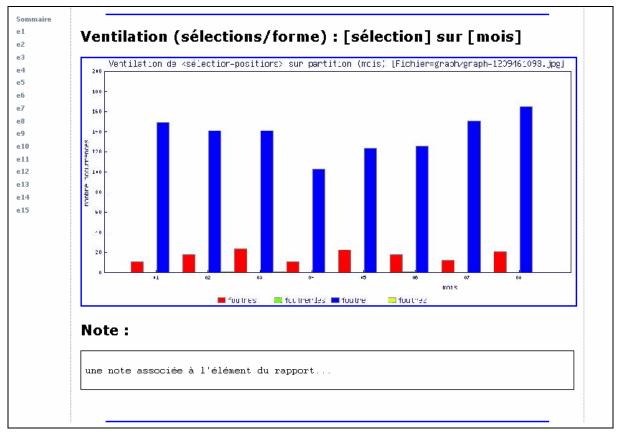


Figure 247 : Elément du rapport enregistré

On trouvera un exemple complet de rapport sur la page web du *Trameur*.

30 Fonctionnalités complémentaires

30.1 Graphes døannotation

30.1.1 Intégration dans le *Trameur* døun module dømportation døun graphe døannotations sur la *Trame*

(cf « Perspective fonctionnelle de la phrase : la relation thème-rhème » F. Daoust 2009)

Illustration sur l\u00exemple de Fran\u00e7ois :

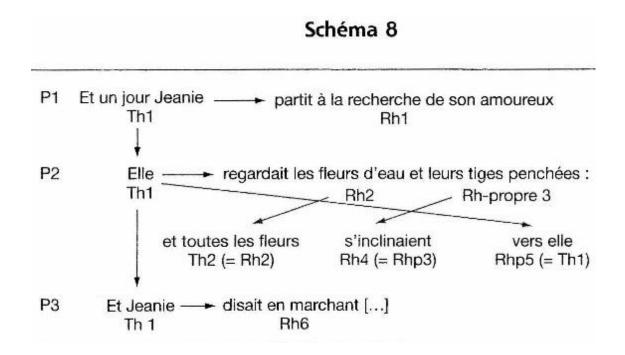
Et un jour Jeanie partit à la recherche de son amoureux. Elle regardait les fleurs d'eau et leurs tiges penchées : et toutes les fleurs s'inclinaient vers elle. Et Jeanie disait en marchant ...

Les données de test utilisées ici (fichiers texte de travail et annotation+graphe) sont disponibles dans le dossier textes du répertoire de travail du *Trameur* après installation.

Etape 1

Ce texte est chargé dans le *Trameur* (création døune base textométrique) : une *Trame* est créée, pas de *Cadre* explicite ici.

Le schéma suivant décrit le système døannotation et son graphe (cf Daoust 2009):



Le fichier suivant décrit un exemple de fichier à importer décrivant le système døannotation (positions sur la *Trame*) et le graphe associé :

```
<?xml version="1.0" encoding="iso-8859-1"?>
<TET>
    <teiHeader>
         <fileDesc>
             <titleStmt>
                 <title>Export Graphe Annotation sur la Trame d'un texte via Le Trameur
8.0b074</title>
             </titleStmt>
             <publicationStmt>
                 ...
             </publicationStmt>
             <sourceDesc>
                 <bib>> ... </bib>>
             </sourceDesc>
         </fileDesc>
         <encodingDesc>
             ...
         </encodingDesc>
    </teiHeader>
<t.ext.>
<body>
<graph type="directed" name="annotation-structurelle">
<node positions="1 3 5 7" labelnode="Th1"/>
<node positions="9_11_13_15_17_19_21" labelnode="Rh1"/>
<node positions="24" labelnode="Th1"/>
<node positions="26_28_30_32_35_37_39_41_43" labelnode="Rh2"/>
<node positions="37_39_41_43" labelnode="Rh-propre-3"/>
<node positions="47_49_51_53" labelnode="Th2"/>
<node positions="55_58" labelnode="Rh4"/>
<node positions="60_62" labelnode="Rhp5"/>
<node positions="65_67" labelnode="Th1"/>
<node positions="69_71_73" labelnode="Rh6"/>
<arc from="1_3_5_7" to="9_11_13_15_1</pre>
                              to="9_11_13_15_17_19_21"
                                                               labelfrom="Th1"
                                                                                      labelto="Rh1"
labelarc="dépendance à définir"/>
<arc from="1_3_5_7"
                         to="24"
                                   labelfrom="Th1" labelto="Th1" labelarc="dépendance
définir"/>
<arc from="24" to="65_67" labelfrom="Th1" labelto="Th1" labelarc="dépendance à définir"/>
         from="24"
                        to="26_28_30_32_35_37_39_41_43"
                                                               labelfrom="Th1"
                                                                                     labelto="Rh2"
<arc
labelarc="dépendance à définir"/>
<arc from="26_28_30_32_35_37_39_41_43"
                                              to="37 39 41 43" labelfrom="Rh2"
                                                                                      labelto="Rh-
propre-3" labelarc="dépendance à définir"/>
<arc from="26_28_30_32_35_37_39_41_43" to="47_49_51_53" labelfrom="Rh2"</pre>
                                                                                     labelto="Th2"
labelarc="dépendance à définir"/>
          from="37 39 41 43"
                                    to="55 58"
                                                      labelfrom="Rh-propre-3"
labelarc="dépendance à définir"/>
<arc from="24" to="60_62" labelfrom="Th1" labelto="Rhp5" labelarc="dépendance à définir"/>
<arc from="65_67" to="69_71_73" labelfrom="Th1" labelto="Rh6" labelarc="dépendance à</pre>
définir"/>
</graph>
</body>
</text>
</TEI>
```

- Chaque élément node décrit les positions visées de la *Trame* (via loattribut positions), et la valeur du label pour ce nò ud de positions (via loattribut labelnode). Quand le label du nò ud noest pas spécifié explicitement, une valeur par défaut lui est attribuée (les positions de ses constituants).
- Chaque élément arc décrit les liens entre les éléments node : løattribut from porte les positions de la *Trame* du nò ud source de løarc, løattribut to porte les positions de la *Trame* du nò ud cible de løarc, løattribut labelfrom porte la valeur du label du nò ud source, løattribut labelto porte la valeur du label du nò ud cible et løattribut labelarc porte la valeur du label associé à løarc.

CLESTHIA>SYLED>CLA²T

Les positions de la *Trame* sont désignées en les listant une à une et en respectant ce format : positions="26 27 35", ici 3 positions discontinues de la *Trame* sont désignées.

Etape 2 : importation de løannotation et du graphe associé dans Le Trameur

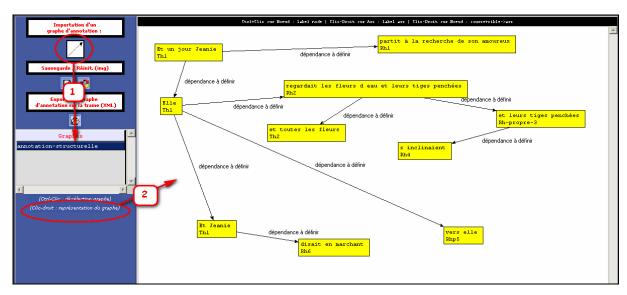


Figure 248: Importation d'annotation+graphe

Une fois chargée, le graphe est disponible en mémoire dans le *Trameur* (1). On peut aussi visualiser le graphe de cette annotation (2).

30.1.1 Génération de graphes d@annotations

Nous allons montrer ci-dessous comment générer directement dans le *Trameur* un tel graphe en reconstruisant le graphe vu dans la partie précédente.

Etape 1

Le texte précédent est chargé dans le *Trameur* (création døune base textométrique) : une *Trame* est créée, pas de *Cadre* explicite ici.

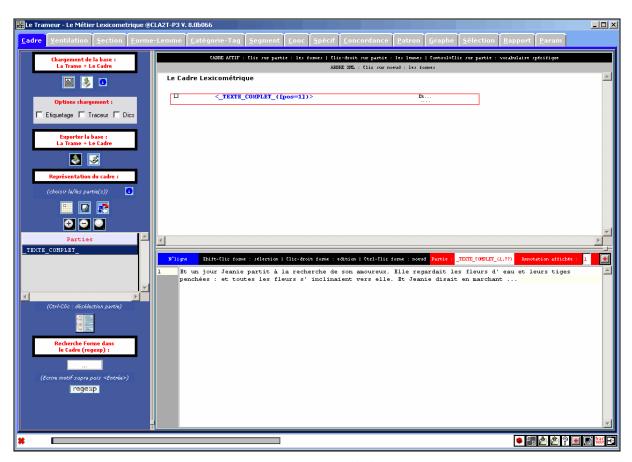


Figure 249 : Chargement de la base, accès à la *Trame* par le *Cadre*

Løéditeur de la *Trame* (accès par le *Cadre* ici), permet de définir des nò uds : le raccourci clavier *Control-Clic* sur un item de la *Trame* permet de le sélectionner, dans løexemple suivant, les 4 premiers items sont sélectionnés, ils vont être utilisés pour définir le premier nò ud du graphe.



Figure 250 : Sélection des items constituant un nò ud du graphe

On peut ensuite générer le nò ud (nœud1) (double-clic dans lééditeur) et le donner à voir dans léonglet **GRAPHE** :



Figure 251: Construction du nò ud

Control-Clic sur le nò ud permet de lui associer un label (valeur qui sera associée à chaque item constituant le nò ud défini pour løannotation en cours de construction):

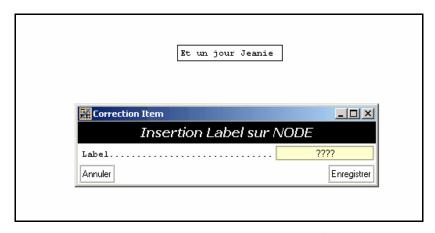


Figure 252 : Insertion d'un label sur le nò ud



Figure 253: Le nò ud a un label

On peut ensuite réitérer løpération de création døun nò ud (nœud2):

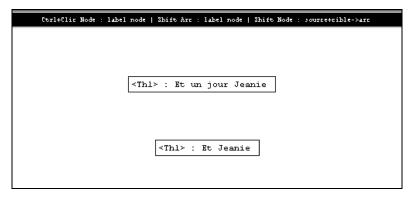


Figure 254 : De nouveaux nò udsí

Pour créer un arc entre nœud1 et nœud2, on sélectionne le nò ud source (*clic-droit* sur nœud1 qui devient rouge) puis on sélectionne le nò ud cible (*clic-droit* sur nœud2 qui devient bleu). Løarc est ensuite généré :

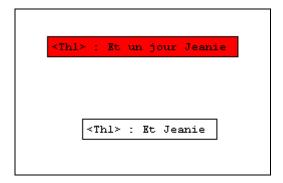


Figure 255 : La source d'un arc

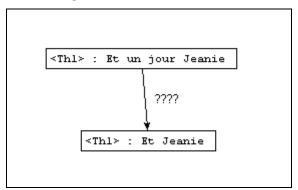


Figure 256 : La cible de l'arc... l'arc est défini

Control-Clic sur løarc permet de lui associer un label :

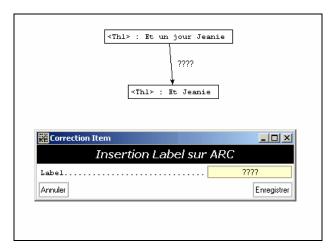


Figure 257: Ajout d'un label sur l'arc

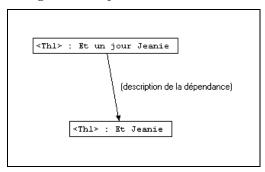


Figure 258: L'arc a son label

Une fois le graphe complet construit,

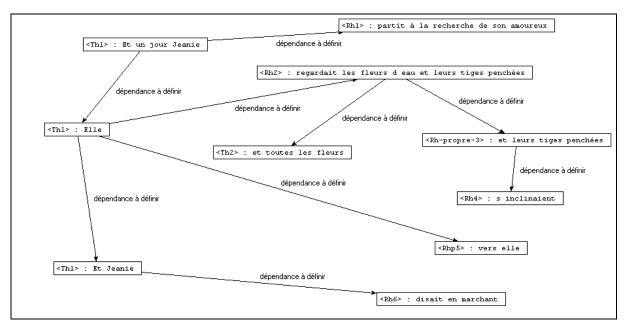


Figure 259: Graphe complet construit

il est possible de le « générer », ce processus permet de créer un fichier dont le format est similaire à celui présenté *supra* :

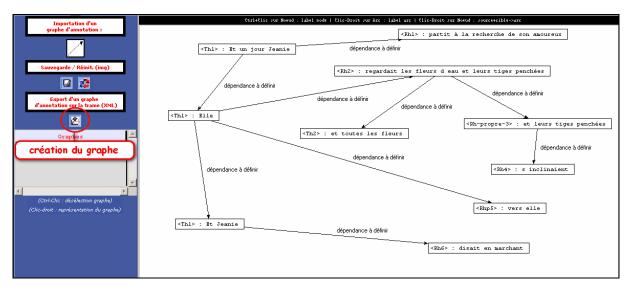


Figure 260 : Sauvegarde du graphe

Ce fichier est ensuite ré-importable par le processus décrit dans la partie 1.

30.1.3 Exploitation des graphes d@annotation

Chantier en cours.

30.2 Base textométrique de textes alignés

30.2.1 Mise au jour døalignement dans la carte des sections

Objectif: Mise au jour dœun alignement par sections dans la représentation graphique dœun corpus aligné via la *Carte des sections*

Corpus exemple : plusieurs traductions alignées du discours dønvestiture de Barak Obama en 2009.

- **volet-en**: la transcription parue sur le site du New York Times http://www.nytimes.com/2009/01/20/us/politics/20text-obama.html
- volet -fr-1 : la traduction du discours mise en ligne sur le site du journal Le Monde (URL non accessible hors abonnement)
 - volet -fr-2 : la traduction du discours mise en ligne sur le site du journal Libération (AFP)

http://www.liberation.fr/monde/0101313402-point-par-point-le-discours-d-investiture-debarack-obama-nouveau-president-des-etats-unis

- **volet -fr-3** : la traduction du discours mise en ligne sur le site de RFI http://www.rfi.fr/actufr/articles/109/article_77584.asp
- \bullet volet -fr-4 : la traduction du discours via Google de volet-en <u>http://translate.google.fr/</u>

Ces 5 volets sont disponibles dans un fichier au format TMX.

On présente ci-dessous une illustration des différents paramétrages donfichage de la carte des sections sur un corpus exemple décrit ci-dessus.

Méthode n°1: Lecture du fichier alignement au format TMX

- Le *Trameur* peut lire directement le fichier au format TMX contenant les 5 volets
- Une fois chargé, on peut visualiser cet alignement *via* la carte des sections en choisissant comme délimiteur de partie l'élément **tu** (bloc de traduction contenant les différents volets de la même zone textuelle (phrases par exemple))
- L'alignement à gauche des sections permet de visualiser globalement tous les blocs et leurs différents volets (par section)

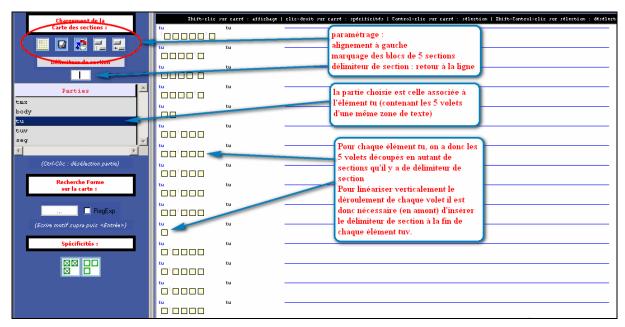


Figure 261: Alignement, partition tu (1)

Le marquage de blocs de sections est ici maintenu.

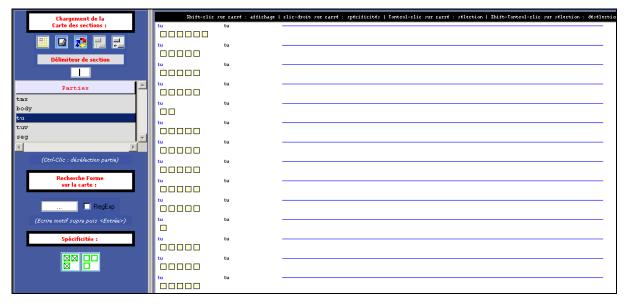


Figure 262 : Alignement, partition tu (2)

Le marquage de blocs de sections est ici inhibé.

