

Analyses linguistiques

max.silberztein@univ-fcomte.fr

Analyser un texte = Annoter le texte

Elisabeth Borne doit rencontrer les responsables de la majorité sénatoriale mercredi pour passer en revue leurs convergences et lignes rouges, avant l'ouverture des débats en séance publique jeudi.

- j'ai trouvé une phrase
- j'ai trouvé des groupes nominaux
- j'ai trouvé un nom propre
- j'ai trouvé des dates
- ...
- je peux traduire la séquence de texte ainsi :

Elisabeth Borne is due to meet with the leaders of the senatorial majority on Wednesday to review their convergences and red lines, before the opening of the debates in public session on Thursday.

Les annotations sont rangées dans la Structure d'Annotation du Texte (TAS)

- La TAS contient toutes les annotations potentielles associées à un texte
- Toutes les unités linguistiques sont représentées par une annotation, y compris les ALU :
 - les mots simples, ex. « table »
 - les affixes, ex. « re » (dans « relire »)
 - les mots composés, ex. « pomme de terre »
 - Les expressions discontinues, ex. « prendre xxx en compte »
- Une séquence de texte peut être annotée plusieurs fois

Alimenter la structure d'annotations

On ajoute des annotations pour décrire des séquences de texte :

- les formes simples, ex.: *mangerions* => <manger,V+tr+C+1+p>
- les formes agglutinées, ex. : *du* => <de,PREP> <le,DET+m+s>
- les mots composés, ex. : *pommes de terre* => <pomme de terre,N+Conc+f+p>
- les expressions discontinues, ex. :
pris cet avis en compte => <prendre en compte,V+PP+m+s>

Types d'ambiguïté

- entre deux entrées lexicales, ex. :

... Il vole à Lyon ...

<voler,VERBE+HVCàH> =: *Léa vole un cendrier à Luc*

<voler,VERBE+NVL> =: *L'avion vole vers Lyon*

- entre une entrée lexicale et une forme fléchie, ex. :

... place Leclerc sur la carte ...

<place,NOM+FLX=CRAYON> → *la place Leclerc*

<placer,VERBE+FLX=AIDER> → *Léa la place sur la table*

Types d'ambiguïté

- entre une entrée lexicale et une forme agglutinée, ex. :

... Léa veut relever les dossiers ...

<re,PREF><lever,VERBE+DRV=RE> → *Luc peut relever la main*

<relever,VERBE+FLX=AIDER> → *Léa veut relever les copies*

- entre formes fléchies d'une même ALU, ex. :

... aide mon ami ...

<aider,VERBE+FLX=AIMER+PR+1+s> → *Je l'aide ;*

<aider,VERBE+FLX=AIMER+PR+3+s> → *on m'aide ;*

<aider,VERBE+FLX=AIMER+S+1+s> → *Je l'aide ;*

<aider,VERBE+FLX=AIMER+S+3+s> → *on m'aide ;*

<aider,VERBE+FLX=AIMER+IP+2+s> → *aide-le*

Types d'ambiguïté

- entre deux formes fléchies de deux ALU, ex. :

... portions rapidement ...

<porter,VERBE+FLX=AIMER> → *Nous portions les sacs*

<portion,NOM+Abst> → *Les portions alimentaires*

- entre un mot simple et un mot contracté, ex. :

... du café ...

<du,DET+Partitif> → *Elle mange du pain*

<de,PREP><le,DET> → *Il revient juste du café*

Types d'ambiguïté

- entre un mot composé et une séquence de mots simples, ex. :

... il y a une table ronde en salle A32 ...

<table ronde,NOM+Abst> → *Organisons une table ronde*

<table,NOM+Conc>

<rond,A+FLX=COUSIN> → *Achète une table ronde*

- entre deux mots composés, ex. :

... Les pieds noirs ont manifesté ...

<pied noir,NOM+Hum+p> → (colons français en Algérie)

<pied noir,NOM+Hum+p> → (Amérindiens, état du Montana)

Types d'ambiguïté

- entre une expression et une séquence de mots simples, ex. :

... il a bu la tasse ...

boire la tasse, EXP+C1D → *Il a bu la tasse dans la piscine*

boire, VERBE

le, DET

tasse, NOM+Conc

- entre deux expressions, ex. :

... Léa a tiré les marrons du feu ...

tirer les marrons du feu, EXP → (récupérer les bénéfices)

tirer les marrons du feu, EXP → (travailler pour le bénéfice de quelqu'un d'autre)

Types d'ambiguïté

- Ambiguïtés entre structures syntaxiques différentes

L'infirmière ferme la porte

(N₀ *l'infirmière*) (Verbe *ferme*) (N₁ *la*) (Verbe *porte*)

(N₀ *l'infirmière*) (Verbe *ferme*) (N₁ *la porte*)

Luc recouvre la bouteille de plastique

(N₀ *Luc*) (Verbe *recouvre*) (N₁ *la bouteille de plastique*)

(N₀ *Luc*) (Verbe *recouvre*) (N₁ *la bouteille*) (N₂ *de plastique*)

Représenter les ambiguïtés : la structure d'annotation des textes (TAS)

- Une annotation est un triplet (position, longueur, analyse), ex.:
 - une annotation lexicale :
(17, 5, <ce,DET+Dem+f+s>)
 - une annotation morphologique :
(329, 6, <CXXXII,CR+Val="132">)
 - une annotation syntaxique :
(3213, 21, <DATE>)
- Les ambiguïtés non résolues sont représentées par des annotations parallèles

NooJ - [La femme de trente ans.not]

File Edit Lab Project Windows Info TEXT

- 1 + / 926 TUs

Characters
Tokens
Digrams
Annotations
Ambiguities

Language is "French (France) (fr)".
Text Delimiter is: \n (NEWLINE)
Text contains 926 Text Units (TUs).
81359 tokens including:
50000 words from

☒ Show Text Annotation Structure

Au commencement du mois d'avril 1813, il y eut un dimanche dont la matinée promettait un de ces beaux jours où les Parisiens voient pour la première fois de l'année le ciel sans boue et leur ciel sans nuages. Avant midi, un cabriolet à pompe attelé de deux chevaux fringants déboucha dans la rue de Rivoli par la rue Castiglione, et s'arrêta derrière plusieurs équipages stationnés à la grille nouvellement ouverte au milieu de la terrasse des Feuillants. Cette leste voiture était conduite par un homme en apparence soucieux et maladif; des cheveux grisonnants couvraient à peine son crâne jaune et le faisaient vieux avant le temps; il jeta les rênes au laquais à cheval qui suivait sa voiture, et descendit pour prendre dans ses bras une jeune fille dont la beauté mignonne attira l'attention des oisifs en promenade sur la terrasse. La petite personne se laissa complaisamment saisir par la taille quand elle fut debout sur le bord de la voiture, et passa ses bras autour du cou de son guide, qui la posa sur le trottoir, sans avoir chiffonné la garniture de sa robe en reps vert. Un amant n'aurait pas eu tant de soin. L'inconnu devait être le père de cette enfant qui, sans le remercier, lui prit familièrement le bras et l'entraîna brusquement dans le jardin. Le vieux père remarqua les regards émerveillés de quelques jeunes gens, et la tristesse empreinte sur son visage s'effaça pour un moment. Quoiqu'il fût arrivé depuis longtemps à l'âge où les hommes doivent se contenter des trompeuses jouissances que donne la vanité, il se mit à sourire.

"L'on te croit ma femme", dit-il à l'oreille de la jeune personne en se redressant et marchant avec une lenteur qui la désespéra.

Il semblait avoir de la coquetterie pour sa fille, et jouissait peut-être plus qu'elle des oeillades que les curieux lançaient sur ses petits pieds chaussés de brodequins en prune puce, sur une taille délicieuse dessinée par une robe à guimpe, et sur le cou frais qu'une collerette brodée ne cachait pas entièrement. Les mouvements

461	467	473	481	487	496	500
ce,DET+z1+f+s	leste,A+z1+m+s	voiture,N+z1+f+s	être,V+z1+I+3+s	conduite,N+z1+f+s	par,N+z3+m+s	un,DET+
	leste,A+z1+f+s	voiturer,V+z2+PR+1+s		conduire,V+z1+PP+f+s	par,PREP+z1	un,N+z1
	lester,V+z1+PR+1+s	voiturer,V+z2+PR+3+s				un,N+z1
	lester,V+z1+PR+3+s	voiturer,V+z2+S+1+s				
	lester,V+z1+S+1+s	voiturer,V+z2+S+3+s				
	lester,V+z1+S+3+s	voiturer,V+z2+IP+2+s				
	lester,V+z1+IP+2+s					

Cancel

Text Annotation Structure

- une séquence de mots peut être annotée :
<pomme de terre,N+Conc+f+s>
- un mot peut être annoté par une séquence d'annotations :
du => **<de,PREP> <le,DET+m+s>**
- une annotation peut être discontinue :
prendre qqchose en compte =>
<prendre en compte,V>

Annotation de la forme *du*

The screenshot shows the NooJ software interface with a text document titled "La femme de trente ans.not". The main text area displays a paragraph from Balzac's "La femme de trente ans". The sentence "autour du cou de son guide" is highlighted in blue. Below the text, a table shows the morphological analysis of the words in the highlighted phrase. A red circle highlights the analysis for "du".

974	981	981.01	988	988.01	991	995
autour,N+z1+m+s	du,DET+z1+m+s		cou,N+z1+m+s	de,DET+z1	son,DET+z1+m+s	guide,N+z1+f+s
autour,PREP+z1	de,PREP	le,DET+m+s		de,PREP+z1	son,N+z1+m+s	guide,N+z1+m+s
				de,PREP	un,DET+f+p	guider,V+z1+PR+1+s
					un,DET+m+p	guider,V+z1+PR+3+s

Annotation de *ciel sans nuages*

NooJ - [La femme de trente ans.not]

File Edit Lab Project Windows Info TEXT

- 1 + / 926 TUs

Characters
Tokens
Digrams
Annotations
Ambiguities

☒ Show Text Annotation Structure

Language is "French (France) (fr)".
Text Delimiter is: \n (NEWLINE)
Text contains 926 Text Units (TUs).
81359 tokens including:
60700 word forms

Au commencement du mois d'avril 1813, il y eut un dimanche dont la matinée promettait un de ces beaux jours où les Parisiens voient pour la première fois de l'année leurs pavés sans boue et leur ciel sans nuages. Avant midi, un cabriolet à pompe attelé de deux chevaux fringants déboucha dans la rue de Rivoli par la rue Castiglione, et s'arrêta derrière plusieurs équipages stationnés à la grille nouvellement ouverte au milieu de la terrasse des Feuillants. Cette leste voiture était conduite par un homme en apparence soucieux et maladif, des cheveux grisonnants couvraient à peine son crâne jaune et le faisaient vieux avant le temps; il jeta les rênes au laquais à cheval qui suivait sa voiture, et descendit pour prendre dans ses bras une jeune fille dont la beauté mignonne attira l'attention des oisifs en promenade sur la terrasse. La petite personne se laissa complètement

