

LING3822 : SYNTAXE

LABORATOIRES D'OBSERVATION DE LA LANGUE

au moyen de l'analyseur linguistique

NooJ

®

Auteure :
Gisèle Chevalier
chevalg@umoncton.ca



CRLA, Université de Moncton
© Automne 2005

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Ce cahier est un premier pas vers l'élaboration d'un matériel d'accompagnement pour l'enseignement/l'apprentissage de la syntaxe destiné à des étudiants qui ne se sentent pas de vocation pour la linguistique, mais qui doivent obligatoirement aborder ce domaine sans toujours maîtriser les notions grammaticales préalables.

Les cinq laboratoires que je mets en ligne dans l'espoir de trouver une éventuelle collaboration, visent à montrer comment appréhender NooJ et obtenir des « résultats » assez rapidement. C'est un premier jet, improvisé en quelques heures, qui, je l'espère mènera à une trousse bien structurée dans quelques mois.

Vous pouvez télécharger le programme NooJ gratuitement et vous joindre à la communauté NooJ en allant sur le site suivant :

<http://perso.wanadoo.fr/rosavram/pages/noojpag.html>

Les activités exploitent le texte La dame de trente ans, qui vient avec le logiciel et les ressources lexicales du français.

Le développement de ce matériel est fondé sur le manuel d'utilisation de INTEX, et des des chapitres disponibles en novembre 2005 des manuels complémentaires en préparation pour l'utilisation de NooJ, celui que prépare Michel Roux, pour le français, et celui de Max Silberztein, pour l'anglais. (Consulter l'onglet (doc & help) sur le site NooJ).

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Premier laboratoire

NooJ

NooJ et les données numériques sur un texte

Préambule

La présente police de caractère indique les actions à faire.
Les apostrophes ‘contiennent un choix ’ dans un menu déroulé.
La flèche → indique le résultat des actions prises : apparition d’une nouvelle fenêtre, ou modification dans la fenêtre active ou la barre de menu.

Le point • introduit des commentaires sur les informations que produisent les fonctions que l’on explore ou vous invite à faire des observations particulières.

Les paragraphes à fond verdâtre contiennent des suggestions concernant la navigation dans l’accomplissement de certaines tâches répétitives.

1. Ouverture du programme NooJ

Programme >> NooJ

→ Une boîte apparaît

Run

OK pour tout message, s’il y a lieu, jusqu’à l’ouverture de NooJ.

2. Choisir la langue de travail

Info >> Preferences

- General : fr
- Lexical Analysis : cocher
 - **delaf.nod** : l’extension ‘.nod’ correspond à NooJ Dictionary)
 - **élision.nod** : reconnaissance des élisions *du, de la, des, au, etc.*
- Syntactic analysis : ne pas modifier.

OK

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> ‘delaf.nod’ et ‘élision’.

3. Ouvrir un texte à analyser

Project >> run project >> la femme de trente ans.nop

- REM : ce texte est exempt de droits d'auteurs. Son utilisation est libre.
- REM : .nop signifie NooJ Project

➔ Plusieurs fenêtres apparaissent.

- La première fenêtre qui apparaît donne des informations sur l'auteur du projet, sur la source du texte ouvert, et des données statistiques sur le texte :
 - nombre de paragraphes
 - le nombre de « tokens » : nombre de formes différentes qui apparaissent au moins une fois, et indépendamment de leur fréquence;
 - le genre de tokens;
 - les 'annotations', soit le nombre de mots étiquetés quant à la catégorie grammaticale.
- La 2^e fenêtre : 'Locate pattern' (la fermer pour l'instant)
- La 3^e fenêtre : C'est le texte à traiter : La femme de trente ans

- La grande fenêtre Windows indique PROJECT MODE en rouge, dans la marge inférieure droite.

Dans NooJ, plusieurs fenêtres s'ouvrent successivement.

