

# Annotations *Rhapsodie* pour le *Trameur*

10/06/2015 19:03:24

Serge Fleury

(document de travail)

## Références

- **Le Trameur, manuel d'utilisation**

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

Dans cette documentation, la partie « *Relations de dépendance entre les items de Trame (via leurs annotations)* » présente l'état des développements actuellement disponibles dans le *Trameur* pour l'exploitation des relations de dépendance.

- **Le Trameur. Propositions de description et d'implémentation des objets textométriques**

<http://www.tal.univ-paris3.fr/trameur/trameur-propositions-definitions-objets-textometriques.pdf>

Ce document met au jour une description des objets textométriques et les méthodes mises en œuvre 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.

- **Description du format tabulaire du TreeBank Rhapsodie** (Ilaine Wang, Rachel Bawden). Version : morpho-syntaxe, micro-syntaxe, macro-syntaxe, prosodie) May 10, 2015

- **Annotations Rhapsodie pour le Trameur**

<http://www.tal.univ-paris3.fr/trameur/bases/rhapsodie2trameur.pdf>

<http://www.tal.univ-paris3.fr/trameur/bases/rhapsodie2trameur-v4.pdf>

<http://www.tal.univ-paris3.fr/trameur/bases/rhapsodie2trameur-v6.pdf>

Ces documents présentent les différentes étapes du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique ; présentation des différents processus de traitements des annotations de dépendance.

Le contenu du document actuel complète et enrichit la présentation faite dans les 2 références précédentes.

# 1. Préambule

Ce document commence par décrire le processus de transcodage des données issues du projet *Rhapsodie* (<http://projet-rhapsodie.fr/>) sous la forme d'une base textométrique importable dans le *Trameur*. Il présente ensuite différentes fonctionnalités mises en œuvre pour traiter ce type de données via le *Trameur*.

# 2. Données Rhapsodie

Les données traitées sont disponibles sur la page du projet *Rhapsodie*.

Par exemple ici :

- Téléchargement des fichiers de codage microsyntactique version bêta 10/13 ([zip](#))
- Tutoriel codage microsyntactique ([pdf](#))

Le fichier complet d'annotations a l'allure suivante (lecture partielle ici dans un tableur) :

Ce fichier tabulé est composé de 63 colonnes.

Ces données sont constitués par un certain nombre de textes (l'identifiant du texte (Text\_id) est visible dans la première colonne), chacun d'eux est segmenté en « unité englobante » / arbre (UE, seconde colonne), chacune d'elle est segmentée en token (troisième colonne), chacun d'eux est annoté (les autres colonnes).

Les données issues de Rhapsodie sont décrites dans le tableau suivant :

source : Description du format tabulaire du TreeBank *Rhapsodie* (Ilaine Wang, Rachel Bawden)

1	Text ID	le nom du texte (D0001, M2006 etc.)
2	Tree ID	le numéro de l'arbre dans le texte
3	Token ID	le numéro du token dans l'arbre
4	Token	la forme du token. Des lexèmes composés de plusieurs mots orthographiques ont été segmentés en tokens individuels. Un token est donc un segment de la transcription compris entre deux blancs ou un blanc et un signe de ponctuation. Tous les caractères qui ne sont pas des lettres (les espaces, les tirets et les apostrophes) sont considérés comme des tokens individuels aussi.
5	Speaker	l'identifiant du locuteur. En cas de chevauchement, on peut avoir plusieurs locuteurs (annotés alors \L1-\L3 par exemple).
6	Word span	la position du token dans le mot forme. La valeur est soit B (begin) pour le premier token d'un mot, soit I (inner) pour les tokens qui en sont pas les premiers tokens du mot.
7	Word form	le mot-forme auquel appartient le token. Dans le cas d'un mot forme comprenant plusieurs tokens, le mot-forme est uniquement marqué pour le premier token.
8	Lemma	le lemme du lexème auquel appartient le token. Dans le cas où il y a plusieurs tokens qui font partie du même lexème, le lemme n'est pas répété : le lemme est écrit dans cette colonne pour la ligne correspondant au premier token du lexème
9	POS	la catégorie morpheo-syntaxique du mot auquel appartient le token parmi N, V, Adj, Adv, I, P, D, Cl, Pro, CS, Qu.
10	Mood	le mode pour les verbes parmi 'indicative', 'subjunctive', 'imperative', 'past_participle' et 'present_participle'. Dans le cas où la forme est ambiguë, les deux possibilités de mode sont indiquées (ex : indicative/subjunctive).
11	Tense	le temps grammatical du verbe parmi 'present', 'future', 'conditional', 'imperfect' et 'perfect'. Le temps est marqué uniquement pour les verbes qui ont pour mode 'indicative'.
12	Person	la personne grammaticale pour les verbes et les pronoms personnels, ('1', '2' ou '3'). En cas d'ambiguïté, les personnes possibles sont toutes écrites séparées par des barres obliques (ex : '1/2/3').
13	Number	le nombre grammatical ('sg' ou 'pl' ou 'sg/pl' en cas d'ambiguïté) pour les verbes conjugués, les noms, les adjectifs, les pronoms et certains mots qu- (quel, quels, laquelle etc.).
14	Gender	le genre grammatical ('masc', 'fem' ou 'masc/fem' en cas d'ambiguïté) pour les noms, les adjectifs, les participes passés et certains mots qu- (quel, quels, laquelle etc.).
15	ID dep	le numéro du gouverneur par dépendance. Le numéro du gouverneur correspond à la colonne Token ID. Dans le cas où un gouverneur est constitué de plusieurs tokens, c'est le Token ID du premier token qui est pris comme numéro de gouverneur. Ce principe tient aussi pour les autres types de liens de dépendance.
16	Type dep	le type de lien de dépendance correspondant à ID dep.
17	ID plain	le numéro du gouverneur par dépendance 'primitive'.
18	Type plain	le type de lien de dépendance (primitif), correspondant à ID plain (correspondant aux liens pred, root, sub, dep, obj, obl, ad). N.B. Il ne peut y avoir qu'un seul type de dépendance primitive et un seul gouverneur primitif par token.
19	ID para	le numéro du gouverneur par lien paradigmatique.
20	Type para	le type de lien paradigmatique (parmi les types para_disfl, para_coord, para_intens, para_dform, para_reform, para_hyper, para_negot) N.B. Il ne peut y avoir qu'un seul type de dépendance paradigmatique et un seul gouverneur

		paradigmatique par token.
21	Gov inher	le numéro du gouverneur par lien hérité.
22	Type inher	le type de lien hérité (parmi pred_inherited, root_inherited, sub_inherited, dep_inherited, obj_inherited, obl_inherited, ad_inherited). N.B. Il ne peut y avoir qu'un seul type de dépendance par token, mais il peut y avoir plusieurs gouverneurs par dépendance héritée. Dans ce cas, les numéros des gouverneurs sont séparés par une virgule. Ex :
23	Gov junc	le numéro du gouverneur par lien `junc' (de jonction)
24	Type junc	le type de lien junc - il n'y en a qu'un seul, donc ceci correspond toujours à junc. Cette colonne est ici pour l'uniformité du tableau.
25	Gov junc-inher	le numéro du gouverneur par lien `junc inherited' (de jonction héritée)
26	Type junc-inher	le type de lien junc inherited - il n'y en a qu'un seul, donc ceci correspond toujours à junc_inherited. Cette colonne est ici pour l'uniformité du tableau.
27	Layer	l'appartenance à un entassement. Dans cette annotation, les différents niveaux d'entassement sont écrasés.
28	IU	l'appartenance à une unité illocutoire (UI, ou Illocutionary Unit en anglais). Cette information est bien fournie au format BILOU, mais on note que pour cette unité, la valeur O n'est pas utilisée du fait que tous les tokens font nécessairement partie d'une UI.
29	Nucleus	l'appartenance à un noyau.
30	Prenucleus	l'appartenance à un pré-noyau.
31	Gov prenuceus	l'appartenance à un pré-noyau régi, c'est-à-dire qui fait partie de la même unité rectionnelle que l'UI à laquelle il est rattaché.
32	Innucleus	l'appartenance à un in-noyau.
33	Gov innucleus	l'appartenance à un in-noyau régi.
34	Postnucleus	l'appartenance à un post-noyau.
35	Gov postnucleus	l'appartenance à un post-noyau régi.
36	IU parenthesis	l'appartenance à une UI parenthétique.
37	IU graft	l'appartenance à une greffe, qui correspond généralement à un discours rapporté ou bien à des séquences qui contribuent à remplir la position syntaxique d'un élément recteur.
38	IU embedded	l'appartenance à une unité enchâssée qui ne correspond pas à une UI (alors que c'est le cas de la greffe).
39	Associative nucleus	l'appartenance à une liste fermée de marqueurs discursifs établie pour le projet.
40	Intro IU	l'appartenance à une unité qui permet d'introduire une UI.
41	Period	l'appartenance à une période intonative. Cette information est donnée au format BILOU (décrit dans l'introduction de la section précédente) sans le O comme c'était le cas pour l'UI (28), puisque, de la même manière, tout token fait partie d'une période intonative. On y trouve toutefois un détail supplémentaire : en plus de marquer la position du token dans la période, chaque valeur peut également indiquer s'il s'agit d'une période tronquée à gauche (-B, -I, -L, -U), à droite (B-,I-,L-,U-), des deux côtés (-B-,-I-,-L-,-U-).
42	Period tone	le profil tonal (ou contour intonatif) de la période intonative dans laquelle se trouve le token.
43	Package	délimitation des paquets intonatifs en unités au format BILOU (sur le schéma des périodes et toujours sans le O puisque tout token fait forcément partie d'un paquet intonatif)
44	Package_type	le type de paquet intonatif dans lequel le token se trouve parmi filled-dis, filled-