191	201	206
leur,DET+z1+m+s	ciel sans nuage,N+NPN+m+s	
leur,DET+z1+f+s	ciel,N+z1+m+s	sans,PREP+z1
leur,PRO+z1+3+m+p	ciel,INTJ+z1	nuage,N+z1+m+p
		nuager,V+z3+PR+2+s

Cancel

Annotation d'expressions discontinues

NooJ - [zzz.not]

File Edit Lab Project Windows Info TEXT

- 1 + / 1 TUs

Characters
Tokens
Digrams
Annotations
Unknowns

☒ Show Text Annotation Structure

Language is "French (France) (fr)".
Text Delimiter is: \n (NEWLINE)
5 tokens including:
5 word forms

Luc accéléra soudain le mouvement

4

13

21

24

accélérer,V+C1D+N0Hum+N0NHum+DET1=le+N1=mouvement+Passif+PS+3+s

accélérer,V+z1+PS+3+s

soudain,A+z1+m+s

soudain,ADV+z1

DET

le,PRO+z1+3+m+s

le,DET+z1+m+s

N

mouvement,N+z1+m+s

Cancel

Analyse lexicale =
représenter les ambiguïtés dans la phrase :

Pierre mange une pomme de terre cuite

Pierre,PRENOM+masculin

manger,V+Tr+FLX=MANGER

un,DETERMINANT+indéfini+FLX=UNUNEDES

une,N+Abstrait+féminin+FLX=TABLE

pomme,N+concret

pomme,ADJ+couleur

pomme,ADJ+humain

pommer,V+Intr

pomme de terre,N+Concret

terre,N+Concret

terre,N+matériau

terre cuite,N+Concret

terre cuite,N+matériau

terrifier,V+Pron+FLX=AIMER

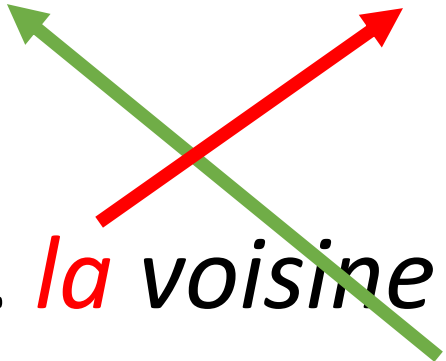
cuire,V+Tr+FLX=CUIRE

cuite,N+Abstrait+féminin+FLX=TABLE

cuisiner,V+Pron+FLX=AIMER

Levée d'ambiguïté à l'aide de grammaires syntaxiques

Luc la rend à la voisine



- un déterminant, ex. *la voisine*
- un pronom préverbal, ex. *Luc la rend*
- un nom abstrait, ex. *Le violoniste donne le la*

Pour entrer ces noeuds, on utilise le caractère “/”, exactement comme dans les expressions flexionnelles :

la/<PRO> signifie qu’on impose que le mot “la” soit un pronom → on supprime toutes les autres annotations

la/<DET> signifie qu’on impose que le mot “la” soit un déterminant → on supprime toutes les autres annotations

Levée d'ambiguïté à l'aide de grammaires syntaxiques

Luc la rend à la voisine

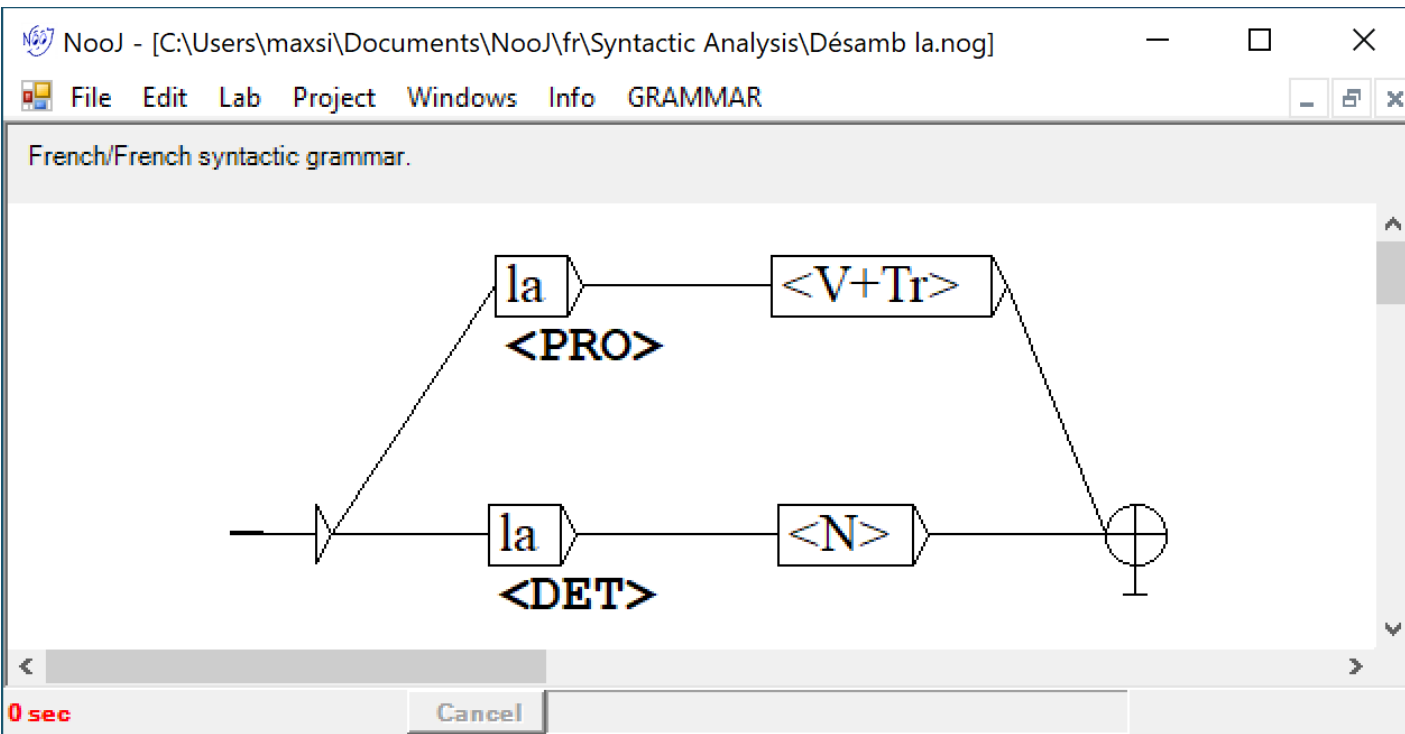
- Si le mot “la” est suivi d’un <V+Tr>, alors c’est un <PRO>
- Si le mot “la” est suivi par un <N>, alors c’est un <DET>

Pour entrer ces noeuds, on utilise le caractère “/”, exactement comme dans les expressions flexionnelles :

la/<PRO> signifie qu’on impose que le mot “la” soit un pronom → on supprime toutes les autres annotations

la/<DET> signifie qu’on impose que le mot “la” soit un déterminant → on supprime toutes les autres annotations

Levée d'ambiguïté à l'aide de grammaires syntaxiques



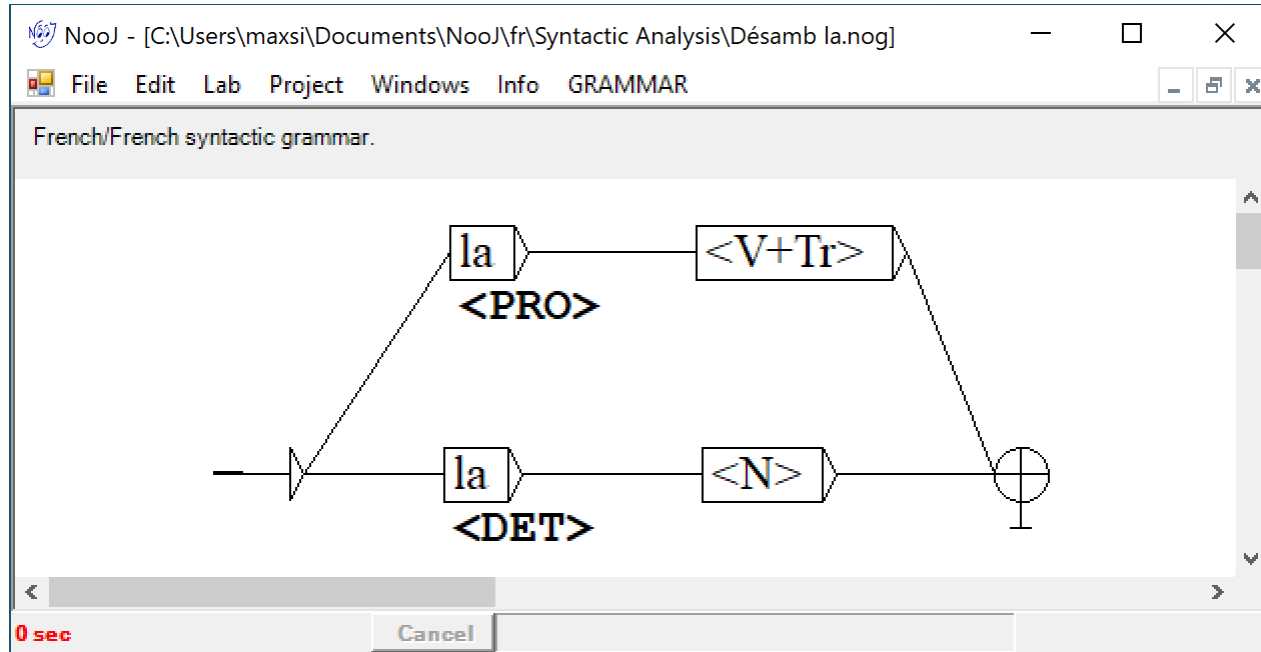
- Si le mot “la” est suivi d’un verbe transitif <V+Tr>, alors c’est un <PRO>
- Si le mot “la” est suivi par un nom, alors c’est un <DET>

On utilise le caractère “/”, exactement comme dans les expressions flexionnelles :

la/<PRO> : on impose que le mot “la” soit un pronom → on supprime toutes les autres annotations

la/<DET> : on impose que le mot “la” soit un déterminant → on supprime toutes les autres annotations

Levée d'ambiguïté à l'aide de grammaires syntaxiques



Trois cas :

... *la maison* ... → *la* est un déterminant

... *la mange* ... → *la* est un pronom

... *la place* ... → *la* peut être un déterminant ou un pronom

Exemples de grammaires de levée d'ambiguïté

- NooJ donne la priorité aux séquences reconnues les plus longues, ex. :

c | c' / <PRO> est / <V>

→ on ne sait pas ce que représente le mot « c », sauf s'il est dans la séquence « c'est » ; dans ce cas, on lève les ambiguïtés sur les deux mots

- La négation est utile pour lever des ambiguïtés, ex. :

<V> / <!V+Pronominal> | (me | te | se) <V+Pronominal>

→ un verbe ne peut être pronominal que s'il est précédé par *me*, *te* ou *se*

(je | j') <V+1+s> / <V+1+s> | <V> / <!V+1+s>

→ un verbe ne peut être conjugué à la première personne que s'il est précédé par *je*

de / <PREP> | de (très | <E>) <A+p> <N+p>

→ *de* est une préposition sauf si précédé d'un adjectif pluriel

La TAS représente les ambiguïtés (avant)

NooJ - [[Modified] C:\Users\Max\Documents\NooJ\fr\Projects\Marie.not] looJ\fr\Projects\Marie.not

File Edit Lab Project Windows Info TEXT TEXT

Characters Tokens Digrams Language is "French(fr)".
Text Delimiter is: \n (NEWLINE)
8 tokens including:
7 word forms

✓ Show Text Annotation Structure

Marie donne la pomme de terre cuite.

