

# Analyses linguistiques

[max.silberztein@univ-fcomte.fr](mailto: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.

- C'est une phrase
- Où sont les **groupes nominaux**
- Où sont les **noms propres**
- Où sont les compléments
- ...

# 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.

- C'est 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 »
- Chaque analyse va ajouter ou supprimer des annotations

# Alimenter la structure d'annotations

On ajoute des annotations pour décrire des séquences de texte :

- formes simples, ex.: *mangerions* => <manger,V+tr+C+1+p>
- formes agglutinées, ex. : *du* => <de,PREP> <le,DET+m+s>
- 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+FR> → (colons français en Algérie)

<pied noir,NOM+Hum+p+USA> → (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> → *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*

<l'infirmière ferme,N0> <la,N1> <porte,VERBE>

<l'infirmière,N0> <ferme,VERBE> <la porte,N1>

*Luc recouvre la bouteille de plastique*

<Luc,N0> <recouvre,VERBE> <la bouteille de plastique,N1>

<Luc,N0> <recouvre,VERBE> <la bouteille,N1> <de plastique,N2>

# 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:  
50722 word forms

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 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	477	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+z1
	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 contains a paragraph from Balzac's "La femme de trente ans". The phrase "autour du cou de son guide" is highlighted in blue. Below the text, a table displays the morphological annotations for each word in the phrase. A red circle highlights the annotations for "du", and a red vertical bar highlights the annotations for "autour".

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*

The screenshot shows the NooJ software interface with a text document titled "[La femme de trente ans.not]". The main window displays a paragraph of text from Balzac's "La femme de trente ans". The phrase "leur ciel sans nuages" is highlighted in blue. Below the text, a table shows the morphological analysis of the words. A red oval highlights the analysis for "ciel sans nuages".

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

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	nuage,N+z1+m+p
leur,PRO+z1+3+m+p	ciel,INTJ+z1		nuager,V+z3+PR+2+s

Cancel



# Annotation d'expressions discontinues

The screenshot shows the NooJ software interface with the following components:

- Window Title:** NooJ - [zzz.not]
- Menu Bar:** File, Edit, Lab, Project, Windows, Info, TEXT
- Toolbar:** Includes a minus sign, a box with '1', and a plus sign, followed by '/ 1 TUs'.
- Left Panel:** A list of annotation types: Characters, Tokens, Digrams, Annotations, and Unknowns. The 'Annotations' option is selected. A checkbox labeled 'Show Text Annotation Structure' is checked.
- Main Text Area:** Displays the sentence 'Luc accéléra soudain le mouvement'.
- Annotation Diagram:** A diagram below the text showing the morphological and syntactic analysis of the words. It consists of four vertical columns representing different parts of the sentence, with arrows indicating the flow of information and dependencies between them.

Column 1 (Index 4)	Column 2 (Index 13)	Column 3 (Index 21)	Column 4 (Index 24)
accélérer, V+C1D+N0Hum+N0NHum+DET1=le+N1=mouvement+Passif+PS+3+s	soudain, A+z1+m+s	DET	N
accélérer, V+z1+PS+3+s	soudain, ADV+z1	le, PRO+z1+3+m+s	mouvement, N+z1+m+s
		le, DET+z1+m+s	
- Bottom Panel:** A 'Cancel' button.

Exercice : Construire le dictionnaire correspondant à la phrase :

*Pierre a volé un jeu de cartes vertes du bureau général*

Exercice : Construire le dictionnaire correspondant à la phrase :

*Pierre a volé un jeu de cartes vertes du bureau général*

Pierre,PRENOM+masculin

pierre,N+concret+féminin

voler,V+Transitif+FLX=AIMER

voler,V+Intransitif+FLX=AIMER

un,DETERMINANT+indéfini

jeu,N+abstrait

jeu de cartes,N+abstrait

de,PREP

carte,N+concret+féminin

vert,ADJ

carte verte,N+Abstrait+féminin

du,DET+partitif

le,DET+défini

bureau,N+concret

bureau,N+lieu

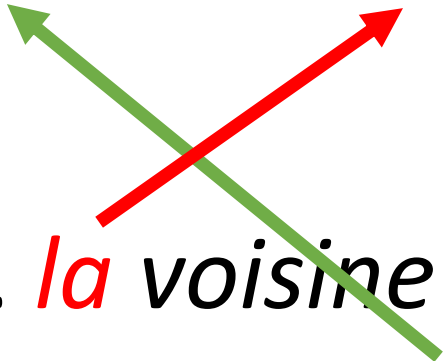
bureau général,N+organisation

général,N+humain

général,ADJ

# 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*

la/<**PRONOM**> signifie qu'on impose que le mot "la" soit un pronom → on supprime toutes les autres annotations