L'option 'Cascade' de Windows dans la barre de menus, ou les titres des fenêtres vous aideront à naviguer d'une fenêtre à l'autre.

1 clic dans la fenêtre du texte

➔ le mot TEXT apparaît alors dans la barre de Menu, à droite de WINDOWS.

2 clics sur 'annotations' dans le cadre blanc de la bande titre grise

➔ Fenêtre 'untitled' apparaît, ainsi que 'Vocabulary contains 17707 entries' dans la zone grise, juste en-dessous.

- NooJ présente la liste des mots par ordre de fréquence (freq). On peut cliquer sur 'entry' de la barre, pour obtenir un classement par ordre alphabétique et revenir à l'ordre par fréquence en cliquant sur **freq**.
- OBSERVATION :
 - de quelles catégories sont les mots les plus fréquents?
 - quels sont les premiers mots de classe majeure, et à quel rang (10^e, 20^e, 30^e mot de la liste?) commencent-ils à apparaître, de quelles catégories sont-ils?
 - quelle est la catégorie des mots de freq. de 100, de 50?

5. Fermer NooJ

Project >> close project

File >> exit

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Deuxième laboratoire

NooJ

NooJ et les statistiques lexicales sur un texte

1. Ouvrir NooJ

DEMARRER >> NooJ

Info >> Preferences >> General : fr >> Lexical Analysis : 'delaf.nod' et 'élision' >>> OK

Pour ne pas perdre de vue les fenêtres que NooJ rapetisse occasionnellement : agrandir la fenêtre de NooJ en pointant le curseur sur le bas de fenêtre et en tirant la fenêtre vers le bas, sans qu'elle ne touche la barre des tâches de Windows.

2. Ouvrir le texte à analyser

Project >> Run project >> la dame de trente ans.nop

1 clic dans fenêtre du texte

2 clics sur 'annotations'

➔ Remarquez que le mot TEXT dans la barre de menu est remplacé par DICTIONARY.

Pour mieux visualiser la suite :

Clic sur WINDOWS dans la barre de menus >> TILE VERTICALY

Ajuster les fenêtres VOCABULARY et LA DAME DE TRENTÉ ANS de façon à ce qu'elle ne se touchent pas et ne remplissent pas complètement la fenêtre de NooJ.

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

3. Exploration du tableau de vocabulaire du texte « La femme de trente ans »

1 clic sur **DICTIONARY** >> **view as a table** :



➔ Une table de style « chiffrier » apparaît. C'est la liste des mots contenus dans le texte, et leurs propriétés morphologiques, syntaxiques et sémantiques décrites dans les différentes colonnes.

Agrandir la fenêtre en cliquant sur le carré du milieu de la barre de menu dans le coin droit supérieur.

Réduire la largeur de la colonne S-lemme en cliquant sur la barre verticale entre S-lemme et catégorie, et en « draguant » la barre vers la gauche, de façon à avoir toutes les colonnes à l'écran.

- **OBSERVATIONS** : à partir des colonnes du tableau de vocabulaire, comparez les informations prescrites dans un dictionnaire catégoriel selon la grammaire générative avec les informations que fournit le dictionnaire de NooJ.
- Information morphologique : Genre. Nombre. Personne. Temps.
- Information grammaticale : V, N, A, ADV, PREP, PRON, CONJS, INTER, etc.
- Information sémantique : les contraintes de sélection
- Information structurale : la structure des noms composés
- et il est possible d'ajouter des catégories, comme les domaines (médical, transport, etc.), le cadre de sous-catégorisation.

REM : NooJ est basé sur les principes de la grammaire distributionnelle de Harris et de nombreux ouvrages ont été consacrés à la description exhaustive des mots du français, et de nombreuses autres langues.

4. Observer la description des unités particulières

Aller aux mots suivants en utilisant l'ascenseur, à droite de l'écran, et remarquez entre autre :

- abaisse z 3
- abattît z 2

Pourquoi z2 ou z3? Connaissez-vous ces mots?