0	6	12	15	21	24	30
marier, V+Aux=a+Pers=1+Nb=s+Temps=PR	donne, N+Genre=f+Nb=s	la, DET+Genre=f+Nb=s+Spec=def	pomme de terre, N+Cmp=N_GP+Genre=f+Nb=s			
marier, V+Aux=a+Pers=3+Nb=s+Temps=PR	donner, V+Aux=a+Pers=1+Nb=s+Temps=PR	le, PRO+Dist=clit+Pers=3+Nb=s+Genre=f+Fonc=acc	pomme de terre, N+Cmp=NDN+Sem=Conc+Genre=f+Nb=s+DOM=Alim			
marier, V+Aux=a+Pers=1+Nb=s+Temps=S	donner, V+Aux=a+Pers=3+Nb=s+Temps=PR	la, N+Genre=m+Nb=s	pommer, V+Aux=a+Pers=3+Nb=s+Temps=PR	de, DET+Genre=m+Nb=p	terre cuite, N+Cmp=NA+Sem=Conc+z1+Genre=f+Nb=s+PARTITIF	
marier, V+Aux=a+Pers=3+Nb=s+Temps=S	donner, V+Aux=a+Pers=1+Nb=s+Temps=S	la, N+Genre=m+Nb=p	pommer, V+Aux=a+Pers=1+Nb=s+Temps=S	de, DET+Genre=f+Nb=p	terror, V+Aux=a+Pers=1+Nb=s+Temps=PR	cuit, A+Genre=f+Nb=s
marier, V+Aux=a+Pers=2+Nb=s+Temps=IP	donner, V+Aux=a+Pers=3+Nb=s+Temps=S		pommer, V+Aux=a+Pers=3+Nb=s+Temps=S	de, PREP+Gvn=pv	terror, V+Aux=a+Pers=3+Nb=s+Temps=PR	cuite, N+Genre=f+Nb=s
Marie, N+PRENOM+Genre=f+Nb=s	donner, V+Aux=a+Pers=2+Nb=s+Temps=IP		pommer, V+Aux=a+Pers=2+Nb=s+Temps=IP		terror, V+Aux=a+Pers=1+Nb=s+Temps=S	cuire, V+Aux=a+Genre=f+Nb=s+Temps=PP
Marie, N+PRENOM+Genre=m+Nb=s			pommer, V+Aux=e+Pers=1+Nb=s+Temps=PR		terror, V+Aux=a+Pers=3+Nb=s+Temps=S	
			pommer, V+Aux=e+Pers=3+Nb=s+Temps=PR		terror, V+Aux=a+Pers=2+Nb=s+Temps=IP	
			pommer, V+Aux=e+Pers=1+Nb=s+Temps=S		terre, N+Genre=f+Nb=s	
			pommer, V+Aux=e+Pers=3+Nb=s+Temps=S			
			pommer, V+Aux=e+Pers=2+Nb=s+Temps=IP			
			pomme, N+Genre=f+Nb=s			
			pommer, V+Aux=a+Pers=1+Nb=s+Temps=PR			

0.4 sec Cancel

La TAS représente les ambiguïtés (après)

NooJ - [[Modified] C:\Users\Max\Documents\NooJ\fr\Projects\Marie.not]

File Edit Lab Project Windows Info TEXT

- 1 + / 1 TUs

Characters Tokens Digrams

Language is "French(fr)".
Text Delimiter is: \n (NEWLINE)
8 tokens including:
7 word forms

☒ Show Text Annotation Structure

Marie donne la pomme de terre cuite.

0	6	12	15	21	24	30
Marie,PRENOM+f+s	donne,N+Genre=f+Nb=s	la,DET+Genre=f+Nb=s+Spec=def	pomme de terre,N+Cmp=NDN+Sem=Conc+Genre=f+Nb=s+DOM=Alim			
	donner,V+Aux=a+Pers=3+Nb=s+Temps=PR		pomme,N+Genre=f+Nb=s	de,PREP+Gvn=pv	terre cuite,N+Cmp=NA+Sem=Conc+z1+Genre=f+Nb=s+PARTITIF	
	donner,V+Aux=a+Pers=2+Nb=s+Temps=IP				terre,N+Genre=f+Nb=s	cuire,V+Aux=a+Genre=f+Nb=s+Temps=PP

Cancel

Analyse syntaxique en constituants

NooJ - [analyse syntaxique.not]

File Edit Lab Project Windows Info TEXT

- 1 + / 2 TUs

☒ Show Text Annotation Structure

Characters
Tokens
Digrams
Annotations
Unknowns

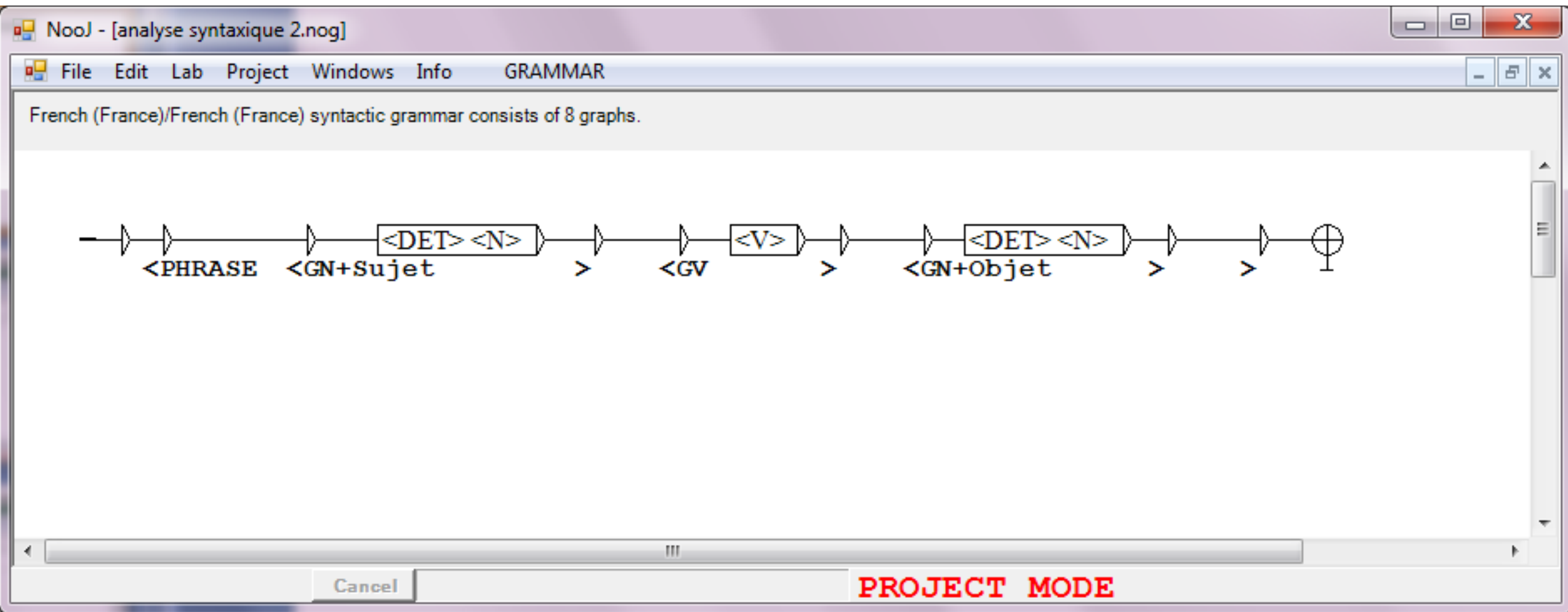
Language is "French (France) (fr)".
Text Delimiter is: \n (NEWLINE)
Text contains 2 Text Units (TUs).
6 tokens including:
e ...nd f...e

Le chat mange le rat.

0	3	8	14	17
le,DET+z1+m+s	chat,A+z1+m+s	manger,V+z1+PR+1+s	le,DET+z1+m+s	rat,A+z1+m+s
le,PRO+z1+3+m+s	chat,N+z1+m+s	manger,V+z1+PR+3+s	le,PRO+z1+3+m+s	rat,A+z1+f+s
		manger,V+z1+S+1+s		rat,N+z1+m+s
		manger,V+z1+S+3+s		
		manger,V+z1+IP+2+s		

Cancel PROJECT MODE

Une grammaire annote les constituants



Une grammaire annote les constituants

Nool

File Edit Lab Project Windows Info GRAMMAR

Structure of analyse syntaxiq... Expand Collapse Refresh

- Main
 - GN
 - DET
 - DetDef
 - DetDefPlur
 - DetDefSing
 - DetIndef
 - GV

analyse syntaxique.nog

French (France)/French (France) syntactic grammar consists of 8 graphs.