la/<**DETERMINANT**> 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*

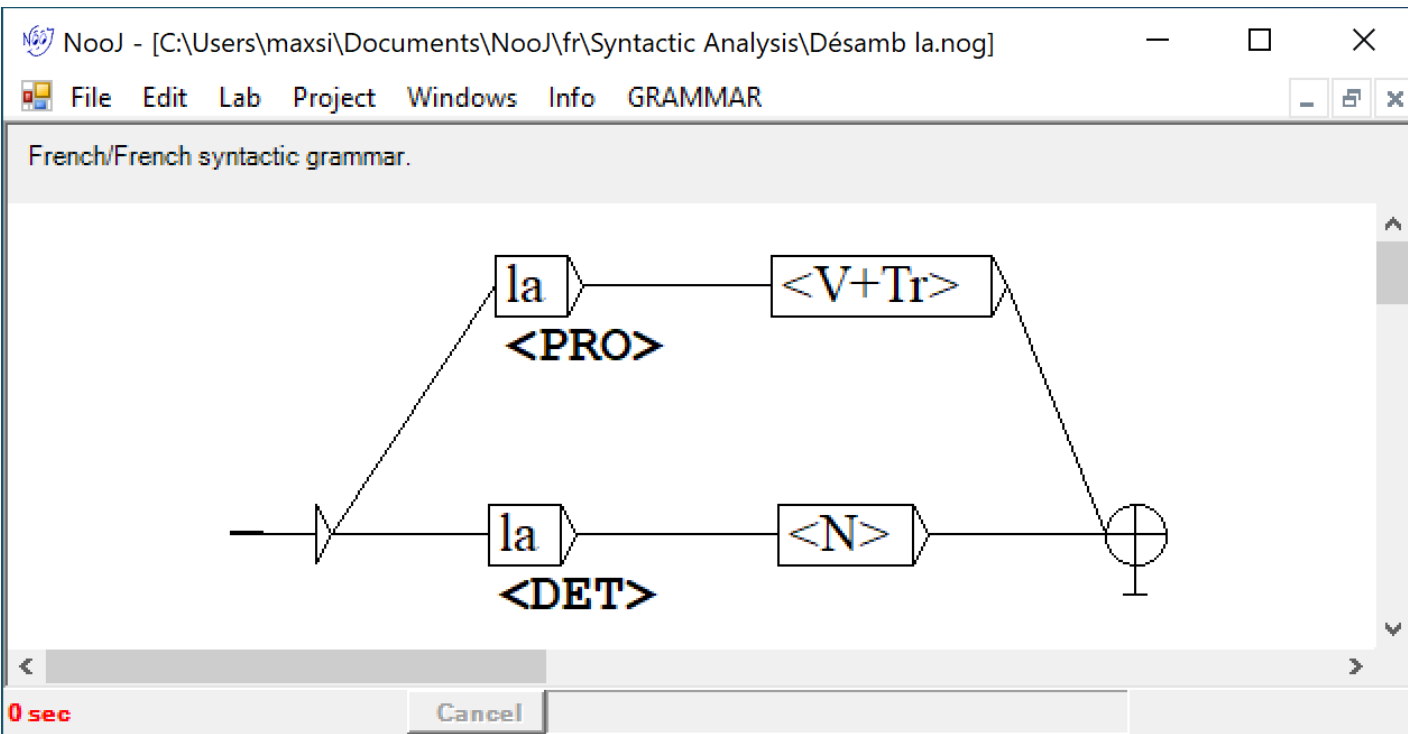
- Si le mot “la” est suivi d'un **<V+Transitif>**, alors c'est un **<PRONOM>**
- Si le mot “la” est suivi par un **<N>**, alors c'est un **<DETERMINANT>**

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



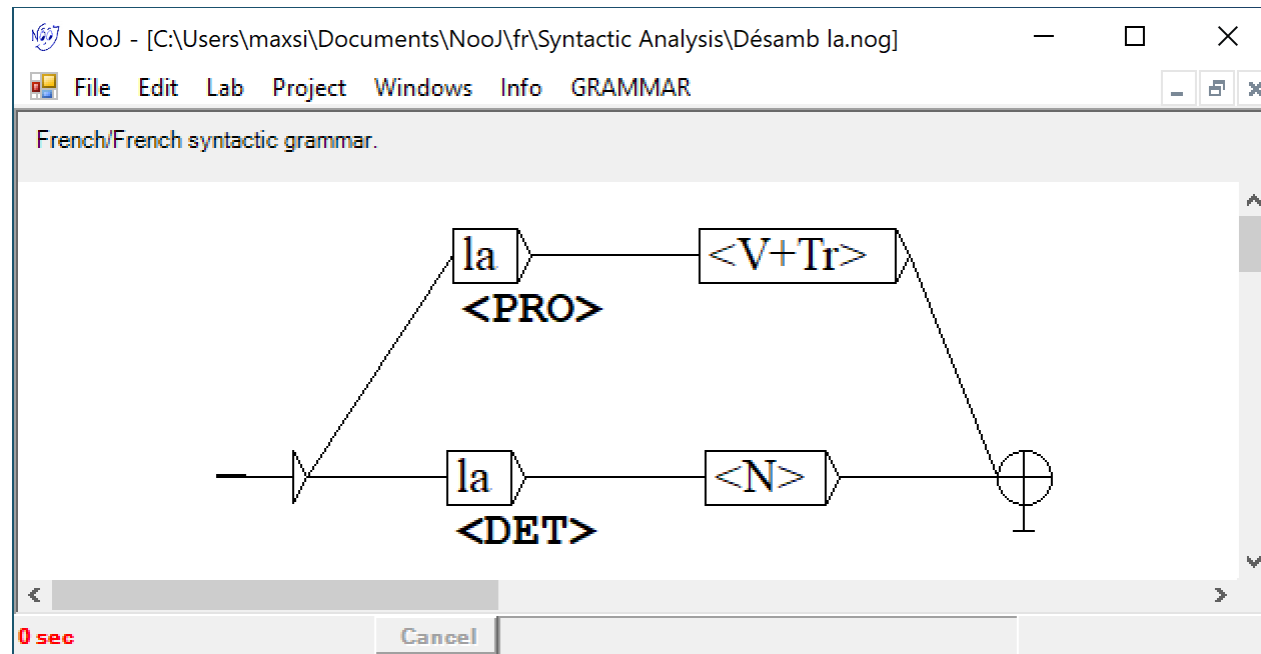
- Si le mot “la” est suivi d’un verbe transitif <V+Tr>, alors c’est un <PRO>
- Si le mot “la” est suivi par un nom, alors c’est un <DET>

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

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

$c \mid c' / \langle \text{PRO} \rangle \text{ est} / \langle \text{V} \rangle$

→ 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. :

$\langle \text{V} \rangle / \langle \text{!V+Pronominal} \rangle \mid (\text{me} \mid \text{te} \mid \text{se}) \langle \text{V+Pronominal} \rangle$

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

$(\text{je} \mid \text{j}') \langle \text{V+1+s} \rangle / \langle \text{V+1+s} \rangle \mid \langle \text{V} \rangle / \langle \text{!V+1+s} \rangle$

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

$\text{de} / \langle \text{PREP} \rangle \mid \text{de} (\text{très} \mid \langle \text{E} \rangle) \langle \text{A+p} \rangle \langle \text{N+p} \rangle$

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

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

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	terrer, 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	terrer, 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		terrer, 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		terrer, V+Aux=a+Pers=3+Nb=s+Temps=S	
			pommer, V+Aux=e+Pers=3+Nb=s+Temps=PR		terrer, 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

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 ... and f ...