Inconvénient : on nøa pas la vue globale permettant de contraster les 4 volets

Méthode n°2 : Lecture du fichier alignement au format Lexico/*Trameur*

- Reconstruction des 4 volets :
 - A partir du TMX précédent, chargement du TMX dans *mkAlign*²³
 - o sauvegarde de chacun des volets (format txt) : les éléments tuv du fichier TMX étant potentiellement une section
- Concaténation des 4 volets dans un même fichier et insertion de balise : <volet=en>...<volet=fr1> etc.
- Chargement dans le *Trameur* + carte des sections + alignement à gauche

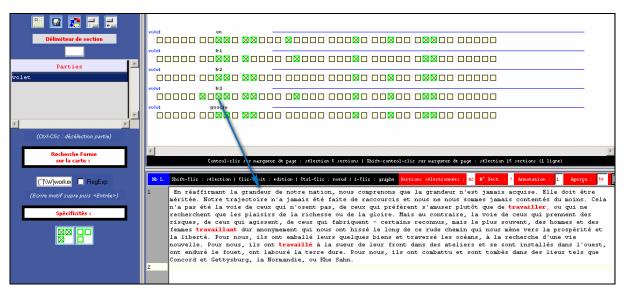


Figure 263: Alignement, partition volet (1)

Marquage des blocs de sections (projection de la famille de mots « travail / work » sur la carte des sections).

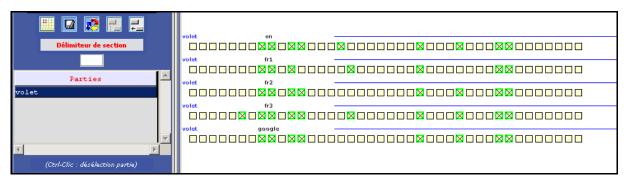


Figure 264: Alignement, partition volet (2)

Inhibition du marquage des blocs de sections.

Ici, les 4 volets alignés sont visualisables globalement (par section (tuv dans le TMX de départ))

230

²³ http://tal.univ-paris3.fr/mkAlign/

30.2.2 Exploration døune base textométrique de textes alignés

Cette section présente les fonctionnalités permettant de dœxplorer une base intégrant 2 (ou plusieurs) textes alignés. Cette mise à jour reprend une fonctionnalité déjà disponible dans $mkAlign^{24}$ permettant notamment de mettre à jour des équivalents traductionnels dans des textes alignés par un calcul de cooccurrence réalisé en parallèle sur 2 volets de løalignement traité.

30.2.2.1 Chargement døune nouvelle base intégrant 2 textes bilingues alignés

Pour illustrer cette fonctionnalité, nous commençons par choisir un exemple de texte bilingue à charger par la suite dans le *Trameur*. Nous travaillerons ici sur un texte de convention européenne disponible en français et en anglais, ces 2 volets étant alignés (ces 2 fichiers sont disponibles dans le programme *mkAlign*).

```
Expertis = compo-
convention de sawvegarde des droits de l'homme et des libertés fondamentales
calle qu'amendée par le protocole n° 11
come, 4.xt.1950
les gouvernements signataires, membres du conseil de l'europe,
considérant que cette déclaration tend à assurer la reconnaissance et l'application universelles et effectives des droits qu'i y sont énoncés;
considérant que cette déclaration tend à assurer la reconnaissance et l'application universelles et effectives des droits qu'i y sont énoncés;
considérant que cette déclaration tend à assurer la reconnaissance et l'application universelles et effectives des droits qu'i y sont énoncés;
considérant que cette déclaration tend à assurer la reconnaissance et l'application universelles et effectives des droits qu'i y sont énoncés;
considérant que cette déclaration tend à assurer la reconnaissance et l'application universelle es et effectives des droits de l'homme dont la se internation plus étroite entre ses membres et que l'un des sevons d'attent en concession et de la paix dans le monde et dont le maintien repose easentiellement
seu un régine politique véritablement démocratique, d'une part, et, d'autre part, sur une conception commun d'idéal et de traditions politiques, de respect de la liberté et de préémi inence du dort, à prendre les pressières neurres propres à assurer la garantie collective de certains des droits énoncés dans la déclaration universelle,
sont convenus de ce qui suit;

virticle 1 – obligation de respector les droits de l'homme 1
les houtes parties contractances reconnaissent à toute personne relevant de leur juridiction les droits et libertés définis au titre i de la présente convention:
titre 1 – droits à l'ibertés 1
le droit de toute personne à la vie 1
le droit de toute personne à la vie 1
le droit de toute personne à la vie est puni de cette peine par la loi. la mort me peut être infligée à quiconque intentionnellement, sauf en exécution d'une sentence capitale prononcée par un tris
sunai au cas où le délit est puni de cette peine par la lo
```

Figure 265 : fichier 1, volet français (extrait)

Figure 266 : fichier 1, volet anglais (extrait)

Les 2 textes sont structurés en partie (via un système de balisage). Løalignement est matérialisé dans les 2 volets par le caractère # qui induit les alignements respectifs des différents paragraphes.

-

²⁴ http://www.tal.univ-paris3.fr/mkAlign/

Dans *mkAlign*, les 2 fichiers sont chargés un à un (le volet FR et le volet EN) dans les 2 volets de l'alignement.



Figure 267: Edition d'un alignement dans mkAlign

Pour le *Trameur*, la base textométrique est construite ici par concaténation des 2 fichiers en un seul fichier, en intégrant en outre un marqueur de partie supplémentaire pour contraster les 2 volets (<volet="FR"> et <volet="EN"> au début de chacun des 2 volets).

```
Consideration of assuregated des decits de l'homme et des libertés fondamentales et les quivements par le protocole n'il comme, xi.1150 considerant la déclaration suns de nouvell de l'europe, considérant la déclaration tend à assurer la recommanissance et l'application universalle set office, vivale de decits qui vende la descate on universalle des fortice du l'homme, proclames par l'assemblée générale des nations unies le 10 décembre 1949; considérant que cette déclaration tend à assurer la recommanissance et l'application universalles et effectives de decits qui vende de course de l'application universalle set et l'application universalle set et des surprises de course qui vende de course qui vende de course de la paix dans le monde et dont et la savequaté et le développe considérant que les but du nouvel la descate du l'application universalles et décente de la paix dans le monde et dont et la savequaté et le développe considérant que les but de l'application d'une sentence de ce qui alle l'application de l'application de l'application de l'application de l'application de l'application d'application d'applicat
```

Figure 268 : Préparation d'une base textométrique de textes bilingues alignés

La figure précédente donne à voir le fichier construit en mettant au jour les 2 volets (2 zones différentes du même fichier : le volet FR suivi du volet EN). Une fois chargée dans le *Trameur* (le caractère # ayant été ajouté à la liste des délimiteurs), la *Carte des sections* a løallure suivante (avec le caractère # comme délimiteur de sections):

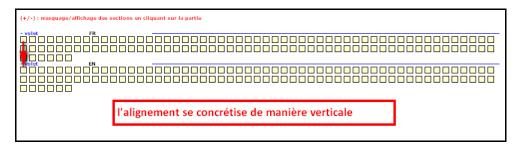


Figure 269 : Carte des sections d'un alignement dans le Trameur

On choisit d'afficher la partition VOLET pour contraster les 2 volets et on aligne à gauche les sections. Dans *mkAlign*, la *Carte des sections* de løalignement a løallure suivante :

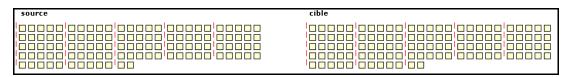


Figure 270: Carte des sections d'un alignement dans mkAlign

Løalignement se lit de manière horizontale. Contrairement à *mkAlign*, l'alignement produit *via* la *Carte des sections* dans le *Trameur* se lit de manière verticale ; dans la figure précédente, le premier carré FR a son "miroir" au « bout de la flèche »í Si on projette une forme sur la carte, on obtient ceci :

Dans mkAlign:



Figure 271: Projection d'une forme sur carte des sections dans mkAlign

Le mot « conseil » est projeté sur le volet FR (à gauche), la zone miroir est matérialisée dans le volet EN par des carrés cerclés de rouge.

Dans le *Trameur* :

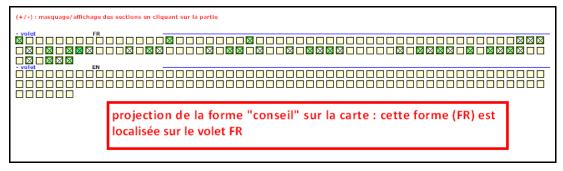


Figure 272: Projection d'une forme sur la carte des sections dans Le Trameur

Le mot « conseil » est projeté sur la carte, seul le volet FR (en haut) met au jour la présence de la forme visée dans les sections cochées, la zone miroir nœst pas directement matérialisée dans le volet EN.

Dans le *Trameur*, le calcul des spécificités des sections contenant une forme projetée tient compte du vocabulaire de l'ensemble du corpus (ici les 2 volets). Alors que dans *mkAlign*, ce calcul se fait sur le volet FR et en parallèle sur le volet EN (les 2 volets étant dissociés).

La nouvelle fonctionnalité introduite dans le *Trameur* consiste à "dissocier" les 2 volets de løalignement et à reproduire un calcul en parallèle sur les 2 volets (*i.e* comme dans *mkAlign*):

- Calcul des cooccurrences sur les sections contenant la forme projetée sur le volet FR
- En parallèle, calcul des cooccurrences sur les sections miroirs des précédentes

Pour cela, on dispose d'un nouveau menu dans l'onglet "Section" :

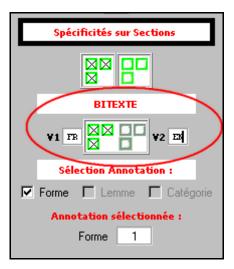


Figure 273: Paramétrage d'un bitexte

Dans la zone BITEXTE : il est possible de considérer la *Carte des sections* comme une concaténation de 2 volets alignés.

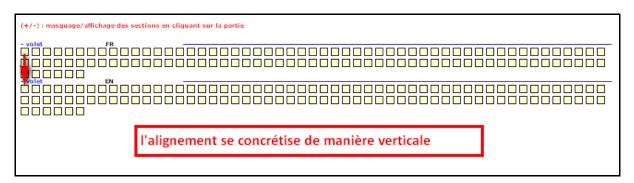


Figure 274: Mise au jour du bitexte via la carte des sections

Cøst le cas ci-dessus, pour la partition VOLET, on dispose bien døune partie VOLET=FR et døune partie VOLET=EN, les sections de chacun de ces volets étant alignées verticalement.

Pour amorcer le calcul, on paramètre donc les 2 volets à contraster : le volet V1 est celui contenant la forme projetée (ici FR est renseignée dans la zone de saisie V1) et le volet V2 est le miroir (ici EN dans la zone de saisie V2).

Le bouton du milieu permet de lancer le calcul qui produit ceci : (2 extraits : volet SOURCE et volet CIBLE)

```
Specificites du vocabulaire (Annotation Forme) sur les sections SOURCE contenant le motif : conseil
  Nombre d'occurrences du texte global = 11099
Nombre d'occurrences dans la partie visee = 5314
Seuil : 10
     ----- (cible infra)
(c)Die inital : (c)Die inital 
 conseil
                                                                                                                                                                                                                   26.4
 europe
                                                                                                                                                                                                                   19.6
                                                                                                                                                                                                                                                                                192
                                                                                                                                                                                                                                                                                                                                            152
                                                                                                                                                                                                                                                                               95
47
                                                                                                                                                                                                                                                                                                                                            85
47
protocole
                                                                                                                                                                                                                   18.1
secrétaire
déclaration
                                                                                                                                                                                                                   16.1
                                                                                                                                                                                                                   15.0
                                                                                                                                                                                                                                                                                49
                                                                                                                                                                                                                                                                                                                                             48
  ratification
                                                                                                                                                                                                                                                                               231
50
32
                                                                                                                                                                                                                                                                                                                                             163
                                                                                                                                                                                                                    12.9
                                                                                                                                                                                                                                                                                                                                            47
31
général
                                                                                                                                                                                                                   12.6
  membres
                                                                                                                                                                                                                   9.7
8.9
présent
                                                                                                                                                                                                                                                                                83
                                                                                                                                                                                                                                                                                                                                             65
vigueur
                                                                                                                                                                                                                                                                                                                                            33
23
22
21
                                                                                                                                                                                                                                                                                23
suit
acceptation
approbation
                                                                                                                                                                                                                                                                                22
                                                                                                                                                                                                                   7.6
7.4
7.3
                                                                                                                                                                                                                                                                               25
32
91
signature
mois
                                                                                                                                                                                                                                                                                                                                           24
29
67
19
qui
                                                                                                                                                                                                                                                                                19
19
   erritoires
                                                                                                                                                                                                                    7.1
                                                                                                                                                                                                                                                                                18
                                                                                                                                                                                                                                                                                                                                             18
 entrera
                                                                                                                                                                                                                    6.8
                                                                                                                                                                                                                                                                                18
17
  jour
    out
                                                                                                                                                                                                                    6 1
                                                                                                                                                                                                                                                                                67
                                                                                                                                                                                                                                                                                                                                             50
    otification
                                                                                                                                                                                                                                                                                                                                             24
  date
```

Figure 275 : Spécificités volet SOURCE de l'alignement

	CIBLE		
Specificites du vocabulaire (Annotati	on Forme) sur les sections CIBL	E miroir de cell	es contenant le motif : conseil
Nombre d'occurrences du texte global	= 10453		
Nombre d'occurrences dans la partie v	risee = 5047		
Seuil : 10			
(Specificites positives en haut de li	ste, negatives en bas)		
Forme	Ind-Specif	Fq-Totale	Fq-Partie
council	25.2	76	76
europe	24.8	75	75
secretary	15.9	47	47
protocol	15.7	87	77
declaration	15.6	46	46
ratification	13.3	44	43
general	10.7	55	49
member	9.3	31	30
day	8.6	24	24
signature	8.6	24	24
into	8.4	28	27
of	8.1	873	496
acceptance	8.0	22	22
approval	7.6	21	21
date	7.3	32	29
at	6.7	54	43
first	6.3	21	20
signed	6.1	16	16
month	6.1	16	16
notification	6.1	16	16
force	6.1	31	27
territories	6.1	16	16
being	5.7	26	23
deposit	5.7	15	15
territory	5.1	33	27
enter	5.1	17	16
after	4.9	26	22
expiration	4.8	12	12
signatory	4.8	12	12
instrument	4.8	16	15
deposited	4.5	15	14
fundamental	4.4	18	16
rome	4.2	10	10
instruments	4.2	10	10
receipt	4.2	10	10
terms	4.2	10	10
addressed	4.2	10	10
this	4.0	118	74

Figure 276 : Spécificités volet CIBLE de l'alignement

Le calcul a produit le calcul des cooccurrences sur les sections du volet FR contenant la forme projetée (le mot « conseil »), *i.e* le volet SOURCE ; en parallèle ce calcul a aussi été réalisé sur le volet CIBLE : les sections miroir des précédentes dans lesquelles on trouve par

exemple lœ́quivalent traductionnel de « conseil » mis au jour par le calcul. C'est à dire la même chose que *via mkAlign* :