REM : Dans NooJ acadien, on pourrait utiliser z2 pour les régionalismes, et z3 pour les emprunts, en chiac.

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

- agent de change sous Sem (sémantique), il y a la note Hum, sous Structure, NDN, c.a.d. Nom composé d'un nom, de la préposition « de » suivie d'un nom.
- air froid sous Sem (sémantique), il y a la note Conc (concret), et sous Structure, NA, c'est-à-dire, que c'est un mot composé d'un nom suivi d'un adjectif.

1 clic sur le titre de la colonne Sem

➔ les entrées sont ordonnées selon les informations sémantiques (en commençant par l'absence de données)

Aller à la fin du texte en descendant l'ascenseur dans la barre verticale droite, jusqu'à la fin.

- Noter les annotations sémantiques en remontant tranquillement l'ascenseur: Hum, Coll, Conc, Anl, Abst, etc.
 - *bon goût* et *grand air*, ont le trait [Abst]
 - *air froid* a le trait [Conc]
- rem : air est-il concret ou abstrait, au bout du compte?!

1 clic sur l'entête de la colonne Struct

➔ les entrées sont ordonnées alphabétiquement selon les codes de la structure.

- Quels mots ont une information dans la colonne structure?
- D'après les entrées du dictionnaire du texte de Balzac, quelle est la signification des annotations dans le dictionnaire.

XX	XN	XA
VX	VN	PN
NX4		
NDN	NAN	

1 clic sur genre

➔ Les mots féminin sont regroupés, puis les mots masculins (quand il y a une valeur) ET ainsi de suite pour toutes les entêtes de colonnes.
REM : *multi* veut dire que le mot pour prendre de multiples catégorisations, c.a.d. qu'il est morphologiquement ambigu.

5. Fin du deuxième laboratoire

Project >> close project

File >> exit

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Troisième laboratoire

NooJ

L'analyse lexicale du texte

Ouvrir NooJ choisir la langue de travail et préparer le matériel (texte et ressources lexicales)

DEMARRER >> NooJ

Info >> Preferences >> General : fr >> Lexical Analysis : 'delaf.nod' et 'élision'>>OK

Project >> Run project >> la dame de trente ans.nop

1 clic dans fenêtre du texte

1. Générer le tableau de concordance pour un mot

Dans la fenêtre du texte

2 clics sur 'tokens'

➔ Tokens in La femme de trente ans dans la bande bleue

1 clic sur le premier mot >> 1 clic sur un bouton de couleur (au bas de la boîte)

➔ Concordance for La femme de trente ans indique le nom de la fenêtre dans la bande bleue : Selected tokens 3/3 (dans la bande grise du bas de la boîte)

- NooJ présente la « séquences » recherchée au centre du tableau, et un court contexte à gauche et à droite de chaque occurrence ('token').
- Dans le coin inférieur droit : le nom de la séquence recherchée et le nombre d'occurrences trouvées.

Faire en sorte que la fenêtre TEXTE et la fenêtre Concordance soient côte à côte.

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Dans la fenêtre Concordance

1 clic sur un des mots de la colonne Seq.

→ le mot est surligné dans la fenêtre du texte et donne ainsi à voir un contexte large.

1 clic sur le bouton 'clear concordance' dans le coin supérieur gauche.

→ les données sont effacées : RE = «token recherchée » 0/0

- le bouton 'exporter' permet de conserver les données dans un fichier .txt, pour exploitation dans un autre logiciel, de traitement de texte, chiffrier ou base de données. Ce fichier sera enregistré dans le classeur NooJ qui a été créé lors de l'ouverture du programme. Au labo, ce classeur est enregistré sur le serveur P, à l'adresse de votre compte.

2. Colorier le texte

Ne pas colorier le texte s'il y a plus de 200 occurrences, car le système, dans son état actuel risque de « geler », et vous obliger de repartir à zéro. Démarrer >> NooJ, etc.