		pause, included, lone, lone-dis-strong, motherless, motherless-dis-weak, silent-pause, ou tail.
45	Package tone	le profil tonal (ou contour intonatif) du paquet intonatif dans lequel se trouve le token.
46	Group	délimitation des groupes rythmiques en unités au format BILOU (toujours sans le O puisque tout token fait forcément partie d'un groupe rythmique)
47	Group_type	le type de groupe rythmique dans lequel le token se trouve parmi dis-strong, dis-weak, filled-dis, filled-pause, silent-pause, strong, tail, ou weak.
48	Group tone	le profil tonal (ou contour intonatif) du groupe rythmique dans lequel se trouve le token.
49	Foot	délimitation des unités de pieds métriques au format BILOU (toujours sans le O)
50	Foot_type	le dernier pied métrique du token. à l'instar du groupe rythmique, il peut être annoté dis-strong, dis-weak, filled-dis, filled-pause, silent-pause, strong, tail, ou weak.
51	Foot tone	le profil tonal (ou contour intonatif) du dernier pied métrique du token.
52	Syllable	délimitation des unités syllabiques au format BILOU (toujours sans le O) analysées dans Syllable tone. Etant donné que l'on n'étudie dans ce tabulaire que la dernière syllabe de chaque token, on note simplement U pour un token constitué d'un mot ou d'une syllabe, mais on regroupe plusieurs mots constituant un seul token (d+'+abord), ou encore plusieurs tokens constituant une unique syllabe (ex : de la [dla]).
53	Syllable tone	le profil tonal (ou contour intonatif) de la dernière syllabe du token.
54	Prominence initial	le degré de proéminence de la première syllabe du token. Une proéminence peut être annotée W pour Weak ou S pour Strong. Dans le cas où cette syllabe n'est pas proéminente, elle peut avoir comme valeur 0, _ (pause) ou encore % (syllabe inaudible ou non transcrite en raison d'un chevauchement).
55	Prominence final	le degré de proéminence de la dernière syllabe du token. Elle peut avoir les mêmes valeurs que pour la proéminence initiale.
56	Hesitation	marquée H pour la particule \"euh\" ou bien pour une syllabe hésitante, _ (pause) ou bien % (mot inaudible ou non transcrit en raison d'un chevauchement).
57	Pause length	la durée de la pause (en s), indiquée au niveau du token qui la précède. Autrement dit, si cette donnée est vide pour un token donné, alors il n'est pas suivi d'une pause. à noter que dans le cas précis d'un chevauchement et si les paroles du locuteur principal sont transcrites en premier, la longueur de la pause est indiquée sur le dernier mot, à la _n du chevauchement mais on a ajouté le signe dièse sur la ligne du mot après lequel on observe réellement une pause. Ce choix a été fait afin de préserver un alignement pertinent avec les différentes unités d'analyse, notamment la période.
58	Tmin	le temps de début de chaque token (ou le cas échéant, du chevauchement) au sein de son échantillon.
59	Tmax	le temps de fin de chaque token (ou le cas échéant, du chevauchement) au sein de son échantillon.
60	Syllable length	la durée de la dernière syllabe du token (en ms).
61	Syllable length avg	la durée moyenne de cette syllabe (en ms).
62	Pitch	la hauteur (en demi-tons) de cette syllabe.
63	Pitch avg	la hauteur moyenne (en demi-tons), calculée sur un empan de quelques syllabes précédant et suivant la dernière syllabe du token.

Les annotations sont réutilisées telles quelles par le processus de transcodage vers *le Trameur*. Les annotations de dépendance (16 à 26) qui sont réutilisées 2 à 2 (*cf* jeu de couleur) pour construire respectivement une seule annotation (de type relation) qui est réécrite par exemple sous la forme : `Type_rection(Gov-rection)`. De même, les annotations prosodiques (`Group`, `Period`, `Package`, `Foot`, `Syllable`) seront utilisées pour construire d'autres annotations par croisement d'annotations sur chacun des groupes d'annotation visés (*cf infra*).



Les annotations de dépendance sont fusionnées deux à deux sous la forme d'une relation marquée de la manière suivante :

RELATION(CIBLE) (cf documentation *Trameur*)

RELATION est une chaîne portant le nom de la relation visée

CIBLE est une valeur numérique pointant vers une position de la *Trame*

Pour certaines annotations prosodiques (*Period*, *Group*, *Package*, *Foot*, *Syllable*), une annotation supplémentaire est créée par concaténation d'autres informations :

```

typePeriod      =   Period + Period_tone
typeGroup1      =   Group + Group_type
typeGroup2      =   Group + Group_type + Group_tone
typePackage1    =   Package + Package_type
typePackage2    =   Package + Package_type + Package_tone
typeFoot1       =   Foot + Foot_type
typeFoot2       =   Foot + Foot_type + Foot_tone
typeSyllable    =   Syllable + Syllable_tone
    
```

Les annotations créées (par croisement de colonnes) seront portées par le token associé à un marquage de type B ou U dans le groupe d'annotation visé.

Par exemple, si un token porte la valeur B (ou la valeur U) dans l'annotation *Package*, la valeur *included* dans l'annotation *Package\_type* et la valeur *Hm* dans l'annotation *Package\_tone* alors l'annotation *typePackage1* portera la valeur *B\_included* et l'annotation *typePackage2* portera la valeur *B\_included\_Hm*.

**Cette annotation construite par croisement de colonnes permet dans cet exemple de mettre au jour les packages ayant telle type (*included*) et les packages ayant tel type (*included*) et telle valeur (*Hm*).**

Les autres tokens porteront une « valeur neutre » notée – dans les annotations *typePackage1*, *typePackage2*.

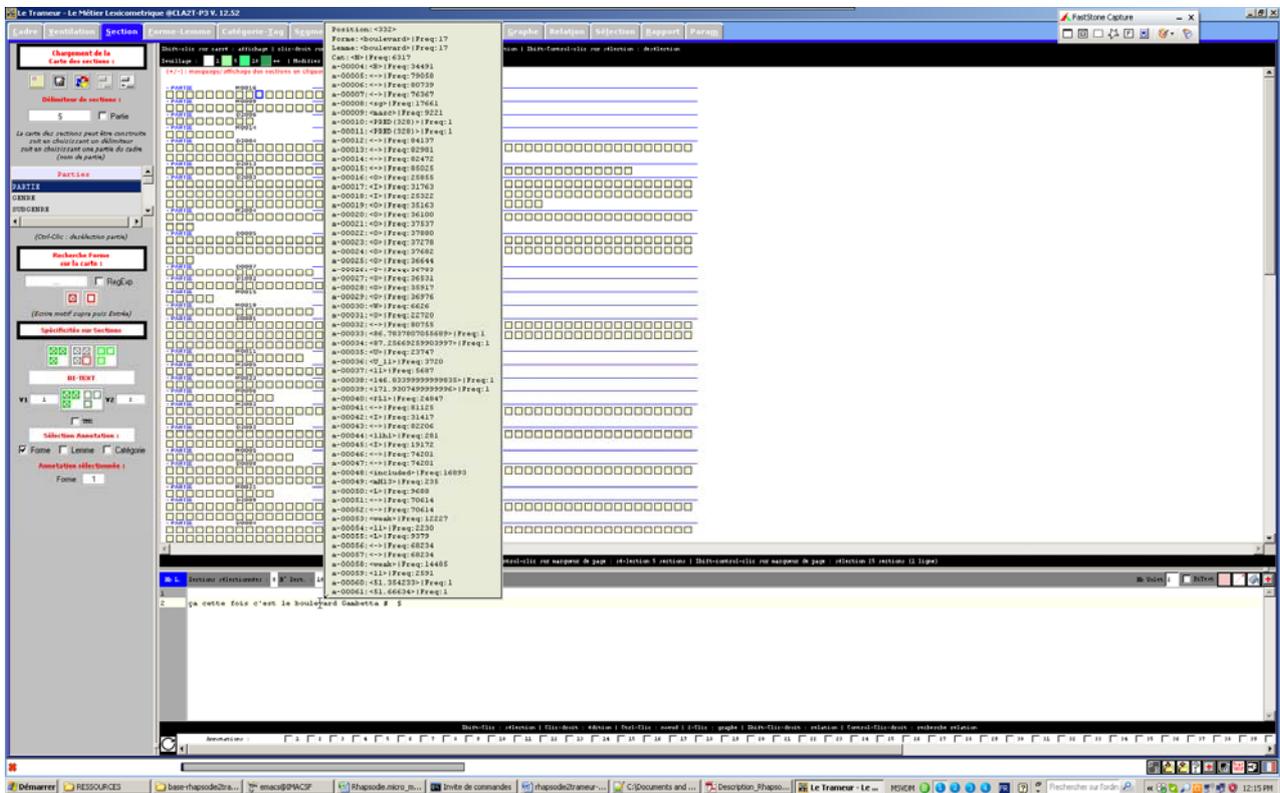
Chaque item de la base textométrique est associé *in fine* à 61 annotations décrites dans le tableau suivant :

n°Annotation	Annotation trameur	Contenu Rhapsodie
1	Forme	Token
2	Lemme	Lemma
3	Cat	POS
4	a-00004	Word_span
5	a-00005	Mood
6	a-00006	Tense
7	a-00007	Person
8	a-00008	Number
9	a-00009	Gender