Specificites du vocabulaire sur les section	ns SOURCE contenant le m	otif : <consei< th=""><th>I></th><th>Specificites du vocabulaire sur les sections</th><th>CIBLE associees aux se</th><th>ections SOURCI</th><th>E contenant</th></consei<>	I>	Specificites du vocabulaire sur les sections	CIBLE associees aux se	ections SOURCI	E contenant
Nombre d'occurrences du texte global = 110;	99		-	Nombre d'occurrences du texte global = 10453			
Nombre d'occurrences du teste gibbal - 110.				Nombre d'occurrences dans la partie visee = !	047		
Seuil : 10	- 5514			Seuil : 10	,047		
(Specificites positives en haut de liste,	negatives on hes)			(Specificites positives en haut de liste, ne	vativas en has)		
Le fichier construit : C:/SERGEF~1/SF-WORK		OMIZEN 1 /CONTZEN	-1/FD/ownert	Le fichier construit : C:/SERGEF~1/SF-WORK/O		OMITEN 1 /CONTENN.	.1/FD/ownert
Be lichief consciute : c./BERGEF-1/SF-WORK	/0/-DEVEN/INABIO-1/COIPUS/C	ONVEN-1/CONVEN	-I/FK/Exporc	he fichief consciute : c./bakbar-1/5r work/o	/ DEVEN/INCREIS-I/COIPUS/C	ONVEN-I/CONVEN-	1/FK/exporc
Forme	Ind-Specif	Fq-Totale	Fq-Part	Forme	Ind-Specif	Fq-Totale	Fq-Part
conseil	26.4	79	79	council	25.2	76	76
europe	26.1	78	78	europe	24.8	75	75
du	19.6	192	152	secretary	15.9	47	47
protocole	18.1	95	85	protocol	15.7	87	77
secrétaire	16.1	47	47	declaration	15.6	46	46
déclaration	15.0	49	48	ratification	13.3	44	43
ratification	13.4	44	43	general	10.7	55	49
le	12.9	231	163	member	9.3	31	30
général	12.6	50	47	day	8.6	24	24
membres	9.7	32	31	signature	8.6	24	24
présent	8.9	83	65	into	8.4	28	27
vigueur	8.5	36	33	of	8.1	873	496
suit	8.4	23	23	acceptance	8.0	22	22
acceptation	8.0	22	22	approval	7.6	21	21
approbation	7.7	21	21	date	7.3	32	29
signature	7.6	25	24	at	6.7	54	43
mois	7.4	32	29	first	6.3	21	20
qui	7.3	91	67	force	6.1	31	27
territoires	7.1	19	19	signed	6.1	16	16
premier	7.1	19	19	month	6.1	16	16
entrera	6.8	18	18	notification	6.1	16	16
n"	6.8	18	18	territories	6.1	16	16
jour	6.4	17	17	deposit	5.7	15	15
tout	6.1	67	50	being	5.7	26	23

Figure 277 : Spécificités des volets SOURCE et CIBLE via mkAlign

30.2.2.2 Chargement døune nouvelle base intégrant 2 textes alignés

Le même type de calcul est possible sur une base intégrant 2 textes alignés døune même langue, par exemple deux traductions alignées døun même texte. Nous considérons ici 2 traductions²⁵ de løouvrage døHemingway « Le vieil homme et la mer » : celle de J. Dutourd (1952) et celle de F. Bon (2012).

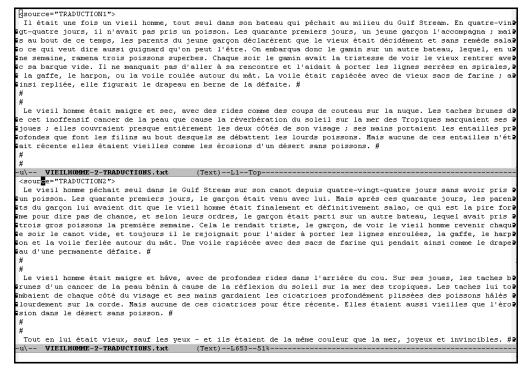


Figure 278 : Base textométrique de 2 traductions alignées

Comme précédemment, une base regroupant les 2 traductions est obtenue en concaténant les 2 fichiers. La figure précédente donne à voir le début des 2 volets dans la base ainsi constituée. Une fois le fichier chargé, la *Carte des sections* a cette allure :

_

²⁵ http://www.tal.univ-paris3.fr/mkAlign/#p7b

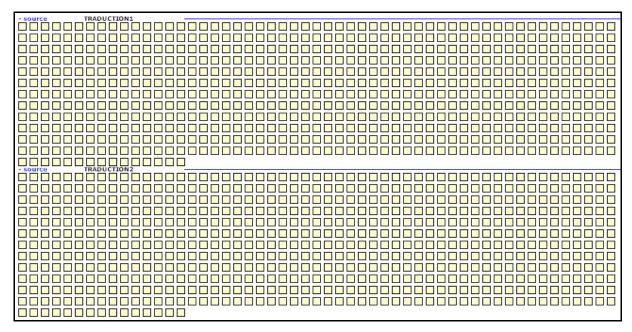


Figure 279 : Carte des sections d'un alignement de traductions dans le Trameur

Le volet «TRADUCTION1 » correspond à la traduction de Dutourd, le volet «TRADUCTION2 » à celle de Bon. La projection de la forme « vieil » sur la carte produit cette ventilation :

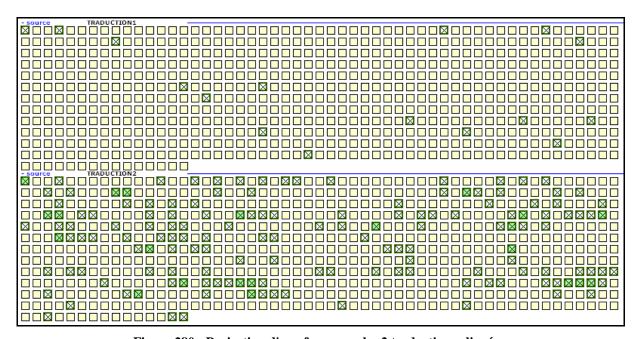


Figure 280 : Projection d'une forme sur les 2 traductions alignées

Cette forme est présente dans les 2 volets et massivement dans le second. Le calcul des cooccurrents présenté *supra* se fait en considérant les 2 volets comme 2 parties disjointes. Ici, la partie TRADUCTION2 est considérée comme le volet SOURCE et la partie TRADUCTION1 comme le volet CIBLE. Le résultat produit est le suivant :

CLESTHIA>SYLED>CLA²T

Specificites du vocabulaire (Annotat	ion Forme) sur les sections SC	OURCE contena	nt le motif : vieil
Nombre d'occurrences du texte global =	. 20257		
Nombre d'occurrences du texte global - Nombre d'occurrences dans la partie vi			
Seuil : 10	.see - 10191		
seull . 10	SUIRCE		
	550162		
(Specificites positives en haut de lis			
Le fichier construit :	oc, negacites en bas,		
C:/DOCUME~1/FLEURY~1/MESDOC~1/MYDROP~1	/WIP/WIP-TR-1/ALIGNE-1/ALIGNE-	l/export/results	necif-source-1414521180.tx
	.,,	-, -m p ,	F
Forme	Ind-Specif	Fq-Totale	Fq-Partie
vieil	**	185	185
honne	**	211	189
Le	11.6	125	80
le	8.8	1005	434
requin	6.9	40	29
sur	4.7	175	84
ses	4.6	106	55
lame	4.3	10	9
était	4.3	251	113
étaient	4.0	40	24
yeux	3.9	28	18
nageoires	3.9	9	8
avaient	3.9	32	20
avait	3.8	177	81
parlaient	3.7	6	6
frappa	3.6	13	10
dents	3.5	8	7
continuait	3.3	10	8
Parfois	3.3	5	5
harpon	3.2	20	13
ainsi	3.1	7	6
surface	3.1	29	17
et	3.0	989	379
leurs	3.0	19	12
son	3.0	157	69
poissons	3.0	34	19
eau	2.9	73	35
chair	2.9	15	10
salée	2.8	4	4
élevait	2.8	4	4
cou	2.8	4	4

Figure 281 : Spécificités sur volet SOURCE (TRADUCTION2) : sections contenant le motif "vieil"

Specificites du vocabulaire (Annotat	CIBLE	miroir de cell	es contenant le motif : viei
Nombre d'occurrences du texte globa:	1 = 29982		
Nombre d'occurrences dans la partie			
Seuil : 10			
(Specificites positives en haut de :	liste, negatives en bas)		
Forme	Ind-Specif	Fq-Totale	Fq-Partie
zieux	32.8	269	186
Le	11.3	181	104
requin	7.7	35	27
rieil	7.0	16	15
le	5.6	889	358
avait	4.9	163	78
5a	4.7	123	61
la	4.5	826	326
iu	4.1	209	93
lame	3.8	6	6
L	3.7	525	209
surface	3.7	15	11
eau	3.7	80	40
narpon	3.7	19	13
ses	3.6	98	47
étaient	3.6	43	24
Les	3.5	46	25
frappa	3.5	8	7
jaillissaient	3.4	5	5
chair	3.4	12	9
grosses	3.4	5	5
était	3.2	251	104
plusieurs	3.1	7	6
gueule	3.0	9	7
nâchoires	3.0	9	7
aperçut	3.0	17	11
:ôté	3.0	17	11
rit	3.0	11	8
laissait	2.9	4	4
faux	2.9	4	4
enfonça	2.9	4	4
ooissons	2.9	33	18
pièce	2.9	4	4

Figure 282 : Spécificités sur volet CIBLE (TRADUCTION1) : sections MIROIR des sections contenant le motif "vieil"

Dans la traduction de F. Bon, on trouve plutôt la séquence « le vieil homme » et dans celle de J. Dutourd, la séquence « le vieux ».

30.2.2.3 Base textométrique : alignement au format TMX

Un alignement au format TMX permet lui aussi de réaliser un calcul contrastif de cooccurrences sur 2 des volets intégrés dans ce type d'alignement. Nous prendrons ici comme exemple de fichier TMX, un alignement intégrant 4 volets de l'ouvrage õLe vieil homme et la merö : le volet anglais original (noté en), les 2 volets traduits en français déjà mentionnés (noté fr1 et fr2) et un volet traduit en espagnol (noté es).

Le fichier a l'allure suivante (affichage partiel du fichier dans un navigateur, puis dans un éditeur (code source)):

Volet	Anglais	Français (Dutourd)	Français (Bon)	Espagnol
1	He was an old man who fished alone in a shiff in the Golf Stream and he had gone sighty-four days now without taking a fish. In the first forty days a Boy had been with him. But after forty days without a fish the Boy's passets had took he that the hold man was now definitely and finally maken, which is the worst form of whichly, and the boy had gone as their souther four which in caught these gone fish the first week. It made the boy mad to me at their onders in another boat which mamping and the single with Gone to halp find charp with the solid man come in each day with his skiff mappy and he always with Gone to halp find charp without the solid lively with Gone to halp find the bays with the solid lively with Gone to halp find the part was promoted to the grift and happens and the sail that single or the grift and happens and the sail that find or permanent defeat.	Il statt une fois un visit homme, tout seul dans son hatane qui păthait su milies de Guif Stream. En quatem-vingu-quates [Jone, 1] in Fesati pas puis un poisson. Les quaestie pestieses [Jone, un jeune gasona l'accompagna mais au hout de ce tempe, les pasents de jeune garon déclarément que la visue statt décidement et sans estedé salor se qui vest dies aussi puipaned qu'en peut l'étre. On anhaequa donc le grain seur mutres battan, isquel, en une session, sement troit poissons arquebles. Châque sois le passi svuit la tractions de voir la videre entres au le passi svuit la tractions de voir la videre entres au le passi svuit la tractions de voir la videre entres au la passi svuit la tractions de voir la videre entres au la passi svuit la traction de voir la videre entres au la passi svuit la traction de voir la videre entres au la passi svuit la traction de visit la videre entre au la passi svuit la traction de visit la videre entre au la passi svuit la passi de visit la videre de la défait a.	ie vieil himme päthait seul dans is Guif Steam sur son cannot dépuis quater-vingt-quater jours ann avez près un poisson. Les quaemnts premises jours, le gardon était vant avez lei, Misi appés coa quaemnts possenses pour les passens de queson ini avaient dit que le vieil himme était finalement et définitiement maillo, es qui set le pies fereme pour dies pas de chance, et main leurs ordens, le garcon était pareit sons pour dies pas de chance, et main leurs ordens, le garcon était pareit seu en notes bienes, léqui aveil pois totais gres poissons in première semaine. Cala le residist teitre. le garcon de voir le vieils pois et l'estait pois contra pareit de la companie de la c	The un visjo que pasable solo en un hote en al Oulf Stream y haste coheart y castro dies que no cogis un per. Thi los prisenes cuestrat diala habit traide consigo a un mochacho. Paro despois de cuarents dias habit traide consigo a un mochacho. Paro despois de cuarents dias in habit pessado les padere del mochacho iz habitan dicho que al viejo attaba definitiva y mestadamente sala, lo cost ser la porte forme de la cala saria, y por orden de seu paces el mochacho habita salido, non coto bote que cognit ches baccos pecca le primera samma. Entraitecta al mochacho ver al viejo especia color de con porte per per per per per per per per per pe
2	The old man was thin and gendt with deep weinkies in the back of his mack. The brown bictobes of the bone-client kin cancer the san beings form its efficacion on the teppic mas were on his sheaks. The blotches are well down the sides of his trees and his hands had the deep-ceased state from haddling heavy fish on the cords. But now of these states were fresh. They were as old as sections in a finhless desert.	Le viail homme stait maigne et set, avec des rides comme des crupé de conteas ser la nuya. Les tabhas brusas de crisoffessif camere de la pase, que couse la sévendestion de soluil ser la mes des Tropiques marquiant les joues ; alles coveraints praèque ettissement les Joses ; alles coveraints praèque ettissement les deux oftés de sen visage ; ser mains portainent les entailles profindes que font les faits en la centificie profindes que font les faits en la centificie profindes que font les faits en la centificie de des la centificie de la centificie de la comme de la centificie de la centificie de la centificie de d'un désert sans poissons.	in visil homes disit majors at hive, avec de porfondes vides dans l'excises de con. Sur ses jouas, les Indeha Brimes fort manner de la pasa bésin a cause de la sefficient de soluti sur la mer des treptques. Des trabes la tombalent de chequi côté de visage et ses mains gérélant les cincipies perfondement pissees des gérélant les cincipies perfondement pissees des persons hills lechément sur la soude. Mais second de ces cientices pour tre second. Mais second de ces cientices pour tre second des la désert seus poisson.	El visjo sea fiaco y despañado, con arrugas porfindas en la parte posterior del ceallo. Las paradas nacionad del hosigo colores de la pial que el sol prodose con sus esfisjos en al mar tropical establen en sos megillas. Essa porca correlam por los indes de se case hasta hastates ablo y ses manos todado las hondas cicircoles que como in- menta de la como de la como como como la paradece pocas. Es como como controles se sectiones como ten visigas como las erestoras de un acido desianto.

Figure 283 : Alignement TMX dans un navigateur (avec feuille de styles)

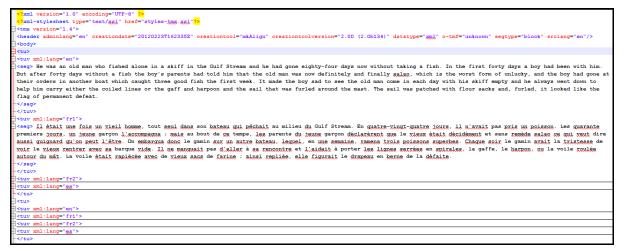


Figure 284 : Alignement TMX (code source)

La figure précédente donne à voir 2 zones de texte avec pour chacune les 4 volets de langue (dans un ordre identique : en, fr1, fr2, es). Pour la première zone de texte dans le code source, seules les 2 premières langues sont affichées dans lœditeur.

Un alignement TMX peut être schématisé ainsi :

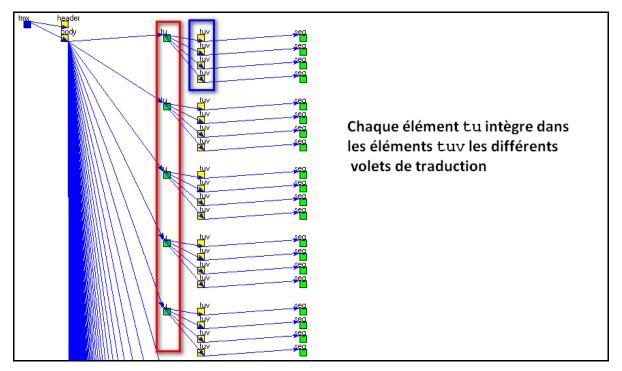


Figure 285 : Alignement TMX, arbre des éléments

Chaque élément tu de l'arbre XML encapsule les différents volets de traduction dans autant d'éléments tuv. L'alignement se lit de manière verticale pour chaque unité de langue (tu) dans la figure précédente.