Show Text Annotation Structure

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

NooJ - [analyse syntaxique 2.nog]

File Edit Lab Project Windows Info GRAMMAR

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

<PHRASE <GN+Sujet <DET> <N> <GV <V> <GN+Objet <DET> <N>

Cancel PROJECT MODE

# Une grammaire annote les constituants

Noo

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> <GV> <GN+Objet>

Cancel PROJECT MODE

# Une grammaire annote les constituants

The screenshot displays the Noo software interface. The main window, titled 'analyse syntaxique.nog', shows a syntactic graph for the French noun phrase 'GN'. The graph starts with a root node 'GN' (represented by a triangle) which branches into three children: 'DET' (a yellow box), '<A>' (a white box), and '<N>' (a white box). The 'DET' node further branches into 'DetDef' (a triangle), which then branches into 'DetDefPlur' and 'DetDefSing'. The '<N>' node branches into 'DetIndef' (a triangle). The graph ends with a terminal node 'GV' (a circle with a cross). The interface includes a menu bar (File, Edit, Lab, Project, Windows, Info), a toolbar (Expand, Collapse, Refresh), and a status bar at the bottom with 'Cancel' and 'PROJECT MODE' buttons.

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.

GN

DET <A> <N> GV

Cancel PROJECT MODE

# Une grammaire annote les constituants

The screenshot displays the Noo! software interface. The main window, titled "analyse syntaxique.nog", shows a syntactic grammar for French determiners. The grammar consists of 8 graphs. The current graph, labeled "DET", is a simple disjunctive structure. It starts with a root node (a triangle with an arrow) that branches into two paths. The upper path leads to a node labeled "DetDef", and the lower path leads to a node labeled "DetIndef". Both "DetDef" and "DetIndef" nodes are highlighted in yellow. These two paths converge at a final node (a circle with a cross), which represents the end of the derivation. The left sidebar shows a tree structure of the grammar, with "DET" selected. The bottom of the window features a "Cancel" button and the text "PROJECT MODE" in red.

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.

DET

DetDef

DetIndef

Cancel PROJECT MODE

# Une grammaire annote les constituants

Noo

File Edit Lab Project Windows Info GRAMMAR

Structure of analyse syntaxiq...

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

analyse syntaxique.nog

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

DetDef

DetDefSing

DetDefPlur

Cancel PROJECT MODE



# Une grammaire annote les constituants

The screenshot displays the Noo software interface, which is used for defining and editing syntactic grammars. The main window, titled "analyse syntaxique.nog", shows a rule named "DetDefSing". The rule is represented as a graph with a start node (a triangle with a vertical line) on the left and an end node (a circle with a vertical line) on the right. A box in the middle contains four alternative paths, each starting with a left angle bracket and ending with a right angle bracket. The paths are:  $\langle le, le, DET+m+s \rangle$ ,  $\langle la, le, DET+f+s \rangle$ ,  $\langle l', le, DET+m+s \rangle$ , and  $\langle l', le, DET+f+s \rangle$ . On the left side of the interface, a tree view shows the structure of the grammar, with "DetDefSing" highlighted in blue. The bottom of the window features a "Cancel" button and the text "PROJECT MODE" in red.

# TEXT > Locate

The screenshot shows the Nool software interface. The main window is titled 'CONCORDANCE' and contains a text editor with the text 'Le chat mange le rat.' and a list of analysis options: Characters, Tokens, Digrams, Annotations, and Unknowns. A 'Show Text Annotation Structure' checkbox is present. Below the text editor is a 'Concordance for Text analyse syntaxique.not' window. This window has a 'Reset' button, a 'Display' field set to 5, and radio buttons for 'characters' and 'word forms'. It also has fields for 'before, and' and 'after' (both set to 5), and checkboxes for 'Matches' and 'Outputs'. The concordance table has columns for 'Text', 'Before', and 'Seq.'. The first row shows the text 'Le chat mange le rat/' followed by a red NLP tag sequence: '<PHRASE<GN+Sujet:GN:DET<le,le,DET+z1+m+s:'. The 'Query' field at the bottom shows '1/1'. To the right, a 'Locate a pattern in' dialog box is open. It has a 'Pattern is:' section with radio buttons for 'a string of characters:', 'a PERL regular expression:', 'a Nool regular expression:', and 'a Nool grammar:'. The 'a Nool grammar:' option is selected, and a file path 'C:\Users\Max\Documents\Nool\fr\Syntactic Ana...' is entered. A red circle highlights the 'Syntactic Analysis' checkbox, which is checked. Below this, there are 'Index' and 'Limitation' sections. The 'Index' section has radio buttons for 'Shortest matches', 'Longest matches', and 'All matches'. The 'Limitation' section has radio buttons for 'Only:' (set to 100 matches), 'All matches', and '1 example per match'. A 'Reset Concordance' checkbox is checked at the bottom of the dialog box.

analyse syntaxique.not

1 / 2 TUs

Characters  
Tokens  
Digrams  
Annotations  
Unknowns

Language is  
Text Delimi  
Text contai  
6 tokens in

Show Text Annotation Structure

Le chat mange le rat.

Concordance for Text analyse syntaxique.not

Reset Display: 5 characters before, and 5 after. Display:  Matches  Outputs

word forms

Text	Before	Seq.
Le chat mange le rat/		<PHRASE<GN+Sujet:GN:DET<le,le,DET+z1+m+s:

Query 1/1

Locate a pattern in

Pattern is:

a string of characters:  
 a PERL regular expression:  
 a Nool regular expression:  
 a Nool grammar:

C:\Users\Max\Documents\Nool\fr\Syntactic Ana... Set

Syntactic Analysis

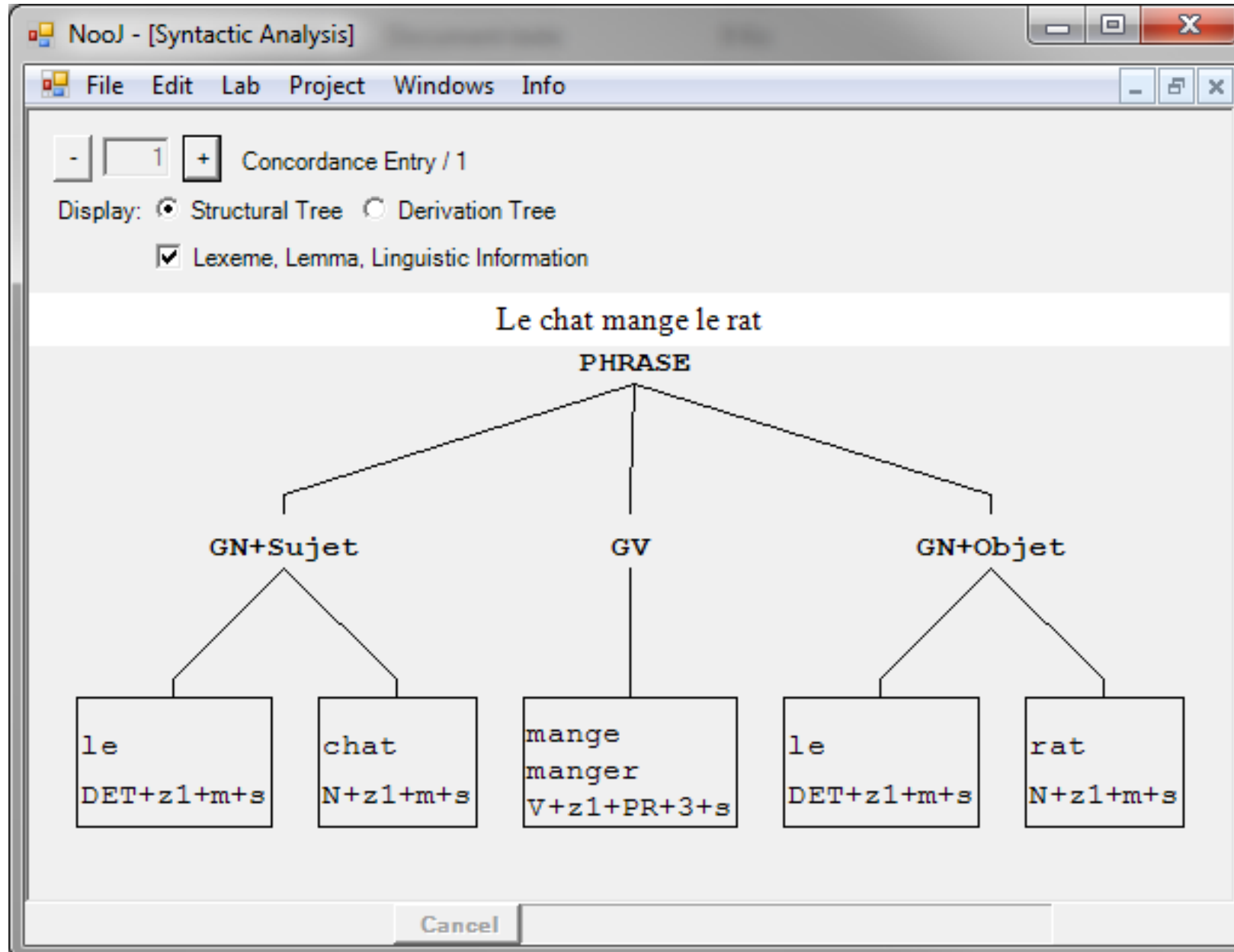
Index Limitation

Shortest matches  Only: 100 matches  
 Longest matches  All matches  
 All matches  1 example per match

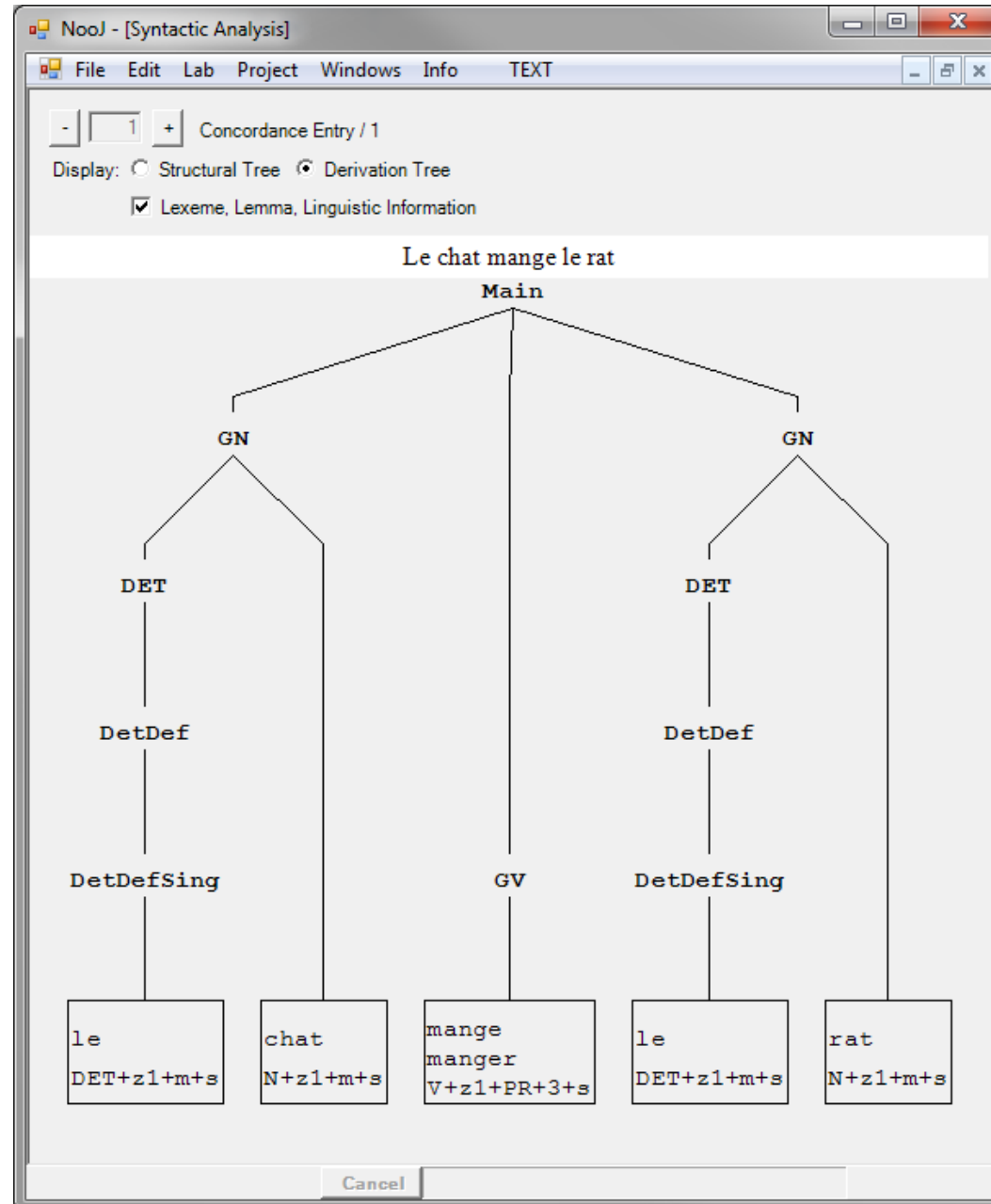
Reset Concordance

Cancel

# CONCORDANCE > Display Syntactic Analysis



# arbre structurel $\neq$ arbre de dérivation



# Types d'ALU : contraction *aux*

The screenshot shows the NooJ software interface. The main window displays 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 "Parse Tree" unselected. Below this, the sentence is analyzed into a syntax tree structure:

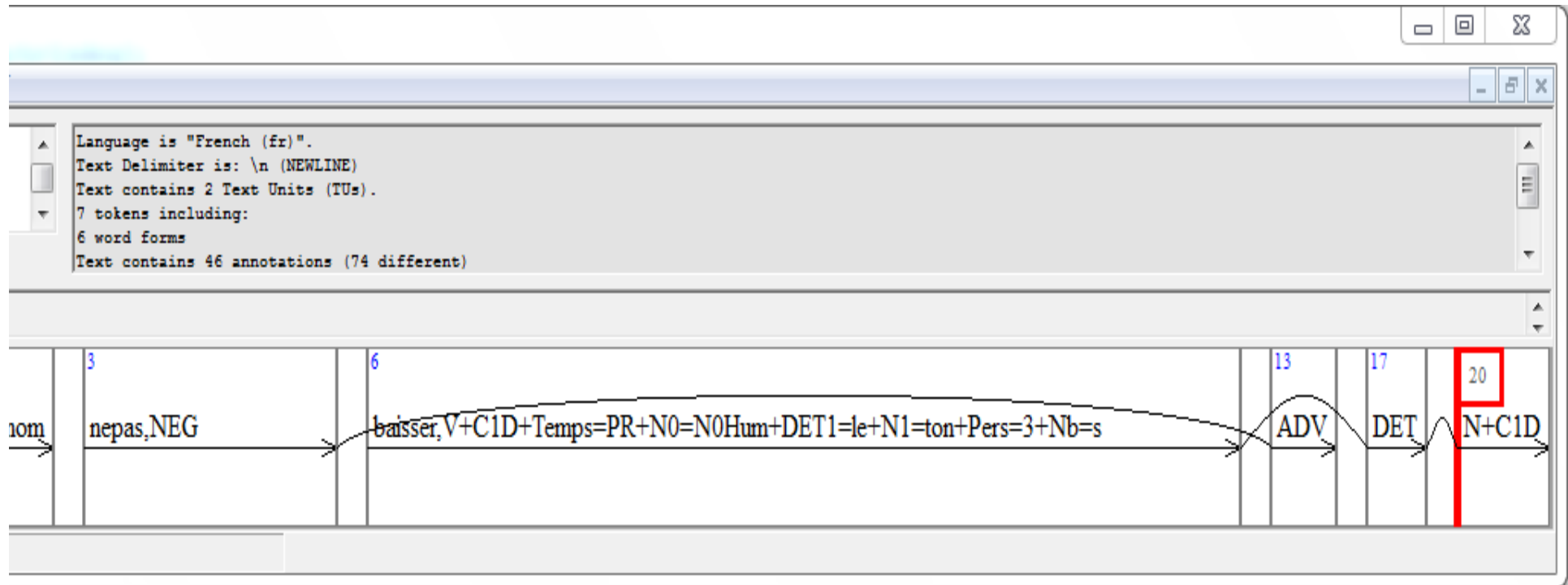
```
graph TD
    PHRASE[PHRASE] --- SUJET[SUJET]
    PHRASE --- VERBE[VERBE]
    PHRASE --- OBJET[OBJET]
    SUJET --- les1[les]
    SUJET --- tambours[tambours]
    VERBE --- battirent[battirent]
    OBJET --- a[à]
    OBJET --- les2[les]
    OBJET --- champs[champs]
```

The tree structure shows the sentence "les tambours battirent aux champs" broken down into its constituent parts: SUJET (SUBJECT) containing "les" and "tambours"; VERBE (VERB) containing "battirent"; and OBJET (OBJECT) containing "à", "les", and "champs".

At the bottom of the dialog box, there is a "Cancel" button.

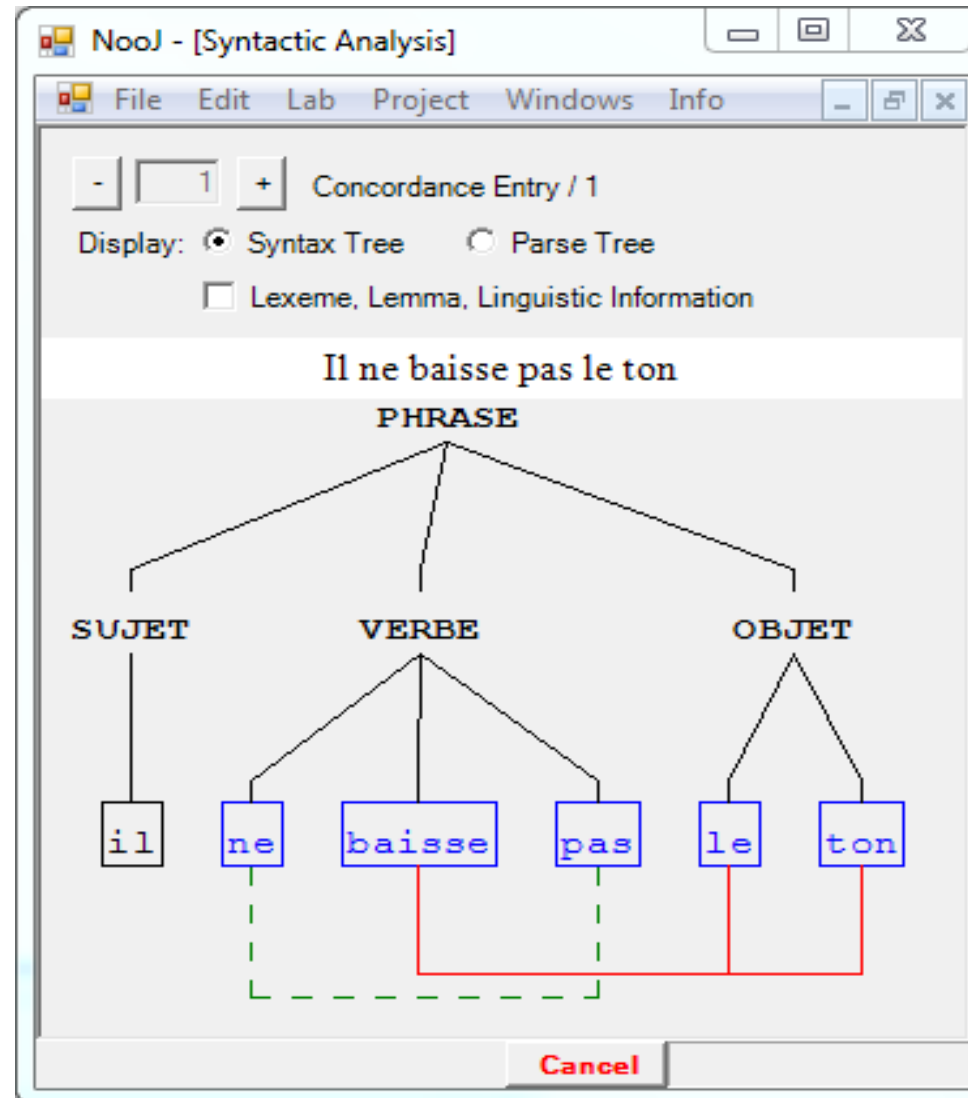
# Types d'ALU : expressions discontinues