<PHRASE>

<GN+Sujet>

GN

<GV>

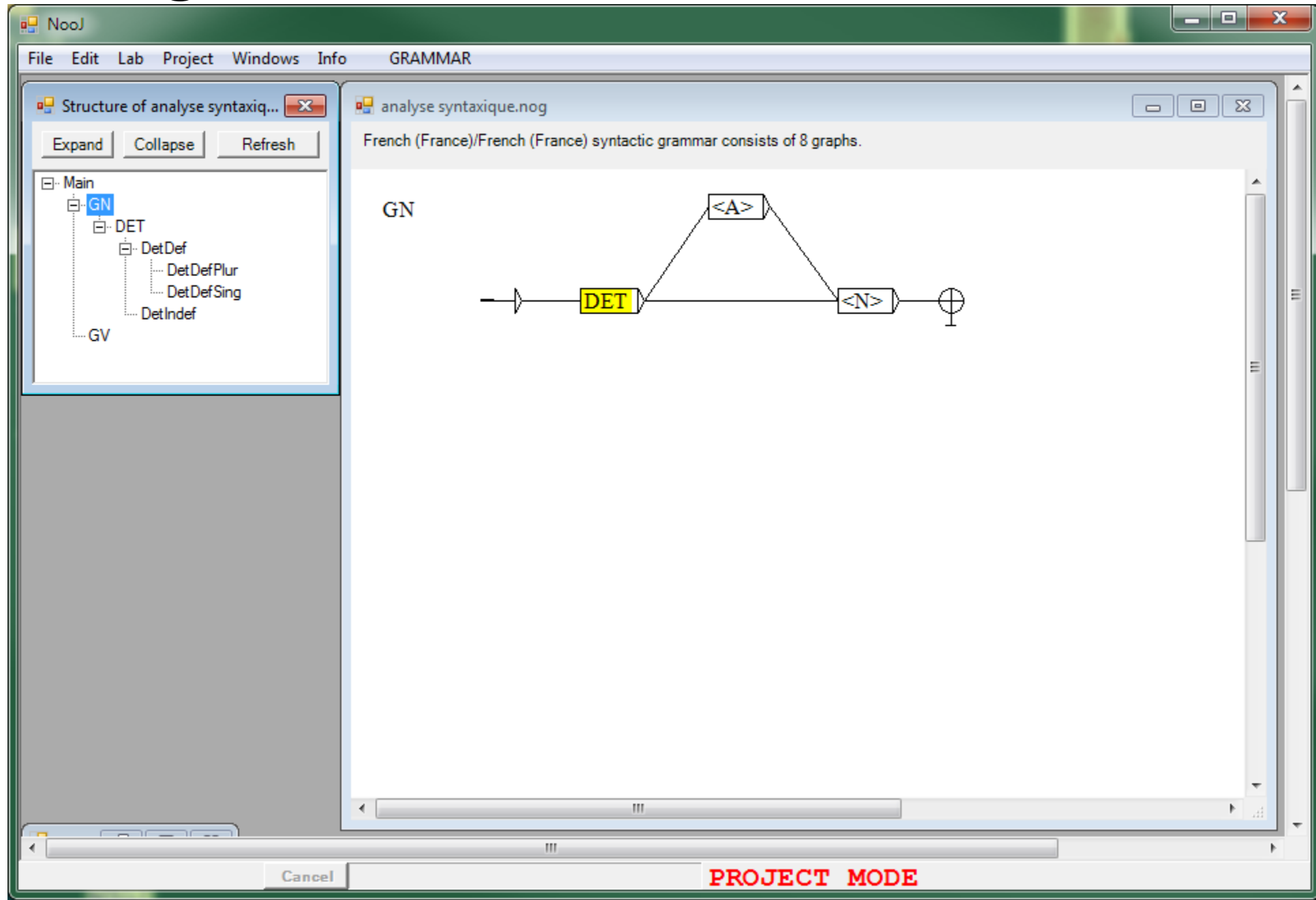
GV

<GN+Objet>

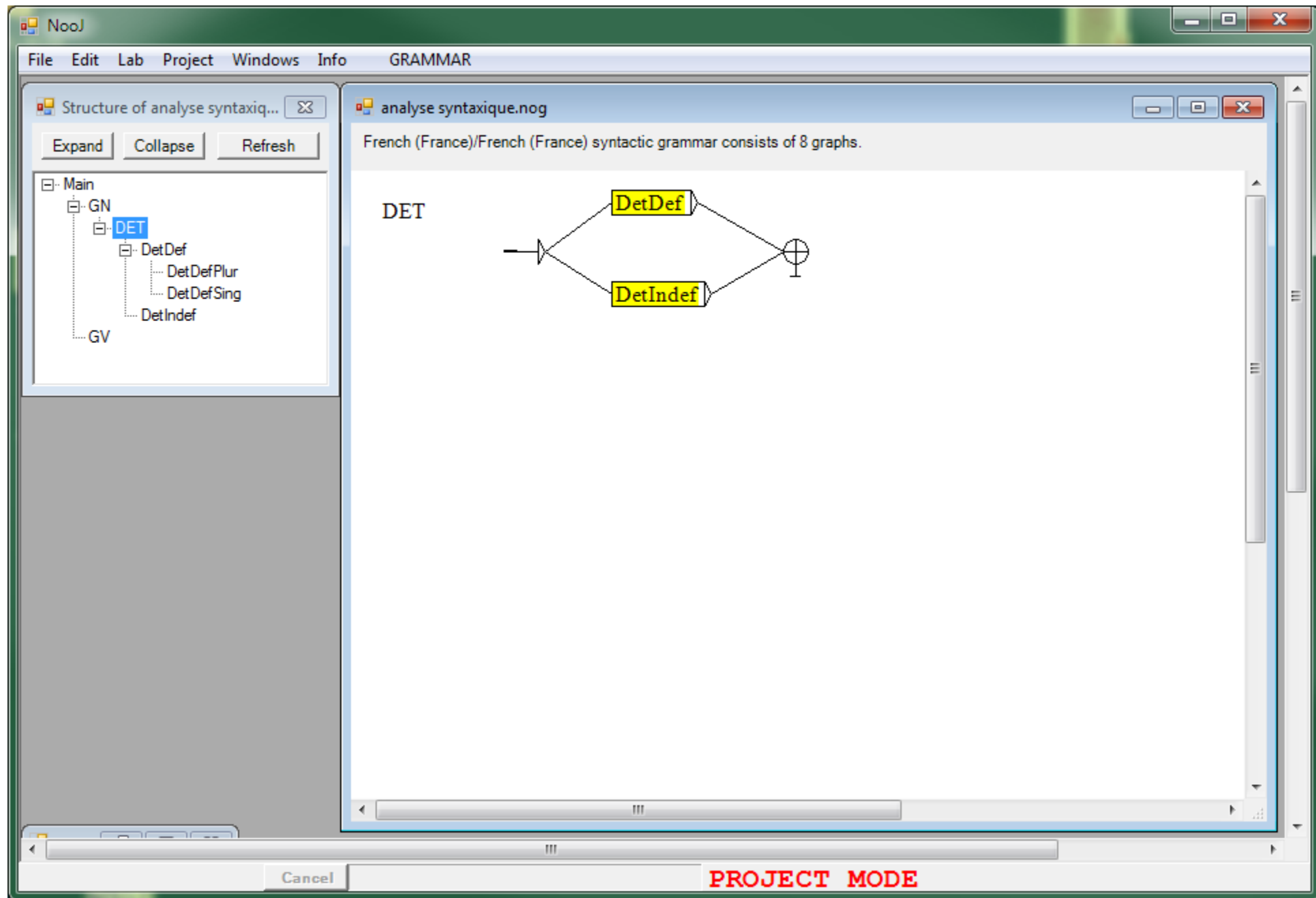
GN

Cancel PROJECT MODE

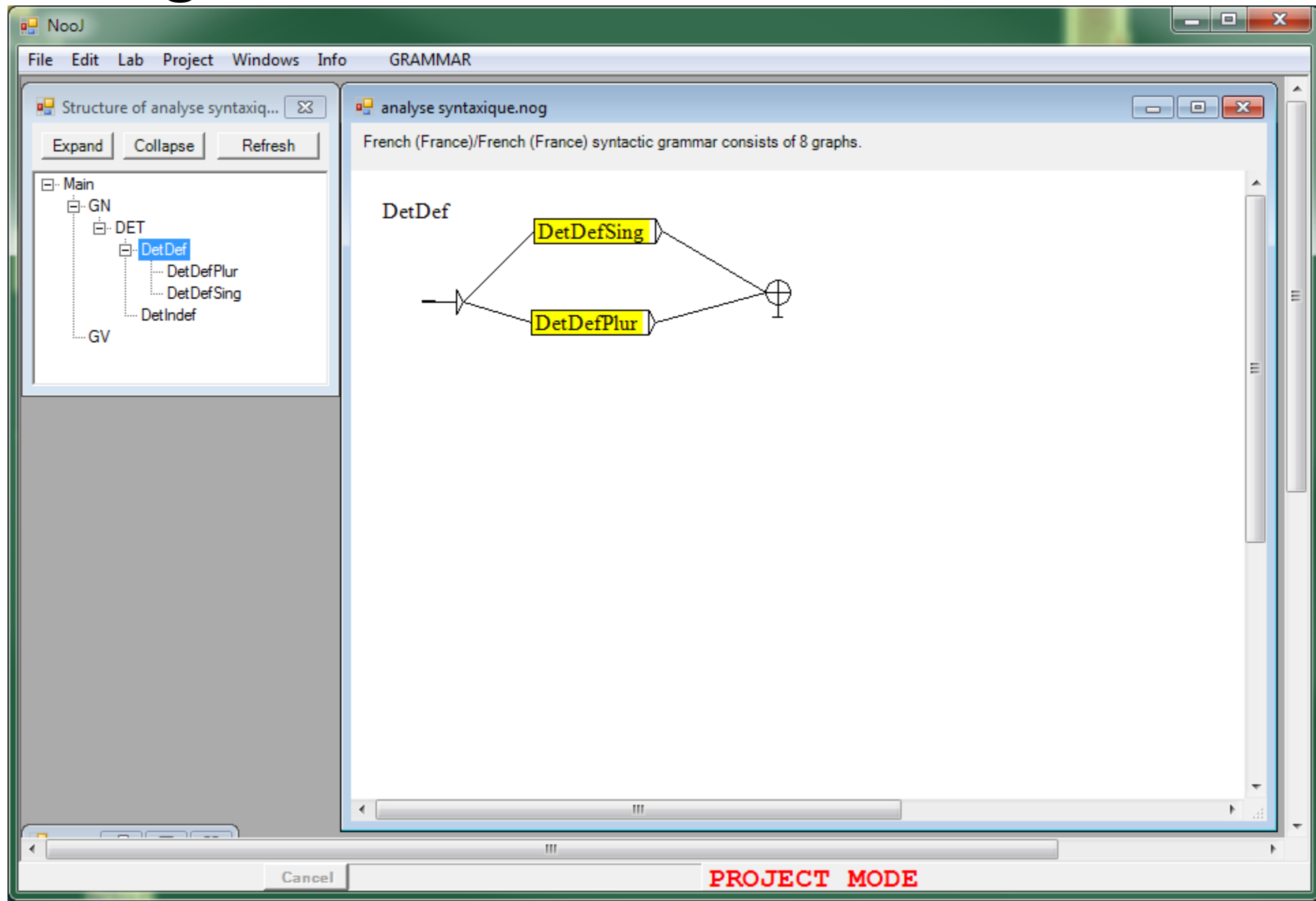
Une grammaire annote les constituants



Une grammaire annote les constituants



Une grammaire annote les constituants



Une grammaire annote les constituants

The screenshot displays the NooJ software interface, which is used for natural language processing. The main window is titled "GRAMMAR" and shows a file named "analyse syntaxique.nog". The interface is divided into two main panels.