Pour contraster 2 volets, il faut donc pouvoir mettre au jour les 2 parties visées : pour la source, tous les éléments tuv d'une langue donnée contenant la forme visée ; pour la cible, les éléments tuv miroirs des précédents dans une autre langue.

Dans le *Trameur*, la *Carte des sections* peut mettre au jour un alignement des différents volets d'un fichier TMX :



Figure 286 : Carte des sections d'un alignement TMX avec le Trameur

Ici la *Carte des sections* est générée à partir de la partition tuv *i.e* l'ensemble des parties tuv de la base textométrique construite à partir de l'alignement TMX. Ces parties intégrées dans le *Cadre* de la base textométrique sont identifiables par leurs positions sur la *Trame*: positions de début et de fin. Dans la carte construite, chaque colonne de sections coïncide avec un volet de l'alignement (1 colonne = 1 langue).

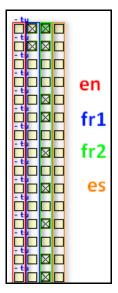


Figure 287 : 1 colonne = 1 volet = 1 langue

Identifier une colonne consiste donc à spécifier le volet visé *i.e* la langue des sections tuv de cette colonne. Les langues sont identifiables par les attributs lang des balises tuv (visibles dans le fichier source *supra*). La *Carte des sections* précédente est construite sur la partition tuv, les volets visés doivent être identifiés par la valeur de la langue du volet source et celle du volet cible. Au final, identifier 2 volets distincts consiste à donner un couple de langues parmi celles encodées dans les balises tuv du fichier TMX. Une section source donnée ayant une section cible miroir dans un voisinage de sections de longueur égal au nombre de volets de l'alignement (*i.e* le nombre de balises tuv dans les éléments tu) et les positions respectives d'une section source et d'une section cible restant inchangées tout au long du fichier TMX. Au final, les couples de sections alignées sont aisément identifiables.

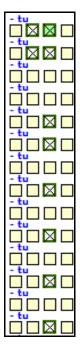


Figure 288 : Distribution d'une forme sur la carte de l'alignement

Une forme étant projetée sur la carte des sections, le calcul de ses cooccurrents consiste à spécifier les 2 volets à contraster. Si on projette la forme õvieilö, a priori, seuls les volets fr2

CLESTHIA>SYLED>CLA²T

et fr1 contiennent des sections avec cette forme (colonne 2 et 3 de la carte), la figure précédente donne à voir un extrait de cette distribution.

Le paramétrage du calcul des cooccurrents consiste à définir la valeur de 2 langues à contraster. Nous allons examiner ci-dessous les 2 couples de volets (fr2,fr1) et (fr2,en).

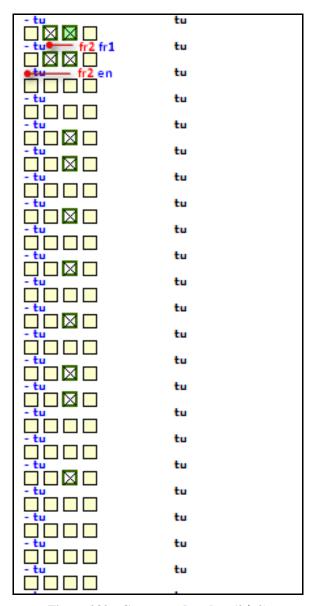


Figure 289 : Contraste de volets (2 à 2)

Pour le couple (fr2, fr1), le paramétrage du calcul est visible ci-dessous : la case à cocher TMX permet de préciser que la base chargée est un alignement au format TMX, les 2 zones de saisie V1 et V2 permettent de spécifier les langues à contraster (attributs lang des éléments tuv), ici :

```
V1 <= xml:lang=fr2 (SOURCE)
V2 <= xml:lang=fr1 (CIBLE)
```

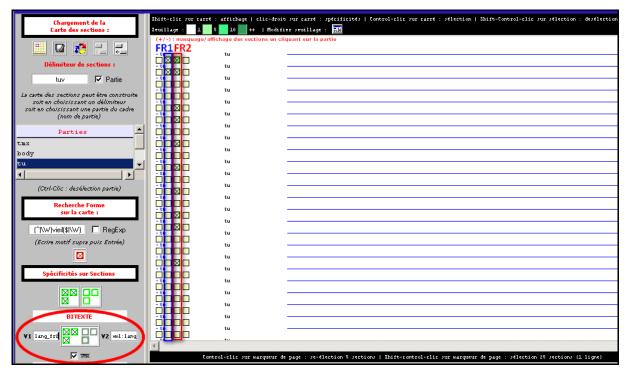


Figure 290 : Contraste de 2 volets (FR2, FR1)

Et le résultat du calcul a løallure suivante :

Nombre d'occurrences du texte globs	1 = 29257		
Nombre d'occurrences dans la partie			
Seuil : 10			
	SOURCE		
	(cible infra)		
(Specificites positives en haut de	liste, negatives en bas)		
Le fichier construit :			
C:/DOCUME~1/FLEURY~1/MESDOC~1/MYDRO	P~1/WIP/WIP-TR~1/ALIGNE~1/ALIGNE~	l/export/results	pecif-source-1415131422.tx
forme	Ind-Specif	Fg-Totale	Fq-Partie
vieil	**	185	185
nomme	**	211	189
Le	11.6	125	80
le	8.8	1005	434
requin	6.9	40	29
sur	4.7	175	84
es	4.6	106	55
lame	4.3	10	9
était	4.3	251	113
étaient	4.0	40	24
reux .	3.9	28	18
nageoires	3.9	9	8
avaient	3.9	32	20
avait	3.8	177	81
parlaient	3.7	6	6
frappa	3.6	13	10
dents	3.5	8	7
continuait	3.3	10	8
Parfois	3.3	5	5
narpon	3.2	20	13
ainsi	3.1	7	6
surface	3.1	29	17
et	3.0	989	379
leurs	3.0	19	12
son	3.0	157	69
ooissons	3.0	34	19
eau	2.9	73	35
chair	2.9	15	10
salée	2.8	4	4
élevait	2.8	4	4

Figure 291 : Spécificités volet SOURCE (FR2)

CLESTHIA>SYLED>CLA²T

	CTRLE		
Specificites du vocabulaire (Annotati	on Forme) sur les sections CIBL	R miroir de cell	es contenant le motif : vieil
Nombre d'occurrences du texte global	= 29982		
Nombre d'occurrences dans la partie v	risee = 10133		
Seuil : 10			
(Specificites positives en haut de li	ste, negatives en bas)		
Forme	Ind-Specif	Fq-Totale	Fq-Partie
vieux	32.8	269	186
Le	11.3	181	104
requin	7.7	35	27
vieil	7.0	16	15
le	5.6	889	358
avait	4.9	163	78
sa	4.7	123	61
la	4.5	826	326
du	4.1	209	93
lame	3.8	6	6
1	3.7	525	209
surface	3.7	15	11
eau	3.7	80	40
harpon	3.7	19	13
ses	3.6	98	47
étaient	3.6	43	24
Les	3.5	46	25
frappa	3.5	8	7
jaillissaient	3.4	5	5
chair	3.4	12	9
grosses	3.4	5	5
- était	3.2	251	104
plusieurs	3.1	7	6
queule	3.0	9	7
mâchoires	3.0	9	7
aperçut	3.0	17	11
côté	3.0	17	11
vit	3.0	11	8
laissait	2.9	4	4
faux	2.9	4	4

Figure 292 : Spécificités volet CIBLE (FR1)

On retrouve ci-dessus les résultats produits précédemment.

De même, pour le couple (fr2, en), le paramétrage du calcul est visible ci-dessous :

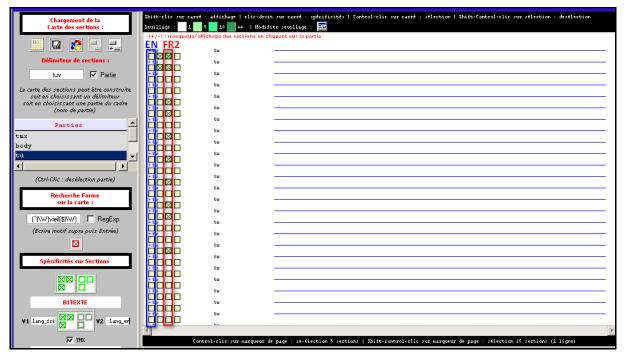


Figure 293 : Contraste de 2 volets (FR2, EN)

Et le résultat produit a løallure suivante :

```
Specificites du vocabulaire (Annotation Forme) sur les sections SOURCE contenant le motif : vieil
Nombre d'occurrences du texte global = 29257
Nombre d'occurrences dans la partie visee = 10191
Seuil : 10
         (Specificites positives en haut de liste, negatives en bas)
Le fichier construit :
C:/DOCUME~1/FLEURY~1/MESDOC~1/MYDROP~1/WIP/WIP-TR~1/ALIGNE~1/ALIGNE~1/export/resultspecif-source-1415131664.txt
                                                          Fq-Totale
vieil
                                                          185
                                                                      185
                                             **
homme
                                                         211
                                                                      189
Le
le
                                             11.6
                                                         125
                                                                      80
                                                         1005
                                             8.8
                                                                      434
                                             6.9
requin
                                                         40
sur
                                                         175
ses
                                             4.6
                                                         106
                                                                      55
lame
                                             4.3
                                                         10
                                                                       9
                                                                      113
                                             4.3
                                                          251
était
étaient
                                             4.0
                                                          40
                                                                      24
yeux
                                             3.9
                                                          28
                                                                       18
nageoires
avaient
                                                          32
                                                                       20
avait
                                             3.8
                                                         177
                                                                      81
parlaient
                                             3.7
                                                          6
                                                          13
                                                                       10
frappa
                                             3.6
dents
                                             3.5
                                                          8
continuait
                                                          10
Parfois
                                             3.3
                                                                       5
harpon
                                             3.2
                                                          20
                                                                      13
ainsi
                                             3.1
                                                                       6
                                                          29
                                                                      17
surface
                                             3.1
et
                                             3.0
                                                          989
                                                                       379
leurs
                                             3.0
                                                          19
                                                                      12
son
                                             3.0
                                                          157
                                                                       69
poissons
                                             3.0
                                                          34
                                                                      19
                                                          73
eau
                                             2.9
                                                                       35
```

Figure 294 : Spécificités volet SOURCE (FR2)

Nombre d'occurrences du texte glob	nal = 26847		
Nombre d'occurrences dans la parti			
Seuil : 10			
(Specificites positives en haut de	liste, negatives en bas)		
Forme	Ind-Specif	Fg-Totale	Fg-Partie
old	49.4	247	200
man	40.6	264	199
The	12.1	194	115
che	7.6	2119	858
shark	7.2	46	33
nis	6.3	430	197
ner	6.0	29	22
yater	4.7	106	56
<i>y</i> ere	4.5	138	69
yide	4.4	13	11
Jas	4.4	435	188
nit	4.3	26	18
olade	4.2	10	9
nad	3.8	203	93
came	3.7	41	24
as	3.6	191	87
by	3.5	29	18
jaws	3.5	17	12
curtles	3.4	8	7
5	3.4	77	39
neck	3.3	5	5
afraid	3.3	5	5
flesh	3.3	5	5
side	3.3	32	19
rose	3.2	14	10
narpoon	3.2	20	13
surface	3.2	16	11
and	3.0	1233	475
closed	3.0	7	6
oirds	3.0	7	6
they	3.0	99	47
Sometimes	3.0	7	6
coming	2.9	15	10

Figure 295 : Spécificités volet CIBLE (EN)

30.2.2.4 Importer une base « døalignements » multi-annotée

Les calculs présentés ci-dessus peuvent être réalisés sur tous les niveaux døannotation disponibles dans la base textométrique chargée dans le *Trameur*.

Nous allons illustrer cette fonctionnalité en construisant une base multi-annotée et intégrant 2 volets alignés. Pour cela, nous procédons de la manière suivante :

- Chargement avec le *Trameur* de la base complète précédente (convention FR et EN) avec étiquetage en Français via *treetagger*
- Export de la base précédente : création døune base multi-annotée (BASE-FR-EN) døun bitexte, le volet FR est « bien » annoté, le volet EN ne løest pas.
- Chargement avec le *Trameur* døune base constituée uniquement du volet EN avec étiquetage en Anglais via *treetagger*
- Export de la base précédente : création døune base multi-annotée du volet EN (BASE-EN)
- Dans un éditeur, suppression dans la base BASE-FR-EN de la *Trame* correspondant au volet EN et insertion « au même endroit » de la *Trame* de la base BASE-EN. Dans la base BASE-FR-EN, les 2 volets sont désormais « bien » annotés (chaque volet ayant son propre jeu déannotations)

La base BASE-FR-EN peut finalement être importée dans le *Trameur*

La base résultante est visible dans la figure suivante (2 extraits de la *Trame* dans lesquels on donne à voir le début des 2 volets) : chaque item de la *Trame* dispose de 3 niveaux døannotation : forme (balise <f>), lemme (balise <l>) et POS (balise (<c>).

```
<items|
<item type="delim" pos="1"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="2"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="3"><f>convention</f><c>NOM</c><1>convention</l></item>
<item type="delim" pos="4"><f> </f><c>DELIM</c><1>BLANK</l></item>
<item type="forme" pos="5"><f>de</f><c>PRP</c><1>de</1></item>
<item type="delim" pos="6"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="7"><f>sauvegarde</f><c>VER_pres</c><1>sauvegarde</1></item>
<item type="delim" pos="8"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="9"><f>des</f><c>PRP_det</c><1>du</1></item>
<item type="delim" pos="10"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="11"><f>droits</f><c>NOM</c><1>droit</1></item>
<item type="delim" pos="12"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="13"><f>de</f><c>PRP</c><1>de</l></item>
<item type="delim" pos="14"><f> </f><c>DELIM</c><1>BLANK</l></item>
<item type="forme" pos="15"><f>1</f><c>VER simp</c><1>1</l></item>
<item type="delim" pos="16"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="forme" pos="17"><f>homme</f><c>NOM</c><1>homme</1></item>
<item type="delim" pos="18"><f> </f><c>DELIM</c><1>BLANK</l></item>
<item type="forme" pos="19"><f>et</f><c>KON</c><1>et</l></item>
<item type="delim" pos="20"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="21"><f>des</f><c>PRP_det</c><1>du</1></item>
<item type="delim" pos="22"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23"><f>libertés</f><c>NOM</c><l>liberté</l></item>
<item type="delim" pos="24"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="25"><f>fondamentales</f><c>ADJ</c><1>fondamental</1></item>
<item type="delim" pos="26"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="27"><f>telle</f><c>PRO_DEM</c><l>tel</l></item>
<item type="delim" pos="28"><f> </f><c>DELIM</c><1>BLANK</1></item>
-u\-- BASE-FR-EN-3-annotations-utf8.xml
                                                 (SGML) -- L37-- 1%--
<item type="delim" pos="23950"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="23951"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="23952"><f>convention</f><c>NN</c><1>convention</l></item>
<item type="dellm" pos="23953"><f> </f><c>DELIM</c><1>BLANK</l></item>
<item type="forme" pos="23954"><f>for</f><c>IN</c><1>For</l></item>
<item type="delim" pos="23955"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23956"><f>the</f><c>DT</c><1>the</l></item>
<item type="delim" pos="23957"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23958"><f>protection</f><c>NN</c><1>protection</1></item>
<item type="delim" pos="23959"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23960"><f>of</f><c>IN</c><1>of</1></item>
<item type="delim" pos="23961"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23962"><f>human</f><c>JJ</c><1>human</1></item>
<item type="delim" pos="23963"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23964"><f>rights</f><c>NNS</c><1>right</1></item>
<item type="delim" pos="23965"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23966"><f>and</f><c>CC</c><1>and</1></item>
<item type="delim" pos="23967"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23968"><f>fundamental</f><c>JJ</c><l>fundamental</l></item>
<item type="delim" pos="23969"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="23970"><f>freedoms</f><c>NNS</c><1>freedom</1></item>
<item type="delim" pos="23971"><f>RETURN</f><c>DELIM</c><1>RETURN</1></item>
<item type="forme" pos="23972"><f>as</f><c>IN</c><1>as</1></item>
<item type="delim" pos="23973"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23974"><f>amended</f><c>VVD</c><1>amend</1></item>
<item type="delim" pos="23975"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23976"><f>by</f><c>IN</c><1>by</1></item>
<item type="delim" pos="23977"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="23978"><f>protocol</f><c>NN</c><1>protocol</l></item>
      BASE-FR-EN-3-annotations-utf8.xml
                                                   (SGML) -- L23991--52%-----
```

Figure 296 : Base textométrique multiannotée d'un alignement de textes bilingues

Une fois importée, les mêmes calculs peuvent être réalisés sur la *Carte des sections* mettant au jour løalignement.