10	a-00010	Type_dep(ID_dep)
11	a-00011	Type_plain(ID_plain)
12	a-00012	Type_junc(ID_junc)
13	a-00013	Type_para(ID_para)
14	a-00014	Type_inherited(ID_inherited)
15	a-00015	Type_junc_inherited(ID_junc_inherited)
16	a-00016	Layer
17	a-00017	IU
18	a-00018	Nucleus
19	a-00019	Prenucleus
20	a-00020	Gov_nucleus
21	a-00021	Innucleus
22	a-00022	Gov_innucleus
23	a-00023	Postnucleus
24	a-00024	Gov_postnucleus
25	a-00025	IU_parenthesis
26	a-00026	IU_graft
27	a-00027	IU_embedded
28	a-00028	Associated_nucleus
29	a-00029	Intro_IU
30	a-00030	Prominence_final
31	a-00031	Prominence_initial
32	a-00032	Hesitation
33	a-00033	Pitch_avg
34	a-00034	Pitch
35	a-00035	Syllable
36	a-00036	typeSyllable
37	a-00037	Syllable_tone
38	a-00038	Syllable_length
39	a-00039	Syllable_length_avg
40	a-00040	Speaker
41	a-00041	Pause_length
42	a-00042	Period
43	a-00043	typePeriod
44	a-00044	Period_tone
45	a-00045	Package
46	a-00046	typePackage1
47	a-00047	typePackage2
48	a-00048	Package_type
49	a-00049	Package_tone
50	a-00050	Group
51	a-00051	typeGroup1

52	a-00052	typeGroup2
53	a-00053	Group_type
54	a-00054	Group_tone
55	a-00055	Foot
56	a-00056	typeFoot1
57	a-00057	typeFoot2
58	a-00058	Foot_type
59	a-00059	Foot_tone
60	a-00060	Tmin
61	a-00061	Tmax

La figure suivante donne à voir, dans *le Trameur*, les annotations disponibles sur un item en contexte (le mot « boulevard ») :



### 3.3 Le Cadre textométrique

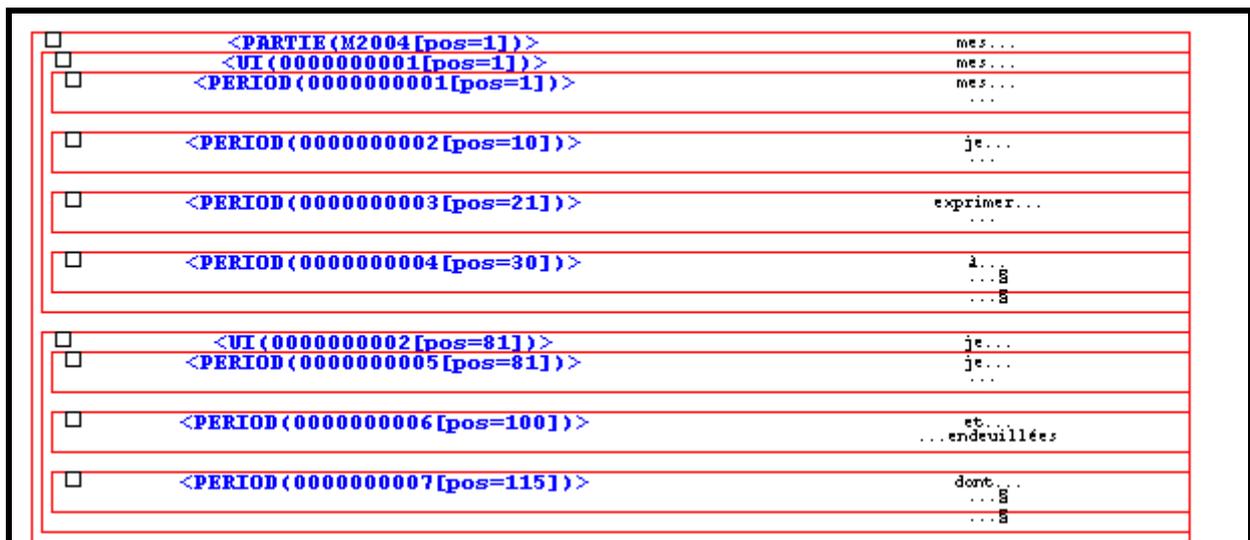
Les différents échantillons initiaux de *Rhapsodie* sont considérés comme autant de parties différentes : la base finale est donc une partition de textes (*Cadre*), chaque partie contient les zones textuelles associées à l'identifiant initial de l'échantillon.

Deux autres partitions permettent de contraster certaines zones textuelles du corpus : les UEs (Unités Englobantes) et les périodes.

#### Découpage des échantillons

- **PARTIE** : partition des échantillons
- **UE** : partition des UEs
- **PERIOD** : partition des périodes

La figure suivante illustre l'organisation des différentes parties et leur structuration logique :



Un échantillon est donc composé d'un ensemble d'UE, chaque UE est composé d'un ensemble de période.

### Catégorisation des échantillons

Chaque échantillon (cf `Text_id` vu précédemment) est catégorisé suivant différents points de vue : genre, type de conversation etc. Les différentes variables permettant de catégoriser les échantillons sont regroupées dans le tableau suivant :

Text_id	Subgenre	Interactivity	Social_context	Event_structure	Channel	Planning_type	Task
D0001	Argumentation	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0002	Argumentation	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0003	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	interview
D0004	Description	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0005	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	interview
D0006	Argumentation	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0007	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D0008	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D0009	Description	interactive	Private	Dialogue	Face_to_Face	spontaneous	Unknown
D0017	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D0020	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D1001	Description	semi-interactive	Private	Dialogue	Face_to_Face	planned	Unknown
D1002	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	lesson
D1003	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	life_story
D2001	Narrative	interactive	Public	Dialogue	Broadcasting	semi-spontaneous	interview
D2002	Argumentation	interactive	Public	Dialogue	Broadcasting	planned	interview
D2003	Description	semi-interactive	Public	Dialogue	Broadcasting	spontaneous	sportscasting
D2004	Narrative	semi-interactive	Public	Dialogue	Broadcasting	spontaneous	life_story
D2005	Narrative	semi-interactive	Public	Dialogue	Broadcasting	planned	interview
D2006	Oratory	semi-interactive	Public	Dialogue	Face_to_Face	planned	Political_speech
D2007	Argumentation	interactive	Public	Dialogue	Broadcasting	spontaneous	Conversation
D2008	Argumentation	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
D2009	Argumentation	semi-interactive	Public	Dialogue	Broadcasting	planned	interview
D2010	Narrative	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
D2011	Argumentation	semi-interactive	Public	Dialogue	Broadcasting	planned	Advertising
D2012	Description	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
D2013	Argumentation	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
M0001	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0002	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_description
M0003	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0004	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0005	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0006	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0007	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0008	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0009	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning



### 3.4 Repérage de Sections

Le processus de transcodage intègre aussi un marquage de sections :

- après chaque UE, un caractère délimiteur de section (§) est introduit pour permettre de construire dans le *Trameur* une représentation cartographique de la base sous la forme d'une *carte des sections* des UE
- Après chaque période, un caractère délimiteur de section (#) est introduit pour permettre de construire dans le *Trameur* une représentation cartographique de la base sous la forme d'une *carte des sections* des périodes

Un extrait de la carte des UE (avec mise au jour des échantillons)

- PARTIE	M2004	
- PARTIE	D2004	
- PARTIE	M0011	
- PARTIE	D0004	
- PARTIE	M0021	
- PARTIE	M2002	
- PARTIE	M2005	
- PARTIE	D2005	

Un extrait de la carte des périodes (avec mise au jour des UE) :