Il ne *baisse* pas *le ton*



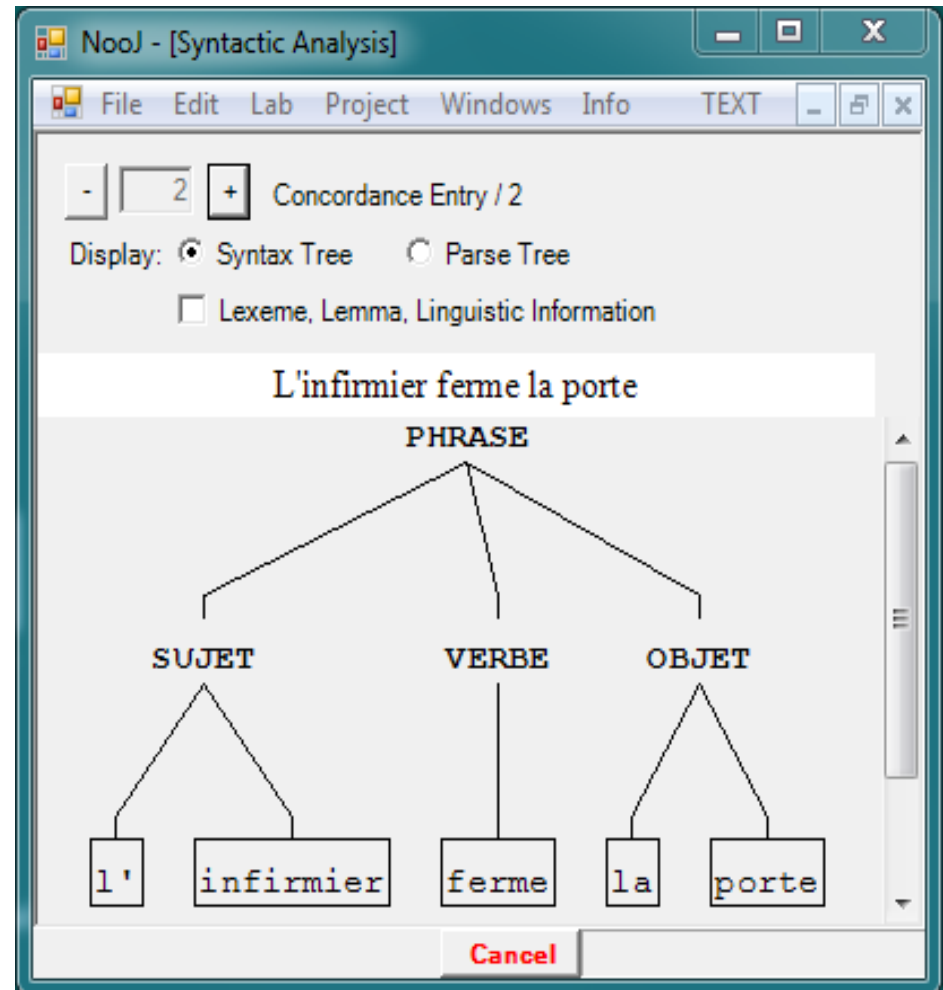
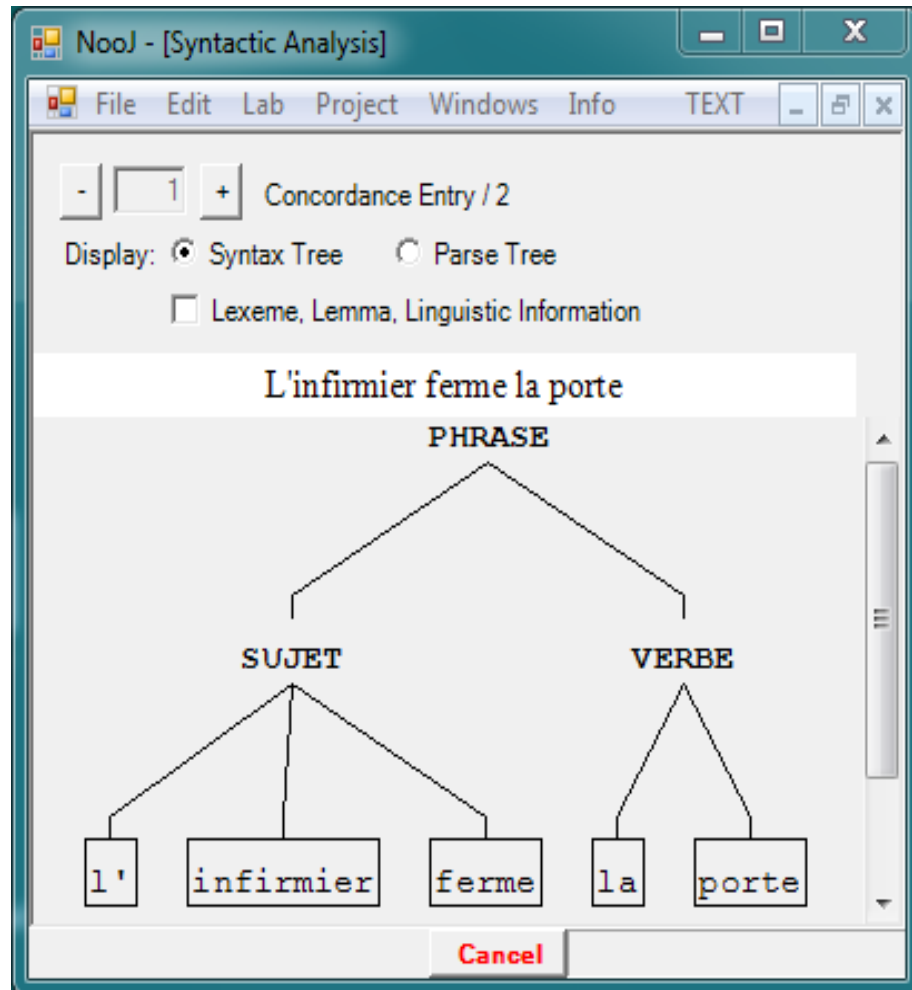
# Types d'ALU : expressions discontinues