On pourrait bien entendu disposer døune base multi-annotée et intégrant 2 volets alignés en procédant døune autre manière : on pourrait par exemple projeter (via un programme ou à la main) des annotations sur la *Trame* døune base intégrant 2 volets alignés.

30.2.2.5 Calculs contrastifs sur une base textométrique de textes alignés

Au delà du calcul des cooccurrents sur 2 volets alignés døune base textométrique *via* la *Carte des sections*, il est possible de mener des calculs textométriques sur chacun des volets en utilisant le *Gestionnaire de Sélection* et notamment les modules permettant de lancer des calculs sur telle ou telle partie de la base de travail.

Sur la base précédente par exemple, les modules de calcul encadrés dans la figure suivante peuvent être lancés sur un des volets de la base après sélection du volet visé dans la liste des parties de la partition volet.

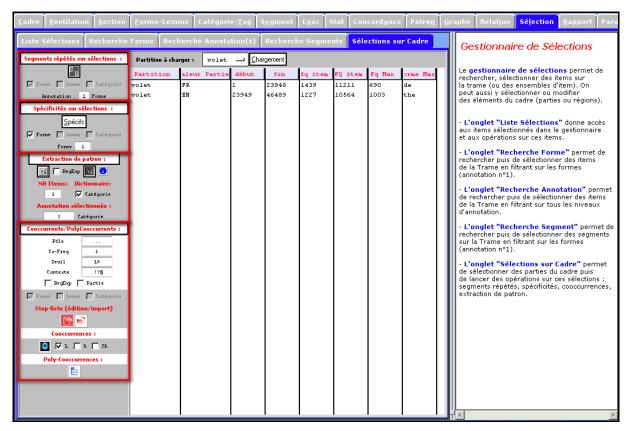


Figure 297: Calculs contrastifs sur une base textométrique d'alignements

30.2.3 Outils et méthodes dœxploration dœun alignement

Dans la suite de ce document, 2 bases textométriques sont utilisées :

- 1. <u>Base bilingue ParTUT2Trameur (français/anglais)</u>. Les 2 volets de cette base disposent de 9 niveaux d'annotation (Alignement de treebank). Cette base permet de tester les traitements spécifiques pour les bases textométriques intégrant 2 textes. <u>Alignement de Treebank pour le Trameur</u> (pdf) : présentation du processus de transcodage de 2 treebanks ParTUT pour construire une base Textométrique.
- 2. <u>Base bilingue "Investiture Obama"</u> (<u>français/anglais</u>). Cette base regroupe le discours original en anglais prononcé par B. Obama le 20 janvier 2009 à Washington, publié sur le site de The New York Times (<u>volet EN</u>), et 4 traductions françaises de ce discours (volets FR0-1-2-3). Les traductions ont été récupérées sur le site officiel de la Maison Blanche (<u>volet FR0</u>), sur les sites des journaux français Le Monde (<u>volet FR1</u>) et Libération (<u>volet FR2</u>), ainsi que sur le site de RFI (<u>volet FR3</u>). Les 5 volets de cette base disposent de 3 niveaux d'annotation (forme, lemme, catégorie). Cette base permet de tester les traitements spécifiques pour les bases textométriques intégrant 2 textes alignés.

30.2.3.1 Edition døun alignement multi-volets

La base *Investiture* étant chargée (par importation), la carte des sections construite via le délimiteur de section § a løallure suivante :

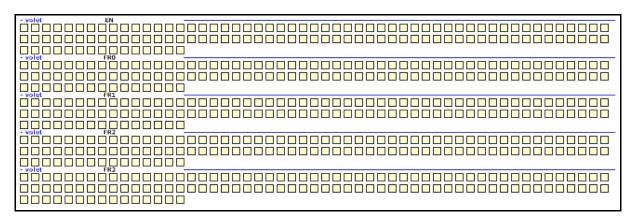


Figure 298 : Carte de sections døun alignement multi-volets

Løalignement des 5 volets se lit de manière verticale comme dans les exemples précédents. Sous la carte des sections, une case à cocher et une zone de saisie permettent de modifier løaffichage des sections. Par défaut, un clic sur une section affiche le contenu de la section dans løditeur de section. Si la case à cocher Bitext est cochée et si on donne le nombre de volets présents dans løalignement chargé et mis au jour dans la carte des sections :



Dans ce cas, lédition de la section cliquée se fait en insérant aussi ses sections miroirs.

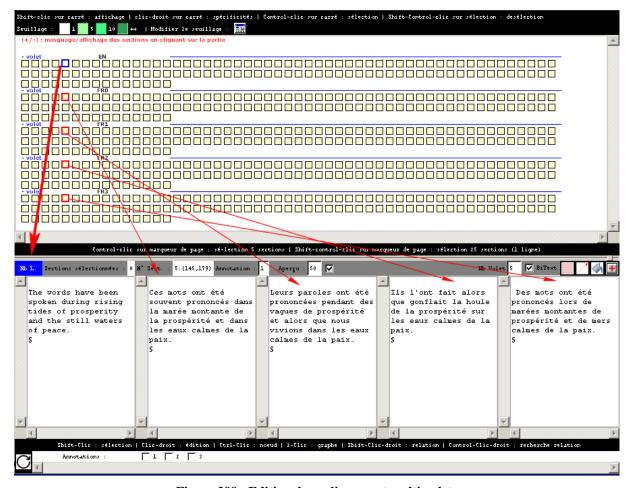


Figure 299: Edition døun alignement multi-volets

Dans la figure précédente, le clic sur la section du volet EN (carré bleu) déclenche la mise au jour des sections miroirs et løaffichage en parallèle de tous les contenus de ces sections alignées.

Remarque : la section cliquée dans la carte est éditée sur la gauche de løéditeur multi-volets, ses miroirs sont ensuite éditées sur sa droite dans løordre døapparition dans la carte.

30.3.3.2 Opérations dans une section multi-volets

Toutes les opérations disponibles dans léditeur de section peuvent désormais être réalisées sur une section multi-volets. Ci-dessous, léaffichage de 3 niveaux déannotations sur les 5 sections éditées conjointement :

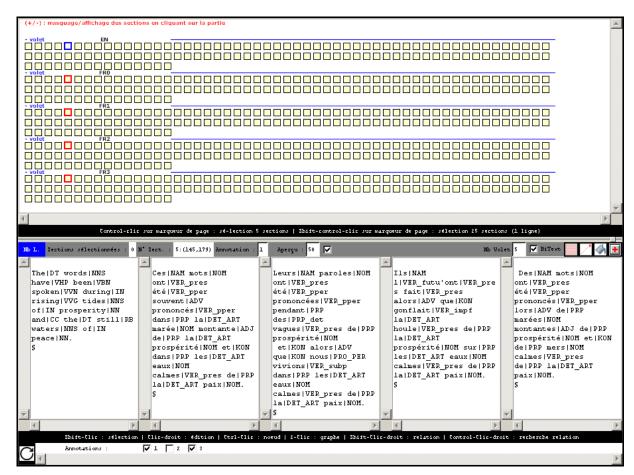


Figure 300 : Opérations dans une section multi-volets

30.3.3.3 Exporter un alignement au format HTML

Quand un alignement multi-volets est chargé, il est possible dœxporter des résultats de recherche ou løalignement complet dans un fichier au format HTML.

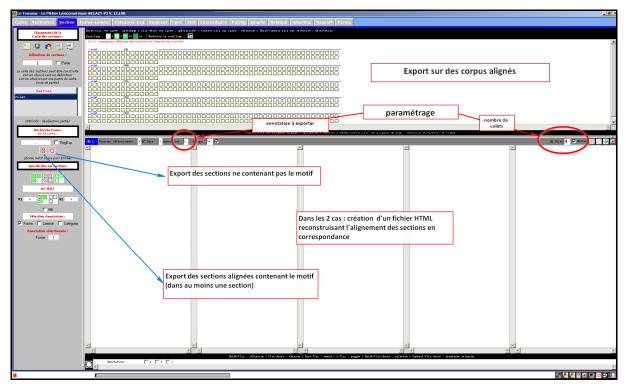


Figure 301 : Export de corpus alignés (HTML)

Ci-dessous, un exemple du résultat produit :



Les motifs cherchés sont matérialisés en rouge gras, les items préalablement sélectionnés via le *Gestionnaire de Sélection* sont matérialisés en jaune.

On trouvera en ligne des exemples de sortie construits sur un corpus aligné :

- (1) Alignement avec présence du motif (annotation n°1 : forme),
- (2) Alignement avec présence du motif (annotation n°2 : lemme),
- (3) Alignement avec absence du motif (annotation n°1 : forme),
- (4) Alignement complet (annotation n°1 : forme).

30.3.3.4 Graphe des relations de dépendances sur une section multi-volets

Cette fois-ci, on considère que la base *parTUT* est chargée dans le Trameur. Cette base intègre un alignement de *treebanks* (*cf* présentation en ligne de cette ressource) : un volet EN et un volet FR. On édite conjointement une section et son miroir :



Figure 302 : Graphe de relations de dépendances dans une section multi-volets (1)

Le bouton « Affichage des relations de dépendance dans la section » (cerclé de rouge cidessus) permet de mettre au jour le graphe de relations sur les 2 zones textuelles alignées ici :



Figure 303 : Graphe de relations de dépendances dans une section multi-volets (2)

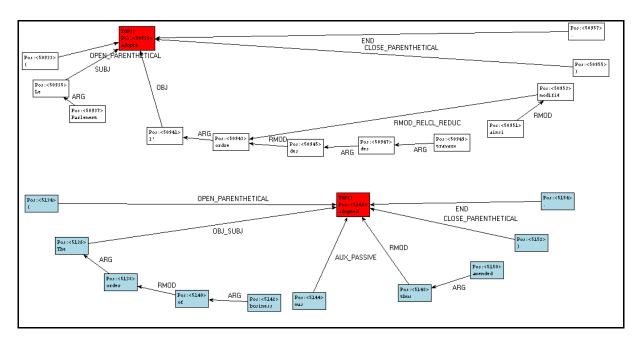


Figure 304 : Graphe de relations de dépendances dans une section multi-volets (3)

30.3 Annotation de données orales

Projet associé: CFPP2000 http://cfpp2000.univ-paris3.fr/index.html

Un moteur de recherche²⁶ permet de lancer des requêtes dans les transcriptions de **CFPP2000** annotées via *treetagger* (similaire à celui déjà en place sur le texte brut des transcriptions²⁷). Les données ont été annotées en utilisant les ressources développées pour <u>le traitement de løoral</u> via *treetagger* par Christophe Benzitoun (ATILF) et intégrées dans le *Trameur*.

Après chargement et annotation d'un fichier dans le *Trameur*, on peut exporter la "base annotée" résultante dans un format XML « proche » du format de départ (accessible via le bouton •): la base résultante est de fait le fichier de départ sur lequel on a projeté des annotations construites par *treetagger*.

Pour le moment les balises auto-fermantes ne sont pas gérées dans le *Trameur*, elles ne sont dans pas conservées dans cet export. Ce type de balise est utilisé dans les textes de transcriptions construits par *Transcriber* par exemple (outil utilisé pour la transcription dans le corpus **CFPP2000**) pour noter des commentaires, distinguer les locuteurs quand plusieurs sont associés à un seul tour de paroles etc. Les figures suivantes permettent de visualiser la perte d'information induite : si on cherche la forme "fonctionnaire" dans les transcriptions brutes puis dans les transcriptions annotées on obtient les 2 résultats suivants :

Recherche sur texte brut:

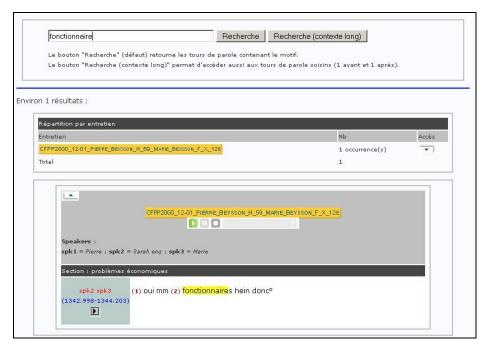


Figure 305: CFPP2000 (1)

Recherche sur texte annoté: (on perd la distinction spk2/spk3)

 $^{^{26}~\}underline{http://cfpp2000.univ-paris3.fr/search-transcription-tt/}$

²⁷ http://cfpp2000.univ-paris3.fr/search-transcription/

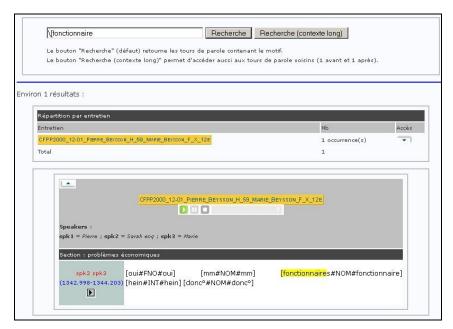


Figure 306: CFPP2000 (2)

Dans le cas de **CFPP2000**, le processus (døannotation puis døexport) intégré dans le *Trameur* permet d'annoter facilement des fichiers de transcription issus de *Transcriber* par exemple et de récupérer ensuite les mêmes fichiers avec annotations morphosyntaxiques (pour le moment, il reste à faire quelques modifications mineures sur le fichier exporté). On peut donc très facilement et à moindre coût réitérer l'étiquetage et produire un nouveau fichier annoté que l'on peut ensuite intégrer dans le moteur de recherche dans les données annotées de **CFPP2000**.

Les éventuelles corrections sur les annotations produites par *treetagger* peuvent se faire dans le *Trameur* (*cf* processus de correction des annotations présentés dans ce document) ou de manière externe (*via* un éditeur XML par exemple).

30.4 Module petiMoteur

Le module *petiMoteur* est disponible dans løonglet **CADRE**. Ce module permet de réaliser un corpus à partir døune liste døURLs (ici une liste de 60 urls pointant vers des pages døaccueil de journaux français en ligne). Une fois le fichier døURLs chargé (1), le module døaspiration (2) construit un fichier concaténant les contenus textuels des différentes URLs. Le fichier produit (text_URLS.txt dans le même dossier que le fichier des URLs) peut ensuite être chargé dans le *Trameur*.