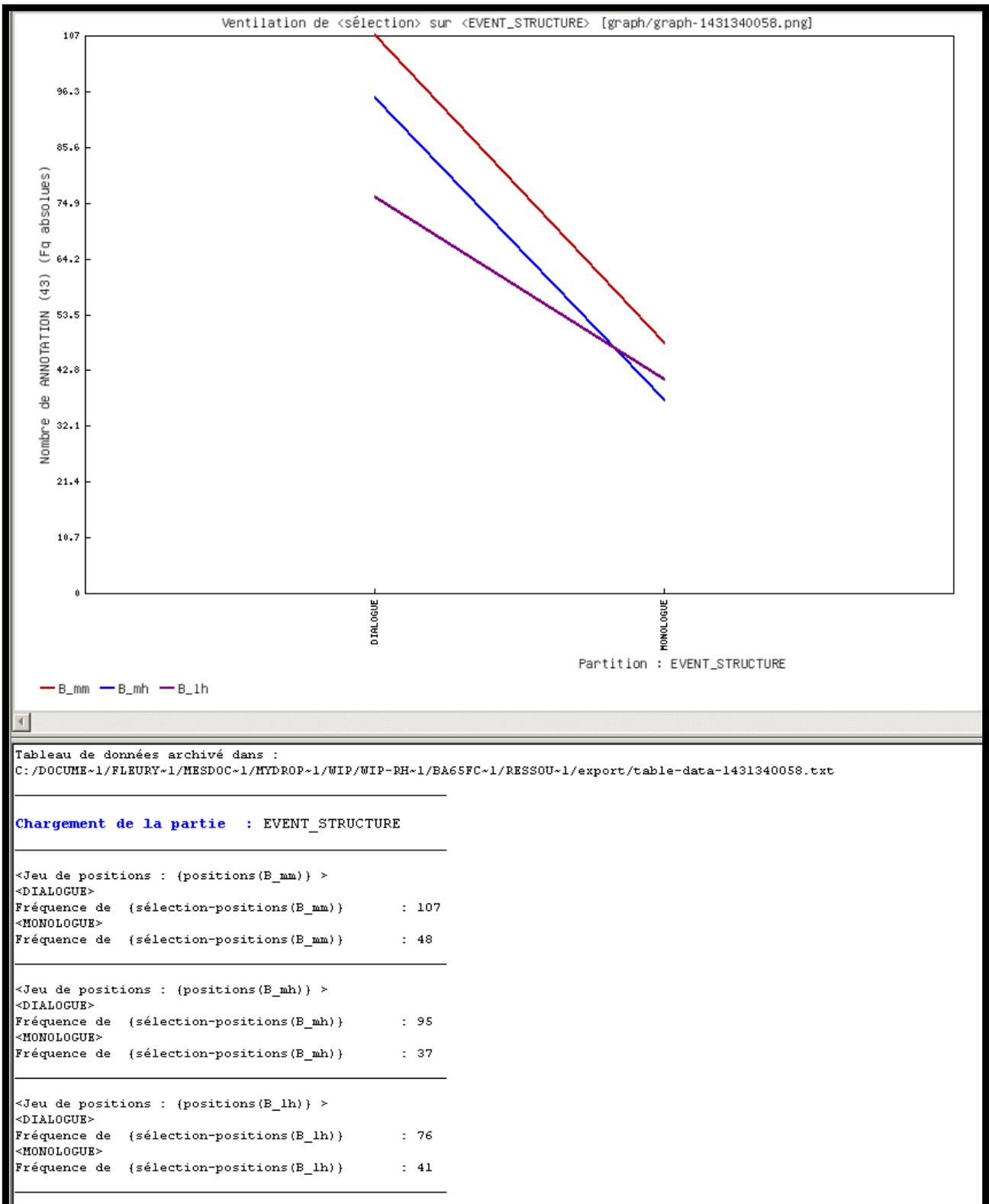
- UI	000000001	
- UI	000000002	
- UI	000000003	
- UI	000000004	
- UI	000000005	
- UI	000000006	
- UI	000000007	
- UI	000000008	
- UI	000000009	
- UI	000000010	
- UI	000000011	
- UI	000000012	
- UI	000000013	
- UI	000000014	

## 4. Ventilation : distribution d'unités

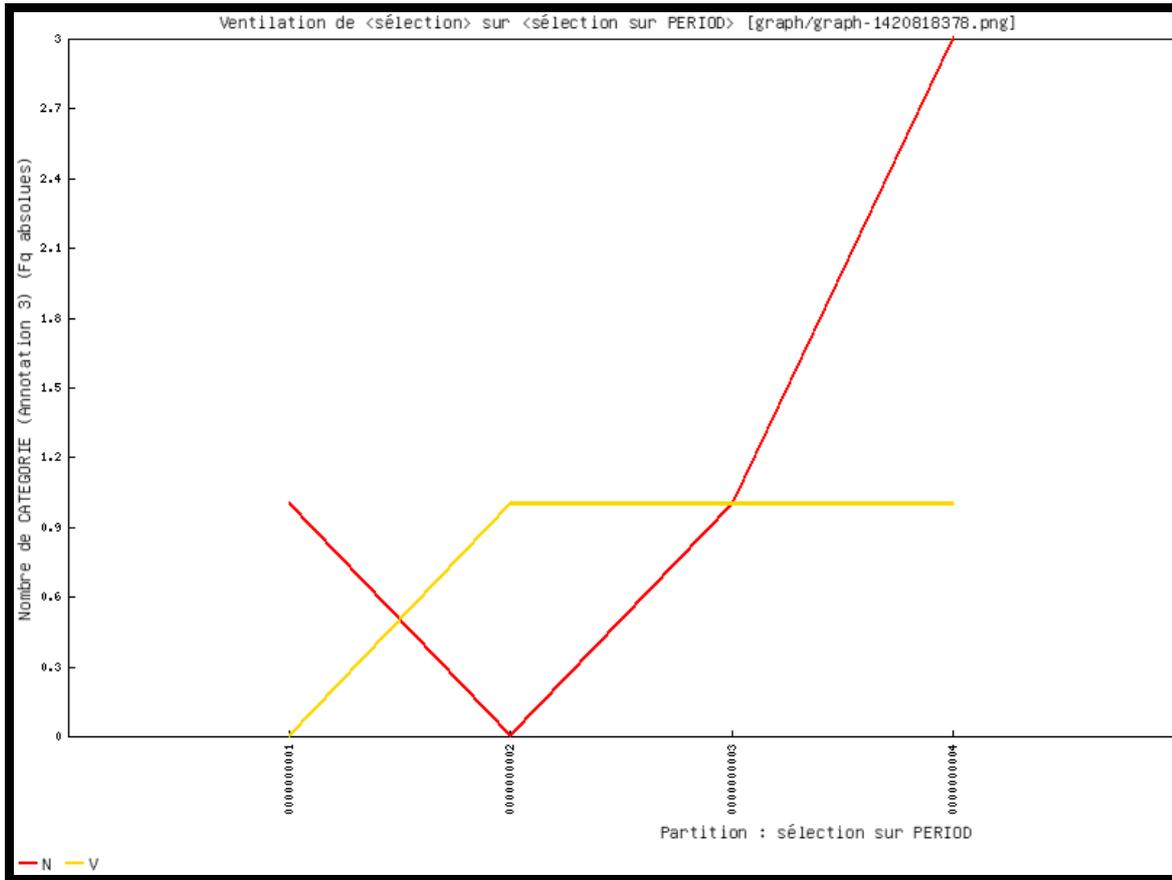
L'introduction de systèmes de parties permet de construire des graphiques de ventilation sur celles-ci.

La figure suivante met au jour un extrait des effectifs des items associés à annotation n°43 `typePeriod` puis la ventilation (en fréquence absolue) de certains d'entre eux (les items sélectionnés) sur la partition `EVENT_STRUCTURE`

Fq	Annotation importée	Nom	Num
82206	-	a-00043	43
155	B_nm	a-00043	43
132	B_mh	a-00043	43
117	B_lh	a-00043	43
110	B_ml	a-00043	43
98	B_mH	a-00043	43
93	B_lH	a-00043	43
83	B_hl	a-00043	43
74	B_lm	a-00043	43
54	B_mL	a-00043	43
50	B_hm	a-00043	43
47	B_mlh2	a-00043	43
46	B_mh12	a-00043	43
43	B_mH12	a-00043	43
43	B_ll	a-00043	43
40	B_mmh2	a-00043	43
40	B_llm2	a-00043	43
38	B_mm12	a-00043	43
35	B_lL	a-00043	43
34	B_mh13	a-00043	43
31	B_H1	a-00043	43
31	B_mmh1	a-00043	43
29	B_llh2	a-00043	43
29	B_mH13	a-00043	43
27	B_hL	a-00043	43
27	B_mm13	a-00043	43
26	B_mlh1	a-00043	43
26	B_llm1	a-00043	43
26	B_lLm2	a-00043	43
23	B_Lh	a-00043	43
22	B_Hm	a-00043	43
20	B_hh13	a-00043	43
20	B_mlh3	a-00043	43
19	B_mh11	a-00043	43
19	B_lmh2	a-00043	43
18	U_mm	a-00043	43