On the left, there is a "Structure of analyse syntaxiq..." panel. It contains a tree view of the grammar structure. The tree is as follows:

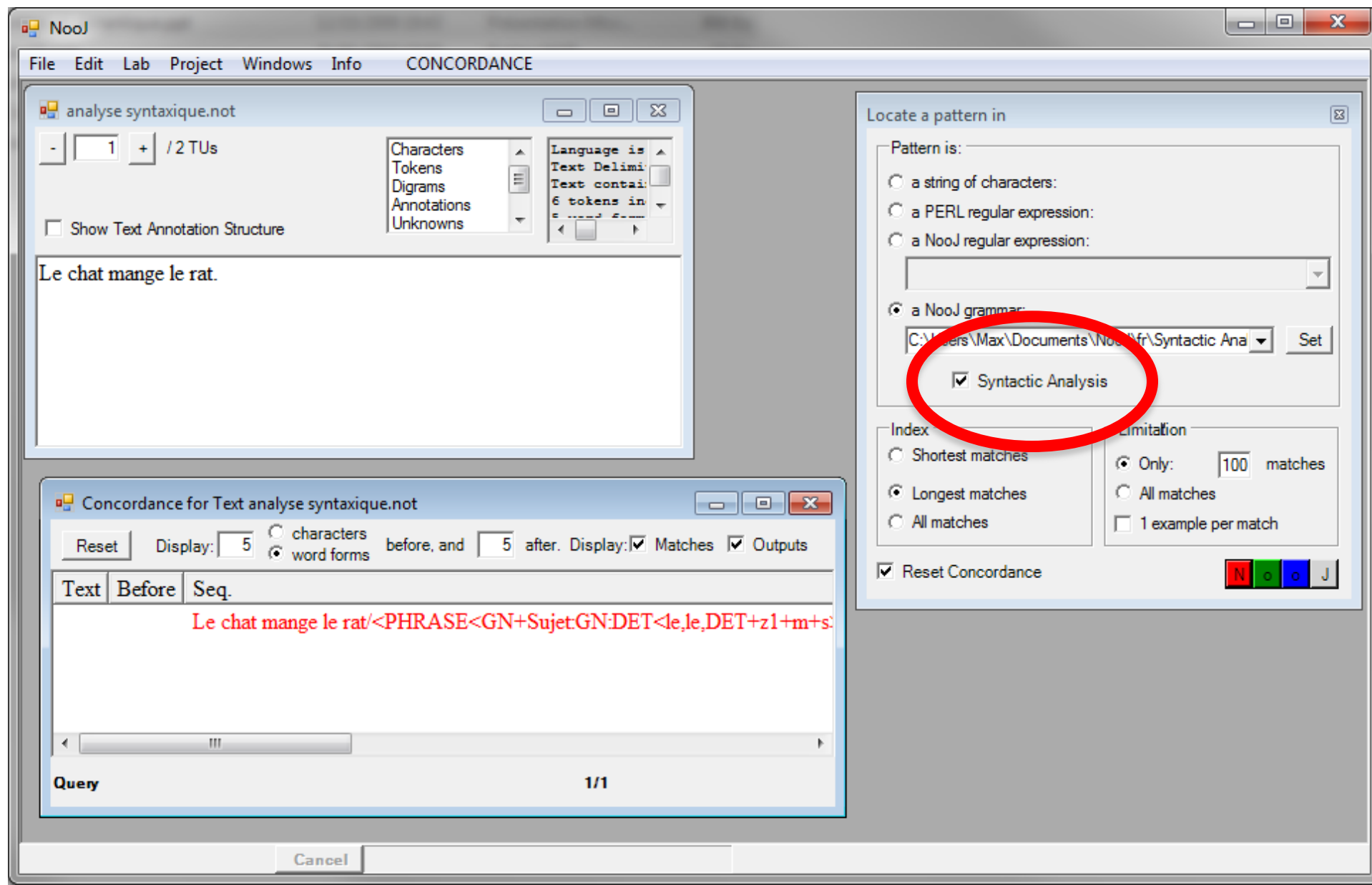
- Main
 - GN
 - DET
 - DetDef
 - DetDefPlur
 - DetDefSing** (highlighted in blue)
 - DetIndef
 - GV

Below the tree view are buttons for "Expand", "Collapse", and "Refresh".

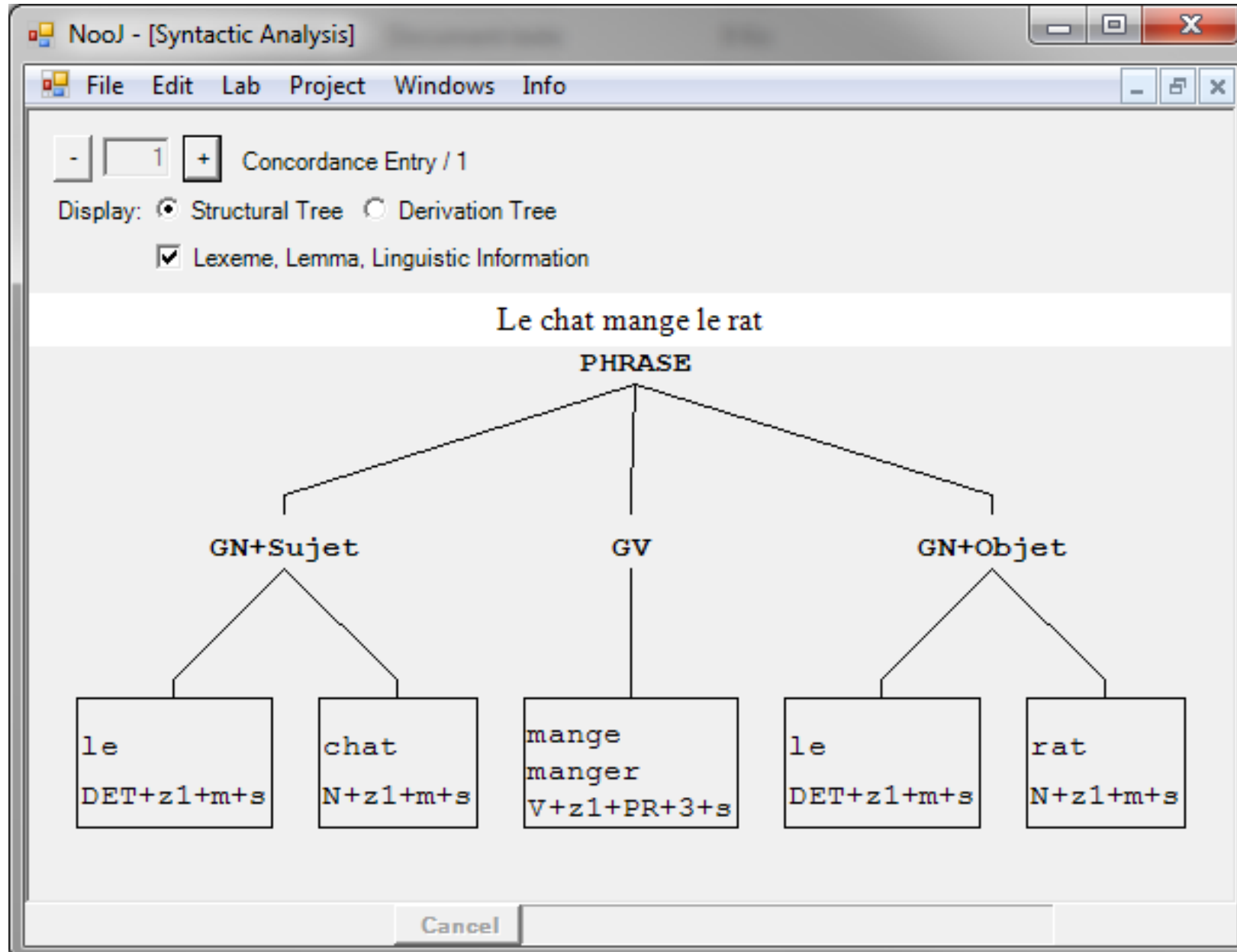
On the right, the main editing area shows the text "French (France)/French (France) syntactic grammar consists of 8 graphs." Below this, a specific graph is displayed, titled "DetDefSing". The graph is a directed graph with a single node and a self-loop. The node is a rectangle containing four strings, each enclosed in angle brackets: $\langle le, le, DET+m+s \rangle$, $\langle la, le, DET+f+s \rangle$, $\langle l', le, DET+m+s \rangle$, and $\langle l', le, DET+f+s \rangle$. The node is connected to a circle with a cross inside, which represents a terminal symbol.

At the bottom of the window, there is a "Cancel" button and a red label "PROJECT MODE".

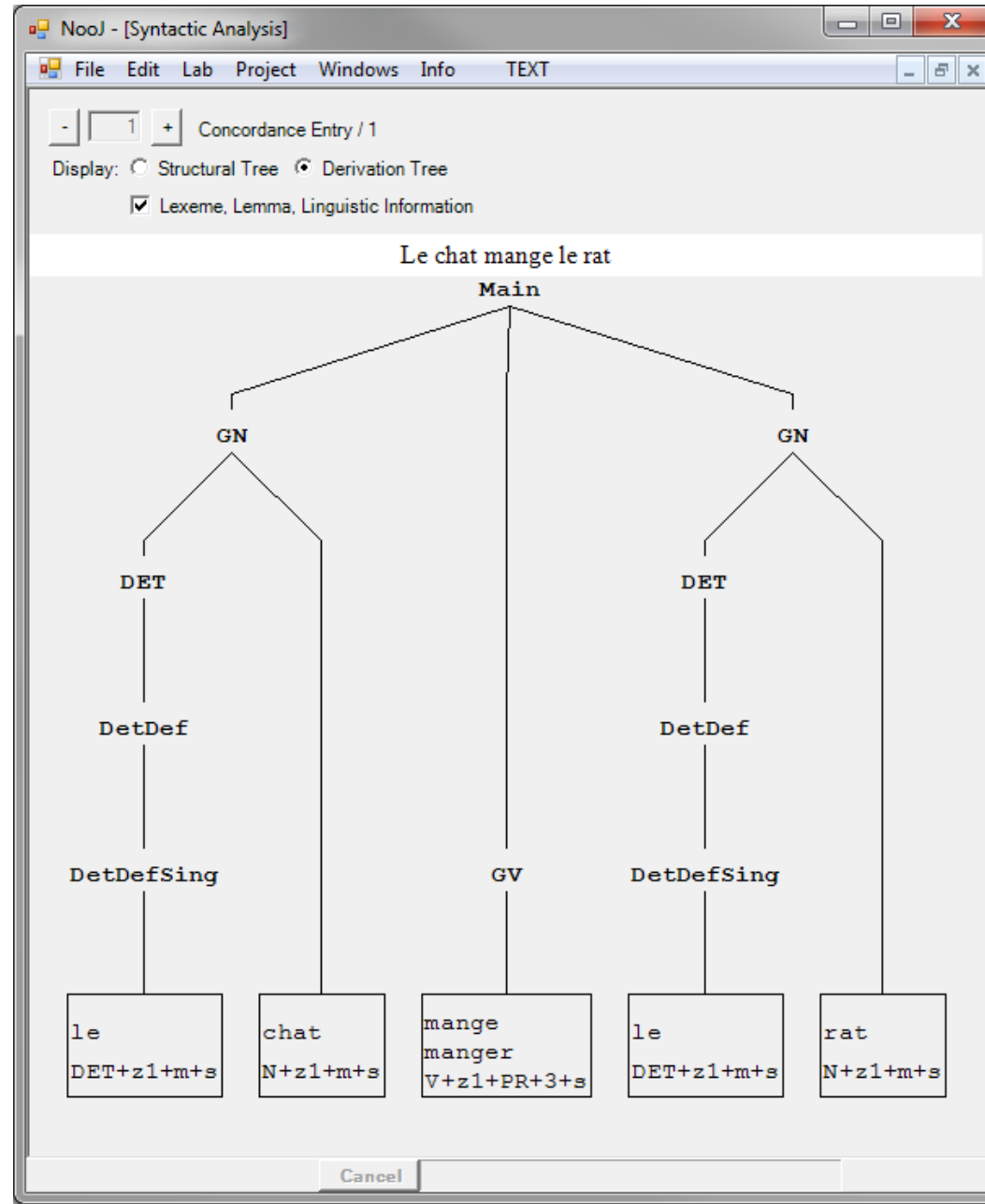
TEXT > Locate



CONCORDANCE > Display Syntactic Analysis



arbre structurel \neq arbre de dérivation



Types d'ALU : contraction *aux*

The screenshot displays the NooJ software interface. The main window shows a document titled "_La femme de trente ans.not" with 24 tokens and 926 TUs. A "Syntactic Analysis" dialog box is open, showing a concordance entry for the sentence "les tambours battirent aux champs". The dialog box has a "Display" section with "Syntax Tree" selected and "Lexeme, Lemma, Linguistic Information" unchecked. The syntax tree is a hierarchical structure representing the sentence. The root node is "PHRASE", which branches into "SUJET", "VERBE", and "OBJET". "SUJET" branches into "les" and "tambours". "VERBE" branches into "battirent". "OBJET" branches into "à", "les", and "champs".