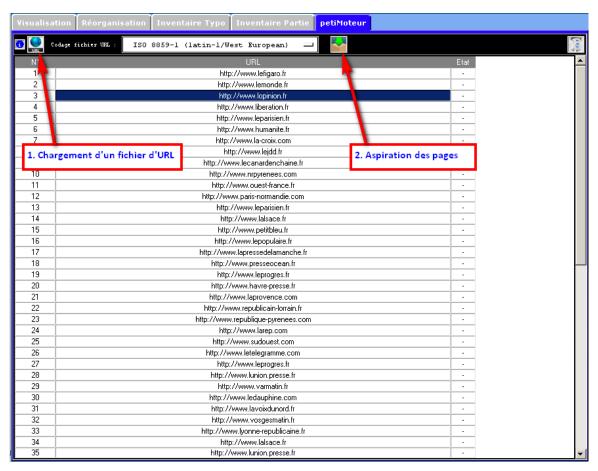


Figure 307: Le petiMoteur

30.5 MOTIFS : repérage et sélection de motifs multi-annotés

Définitions

Le **motif** se présente comme une nouvelle unité textuelle récurrente (A B C, par exemple) composé døunités A, B, C, se situant à différents niveaux (formes graphiques, lemmes, catégories grammaticales, patrons syntaxiques et, éventuellement, schèmes métriques ou prosodiques). La première particularité du motif, en comparaison avec le segment répété ou la cooccurrence, est sa multidimentionnalité : « la notion de motif est conçue comme un moyen de conceptualiser la multidimensionnalité (ou le caractère multi-niveau) de certaines formes récurrentes qui sollicitent à la fois le lexique, les catégories grammaticales et la syntaxe, éventuellement la prosodie, la métrique. » (Longrée et Mellet 2013 : 66)

Mise en ò uvre dans le Trameur

Pour illustrer la fonctionnalité mise en ò uvre dans le *Trameur*, on travaille ci-dessous avec le corpus suivant :

Le corpus **VOEUX** réunit lœnsemble des allocutions du 31 décembre (vò ux aux Français) des présidents de la Cinquième République, de 1959 à 2015. Il contient 58 257 occurrences pour 6 426 formes. Ce corpus est accessible en ligne sur les pages suivantes : (https://sourceforge.net/projects/txm/files/corpora/voeux/ et www.textopol.u-pec.fr [sur demande])

Ce corpus est chargé dans le *Trameur* avec étiquetage via *TreeTagger*. Sur ce corpus, nous allons mettre au jour les séquences textuelles correspondant au motif suivant :

```
<2:nous> <1:avons;3:VER> <3:.*>{1,20} <2:espérer|espoir|confiance|raison|raisons>
```

Chaque item du motif est décrit dans une séquence du type *<description-item>*

La *description døun item* contient des informations à vérifier sur une ou plusieurs couches døannotation pour løtem visé; par exemple sur le second terme du motif on a : <1:avons;3:VER>, ce terme doit permettre de rechercher un VERbe dont la forme graphique est « avons » (løordre des annotations dans la *description døun item* nøa aucune importance).

Le motif précédent décrit donc une séquence textuelle :

- commençant par un item dont løannotation n°2 (lemme) est *nous* (en gros le pronom personnel *nous* (minuscule ou majuscule si les lemmes sont tous codés en minuscule)
- suivi døun item dont løannotation n°1 (forme) est *avons* et dont løannotation n°2 est VER; plus précisément, un item dont løannotation n°1 contient la chaîne de caractère *avons* (par exemple, *avons* ou *savons*í) et dont løannotation n°2 contient la chaîne de caractère *VER*
- suivie de 1 à 20 occurrences doitems de valeur quelconque (via la regexp .*) pour lonnotation n°3

• et se terminant par des items dont løannotation n°2 (lemme) coïncident avec la regexp donnée i.e les unités suivantes : espérer ou espoir, ou confiance ou raison(s)

Ecriture des motifs dans un fichier

Tous les motifs doivent être écrits au préalable dans un fichier au format texte brut. La figure suivante donne un voir un tel fichier ouvert via *Notepad++*:

```
modele-modifs.txt 🗵
           PATTERN OK
           Seul le pole doit etre decrit de maniere "exacte" dans sa forme
                                                        <1:patriotes> <3:ADJ>
          <1:patriotes>
                                                          <1:patriotes> <3:.*>{1,2} <3:ADJ>
           <1:patriotes>
          <1:patriotes> #
                                                         <1:patriotes> <1:ont> <3:VER>{1,2} <3:NOM>
                                                         <2:patriote> <3:VER pres> <3:VER>{1,2} <3:NOM>
           <2:patriote>
          12 <3:DET ART> # <3:DET ART> <3:DET_ART> <
           <3:DET ART> #
                                               <3:DET_ART> <3:ADJ> <3:NOM>{0,2} <3:VER>
          <3:DET_ART> #
                                                 <3:DET_ART> <3:ADJ> <3:NOM>{1,2} <3:VER>
                                                 <3:DET ART> <3:ADJ>{1,2} <3:NOM>{1,2} <3:VER>
          <3:DET ART> #
                                                 <3:DET_ART> <3:ADJ>{1,2} <3:NOM>{1,2} <3:ADJ>{1,2} <3:VER>
          <3:DET ART> #
          <3:DET ART> #
                                                 <3:DET ART> <3:ADJ>{0,2} <3:NOM>{1,2} <3:ADJ>{1,2} <3:VER>
          <3:DET ART> #
                                                 <3:DET ART> <3:ADJ>{1,2} <3:NOM>{1,2} <3:ADJ>{0,2} <3:VER>
          <1:nous>
                                                <1:nous> <2:avoir> <3:.*>{1,20} <2:espérer|espoir|confiance|raison|raisons>
<1:Nous> <1:avons> <3:.*>{1,20} <2:espérer|espoir|confiance|raison|raisons>
          <1:Nous>
                                                  <2:nous> <1:avons> <3:.*>{1,20} <2:espérer|espoir|confiance|raison|raisons>
           <2:nous>
           <2:B> <42:I>{0,100} <42:L>
           <1:patriotes;3:NOM> #
                                                                     <1:patriotes;3:NOM> <3:PRP>{1,2} <3:ADJ>
                                                                     <1:patriotes;3:NOM> <3:PRP>{1,2} <2:rencontrer;3:ADJ>
           <1:patriotes;3:NOM> #
```

Figure 308: Fichier de MOTIFS

Un fichier similaire est fourni dans la distribution standard du logiciel. Ce fichier a løallure suivante :

```
POLE # MOTIF
```

Le POLE est suivi døune tabulation, puis le caractère #, puis de nouveau une tabulation et le MOTIF. On trouve par exemple dans le fichier précédent la ligne suivante :

```
<3:DET_ART> # <3:DET_ART> <3:ADJ>{0,2} <3:NOM> <3:VER>
```

Elle contient la description du pôle (avant le #) puis celle du motif (après le #).

Le POLE correspond au premier terme du MOTIF:

```
3:DET ART
```

ici un item dans løannotation n°3 (catégorie) est DET ART.

On a ensuite la description du MOTIF associé au pôle précédent :

```
<3:DET_ART> <3:ADJ>{0,2} <3:NOM> <3:VER>
```

Chaque item du motif est décrit dans une séquence du type *description-item*; la *description doun item* contient des informations à vérifier sur une ou plusieurs couches d'annotation pour

løtem visé; dans løexemple précédent, seule une couche døannotation est spécifiée pour chaque item du motif. Le motif testé *infra* en utilisera plusieurs.

Le motif commence ici par le pôle précédent (<3:DET_ART>), il est suivi de 0 à 2 occurrences døtem(s) dont løannotation n°3 (catégorie) est ADJ (<3:ADJ>{0,2}), puis døun item dont løannotation n°3 (catégorie) est NOM (<3:NOM>) et il se termine par un item dont løannotation n°3 (catégorie) est VER (<3:VER>).

Actuellement, dans la description du motif, seul le premier terme du motif doit être défini précisément (dans sa valeur). Pour les autres termes du motif, il est possible døutiliser une expression régulière (*cf* exemple initial).

Les termes du motif (sauf le premier et le dernier) peuvent être accompagnés døun marquage de répétition (avec une valeur mini (0 compris) et une valeur maxi).

Chargement du fichier contenant les motifs

Une fois les motifs définis dans le fichier dédié, ce dernier peut être chargé dans le *Trameur* :

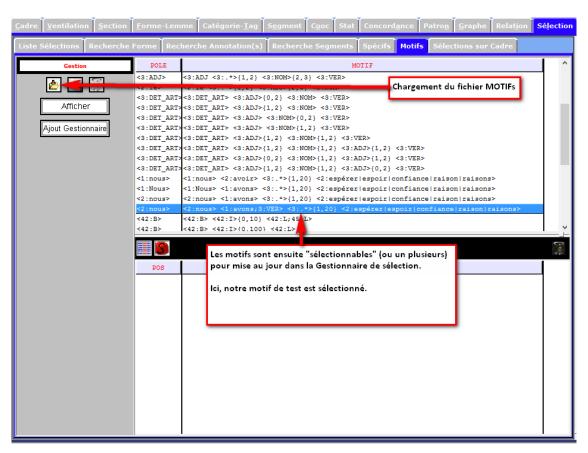


Figure 309 : Fonctionnalités sur les MOTIFS

Remarque : le fichier peut être modifié puis rechargé pour éventuellement prendre en compte de nouveaux motifs. Le motif sélectionné est celui explicité *supra* :

```
<2:nous> <1:avons;3:VER> <3:.*>{1,20} <2:espérer|espoir|confiance|raison|raisons>
```

Rappel : dans ce motif, le second terme intègre une requête sur 2 couches d\(\phi\) annotations pour les items vis\(\epsilon\) :

→ on cherche (1) le lemme « nous » suivi par (2) un item qui doit être un verbe ayant pour forme graphique « avons » etc.

Les différentes contraintes posées sur un item peuvent sœnchaîner les unes derrière les autres et séparées par un point-virgule et dans un ordre quelconque :

```
<i:pattern_sur_i;j:pattern_sur_j;k:pattern_sur_k;...>
```

1. Affichage des occurrences døun motif

Le bouton « Affichage » permet døafficher la séquence textuelle (les formes graphiques) des occurrences du motif :

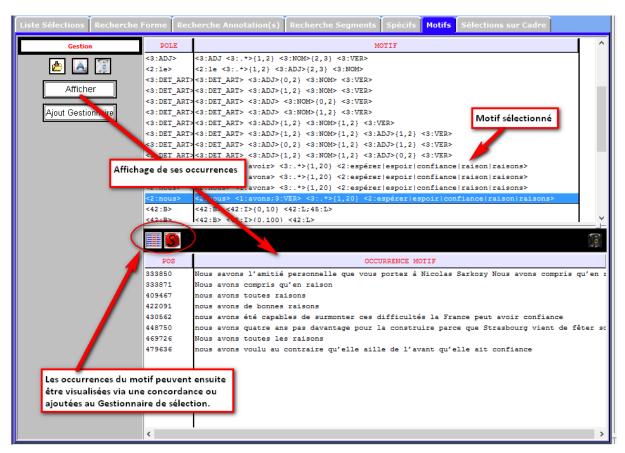


Figure 310: Les occurrences døun MOTIF

Ces occurrences peuvent être affichées en contexte via une concordance : utile pour explorer les différentes annotations de chaque item de la séquence mise au jour.

Elles peuvent aussi être ajoutées au *Gestionnaire de Sélection* (sélection de la séquence visée puis activation du bouton « Sélection »).

2. Sélection d'un motif à "mettre au jour" (*in fine* les occurrences du motif sont ajoutées au gestionnaire de sélection)

Le bouton « **Ajout Gestionnaire** » déclenche directement la recherche des occurrences des motifs sélectionnés et leur insertion dans le *Gestionnaire de Sélection* (sans passer par leur affichage comme précédemment)

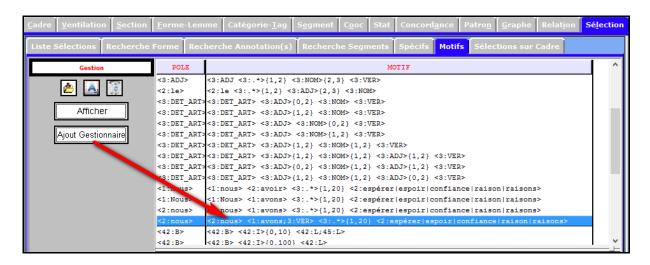


Figure 311 : Sélection døun MOTIF

Affichage du résultat dans le Gestionnaire de Sélection

Une fois les occurrences du motif visé éventuellement ajoutées au *Gestionnaire de Sélection*, il est possible de leur appliquer les traitements disponibles.

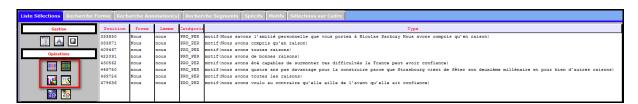


Figure 312: Traitements sur MOTIF

La quatrième colonne donne à voir les séquences textuelles associées au motif (via leur annotation n°1 (forme)) :

REMARQUE: la recherche de motif "traverse" les délimiteurs....

Visualisation en contexte du "motif"

Les items ainsi mis au jour (et les séquences associées : les occurrences du MOTIF) sont ensuite exploitables via les opérations disponibles dans le logiciel. On peut visualiser ces séquences en contexte, via une concordance ou dans une section de la *Carte des Sections* ou encore dans un graphique de ventilation (cf. présentation du *Gestionnaire de Sélection*).

31 Le Trameur en mode console

Le *Trameur* en mode console soutilise dans une fenêtre de commandes :

```
bash-2.02$
bash-2.02$
bash-2.02$
bash-2.02$
bash-2.02$
colored bash-2.02$
bash-2.02$
bash-2.02$
bash-2.02$
colored bash-2.02$
colored bash-2.02$
colored bash-2.02$
colored bash-2.02$

letrameur 5.0 (c) Serge Fleury SYLED/CLA2T 07/06/2007

| Colored bash-2.02$
```

Figure 313 : *Le Trameur* en mode console

Le seul objectif de ce programme est de construire le *Cadre* et la *Trame* døun fichier.

Nous allons montrer sur le même fichier de travail les différents fichiers produits suivant les options définies et visibles dans la figure précédente.

31.1 Fichier de travail:

```
<STRUCTURE="TITRE"> Le dormeur du val

<STRUCTURE="TEXTEPOEME">

<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,

<LIGNE="VERS2"> Accrochant follement aux herbes des haillons

<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,

<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
```