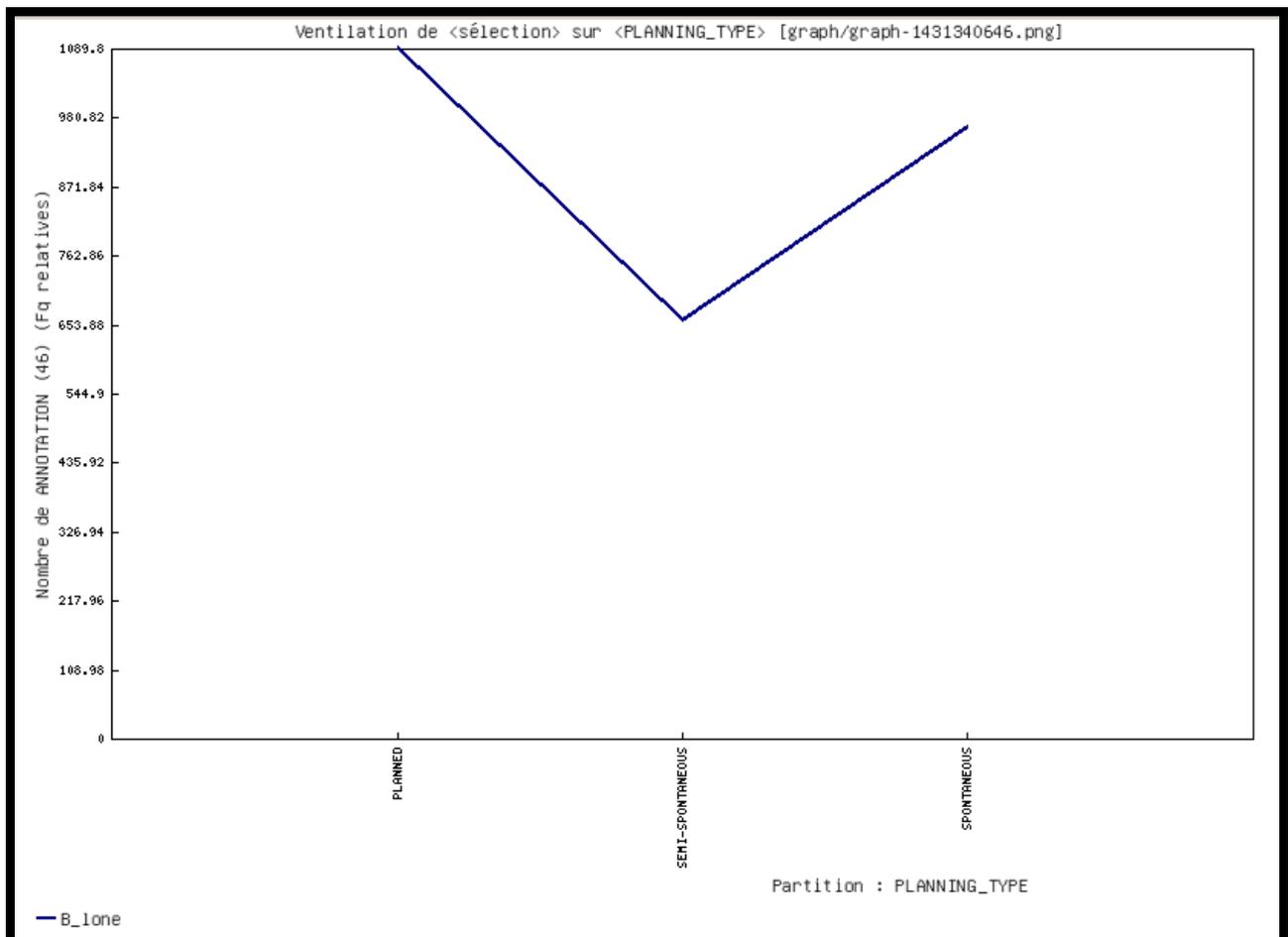


La figure suivante met au jour par exemple la ventilation des items N et V (annotation n°3 : POS) sur la partition PERIOD constituée par les 4 premières périodes de l'UE n°1 de l'échantillon M2004 :



Enfin, les figures suivantes mettent au jour les effectifs de l'annotation n°46 (typePackage1) puis la ventilation (en fréquence relative) de la valeur « B\_lone » sur la partition PLANNING\_TYPE :

Fq	Annotation importée	Nom	Num
74201	-	a-00046	46
3616	B_lone	a-00046	46
2853	B_included	a-00046	46
1548	U_lone	a-00046	46
488	B_motherless	a-00046	46
388	U_filled-pause	a-00046	46
368	B_tail	a-00046	46
253	B_lone-dis-strong	a-00046	46
236	B_filled-dis	a-00046	46
231	U_motherless	a-00046	46
228	U_tail	a-00046	46
202	B_motherless-dis-weak	a-00046	46
139	U_motherless-dis-weak	a-00046	46
129	U_lone-dis-strong	a-00046	46
110	U_filled-dis	a-00046	46
36	U_included	a-00046	46
19	B_filled-pause	a-00046	46
5	U_silent-pause	a-00046	46
4	B_silent-pause	a-00046	46

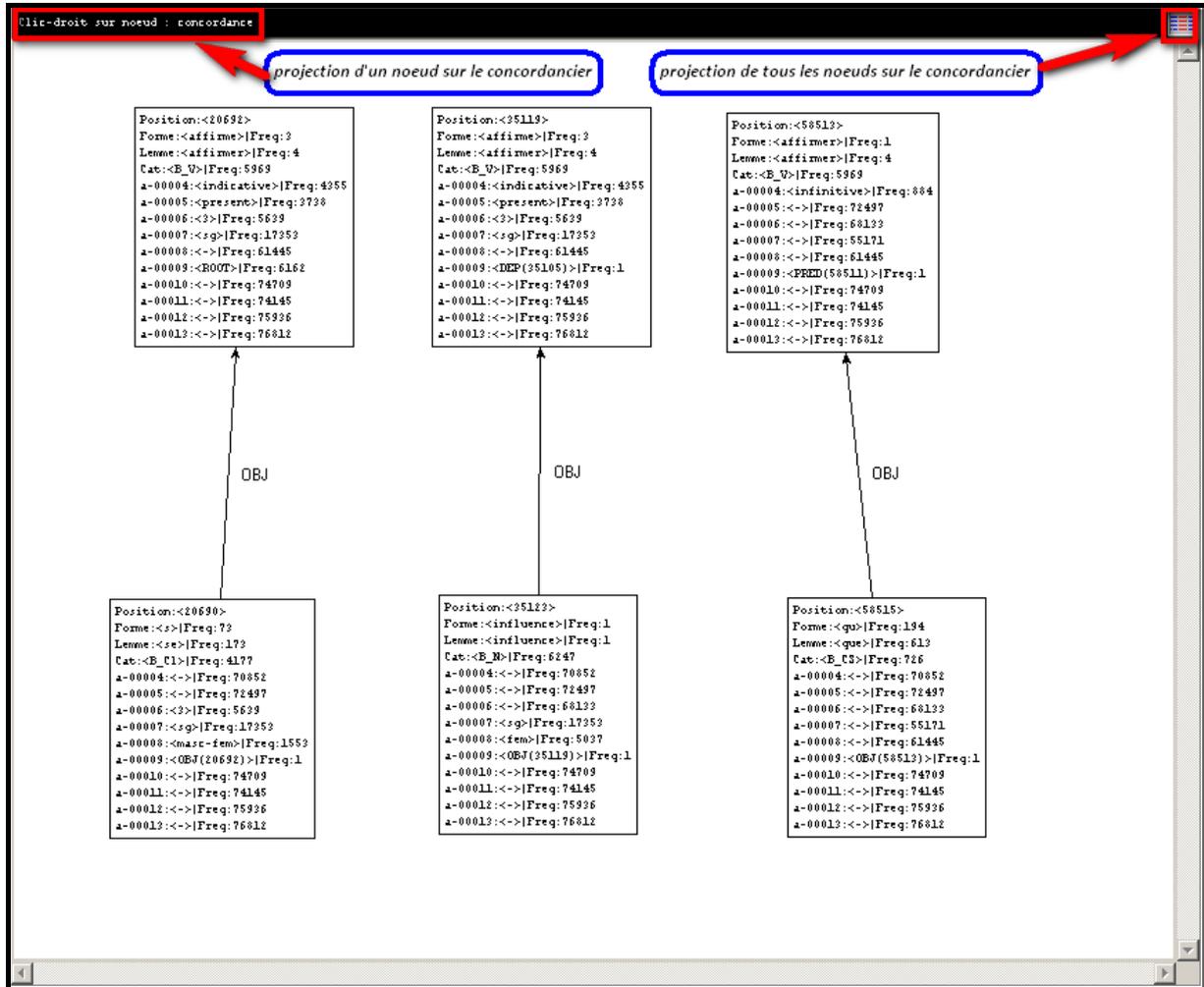






### 5.2 Retour en contexte

Un graphe de relations de dépendance étant produit, chaque nœud du graphe peut-être projeté sur le concordancier (clic-droit sur le nœud). L'ensemble des nœuds peut aussi être projeté globalement sur le concordancier :



La concordance produite dans ce dernier cas a l'allure suivante :

```

-----PARTIE{PARTIE=M2004}-----
§ un nouvel ordre international s'affirme peu à peu
-----PARTIE{PARTIE=M2001}-----
qu'une nation affirme son influence qu'elle pèse dans
-----PARTIE{PARTIE=D2013}-----
Rachida Dati pour oser affirmer qu'il ne s'agit
    
```

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 l'affichage des annotations dans la concordance en matérialisant par exemple la catégorie de chaque item :

```

-----PARTIE{PARTIE=M2004}-----
$ un-B_D nouvel-B_Adj ordre-B_N international-B_Adj s-B_Cl '-I_Cl affirme-B_V peu-B_Adv à-I_Adv peu-I_Adv
-----PARTIE{PARTIE=M2001}-----
qu-B_Qu '-I_Qu une-B_D nation-B_N affirme-B_V son-B_D influence-B_N qu-B_Qu '-I_Qu elle-B_Cl pèse-B_V dans-B_Pre
-----PARTIE{PARTIE=D2013}-----
Rachida-B_N Dati-B_N pour-B_Pre oser-B_V affirmer-B_V qu-B_CS '-I_CS il-B_Cl ne-B_Cl s-B_Cl '-I_Cl agit-B_V
    
```