NooJ

File Edit Lab Project Windows Info

_La femme de trente ans.not

- 24 + / 926 TUs

Characters
Tokens

Language is "French (fr)".
Text Delimiter is: \n (NEWLINE)

☐ Show Text Annotation

Un petit homme asse
l'écuyère, parut tout à
l'homme lui-même; le
épée était à son côté.
place. Aussitôt, les ta
dont l'expression gue
jusqu'à la grosse cais
soldats présentèrent l
premier rang jusqu'au
en rang comme des é
enthousiasmée. Enfin
mouvement avait imp
élan aux aigles, et aux

Syntactic Analysis

- 1 + Concordance Entry / 231

Display: ☒ Syntax Tree ☐ Parse Tree

☐ Lexeme, Lemma, Linguistic Information

les tambours battirent aux champs

PHRASE

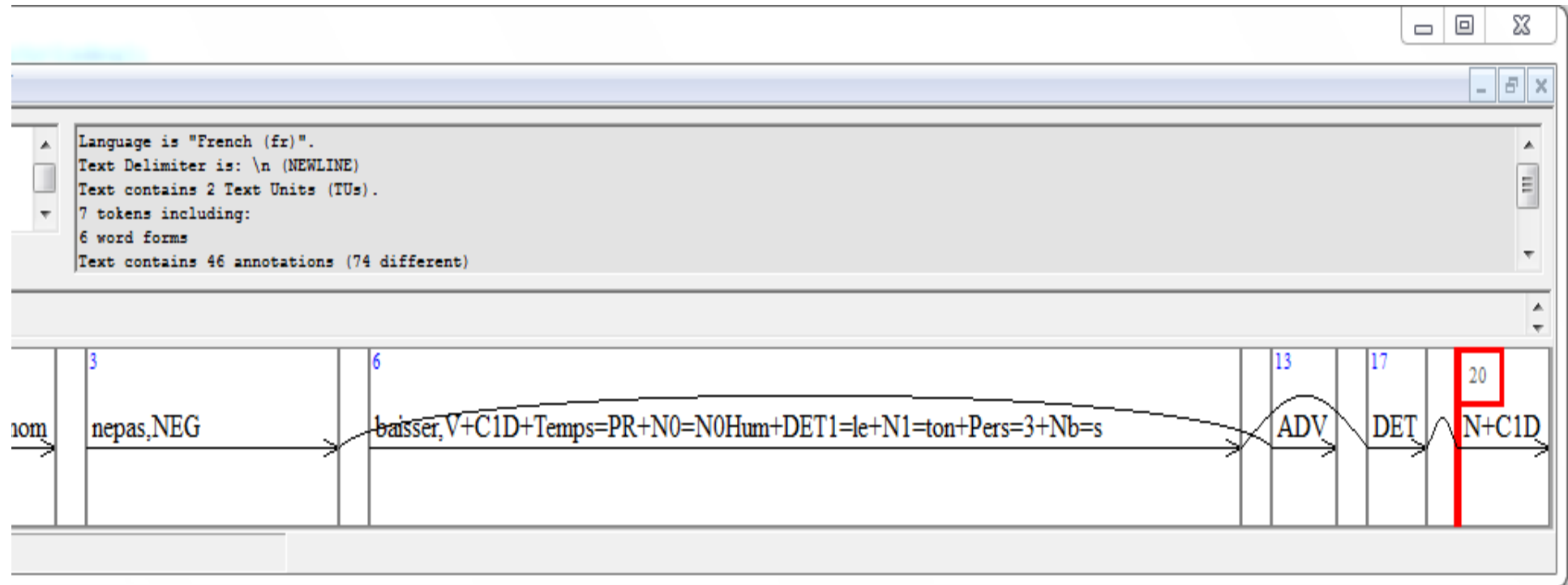
SUJET VERBE OBJET

les tambours battirent à les champs

Cancel

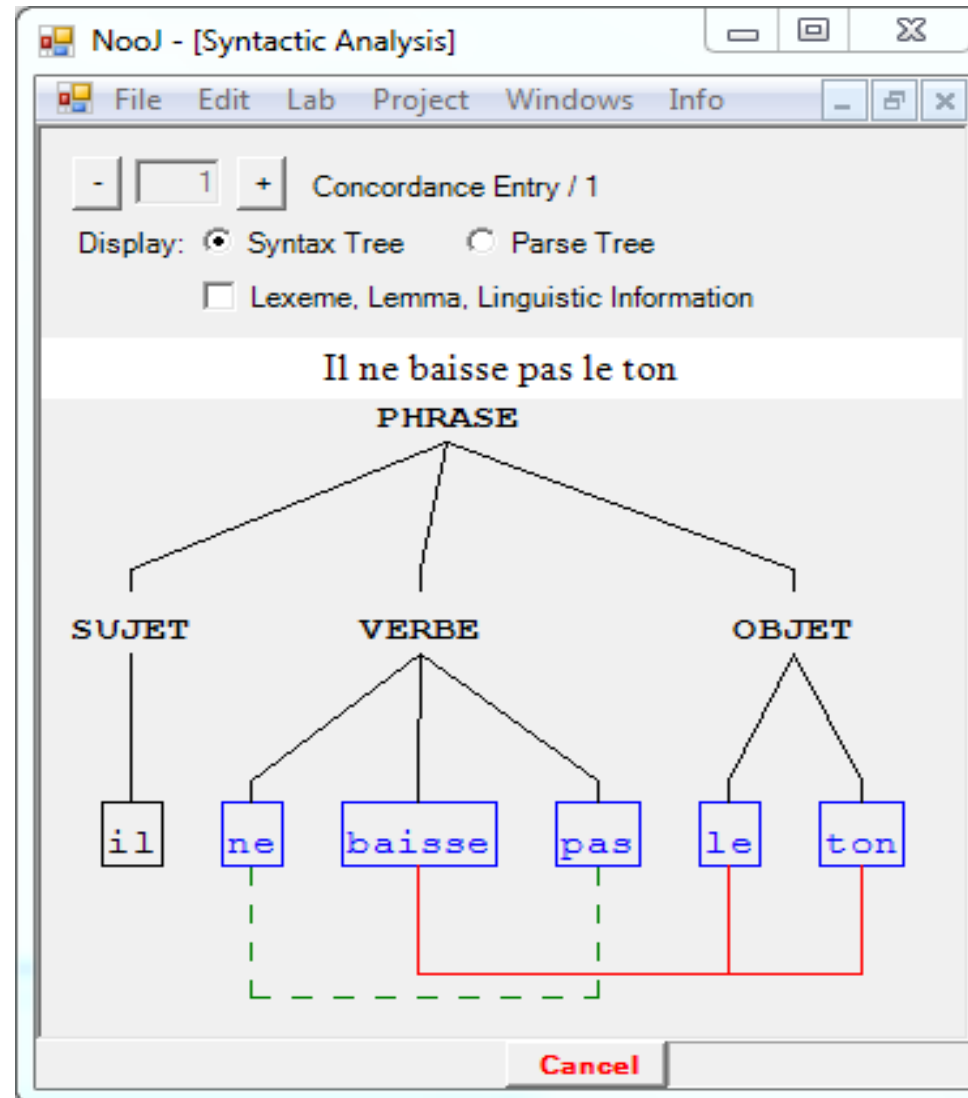
Types d'ALU : expressions discontinues

Il ne *baisse* pas *le* *ton*



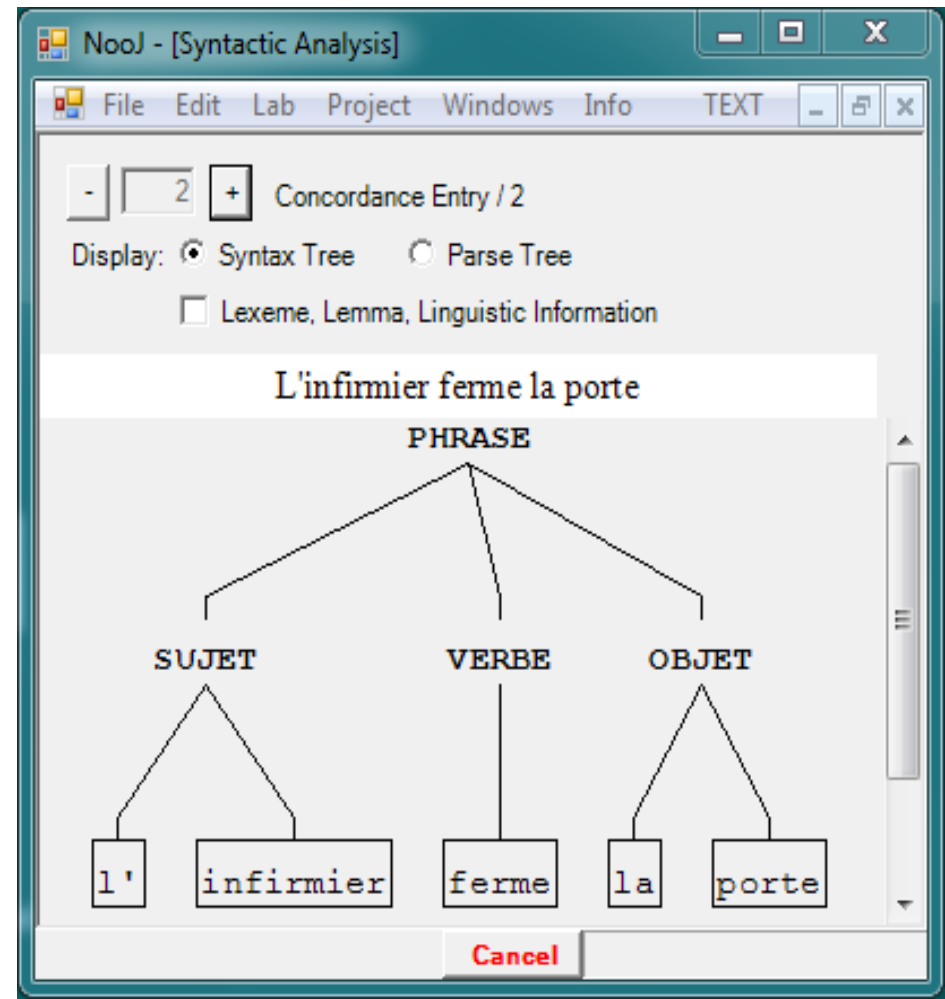
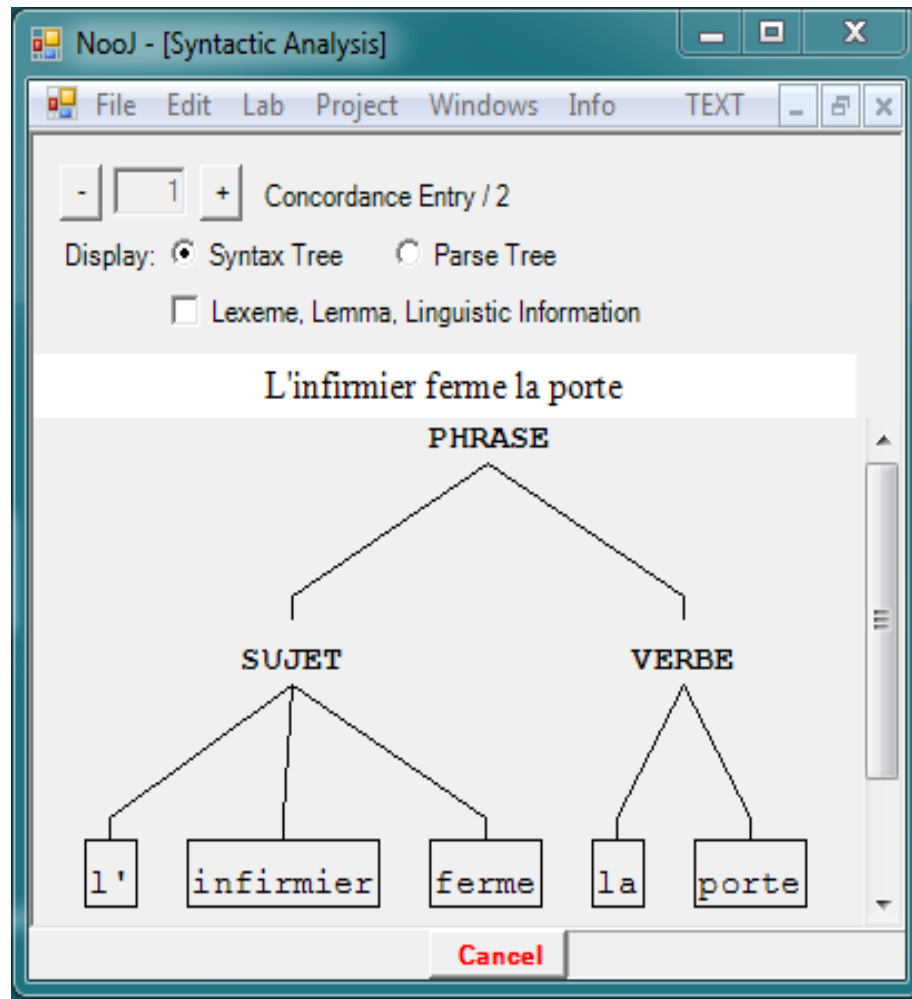
Types d'ALU : expressions discontinues

Il ne *baisse* pas *le ton*



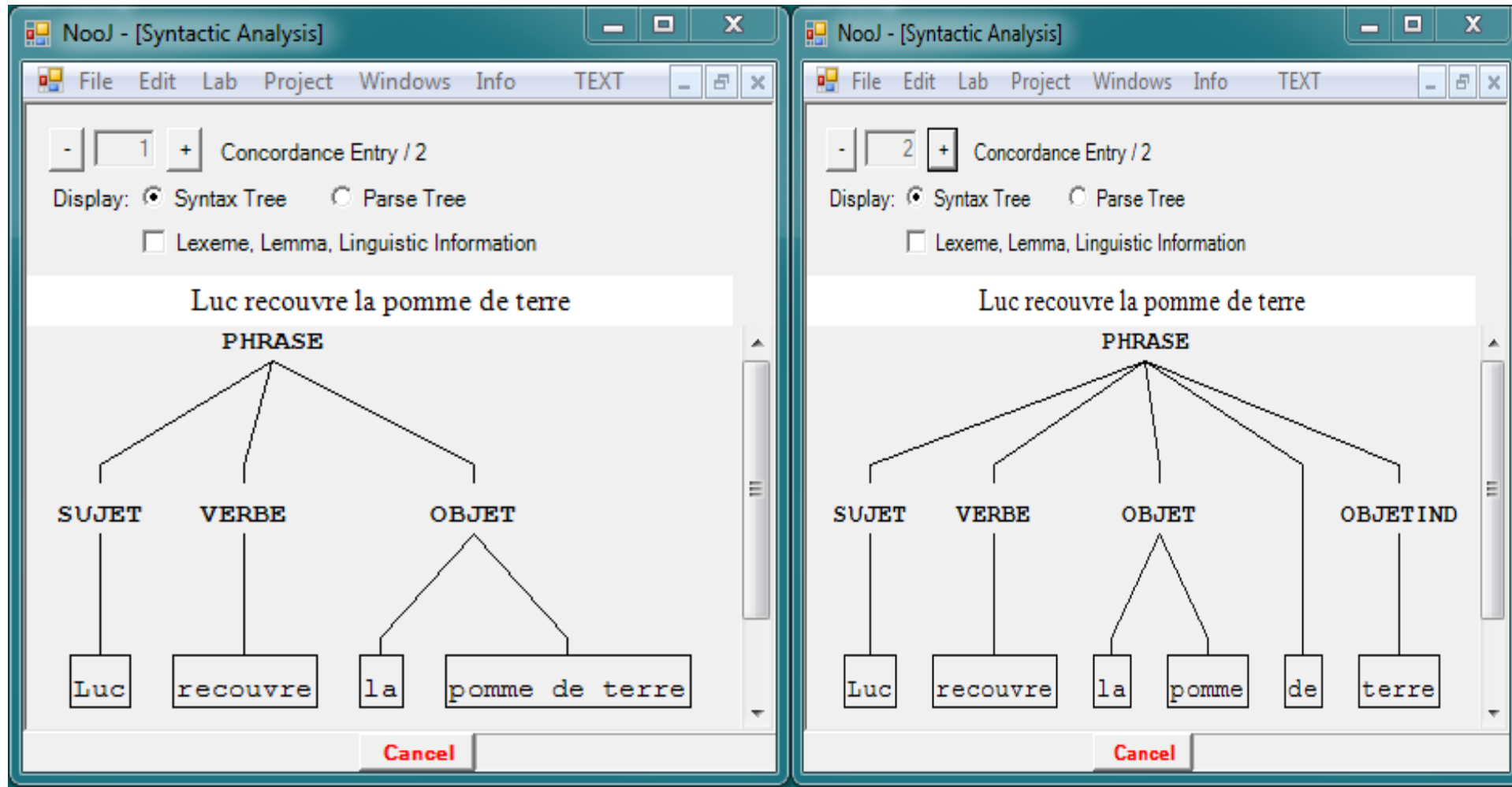
Ambiguïtés structurelles

L'infirmier ferme la porte



Ambiguïtés lexicales \Rightarrow structurelles

Luc recouvre la pomme de terre



Levée d'ambiguïté : CONCORDANCE → Annotate Text

The screenshot shows the NooJ software interface with a concordance table for the sentence "Le chat mange le rat." A modal dialog box titled "NooJ: text was annotated" is displayed, stating "4 annotations were added; 9 annotations were removed" with an "OK" button. The concordance table below shows the following structure:

0	3	8	14	17
PHRASE				