31.2 Le Cadre :

entête

```
<?xml version="1.0" encoding="iso8859-1"?>
<Cadre>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame)
importable et echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
Samedi 26 Mai 2007
10:17:17. Ce document n'est pas encore publie.
</publicationStmt>
<sourceDesc>
Fichier genere par Le Trameur 5.0 (http://www.tal.univ-paris3.fr/trameur).
Fichier traite : dormeurL3.txt
Encodage : iso8859-1
Etiquetage Treetagger : NON
Fichiers construits:
La Trame :
.\dormeurL3-Trame.xml
Le Cadre :
 .\dormeurL3-Cadre.xml
</sourceDesc>
</fileDesc>
</teiheader>
<acces>
<partition nom="structure">
<partie nom="TITRE" ideb="1" ifin="11"/>
<partie nom="TEXTEPOEME" ideb="11" ifin="100"/>
</partition>
<partition nom="LIGNE">
<partie nom="VERS1" ideb="11" ifin="35"/>
<partie nom="VERS2" ideb="35" ifin="49"/>
<partie nom="VERS4" ideb="75" ifin="100"/>
</partition>
</acces>
</Cadre>
```

```
31.3 La Trame:
                                                                                           entête
<?xml version="1.0" encoding="iso8859-1"?>
<Trame>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), impor
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
Samedi 26 Mai 2007
 10:17:17. Ce document n'est pas encore publie.
</publicationStmt>
<sourceDesc>
Fichier genere par Le Trameur 5.0 (http://www.tal.univ-paris3.fr/trameur).
Fichier traite : dormeurL3.txt
Encodage : iso8859-1
Etiquetage Treetagger :NON
Fichiers construits:
La Trame :
 .\dormeurL3-Trame.xml
Le Cadre :
.\dormeurL3-Cadre.xml
</sourceDesc>
</fileDesc>
</teiheader>
<items>
<item type="delim" pos="1"> </item>
<item type="forme" pos="2">Le</item>
<item type="delim" pos="3"> </item>
<item type="forme" pos="4">dormeur</item>
<item type="delim" pos="5"> </item>
<item type="forme" pos="6">du</item>
<item type="delim" pos="7"> </item>
<item type="forme" pos="8">val</item>
<item type="delim" pos="9"> </item>
<item type="forme" pos="10">
</item>
<item type="delim" pos="11"> </item>
<item type="forme" pos="12">C</item>
<item type="delim" pos="13">'</item>
<item type="delim" pos="14"> </item>
<item type="forme" pos="15">est</item>
<item type="delim" pos="16"> </item>
<item type="forme" pos="17">un</item>
<item type="delim" pos="18"> </item>
<item type="forme" pos="19">trou</item>
<item type="delim" pos="20"> </item>
<item type="forme" pos="21">de</item>
<item type="delim" pos="22"> </item>
<item type="forme" pos="23">verdure</item>
<item type="delim" pos="24"> </item>
<item type="forme" pos="25">où</item>
<item type="delim" pos="26"> </item>
<item type="forme" pos="27">chante</item>
<item type="delim" pos="28"> </item>
<item type="forme" pos="29">une</item>
<item type="delim" pos="30"> </item>
<item type="forme" pos="31">rivière</item>
<item type="delim" pos="32">,</item>
<item type="delim" pos="33"> </item>
<item type="forme" pos="34">
</item>
<item type="delim" pos="35"> </item>
<item type="forme" pos="36">Accrochant</item>
<item type="delim" pos="37"> </item>
<item type="forme" pos="38">follement</item>
<item type="delim" pos="39"> </item>
<item type="forme" pos="40">aux</item>
```

```
<item type="delim" pos="41"> </item>
<item type="forme" pos="42">herbes</item>
<item type="delim" pos="43"> </item>
<item type="forme" pos="44">des</item>
<item type="delim" pos="45"> </item>
<item type="forme" pos="46">haillons</item>
<item type="delim" pos="47"> </item>
<item type="forme" pos="48">
</item>
<item type="delim" pos="49"> </item>
<item type="forme" pos="50">D</item>
<item type="delim" pos="51">'</item>
<item type="delim" pos="52"> </item>
<item type="forme" pos="53">argent</item>
<item type="delim" pos="54"> </item>
<item type="delim" pos="55">;</item>
<item type="delim" pos="56"> </item>
<item type="forme" pos="57">où</item>
<item type="delim" pos="58"> </item>
<item type="forme" pos="59">le</item>
<item type="delim" pos="60"> </item>
<item type="forme" pos="61">soleil</item>
<item type="delim" pos="62"> </item>
<item type="delim" pos="63">;</item>
<item type="delim" pos="64"> </item>
<item type="forme" pos="65">de</item>
<item type="delim" pos="66"> </item>
<item type="forme" pos="67">la</item>
<item type="delim" pos="68"> </item>
<item type="forme" pos="69">montagne</item>
<item type="delim" pos="70"> </item>
<item type="forme" pos="71">fière</item>
<item type="delim" pos="72">,</item>
<item type="delim" pos="73"> </item>
<item type="forme" pos="74">
</item>
<item type="delim" pos="75"> </item>
<item type="forme" pos="76">Luit</item>
<item type="delim" pos="77"> </item>
<item type="delim" pos="78">:</item>
<item type="delim" pos="79"> </item>
<item type="forme" pos="80">c</item>
<item type="delim" pos="81">'</item>
<item type="delim" pos="82"> </item>
<item type="forme" pos="83">est</item>
<item type="delim" pos="84"> </item>
<item type="forme" pos="85">un</item>
<item type="delim" pos="86"> </item>
<item type="forme" pos="87">petit</item>
<item type="delim" pos="88"> </item>
<item type="forme" pos="89">val</item>
<item type="delim" pos="90"> </item>
<item type="forme" pos="91">qui</item>
<item type="delim" pos="92"> </item>
<item type="forme" pos="93">mousse</item>
<item type="delim" pos="94"> </item>
<item type="forme" pos="95">de</item>
<item type="delim" pos="96"> </item>
<item type="forme" pos="97">rayons</item>
<item type="delim" pos="98">.</item>
<item type="delim" pos="99"> </item>
</items>
</ra>
```

31.4 La Trame « annotée » (si treetagger est invoqué)

entête

```
<?xml version="1.0" encoding="iso8859-1"?>
<treetagger>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), importable
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
Samedi 26 Mai 2007
10:20:34. Ce document n'est pas encore publie.
</publicationStmt>
<sourceDesc>
Fichier genere par Le Trameur 5.0 (http://www.tal.univ-paris3.fr/trameur).
Fichier traite : dormeurL3.txt
Encodage : iso8859-1
Etiquetage Treetagger :OUI
Fichiers construits :
La Trame :
 .\dormeurL3-Trame.xml
Le Cadre :
.\dormeurL3-Cadre.xml
Le fichier a treetagger :
 .\dormeurL3-2-treetagger.txt
Le fichier issu de treetagger :
.\dormeurL3-etiquette.txt
Le fichier issu de treetagger reconstruisant la Trame initiale :
 .\dormeurL3-Trame-treetager.xml
</sourceDesc>
</fileDesc>
</teiheader>
<items>
<item type="delim" pos="1>><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="2"><f>Le</f><c>ADJ</c><1>Le</1></item>
<item type="delim" pos="3>><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="4"><f>dormeur</f><c>NOM</c><1>dormeur</1></item>
<item type="delim" pos="5»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="6"><f>du</f><c>PRP det</c><l>du</l></item>
<item type="delim" pos="7»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="8"><f>val</f><c>NOM</c><1>val</1></item>
<item type="delim" pos="9»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="10"><f>
</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="11»><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="12"><f>C</f><c>NUM</c><1>C</l></item>
<item type="delim" pos="13"><f>'</f><c>DELIM</c><1>'</1></item>
<item type="delim" pos="14»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="15"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="16»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="17"><f>un</f><c>DET ART</c><l>un</l></item>
<item type="delim" pos="18»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="19"><f>trou</f><c>NOM</c><1>trou</1></item>
<item type="delim" pos="20»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="21"><f>de</f><c>PRP</c><1>de</l></item>
<item type="delim" pos="22»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="23"><f>verdure</f><c>NOM</c><1>verdure</1></item>
<item type="delim" pos="24»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="25"><f>où</f><c>PRO_REL</c><1>où</1></item>
<item type="delim" pos="26»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="27"><f>chante</f><c>VER pres</c><l>chanter</l></item>
<item type="delim" pos="28»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="29"><f>une</f><c>DET_ART</c><1>un</1></item>
<item type="delim" pos="30»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="31"><f>rivière</f><c>NOM</c><1>rivière</1></item>
<item type="delim" pos="32"><f>, </f><c>DELIM</c><1>, </l></item>
<item type="delim" pos="33»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="34"><f>
</f><c>DELIM</c><1>RETURN</1></item>
```

CLESTHIA>SYLED>CLA²T

```
<item type="delim" pos="35»><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="36"><f>Accrochant</f><c>ABR</c><l>Accrochant</l></item>
<item type="delim" pos="37»><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="38"><f>follement</f><c>ADV</c><1>follement</l></item>
<item type="delim" pos="39\gg<f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="40"><f>aux</f><c>PRP_det</c><1>au</1></item>
<item type="delim" pos="41»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="42"><f>herbes</f><c>NOM</c><l>herbe</l></item>
<item type="delim" pos="43»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="44"><f>des</f><c>PRP_det</c><1>du</1></item>
<item type="delim" pos="45»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="46"><f>haillons</f><c>NOM</c><1>haillon</1></item>
<item type="delim" pos="47»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="48"><f>
</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="49»><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="50"><f>D</f><c>ABR</c><1>D</1></item>
<item type="delim" pos="51"><f>'</f><c>DELIM</c><1>'</l></item>
<item type="delim" pos="52»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="53"><f>argent</f><c>NOM</c><1>argent</l></item>
<item type="delim" pos="54»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="delim" pos="55"><f>;</f><c>DELIM</c><1>;</l></item>
<item type="delim" pos="56»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="57"><f>où</f><c>PRO REL</c><l>où</l></item>
<item type="delim" pos="58»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="59"><f>le</f><c>DET ART</c><l>le</l></item>
<item type="delim" pos="60»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="61"><f>soleil</f><c>NOM</c><l>soleil</l></item>
<item type="delim" pos="62»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="delim" pos="63"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="delim" pos="64»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="65"><f>de</f><c>PRP</c><1>de</l></item>
<item type="delim" pos="66»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="67"><f>la</f><c>DET ART</c><1>le</l></item>
<item type="delim" pos="68»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="69"><f>montagne</f><c>NOM</c><1>montagne</1></item>
<item type="delim" pos="70»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="71"><f>fière</f><c>ADJ</c><1>fier</l></item>
<item type="delim" pos="72"><f>,</f><c>DELIM</c><1>,</l></item>
<item type="delim" pos="73»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="74"><f>
</f><c>DELIM</c><1>RETURN</1></item>
<item type="delim" pos="75»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="76"><f>Luit</f><c>VER pres</c><l>luire</l></item>
<item type="delim" pos="77»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="delim" pos="78"><f>:</f><c>DELIM</c><1>:</1></item>
<item type="delim" pos="79»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="80"><f>c</f><c>NOM</c><1>c</l></item>
<item type="delim" pos="81"><f>'</f><c>DELIM</c><1>'</l></item>
<item type="delim" pos="82»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="83"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="84»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="85"><f>un</f><c>DET ART</c><1>un</1></item>
<item type="delim" pos="86»><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="87"><f>petit</f><c>ADJ</c><1>petit</l></item>
<item type="delim" pos="88»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="89"><f>val</f><c>NOM</c><1>val</1></item>
<item type="delim" pos="90»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="91"><f>qui</f><c>PRO_REL</c><l>qui</l></item>
<item type="delim" pos="92»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="93"><f>mousse</f><c>NOM</c><1>mousse</1></item>
<item type="delim" pos="94»><f> </f><c>DELIM</c>1>BLANK</1></item>
<item type="forme" pos="95"><f>de</f><c>PRP</c><1>de</1></item>
<item type="delim" pos="96"><f> </f><c>DELIM</c><1>BLANK</1></item>
<item type="forme" pos="97"><f>rayons</f><c>NOM</c><l>rayon</l></item>
<item type="delim" pos="98"><f>.</f><c>DELIM</c><1>.</l></item>
<item type="delim" pos="99»><f> </f><c>DELIM</c><1>BLANK</1></item>
</items>
</treetagger>
```

32 Bibliographie

- [Daoust, 2006], Daoust F., «Logiciels d'analyse textuelle : vers un format XML-TEI pour l'échange de corpus annotés» in Actes des 8èmes Journées déanalyse statistique des données textuelles, Besançon, 2006. (Disponible en ligne)
- [Fleury, Zimina, 2014], Fleury S. and Zimina M. *Trameur: A Framework for Annotated Text Corpora Exploration*, Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations, August 2014, Dublin, Ireland, pages 57-61, (PDF)
- [Fleury, 2013a], Fleury Serge. Approches textométriques des brouillons, (PDF), séminaire projet ANR ECRITURES, 14 janvier 2013.
- [Fleury, 2013b], Fleury Serge. Le Trameur. Propositions de description et déimplémentation des objets textométriques, (PDF), (texte en cours).
- [Habert & al., 1997], Habert B, Nazarenko A., Salem A. Linguistiques de corpus, Paris, Armand-Colin, 1997
- [Habert, 2005], Habert B. *Instruments et ressources électroniques pour le français*, Paris, Ophrys, 2005
- [Heiden, 2006], Heiden S. « Modèles de données et formats déchange pour léinteropérabilité des outils de textométrie », in Actes des 8èmes Journées déchange statistique des données textuelles, Besançon, 2006. (Disponible en ligne)
- [Lafon, 1984] Lafon Pierre. Dépouillements et statistiques en lexicométrie, Slatkine-Champion, Paris, 1984.
- [Lamalle & al., 2006], Lamalle C., Fleury S., Salem A. « Vers une description formelle des traitements textométriques », in Actes des 8èmes Journées døanalyse statistique des données textuelles, Besançon, 2006. (Disponible en ligne)
- [Lamalle, Salem, 2002], Lamalle C., Salem A. « Types généralisés et topographie textuelle dans løanalyse quantitative des corpus textuels », in *Actes des 6èmes Journées døanalyse statistique des données textuelles*, St Malo, Inria, 2002. (Disponible en ligne)
- [Lebart, Salem, 1994] Lebart L., Salem A. Statistique textuelle, Paris, Dunod, 1994. (Disponible en ligne)
- [Martinez, 2002] William Martinez, Zimina Maria. "Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues", in Actes JADT 2002²⁸, Journées

_

²⁸ http://www.jadt.org/

- Internationales d'Analyse Statistiques des Données Textuelles, St Malo. (<u>Disponible en ligne</u>)
- [Martinez, 2003] William Martinez. Contribution à une méthodologie de léanalyse des cooccurrences lexicales multiples dans les corpus textuels. Thèse de Doctorat en Sciences du Langage, Université de la Sorbonne nouvelle Paris 3, sous la direction déAndré Salem, Paris. (Disponible en ligne)
- [Martinez, Leblanc, 2006] Martinez W., Leblanc J-M. "L'analyse contrastive des réseaux de cooccurrence Le monde dans les discours des présidents de la Cinquième République", in Actes JADT 2006, Journées Internationales d'Analyse Statistiques des Données Textuelles, Besançon. (Disponible en ligne)
- [Pincemin, 2008], Pincemin Bénédicte (CNRS/ICAR). "Modélisation textométrique des textes", in *Actes JADT 2008*, Journées Internationales d'Analyse Statistiques des Données Textuelles, Lyon 2008. (Disponible en ligne)
- [Salem, 1987], Salem André. *Pratique des segments répétés*, Publications de l'INaLF, collection "St.Cloud", Paris, Klincksieck, 1987.
- [Söze-Duval, 2008], Keyser Söze-Duval. *Pour une textométrie opérationnelle*. (Disponible en ligne)
- [Zimina, Fleury, 2015], Zimina Maria, Fleury Serge, "Perspectives de l\u03e9architecture Trame/Cadre pour les alignements multilingues". Nouvelles perspectives en sciences sociales : revue internationale de syst\u00e9mique complexe et d'\u00e9tudes relationnelles, volume 11, num\u00e9ro 1, novembre 2015. http://www.erudit.org/revue/npss/2015/v11/n1/index.html

33 Glossaire

Sources utilisées pour constituer pour ce glossaire : LEBART Ludovic, SALEM André, *Statistique textuelle*, Paris, Dunod, 1994

Ce glossaire est aussi disponible en ligne :

http://tal.univ-paris3.fr/wakka/wakka.php?wiki=Glossaire

Abréviations:

ac Analyse factorielle des correspondances acm Analyse des correspondances multiples cla Classification sp Méthode des Spécificités sr Analyse des segments répétés ling Linguistique stat Statistique sa Segmentation automatique

accroissement spécifique (sp)

spécificité calculée pour une partie d'un corpus par rapport à une partie antérieure

analyse factorielle (stat)

famille de méthodes statistiques d'analyse multidimensionnelle, s'appliquant à des tableaux de nombres, qui visent à extraire des "facteurs" résumant approximativement par quelques séries de nombres l'ensemble des informations contenues dans le tableau de départ.

analyse des correspondances (stat)

méthode d'analyse factorielle s'appliquant à l'étude de tableaux à double entrée composés de nombres positifs. L'AC est caractérisée par l'emploi d'une distance (ou métrique) particulière dite distance du chi-2 (ou c2).

caractère (sa)

signe typographique utilisé pour l'encodage du texte sur un support lisible par l'ordinateur.

caractères délimiteurs / non-délimiteurs (sa)

distinction opérée sur l'ensemble des caractères qui entrent dans la composition du texte, permettant aux procédures informatisées de segmenter le texte en <u>occurrences</u> (suite de caractères non-délimiteurs bornée à ses extrémités par des caractères délimiteurs).