Il ne *baisse* pas le ton



# Ambiguïtés structurelles

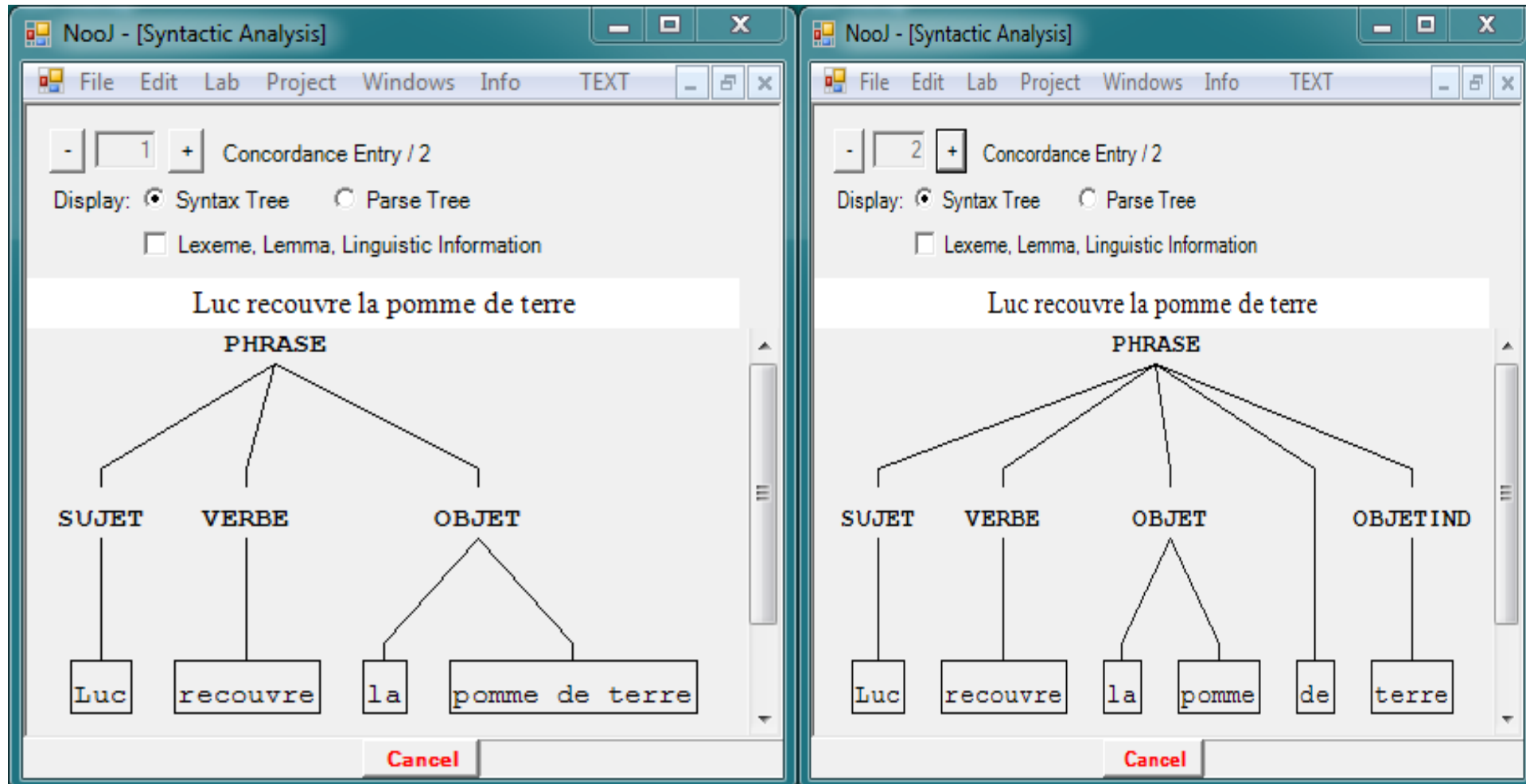
*L'infirmier ferme la porte*





# Ambiguïtés lexicales $\Rightarrow$ structurelles

*Luc recouvre la pomme de terre*



# Représentation de la structure syntaxique dans la TAS

The screenshot shows the NooJ interface with a dialog box indicating that text was annotated. The main window displays the sentence "Le chat mange le rat." and a table representing its syntactic structure. The table has columns for phrase boundaries (0, 3, 8, 14, 17) and rows for different syntactic levels: 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, and rat, N+z1+m+s. Green arrows indicate the hierarchical relationships between these elements. A red box highlights the number 0 in the first column. The interface also includes a menu bar (File, Edit, Lab, Project, Windows), a toolbar with a minus sign, a text input field containing "1", and a plus sign, and a checkbox labeled "Show Text Annotation Structure". A "Cancel" button is visible at the bottom left, and "PROJECT MODE" is displayed in red at the bottom right.