Toutes les zones d'édition du *Trameur* permettent de mettre au jour les annotations de la *Trame* (cf documentation en ligne, partie « 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 d'extraire la relation OBJ (cf graphe infra) :

```

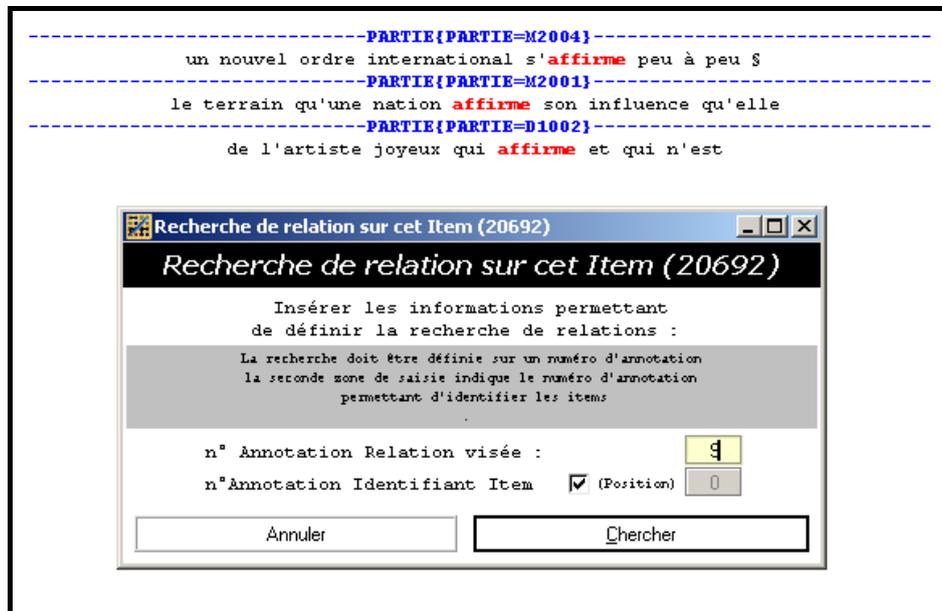
-----PARTIE{PARTIE=M006}-----
$ de la crise aux Antilles il ne devrait pas en être question au sommet social de l'Elysée cet
savamment minütée $ alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et
alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va
d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va 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 touchés 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 congrès $ sept cent quatre-vingt-sept milliards
$ sept cent quatre-vingt-sept milliards de 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 millions d'
Motors et Chrysler ont présenté hier leur plan de restructuration à l'administration Obama $ ils demandent vingt-deux milliards
toujours en garde à vue $ les enquêteurs cherchent maintenant à savoir si elle a été complice dans cette évasion
a été complice dans cette évasion $ elle 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'assise se prononce sur leur demande de supplément d'information $ lundi le
d- ce sont des procédés terroristes $ il le dit à Laurent Doulsant $ direction maintenant Barcelone pour un congrès
Laurent 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 de la zone de relégation après sa victoire sur Caen
sa victoire sur Caen deux zéro $ Nice s'impose à Nancy deux à un $
-----PARTIE{PARTIE=M002}-----
voit euh qui voit ça passer $ et donc elle se dit eh bah elle peut récupérer elle pourra récupérer
rencontrent $ et et donc elle elle tombe $ ils se $ ouais il y a un accident quoi
ouais il y a un accident quoi $ et ils se $ ils tombent tous les deux $ et c'
boulanger voit que la baguette a disparu $ et il se dit que $ ouais moi je pense qu'
va $ alors le euh euh le boulanger dit au policier qu'en fait c'est la fille $
    
```

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

The screenshot displays the Trameur software interface. On the left, there is a sidebar with various options like 'Forme', 'Lemme', and 'Catégorie'. The main area shows a concordance table with text from the 'Rhapsodie' corpus. A dialog box titled 'Sélection des annotations à colorier...' is open, allowing the user to select specific annotations to highlight in the text. The dialog box contains a table with columns for 'Annotation visée' and 'n°Annotation Identifiant Item (Division)'. The user has selected 'B\_N' and '3' in the 'Annotation visée' column and '4' in the 'n°Annotation Identifiant Item (Division)' column. A red arrow points from the dialog box to the corresponding text in the concordance table.

### 5.3 Recherche de dépendance en contexte

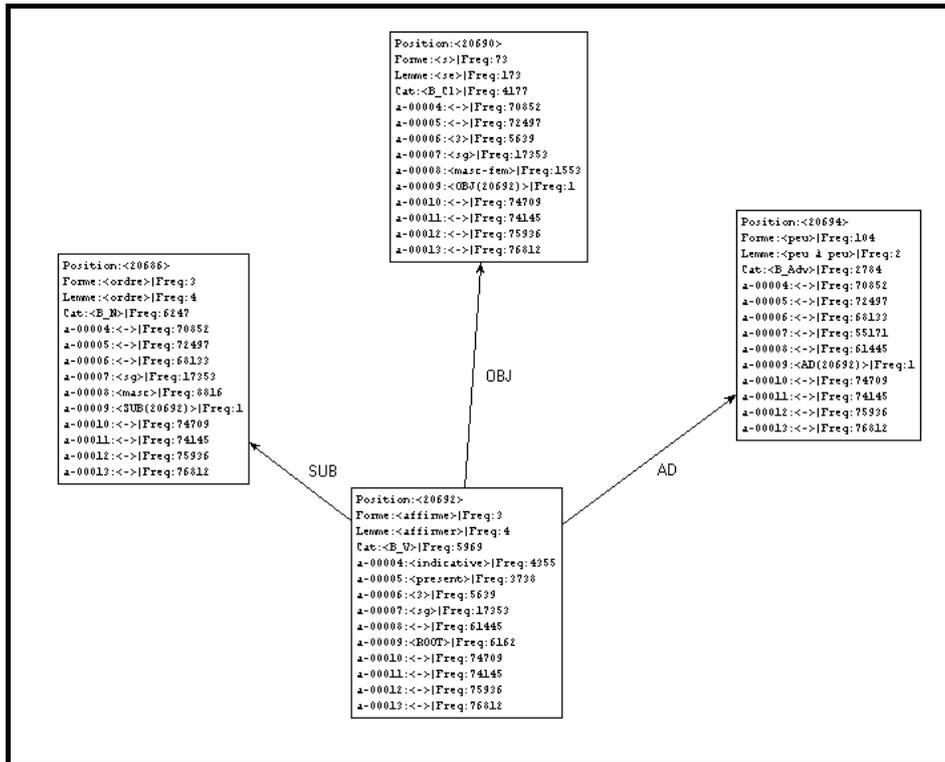
La recherche de dépendance peut aussi être réalisée en contexte, par exemple sur le concordancier. Si on considère la concordance suivante (en haut de la figure) :



Le raccourci clavier `Ctrl-Clic-droit` sur un item de la concordance (ici l'item 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 d'annotation portant ce type d'information : ici le n°9<sup>2</sup>) et comment sont indexer les items sur la trame (ici par leur position).

La recherche conduit à la production d'un graphe mettant au jour toutes les relations sur l'item visé :

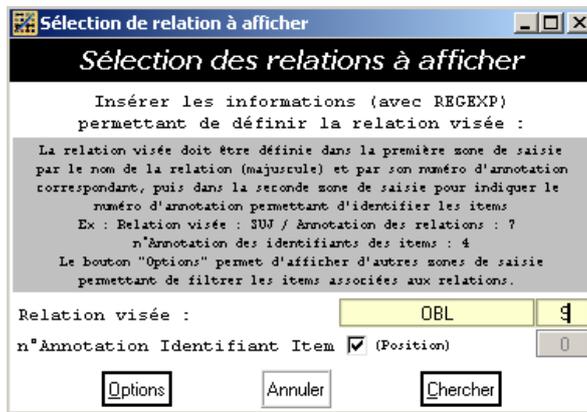
<sup>2</sup> Même remarque que dans la note précédente.