On distingue parmi les caractères délimiteurs:

- les caractères délimiteurs d'occurrence (encore appelés "délimiteurs de forme") qui sont en général : le blanc, les signes de ponctuation usuels, les signes de préanalyse éventuellement contenus dans le texte.
- les caractères délimiteurs de séquences : sous-ensemble des délimiteurs d'occurrence correspondant, en général, aux ponctuations faibles et fortes contenues dans la police des caractères.
- les caractères séparateurs de phrase : (sous-ensemble des délimiteurs de séquence) qui correspondent, en général, aux seules ponctuations fortes.

classification (stat)

technique statistique permettant de regrouper des observations ou des individus entre lesquels a été définie une distance.

classification hiérarchique (cla)

technique particulière de classification produisant par agglomération progressive des classes ayant la propriété

d'être, pour deux quelconques d'entre-elles, soit disjointes, soit incluses.

concordance (sa)

l'ensemble de lignes de contexte se rapportant à une même forme-pôle.

contribution absolue (ou contribution) (ac)

contribution apportée par un élément au facteur. Pour un facteur donné, la somme des contributions sur les éléments de chacun des ensembles mis en correspondance est égale à 100.

contribution relative (ou cosinus carré) (ac)

contribution apportée par le facteur à un élément. Pour un élément donné, la somme des contributions relatives sur l'ensemble des facteurs est égale à 1.

cooccurrence (sa) - (une c.)

présence simultanée, mais non forcément contiguë, dans un fragment de texte (séquence, phrase, paragraphe, voisinage d'une occurrence, partie du corpus etc.) des occurrences de deux formes données.

corpus (ling)

ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique. (lexicométrie) ensemble de textes réunis à des fins de comparaison; servant de base à une étude quantitative.

délimiteurs de séquence (sa)

sous-ensemble des caractères délimiteurs de <u>forme</u> correspondant aux ponctuations faibles et fortes (en général le point, le point d'interrogation, le point d'exclamation, la virgule, le point-virgule, les deux points, les guillemets, les tirets et les parenthèses).

dendrogramme (cla)

représentation graphique d'un arbre de classification hiérarchique, mettant en évidence l'inclusion progressive des classes.

discours/langue

La langue est un ensemble virtuel qui ne peut être appréhendé que dans son actualisation orale ou écrite; "discours" est un terme commode qui recouvre les deux domaines de cette actualisation.

distance du chi-2

distance entre <u>profils</u> de fréquence utilisée en analyse des <u>correspondances</u> et dans certains <u>algorithmes</u> de classification.

éditions de contextes (sa)

éditions de type concordanciel dans lesquelles les occurrences d'une forme sont accompagnées d'un fragment de contexte pouvant contenir plusieurs lignes de texte autour de la forme-pôle. La longueur de ce contexte est définie en nombre d'occurrences avant et après chaque occurrence de la forme-pôle.

éléments d'un segment (sr)

chacune des formes correspondant aux occurrences qui entrent dans sa composition. ex : A, B, C sont respectivement les premier, deuxième et troisième éléments du segment ABC.

éléments actifs (ac ou acm)

ensemble des éléments servant de base au calcul des axes factoriels, des valeurs propres relatives à ces axes et des coordonnées factorielles.

éléments supplémentaires (ou illustratifs) (ac ou acm)

ensemble des éléments ne participant pas aux calculs des axes factoriels, pour lesquels on calcule des coordonnées factorielles qui auraient été affectées à une forme ayant la même répartition dans le corpus mais participant à l'analyse avec un poids négligeable.

énoncé/énonciation (ling)

à l'intérieur du texte un ensemble de traces qui manifestent l'acte par lequel un auteur a produit ce texte.

facteur (ac ou acm)

variables artificielles construites par les techniques d'analyse factorielle permettant de résumer (de décrire brièvement) les variables actives initiales.

forme (sa) ou forme graphique

archétype correspondant aux <u>occurrences</u> identiques dans un corpus de textes, c'est-à-dire aux occurrences composées strictement des mêmes caractères non-délimiteurs d'occurrence.

forme banale (sp)

pour une partie du corpus donnée, forme ne présentant aucune spécificité (ni positive ni négative) dans cette partie .

forme caractéristique (d'une partie)

synonyme de spécificité positive

forme commune

forme attestée dans chacune des parties du corpus

forme originale (pour une partie du corpus)

forme trouvant toutes ses occurrences dans cette seule partie.

fréquence (sa) (d'une unité textuelle)

le nombre de ses occurrences dans le corpus.

fréquence d'un segment (sr) (ou d'une polyforme)

le nombre des occurrences de ce segment, dans l'ensemble du corpus.

fréquence maximale (sa)

fréquence de la forme la plus fréquente du corpus (en français, le plus souvent, la préposition "de").

fréquence relative (sa)

la fréquence d'une unité textuelle dans le corpus ou dans l'une de ses parties, rapportée à la taille du corpus (resp. de cette partie).

gamme des fréquences (sa)

suite notée Vk, des effectifs correspondant aux formes de fréquence k, lorsque k varie de 1 à la fréquence maximale.

hapax

gr. hapax (legomenon), "chose dite une seule fois".

(sa) forme dont la fréquence est égale à un dans le corpus (hapax du corpus) ou dans une de ses parties (hapax de la partie).

identification (stat, ling, sa)

reconnaissance d'un seul et même élément à travers ses multiples emplois dans des contextes et dans des situations différentes.

index (sa)

liste imprimée constituée à partir d'une réorganisation des formes et des occurrences d'un texte, ayant pour base la forme graphique et permettant de regrouper les <u>références</u> relatives à l'ensemble des occurrences d'une même forme.

index alphabétique (sa)

<u>index</u> dans lequel les formes-pôles* sont classées selon l'<u>ordre lexicographique</u> (celui des dictionnaires).

index hiérarchique (sa)

index dans lequel les formes-pôles sont classées selon l'ordre lexicométrique.

index par parties

ensemble d'index (hiérarchiques ou alphabétiques) réalisés séparément pour chaque partie d'un corpus.

lemmatisation

regroupement sous une forme canonique (en général à partir d'un dictionnaire) des occurrences du texte. En français, ce regroupement se pratique en général de la manière suivante :

- les formes verbales à l'infinitif,
- les substantifs au singulier,
- les adjectifs au masculin singulier,
- les formes élidées à la forme sans élision.

lexical (ling)

qui concerne le <u>lexique</u> ou le <u>vocabulaire</u>.

lexicométrie

ensemble de méthodes permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le <u>vocabulaire</u> d'un corpus de textes.

lexique (ling)

ensemble virtuel des mots d'une langue.

longueur (sa) (d'un corpus, d'une partie de ce corpus, d'un fragment de texte, d'une tranche, d'un segment, etc.) le nombre des occurrences contenues dans ce corpus (resp. : partie, fragment, etc.). Synonyme : taille.

On note: T la longueur du corpus; t j celle de la partie (ou tranche) numéro j du corpus.

longueur d'un segment (sr)

le nombre des occurrences entrant dans la composition de ce segment.

occurrence (sa)

suite de caractères non-délimiteurs bornée à ses extrémités par deux caractères délimiteurs de forme.

ordre lexicographique

- pour les formes graphiques :
 - o l'ordre selon lequel les formes sont classées dans un dictionnaire.
 - NB: Les lettres comportant des signes diacrisés sont classées au même niveau que les mêmes caractères non diacrisés, le signe diacritique n'intervenant que dans les cas d'homographie complète. Dans les dictionnaires, on trouve par exemple rangées dans cet ordre les formes : mais, maison, maître.
- pour les polyformes:
 - ordre résultant d'un tri des polyformes par ordre lexicographique sur la première composante. Les polyformes commençant par une même forme graphique sont départagées par un tri lexicographique sur la seconde, etc.

ordre lexicométrique (sa)

- pour les formes graphiques :
 - o ordre résultant d'un tri des formes du corpus par ordre de fréquences décroissantes ; les formes de même fréquence sont classées par ordre lexicographique.
- pour les polyformes:
 - o ordre résultant d'un tri par ordre de longueur décroissante des segments, les segments de même longueur sont départagés par leur fréquence, les segments ayant même longueur et même fréquence par l'ordre lexicographique.

paradigme (ling)

ensemble des termes qui peuvent figurer en un point de la chaîne parlée.

paradigmatique (sa)

qui concerne le regroupement en série des unités textuelles, indépendamment de leur ordre de succession dans la chaîne écrite.

partie

(d'un corpus de textes) fragment de texte correspondant aux divisions naturelles de ce corpus ou à un regroupement de ces dernières.

partition

- d'un corpus de textes :
 - division d'un corpus en parties constituées par des fragments de texte consécutifs, n'ayant pas d'intersection commune et dont la réunion est égale au corpus.
- d'un ensemble, d'un échantillon :
 - division d'un ensemble d'individus ou d'observations en classes disjointes dont la réunion est égale à l'ensemble tout entier.

partition longitudinale (sa)

partition d'un corpus en fonction d'une variable qui définit un ordre sur l'ensemble des parties

périodisation (sa)

regroupement des parties naturelles du corpus respectant l'ordre chronologique d'écriture, d'édition ou de parution des textes réunis dans le corpus.

phrase - (sa) fragment de texte compris entre deux séparateurs de phrase.

polyforme (sr)

archétype des occurrences d'un segment; suite de formes non séparées par un séparateur de séquence, qui n'est pas obligatoirement attestée dans le corpus.

ponctuation

Système de signes servant à indiquer les divisions d'un texte et à noter certains rapports syntaxiques et/ou conditions d'énonciation.

(sa) caractère (ou suite de caractères) correspondant à un signe de ponctuation.

pourcentages d'inertie (ac ou acm)

quantités proportionnelles aux valeurs propres dont la somme est égale à 100. Notées ta.

profil (stat et ac) (d'une ligne ou d'une colonne d'un tableau à double entrée)

vecteur constitué par le rapport des effectifs contenus sur cette ligne (resp. colonne) à la somme des effectifs que contient la ligne (resp. la colonne).

répartition (sa) (des occurrences d'une forme dans les parties du corpus)

nombre des parties du corpus dans lesquelles cette forme est attestée.

section (sr)

portion de texte comprise entre deux délimiteurs de section (exemple : le paragraphe, etc.).

segment (sr)

toute suite d'occurrences consécutives dans le corpus et non séparées par un <u>séparateur</u> de séquence est un segment du texte.

segment répété (sr) (ou polyforme répétée)

suite de forme dont la fréquence est supérieure ou égale à 2 dans le corpus.

segmentaire (sr)

ensemble des termes* attestés dans le corpus.

segmentation

opération qui consiste à délimiter des unités minimales dans un texte.

segmentation automatique

ensemble d'opérations réalisées au moyen de procédures informatisées qui aboutissent à découper, selon des règles prédéfinies, un texte stocké sur un support lisible par un ordinateur en unités distinctes que l'on appelle des <u>unités minimales</u>.

séparateurs de phrases (sa)

sous-ensemble des <u>caractères</u> <u>délimiteurs</u> de séquence* correspondant aux seules ponctuations fortes (en général : le point, le point d'interrogation, le point d'exclamation).

séquence (sa)

suite d'occurrences du texte non séparées par un délimiteur* de séquence.

seuil (stat)

quantité arbitrairement fixée au début d'une expérience visant à sélectionner parmi un grand nombre de résultats, ceux pour lesquels les valeurs d'un indice numérique dépassent ce seuil (de fréquence, en probabilité, etc.).

sous-fréquence (sa) (d'une unité textuelle dans une partie, tranche, etc.)

nombre des occurrences de cette unité dans la seule partie (resp. tranche, etc.) du corpus.

sous-segments (sr)

pour un segment donné, tous les segments de longueur inférieure et compris dans ce segment sont des soussegments. ex : AB et BC sont deux sous-segments du segment ABC.

spécificité chronologique (sp)

spécificité* portant sur un groupe connexe de parties d'un corpus muni d'une partition longitudinale.

spécificité positive (sp)

pour un seuil de spécificité fixé, une forme i et une partie j données, la forme i est dite spécifique positive de la partie j (ou <u>forme caractéristique</u> de cette partie) si sa sous-fréquence est "anormalement élevée" dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou supérieures à la sous-fréquence constatée est inférieure au seuil fixé au départ.

spécificité négative (sp)

pour un seuil de spécificité fixé, une forme i et une partie j données, la forme i est dite spécifique négative de la partie j si sa sous-fréquence est anormalement faible dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou inférieures à la sous-fréquence constatée est inférieure au seuil fixé au départ.

stock distributionnel du vocabulaire (d'un fragment de texte)

le vocabulaire* de ce fragment assorti de comptages de fréquence pour chacune des formes entrant dans sa composition.

syntagmatique (sa)

qui concerne le regroupement des unités textuelles, selon leur ordre de succession dans la chaîne écrite.

syntagme (ling)

groupe de mots en séquence formant une unité à l'intérieur de la phrase.

tableau de contingence (stat)

synonyme de tableau de fréquences ou de tableau croisé: tableau dont les lignes et les colonnes représentent respectivement les modalités de deux questions (ou deux variables nominales), et dont le terme général représente le nombre d'individus correspondant à chaque couple de modalités.

tableau lexical entier (TLE)

tableau à double entrée dont les lignes sont constituées par les <u>ventilations</u> des différentes formes dans les parties du corpus. Le terme générique k(i,j) du TLE est égal au nombre de fois que la forme i est attestée dans la partie j du corpus. Les lignes du TLE sont triées selon l'<u>ordre lexicométrique</u> des formes correspondantes.

tableau des segments répétés (TSR)

tableau à double entrée dont les lignes sont constituées par les <u>ventilations</u> des segments répétés dans les parties du corpus. Les lignes du TSR sont triées selon l'ordre lexicométrique* des segments. (i.e. longueur décroissante, fréquence décroissante, ordre lexicographique).

tableau lexical

tableau à double entrée résultant du TLE par suppression de certaines lignes (par exemple celles qui correspondent à des formes dont la fréquence est inférieure à un seuil donné).

taille (sa) (d'un corpus)

sa longueur* mesurée en occurrences (de formes simples).

terme (sr)

nom générique s'appliquant à la fois aux <u>formes</u> et aux <u>polyformes</u>. Dans le premier cas on parlera de termes de longueur 1. Les polyformes sont des termes de longueur 2,3, etc.

termes contraints / termes libres

un terme S1 est contraint dans un autre terme S2 de longueur supérieure si toutes ses <u>occurrences</u> sont des <u>soussegments</u> de segments correspondant à des occurrences du segment S2. Si au contraire un terme possède plusieurs expansions distinctes, qui ne sont pas forcément récurrentes, c'est un terme libre.

types généralisés (Tgens)

unités de dépouillement définies par l'utilisateur à l'aide d'outils permettant d'effectuer automatiquement des regroupements d'occurrences du texte (ex : les occurrences des formes qui commencent par la séquence de caractère patr : patrie, patriotes, patriotisme, etc.).

unités minimales (pour un type de segmentation)

unités que løn ne décompose pas en unités plus petites pouvant entrer dans leur composition (ex : dans la segmentation en formes graphiques les formes ne sont pas décomposées en fonction des caractères qui les composent)

valeur modale (stat)

valeur pour laquelle une distribution atteint son maximum.

valeurs propres (ac ou acm)

quantités permettant de juger de l'importance des facteurs successifs de la décomposition factorielle. La valeur propre notée la. mesure la dispersion des éléments sur l'axe.a.

valeurs-tests (ac ou acm)

quantités permettant d'apprécier la signification de la position d'un <u>élément supplémentaire</u> (ou illustratif) sur une axe factoriel. Brièvement, si une valeur test dépasse 2 en valeur absolue, il y a 95 chances sur 100 que la position de l'élément correspondant ne puisse être due au hasard.

variables actives

variables utilisées pour dresser une typologie, soit par analyse factorielle, soit par classification. Les typologies dépendent du choix et des poids des variables actives, qui doivent de ce fait constituer un ensemble homogène.

variables supplémentaires (ou illustratives)

variables utilisées a posteriori pour illustrer des plans factoriels ou des classes. Une variable supplémentaire peut-être considérée comme une variable active munie d'un poids nul.

variables de type T

variable dont la fréquence est à peu près proportionnelle à l'allongement du texte. (ex : la fréquence maximale)

variables de type V

variable dont l'accroissement a tendance à diminuer avec l'allongement du texte (ex : le nombre des formes, le nombre des hapax).

ventilation (sa) (des occurrences d'une unité dans les parties du corpus)

La suite des n nombres (n = nombre de parties du corpus) constituée par la succession des sous-fréquences* de

cette unité dans chacune des parties, prises dans l'ordre des parties.

vocabulaire (sa)

ensemble des formes attestées dans un corpus de textes.

vocabulaire commun (sa)

l'ensemble des formes attestées dans chacune des parties du corpus.

vocabulaire de base (sp)

ensemble des formes du corpus ne présentant, pour un seuil fixé, aucune spécificité (négative ou positive) dans aucune des parties , (i.e. l'ensemble des formes qui sont "banales" pour chacune des parties du corpus).

vocabulaire original (sp) (pour une partie du corpus)

l'ensemble des formes originales pour cette partie.

voisinage d'une occurrence (sa)

pour une occurrence donnée du texte, tout segment (suite d'occurrences consécutives, non séparées par un délimiteur de séquence) contenant cette occurrence.





http://www.tal.univ-paris3.fr/trameur/