Dans la fenêtre du texte

1 clic sur **tokens** (pour avoir l'ordre alphabétique)

CTRL+clic sur « main » (43) et « mains » (17) >> 1 clic sur un bouton de couleur

→ la liste des occurrences apparaît

Dans la fenêtre Concordance

1 clic à droite de la souris

→ une boîte de menu contenant deux séries de commandes

1 clic sur 'coloring matching sequence text'



→ les occurrences sont en couleur dans le texte

2 clics sur la 11^e occurrence (« ... *aissa tomber sur la main d'Arthur* ») :

- Voir le résultat dans la fenêtre du texte

1 clic sur le bouton 'Clear Concordance' dans le coin supérieur gauche.

3. Fin du 3^e labo

Project >> close project

File >> exit

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Quatrième laboratoire

NooJ

Les requêtes lexicales au moyen d'expressions naturelles

Démarrer >> NooJ

Project >> run project >> La dame de trente ans.nop

1. Trouver les occurrences de formes du même mot ('Locate Pattern')

Dans la fenêtre du texte (un clic de souris)

CTRL + L (ou 1 clic à droite sur la souris) >> locate pattern

➔ Apparaît Locate pattern in La femme de trente ans

- Cette fonction permet de chercher dans le texte des suites de mots, les différentes formes d'un mot (« main » au singulier et au pluriel), des constructions particulières (auxiliaire + participe passé)

Au cours de ce labo, nous passerons alternativement dans les fenêtres TEXTE, Concordance, Locate pattern.

Pour naviguer plus facilement, choisissez la disposition Cascade dans Windows, et disposez les fenêtres de sorte qu'elles ne se cachent pas l'une derrière l'autre.

-aller dans Windows >> Cascade, puis cliquer dans la bande de titre des fenêtres utiles;

-aller dans Windows >> cliquer sur le nom de la fenêtre que vous voulez en premier plan

-minimiser les fenêtres inactives, et arranger vos fenêtres actives pour qu'elles ne se touchent pas, et cliquer de l'une à l'autre.

Pour toute nouvelle requête, CTRL+L vous ramène à la fenêtre des requêtes.

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

1 clic dans a NooJ regular expression
 taper *main* dans la fenêtre rectangulaire

1 clic sur un bouton de couleur

➔ table de concordance (noter le nombre : 43). RE= main 43/43

Retour à la fenêtre 'Locate pattern' ou TEXTE :

(CTRL + L) taper : *mains* dans la fenêtre rectangulaire >> choisir une autre couleur

➔ table de concordance. RE = main 60/60) (même titre de requête, pas mains)

Dans la fenêtre de concordances

Ascenseur vers la fin

- les occurrences du pluriel se sont ajoutées à celle du singulier

2. Requête par disjonction : SOIT *main* SOIT *mains*

Dans la fenêtre du texte

CTRL+L >> taper (*mains+main*) >> clic sur le bouton **gris**

➔ table de concordance. RE=main 120/120

Dans le tableau de concordance

1 clic à droite sur la souris >> Color matching sequences in text



➔ les occurrences sont « colorées » ou « soulignées » (bouton gris) dans le texte.

Dans Concordance : 'clear concordance'

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

3. Requête par 'lemme'

Dans la fenêtre 'Locate pattern'

Taper <mains> >> clic sur une couleur

➔ table la concordance : RE=main 60/60

- le nom est au singulier, alors que la requête est exprimée au pluriel
- le tableau donne les occurrences de *main* et de *mains*.

Taper <main> >> clic sur gris

➔ table de concordance : RE=main 120/120

Dans la fenêtre Concordance

1 clic à droite sur la souris >> Color matching sequences in text



Dans la fenêtre du texte

1 clic sur la 3^e occurrence avant la fin (...*chot de ses propres mains*)

Agrandir la fenêtre texte

➔ trois occurrences du mot « main » devraient être visibles à votre écran

- Toutes les occurrences sont à la fois soulignées et coloriées, ce qui signifie que la forme au pluriel et singulier entourées des crochets ouvrant (<) et fermant (>) sont interprétées par NooJ comme la même forme de base, appelée le LEMME.