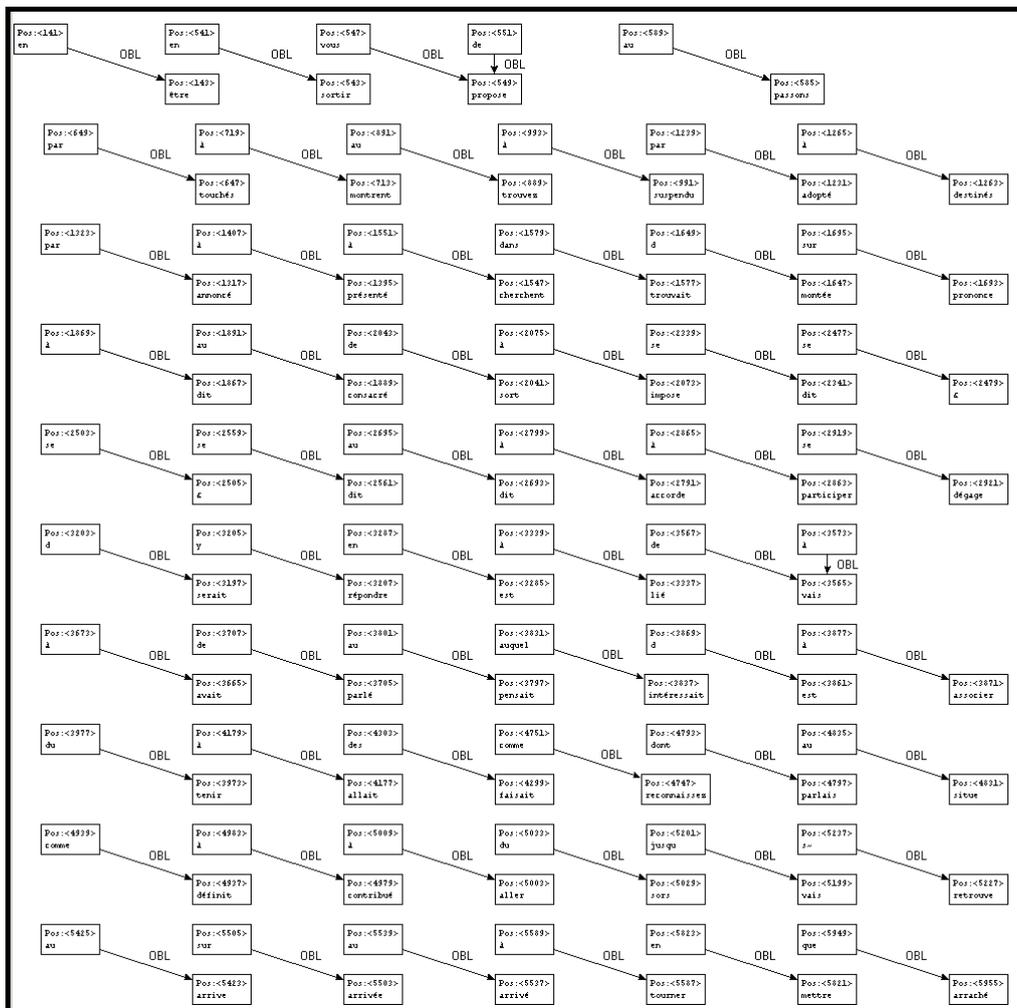
Le graphe donne à voir les 3 relations pointant sur l'item.

### 5.4 Rechercher dans un graphe de dépendance

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

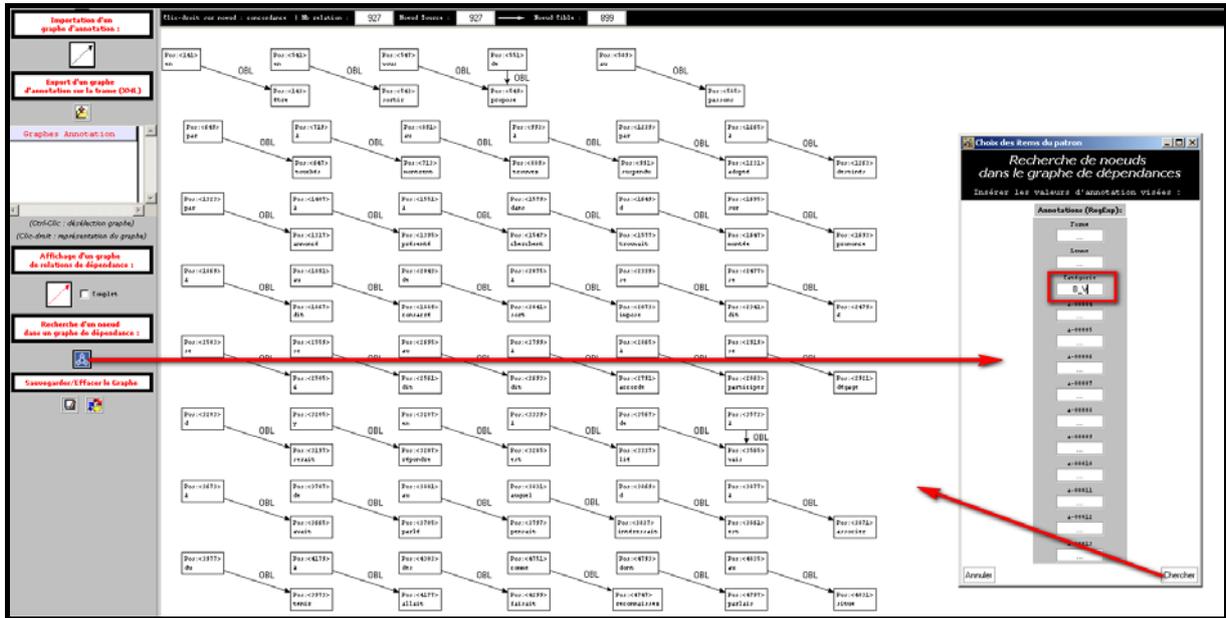


Le résultat produit a l'allure suivante :

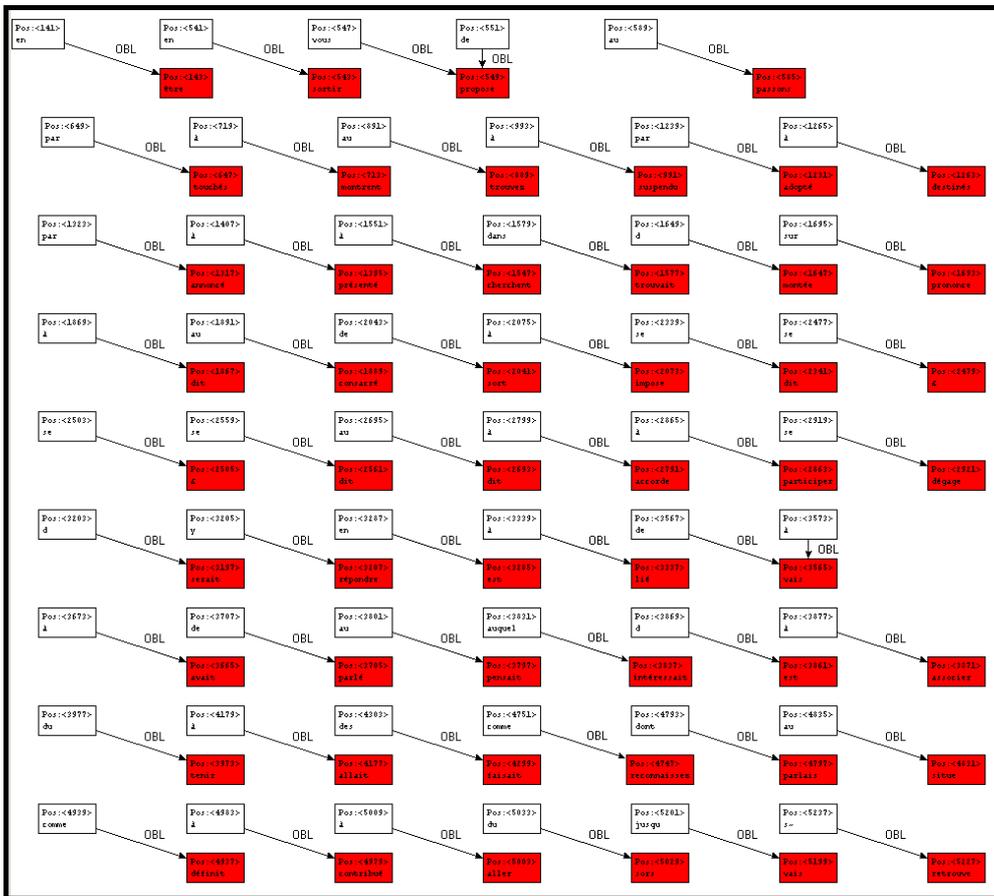


On peut mettre au jour dans ce graphe certains nœuds en filtrant leurs annotations.