GN+Sujet		manger, V+z1+PR+3+s	GN+Objet	
le, DET+z1+m+s	chat, N+z1+m+s	GV	le, DET+z1+m+s	rat, N+z1+m+s

At the bottom of the window, there is a "Cancel" button and the text "PROJECT MODE" in red.

(Rappel : avant l'analyse syntaxique)

The screenshot shows the NooJ interface with the file 'analyse syntaxique.not'. The 'TEXT' tab is active, displaying the text 'Le chat mange le rat.' and a summary of its properties: Language is 'French (France) (fr)', Text Delimiter is '\n (NEWLINE)', Text contains 2 Text Units (TUs), and 6 tokens including 'Le', 'chat', 'mange', 'le', 'rat', and a period.

Below the text, a table displays the morphological analysis for each token. The first column, corresponding to 'Le', is highlighted with a red box and contains the index '0'. The table shows various grammatical features for each token, such as part of speech (DET, PRO, N, V), number (z1, z3), and other syntactic markers.

0	3	8	14	17
le,DET+z1+m+s	chat,A+z1+m+s	manger,V+z1+PR+1+s	le,DET+z1+m+s	rat,A+z1+m+s
le,PRO+z1+3+m+s	chat,N+z1+m+s	manger,V+z1+PR+3+s	le,PRO+z1+3+m+s	rat,A+z1+f+s
		manger,V+z1+S+1+s		rat,N+z1+m+s
		manger,V+z1+S+3+s		
		manger,V+z1+IP+2+s		

At the bottom of the window, there is a 'Cancel' button.

→ L'analyse syntaxique lève les ambiguïtés
(sans grammaires locales)

NooJ - [analyse syntaxique.not]

File Edit Lab Project Windows Info TEXT

- 1 + / 2 TUs

☒ Show Text Annotation Structure

Characters
Tokens
Digrams
Annotations
Ambiguities

Language is "French (France) (fr)".
Text Delimiter is: \n (NEWLINE)
Text contains 2 Text Units (TUs).
6 tokens including:
Le chat mange le rat.