Dans la fenêtre Concordance

clic sur 'clear concordance' (dans le coin gauche supérieur)

- Amusez vous à taper des verbes conjugués à des formes rares, des formes des verbes très irréguliers, des pronoms ou des articles qui ont plusieurs formes...

Fin du 4^e laboratoire

Project >> close project

File >> exit

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Cinquième laboratoire

NooJ

Les requêtes lexicales par catégorie grammaticale

Démarrer >> NooJ

Project >> run project >> La dame de trente ans.nop

1. Recherche de mots d'une catégorie par disjonction : le+la+les+l'

Ctrl+L

Dans la fenêtre 'locate pattern' sélectionner dans le cadre 'Limitation' (en bas à droite)

○ Only : 100 matches

Comme nous travaillons sur une catégorie à autre fréquence dans les textes, limiter le nombre d'occurrences à 200 avant de demander de **colorier les articles dans le texte**, au risque de causer une panne du programme.

Le cas échéant, fermez le programme en utilisant le programme de gestion des tâches : [ctrl+alt+supprimer] >> terminer le programme.

INUTILE d'envoyer le message d'erreur. Revenir à la case départ... Redémarrer tout.

Dans 'NooJ regular expression'

taper : *le+la+les+l'* >> bouton rouge >> 1 clic à droite >> color matching
sequences in text

➔ les 100 premiers articles définis sont en couleur dans le texte.

Dans le tableau de concordance

clic sur l'en-tête de la colonne **seq**

➔ les articles sont ordonnés alphabétiquement

- Tous les articles sont-ils utilisés dans les 100 premières occurrences?
- Essayez la même requête en essayant différentes combinaisons de majuscules et minuscules. NooJ discrimine-t-il la majuscule et la minuscule dans ce contexte?

Organisez les fenêtre de concordance et de texte côte à côte, pour aller facilement de l'une à l'autre

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Dans le tableau de concordance

click sur les la colonne seq → les mots sont ordonnés alphabétiquement

2. Recherche des articles par le lemme : <le>

Dans la fenêtre 'Concordance'

'Clear concordance' (dans le coin supérieur gauche)

→ les occurrences sont toutes effacées

Ctrl+L >> 'a NooJ regular expression' >> taper <le> >> bouton de couleur

→ Table de concordance : Query 100/100 (5054/5054, si on demande 'all matches')

- n'importe quelle forme devrait avoir le même résultat
- remarquez que lorsqu'on utilise les crochets simples, le tableau de concordance ne spécifie pas le nom de la requête
- Toutes les formes sont-elles prises en compte dans la requête?

Dans la fenêtre de 'locate pattern'

o All matches

Taper : <le>+l >>> bouton

→ Concordance : Query : 5991/5991 occurrences

- Est-ce que toutes les occurrences de *la* sont bien des déterminants définis? (trouvez quelques cas contraires, si c'est possible de catégories différentes)
- Comment formuler une requête pour vérifier si les formes *la* et *là* sont ambiguës pour NooJ?

3. Recherche des variantes d'un mot par catégorie

Dans la fenêtre 'Concordance'

'Clear concordance' (dans le coin supérieur gauche)

→ les occurrences sont toutes effacées.

Dans la fenêtre du TEXTE

CTRL + L >> taper <DET> >> bouton gris

→ Query : 15 239 / 15 239

- Qu'est-ce qu'un DET pour NooJ? Trouvez toutes les sous-classes de DET dans la liste de concordance.

Si par malheur vous avez demandé de colorier les occurrences dans le texte, repartez à la case de départ : CTRL+ALT+SUPPRIMER >> terminer maintenant et vous savez maintenant comment revenir au point où NooJ est tombé en panne.