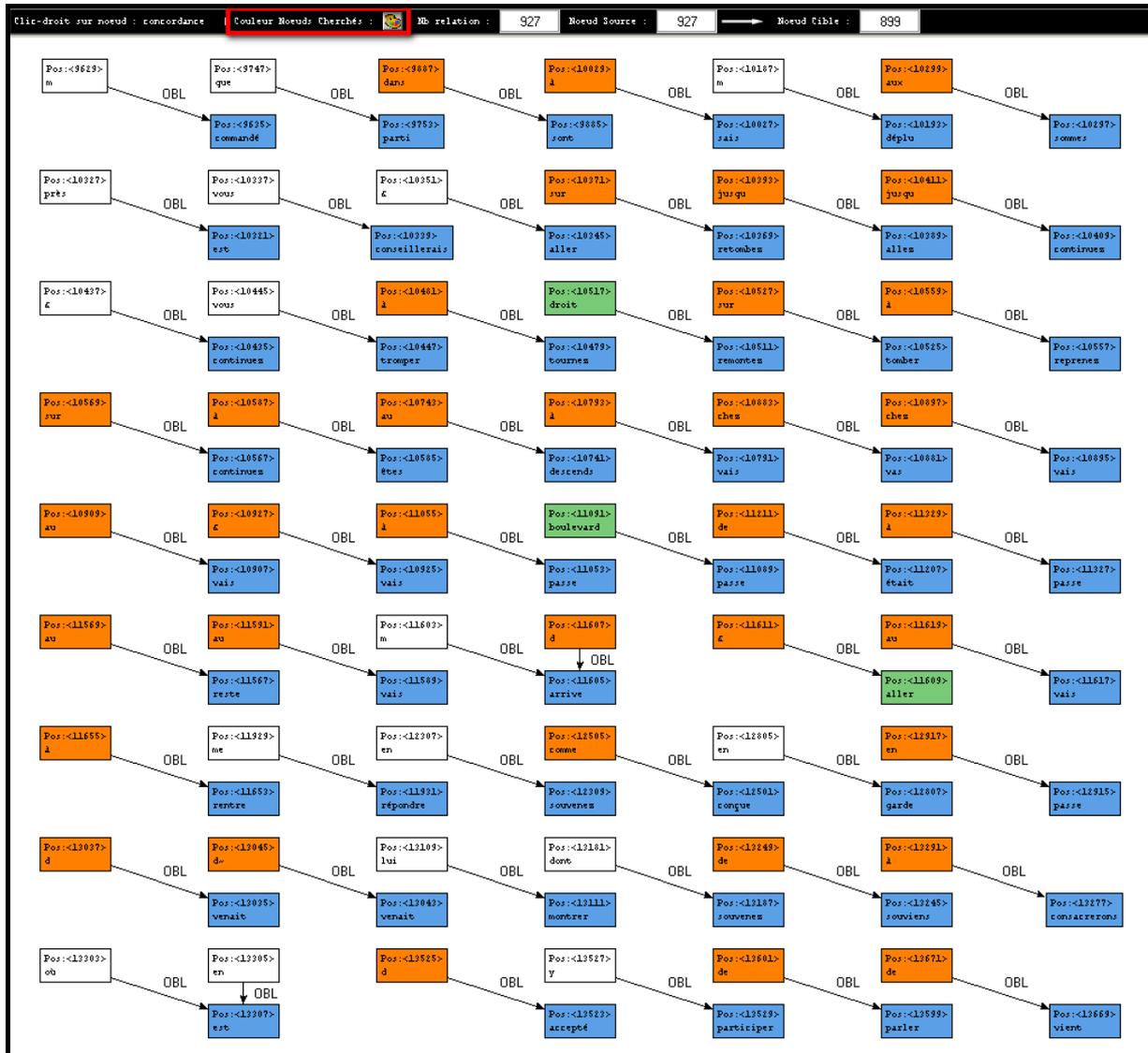
Le processus de filtrage des nœuds du graphe permet de sélectionner certains nœuds sur la base des différentes annotations disponibles. Dans l'exemple suivant, on veut mettre au jour les nœuds de catégorie B\_V :



Le résultat produit a l'allure suivante :



Dans la figure qui suit, 3 filtrages successifs sont réalisés pour marquer successivement les nœuds 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 nœuds visés (verbe en bleu, nom en vert et préposition en orange) :



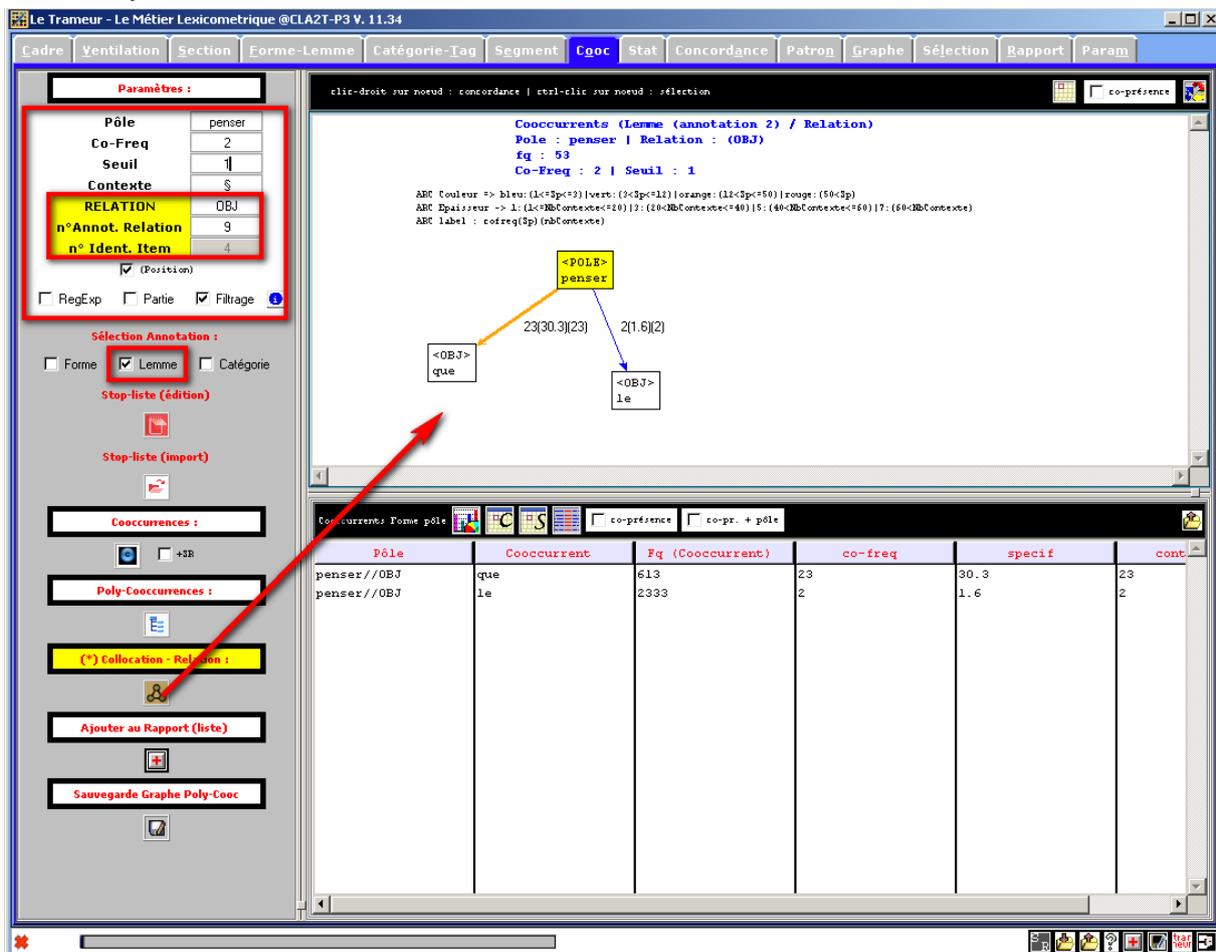
## 6. Recherche de collocation : spécificités sur relation

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 d'une relation entre ce pôle et ses candidats cooccurrents. Pour un item X donné (le pôle), on s'intéresse aux items Y cooccurrents de X et en relation REL avec X (les collocatifs de X) :

$$X \rightarrow \text{REL} \rightarrow Y$$

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



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 : .\* (regex) ; elle vise donc toutes les relations du type : penser-OBJ-y, penser-SUB-y etc.

The screenshot shows the 'Le Trameur' software interface. The main window displays a co-occurrence graph for the word 'penser'. The graph has a central node 'penser' with arrows pointing to various grammatical categories: <OBL> à, <SUB> on, <OBL> en, <OBL> à, <AD> parce que, <SUB> vous, <AD> si, <OBJ> le, <AD> toujours, and <SUB> je. Each arrow is labeled with a co-occurrence frequency and a specificity value in brackets, such as 5(8.2)(5) for 'à' and 31(1)(31) for 'je'.

Below the graph is a table of co-occurrences for the pole 'penser':

Pôle	Coocurrent	Fq (Coocurrent)	co-freq	specif	contextes
penser//OBL	à	187	5	8.2	5
penser//SUB	on	305	4	4.3	3
penser//OBJ	que	613	23	41.7	23
penser//OBL	en	351	2	3.2	2
penser//OBL	à	545	5	8.2	5
penser//OBJ	le	2333	2	2.5	2
penser//AD	si	91	2	2.8	2
penser//AD	parce que	95	3	4.0	3
penser//SUB	vous	437	3	3.2	3
penser//AD	toujours	60	2	2.8	2
penser//SUB	je	786	31	**	31
penser//AD	par	113	4	5.4	3

Une fois le graphe construit, un clic-droit sur un des nœuds 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 nœud *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).

The screenshot shows a concordance table with text excerpts and a detailed view of the word 'vous'. The text excerpts are:

```

-----PARTIE{PARTIE=D2009 }-----
un mot DELIM ou bien vous le penser vraiment DELIM
-----PARTIE{PARTIE=D0006 }-----
dans le quartier de Paris vous en penser quoi ici
-----PARTIE{PARTIE=D2011 }-----
DELIM que^êtreUNKNOWNce que vous en penser de le
    
```

The detailed view of 'vous' shows the following information:

```

Position: <64987>
Forme: <vous> | Freq: 441
Lemme: <vous> | Freq: 437
Cat: <B Cl> | Freq: 4177
a-00004: <-> | Freq: 71023
a-00005: <-> | Freq: 72668
a-00006: <2> | Freq: 988
a-00007: <p1> | Freq: 3992
a-00008: <-> | Freq: 61616
a-00009: <SUB (64991)> | Freq: 1
a-00010: <-> | Freq: 74880
a-00011: <-> | Freq: 74316
a-00012: <-> | Freq: 76107
a-00013: <-> | Freq: 76983
    
```