0	3	8	14	17
PHRASE				
GN+Sujet		manger, V+z1+PR+3+s	GN+Objet	
le, DET+z1+m+s	chat, N+z1+m+s	GV	le, DET+z1+m+s	rat, N+z1+m+s

Cancel PROJECT MODE

Levée d'ambiguïté sémantique

Entrée	E..	AUX	FLX	DRV	CONS	BASE	CLASS	DOM	LEXI	OPER	PREP	SynSem
abaisser	01	AVOIR	CHANTER	ABLE:Z+MENT+RA_E...	T1308+P3008	"baisse"	T3c	LOC	2	"r/d bas qc"		
abaisser	02	AVOIR	CHANTER	MENT:Z	T13g0+P30g0	"baisse"	E3f	TEC	2	"f.iire qc VRS bas"		
abaisser	03	AVOIR	CHANTER	MENT:Z	T1306+P3006	"baisse"	M3c	QUA	2	"f.mvt mns hauteur"		
abaisser	04	AVOIR	CHANTER	MENT:Z	T1306+P3006	"baisse"	M4b	MON	2	"f.mvt mns val"		
abaisser	05	AVOIR	CHANTER	MENT:Z	T1308+P3008	"baisse"	T4b	MED	5	"r/d bas quant"		
abaisser	06	AVOIR	CHANTER	ANT:Z+MENT:Z	T1108+P1000	"baisse"	R2a	PSYt	5	"m.e.état mvs qn abs"		
abaisser	07	AVOIR	CHANTER	MENT:Z	P10a0	"baisse"	M2b	PSY	5	"f.mvt humble A+inf"		Vpro...
abaisser	08	AVOIR	CHANTER		P3001	"baisse"	L3a	VEH	5	"(voie)lc.qp vrs bas"		Vpro...
abaisser	09	AVOIR	CHANTER		P10a0+T11a0	"baisse"	M2b	PSY	5	"f.mvt A état vil"		Vpro...
abalourdir		AVOIR	FINIR	MENT:Z	P1000+T9106	"balourd"	T2b	PSYt	6	"r/d balourd-a"		Vpro...
abandonner	01	AVOIR	CHANTER	DEVERBAL25:Crayo...	T13a0	"abandon"	D2a	DRO	1	"dat qc A qn"		
abandonner	02	AVOIR	CHANTER	DEVERBAL25:Crayon	T13a8+P30a8	"abandon"	D3a	MAR	5	"dat mvt A qc"		
abandonner	03	AVOIR	CHANTER	DEVERBAL25:Crayon	T1300	"abandon"	S3b	EQU	5	"dgrp rênes"		
abandonner	04	AVOIR	CHANTER	ADJ_PP:Z+DEVERBA...	T1300	"abandon"	S4b	PSY	1	"dgrp abs"		
abandonner	05	AVOIR	CHANTER	DEVERBAL25:Crayon	T1300+A10	"abandon"	S4b	SOC	1	"dgrp abs"		
abandonner	06	AVOIR	CHANTER	DEVERBAL25:Crayon	T1300+A10	"abandon"	S2e	SPO	1	"dgrp lutte"		
abandonner	07	AVOIR	CHANTER	DEVERBAL25:Crayon	T13k0	"abandon"	T3f	COM	5	"li.mut CT/PR"		
abandonner	08	AVOIR	CHANTER	ADJ_PP:Z+DEVERBA...	T1307	"abandon"	S3h	SOC	5	"dgrp lc"		
abandonner	09	AVOIR	CHANTER	MENT:Z+DEVERBAL2...	T1101	"abandon"	S2d	LOC	5	"dgrp qn qp"		
abandonner	10	AVOIR	CHANTER	DEVERBAL25:Crayon	T1300+A10	"abandon"	S4b	SOC	5	"dgrp abs"		
abandonner	11	AVOIR	CHANTER		T3100	"abandon"	S2d	SOC	5	"(qc)dgrp qn"		
abandonner	12	AVOIR	CHANTER		P10a0	"don"	P1h	PSY	5	"ger.mens A qc/inf"		Vpro...
abandonner	13	AVOIR	CHANTER	DEVERBAL25:Crayon	P1006	"abandon"	C1a	LANt	5	"loq confidence"		Vpro...
abasourdir	01	AVOIR	FINIR	ADJ_PP2:Z+ANT2+M...	T1108	"sourd"	H1b	PAT	2	"f.som étourdi"		
abasourdir	02	AVOIR	FINIR	ADJ_PP2:Z+ANT2+M...	T3100		P2a	PSY	5	"(abs)f.sent vif à...		
abâtardir	01	AVOIR	FINIR	MENT:Z	T1308+P3008	"bâtard"	T4b	LITt	3	"r/d bâtard abs"		
abâtardir	02	AVOIR	FINIR	ADJ_PP2:Z+MENT	P3000+T3308	"bâtard"	T3a	BIOt	5	"(qc)d bâtard"		Vpro...
abat-carrer		AVOIR	CHANTER	AGE:Z	T1308+P3000	"abat-..."	R3f	PEA	5	"rag qc av abat-ca..."		
abattre	01	AVOIR	BATTRE	ABLE:Z+ANT+AGE:Z...	T13g8		E3f	OBJ	1	"f.iire qc e bas"		
abattre	02	AVOIR	BATTRE		T1308		F3c	MIL	5	"ict mvs aéro"		

Exemple d'ambiguïté sémantique

- Dans LVF, on a 5 entrées pour les différents emplois du verbe *abriter* :

abriter #1 (T11b8, P10b8) : *Luc abrite Léa de la pluie avec un parapluie*

abriter #2 (T1101, P1001) : *Luc abrite des réfugiés chez lui*

abriter #3 (T3100) : *Cet immeuble abrite le service juridique*

abriter #4 (P10b1) : *Luc s'abrite des ennuis derrière un prétexte*

abriter #5 (T13b8) : *On abrite le port des vagues avec des digues*

Structures Syntactiques dans LVF

- A (intransitive), ex. : *On arrive à Lyon*
- N (directe, intransitive), ex. : *Les échecs alternent avec les succès*
- P (pronominale), ex. : *On s'accommode de la situation*
- T (directe, transitive), ex. : *La chaleur accable les estivants*

⇒ 4 grammaires génériques

Levée d'ambiguïté : les emplois transitifs directs

- Dans LVF, on a 5 entrées pour les différents emplois du verbe *abriter* :

abriter #1 (**T11b8**, P10b8) : *Luc abrite Léa de la pluie avec un parapluie*

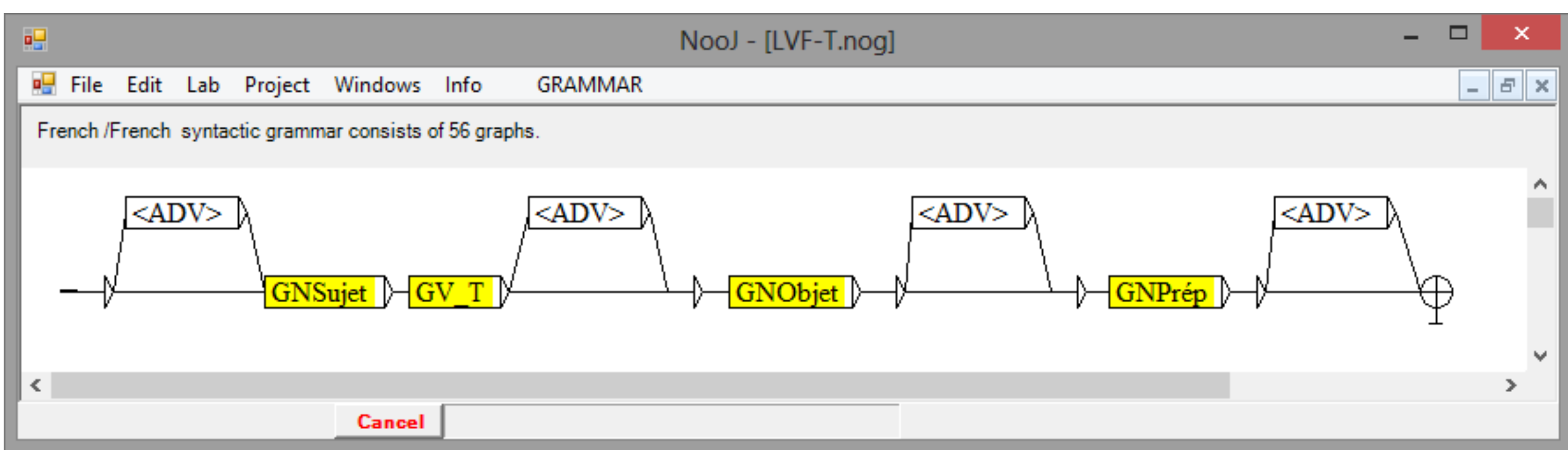
abriter #2 (**T1101**, P1001) : *Luc abrite des réfugiés chez lui*

abriter #3 (**T3100**) : *Cet immeuble abrite le service juridique*

abriter #4 (P10b1) : *Luc s'abrite des ennuis derrière son chef*

abriter #5 (**T13b8**) : *On abrite le port des vagues avec des digues*

Grammaire qui représente les phrases transitives directes



Si on applique la grammaire T \Rightarrow sens 3

centre de Pristina	, il abritait plusieurs partis politiques des minorités ethniques	. L'attentat visait
selon vous, Monsieur	, le bâtiment qui abrite la plus grande quantité de rêves	? L'école? Le
chimiques et bactériologiq...	Ce pays abrite le plus grand réseau mondial de drogue	, et ses banques
froideur presque menaçante	, il abrite des banques	et un inévitable
de produits nucléaires...	Chaque pays abrite ses milieux criminels	. Les principales orga
journée pour 5 livres (6)	: elle abrite un cirque	(sans animaux), un
Alexandra et Lenasia	, elle abrite deux millions cinq cent mille personnes	. Un grand bidonville
'Etat de droit.	Le Cambodge abrite une société structurée	non pas en
ses contradictions internes.	Ce pays montagneux abrite un grand nombre de groupes ethniques	, y compris parmi
terre de Flandre	, la ville qui abrite la Commission européenne	suscite bien des

Levée d'ambiguïté : les emplois pronominaux

- Dans LVF, on a 5 entrées pour les différents emplois du verbe *abriter* :

abriter #1 (T11b8, **P10b8**) : *Luc abrite Léa de la pluie avec un parapluie*

abriter #2 (T1101, **P1001**) : *Luc abrite des réfugiés chez lui*

abriter #3 (T3100) : *Cet immeuble abrite le service juridique*

abriter #4 (**P10b1**) : *Luc s'abrite des ennuis derrière son chef*

abriter #5 (T13b8) : *On abrite le port des vagues avec des digues*

Si on applique la grammaire P \Rightarrow sens 4

première fois aussi	, le premier ministre israélien ne s'abrita pas derrière les slogans traditionnels	- "Jérusalem un
désarmé. En Europe	, il s'abrite derrière les pressions exercées	par le Luxembo
la corruption endémique.	Les responsables politiques s'abritent derrière le carcan traditionnel	des "valeurs a
prend pas parti.	Il s'abrite derrière sa tâche d'homme	de théâtre qui
dissimuler le fond	, il s'abrite en général derrière des considérations de forme,	et remet tout
système financier international.	Il s'abrite derrière le contentieux colonial,	dont il exige
à visage découvert	, ils s'abritent derrière des notables traditionnels	qui les utilisent

Conclusion

- On peut décrire toutes les unités linguistiques atomiques (ALU) à l'aide de dictionnaires et de grammaires : mots simples, affixes, mots composés et expressions
- On peut aussi décrire des structures, ex. compléments de date, déterminants, groupes nominaux, etc. et les représenter dans la TAS
- La TAS permet de représenter toutes les ambiguïtés potentielles
- On peut décrire toutes les séquences valides d'ALU à l'aide de grammaires syntaxiques
- Les grammaires peuvent être utilisées pour supprimer des annotations de la TAS, c'est-à-dire pour lever des ambiguïtés