NooJ - [analyse syntaxique.not]

File Edit Lab Project Windows

1 / 2 TUs

Show Text Annotation Structure

Ambiguities

NooJ: text was annotated

4 annotations were added; 9 annotations were removed

OK

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**

# Analyse linguistique sans analyse syntaxique

The screenshot shows the NooJ software interface with the following components:

- Window title: NooJ - [analyse syntaxique.not]
- Menu bar: File, Edit, Lab, Project, Windows, Info, TEXT
- Toolbar: - 1 + / 2 TUs
- Checkboxes:  Show Text Annotation Structure
- Dropdown menu: Characters, Tokens, Digrams, Annotations, Unknowns
- Text area: Language is "French (France) (fr)".  
Text Delimiter is: \n (NEWLINE)  
Text contains 2 Text Units (TUs).  
6 tokens including:  
e ..... f.....
- Text input field: Le chat mange le rat.
- Table of morphological annotations:

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

→ L'analyse syntaxique lève les ambiguïtés

The screenshot shows the NooJ software interface for syntactic analysis. The window title is "NooJ - [analyse syntaxique.not]". The menu bar includes "File", "Edit", "Lab", "Project", "Windows", "Info", and "TEXT". The toolbar shows a minus sign, a box with "1", and a plus sign, followed by "/ 2 TUs". A dropdown menu is open, listing "Characters", "Tokens", "Digrams", "Annotations", and "Ambiguities". A checkbox labeled "Show Text Annotation Structure" is checked. A text area on the right contains the following information: "Language is 'French (France) (fr)'.", "Text Delimiter is: \n (NEWLINE)", "Text contains 2 Text Units (TUs).", and "6 tokens including: e...". The main text area displays the sentence "Le chat mange le rat." Below this is a table representing the syntactic analysis. The table has five columns with indices 0, 3, 8, 14, and 17. The first column (index 0) is highlighted with a red box. The table contains the following data:

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

Arrows indicate the flow of the analysis: from the first column to the second, from the second to the third, from the third to the fourth, and from the fourth to the fifth. A "Cancel" button is located at the bottom left, and "PROJECT MODE" is displayed in red text at the bottom right.

# Levée d'ambiguïté sémantique

NooJ - [LVF 2014 06 01.dic]  
File Edit Lab Project Windows Info DICTIONARY  
Dictionary contains 25609 entries

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.ire 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.ire qc e bas"		
abattre	02	AVOIR	BATTRE		T1308		F3c	MIL	5	"ict mvs aéro"		

449 sec Cancel

# 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 Syntaxiques 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*

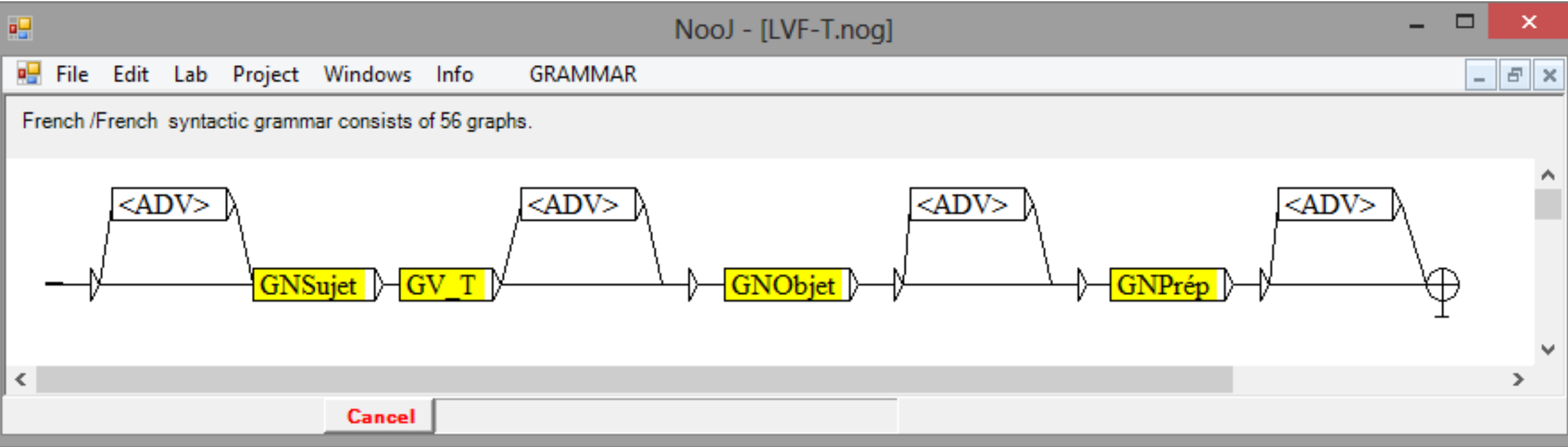
*abriter* #3 (**T3100**) : *Cet immeuble abrite le service juridique*

*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  
selon vous, Monsieur  
chimiques et bactériologiq...  
froideur presque menaçante  
de produits nucléaires...  
journée pour 5 livres (6)  
Alexandra et Lenasia  
'Etat de droit.  
ses contradictions internes.  
terre de Flandre

, il abritait plusieurs partis politiques des minorités ethniques  
, le bâtiment qui abrite la plus grande quantité de rêves  
Ce pays abrite le plus grand réseau mondial de drogue  
, il abrite des banques  
Chaque pays abrite ses milieux criminels  
: elle abrite un cirque  
, elle abrite deux millions cinq cent mille personnes  
Le Cambodge abrite une société structurée  
Ce pays montagneux abrite un grand nombre de groupes ethniques  
, la ville qui abrite la Commission européenne

. L'attentat visait  
? L'école? Le  
, et ses banques  
et un inévitable  
. Les principales orga  
(sans animaux), un  
. Un grand bidonville  
non pas en  
, y compris parmi  
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 désamé. En Europe la corruption endémique. prend pas parti. dissimuler le fond système financier international. à visage découvert	, le premier ministre israélien ne s'abrita pas derrière les slogans traditionnels , il s'abrite derrière les pressions exercées Les responsables politiques s'abritent derrière le carcan traditionnel Il s'abrite derrière sa tâche d'homme , il s'abrite en général derrière des considérations de forme, Il s'abrite derrière le contentieux colonial, , ils s'abritent derrière des notables traditionnels	- "Jérusalem un par le Luxembo des "valeurs a de théâtre qui et remet tout dont il exige qui les utilisent
--	---	--

# Conclusion

- On décrit les unités linguistiques atomiques (ALU) à l'aide de dictionnaires et de grammaires morphologiques : mots simples, affixes, mots composés et expressions
- On décrit les structures syntaxiques, ex. compléments de date, déterminants, groupes nominaux, etc.
- La TAS permet de représenter toutes ces analyses, avec des ambiguïtés potentielles
- Les grammaires syntaxiques peuvent être utilisées pour supprimer des annotations de la TAS, c'est-à-dire pour lever des ambiguïtés lexicales, morphologiques, syntaxiques ou sémantiques.