Fin du 5^e laboratoire

Project >> close project

File >> exit

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Sixième laboratoire

NooJ

Les requêtes au moyen de graphes

Démarrer >> NooJ

Projet >> run project >> la dame de trente ans.nop >> CTRL + L

- Un graphe est un outil qui représente sous forme de transducteurs. Il consiste à décrire une série de trajets à suivre par lesquels une information passe d'un point de départ à un point d'arrivée.
- Le graphe est, entre autres choses, un moyen de formaliser les expressions naturelles et de les sauvegarder pour appliquer les mêmes requêtes à répétition sur un texte ou dans plusieurs textes.
- Les graphes sont produits et sauvegardés dans le classeur « syntactic analysis », que nous avons passé sous silence dans le premier laboratoire.

1. Écrire un graphe pour l'extraction des déterminants articles définis

File >> New >> Grammar (dernier choix du menu déroulant)

➔ Create a new grammar

- (1) choisir la langue : fr
- (2) écrire son nom
- (3) clic sur 'create a syntactic grammar'

➔ Fenêtre 'untitled'. Le choix de la langue est confirmé.

- deux symboles apparaissent : la borne initiale (flèche), la borne finale (±cercle barré sur pied)
- éloigner les deux symboles pour mettre des nœuds entre les deux.

Clic sur borne finale + tenir enfoncé >> glisser vers la droite >> lâcher >> clic dans fenêtre, mais pas sur des nœuds

ctrl+clic au milieu de la fenêtre

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

→ un carré apparaît avec le symbole <E> en bleu, au milieu, des ronds jaunes aux quatre coins..

>> clic sur le <E>

→ le fond bleu disparaît

supprimer le E >> taper le >>> RETOUR >>> taper la >>> RETOUR >>> taper les >>> RETOUR >>> taper l >>> CTRL + ENTER

→ la liste de déterminants apparaît

Si on quitte sans sauvegarder, le signe <E> apparaît. Il faut alors ctrl+click sur le <E> et recommencer, et click+Enter, pour sauvegarder.

Si on fait une erreur, cliquer n'importe où dans la fenêtre puis sélectionner le carré ou le texte en trop (1 clic dessus), et RETOUR.

clic sur borne initiale >> clic sur liste >>

→ la liste de mot est dans un cadre et deux points sont reliés

clic sur liste >>> clic sur borne finale

→ les trois points sont reliés

Attention, toujours aller de gauche à droite. En cas d'erreur, pour effacer les liens indésirables, refaire exactement les mêmes actions, et ce, dans le même ordre.

Prendre l'habitude de cliquer dans la fenêtre avant de commencer une nouvelle action.

CTRL + S

→ une fenêtre apparaît pour la sauvegarde du graphe dans le répertoire des grammaires (.nog = NooJ grammar)

Tapez un nom descriptif >> OK

2. Appliquer le graphe au texte

Dans la fenêtre du texte

CTRL+L >>> o a NooJ grammar

→ Fenêtre « open grammar »

2 clics sur le fichier que vous venez de créer (ou 1 clic >>> OK)

→ retour de la fenêtre 'locate pattern'

clic sur une couleur

→ Concordance : query 100/100

Fin du 6^e laboratoire

Project >> close project

File >> exit

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

LISTE DE CODES UTILISÉS DANS LES DESCRIPTIONS, EXPRESSIONS RATIONNELLES, GRAPHES¹

Codes relatifs aux mots

<E>	: empty string
#	: no space
<WF>	: any word form
<UPP>	: any word form in UPPERCASE
<LOW>	: any word form in lowercase
<CAP>	: any word form in Capital
<L>	: any one-letter word form
<U>	: any one-letter word form in UPPERCASE
<W>	: any one-letter word form in lowercase
<P>	: any delimiter
<DIG>	: any digit
<NB>	: any number, i.e. sequence of digits

o Codes des catégories syntaxiques :

N	nom
V	verbe
A	adjectif
DET	déterminant
ADV	adverbe
PREP	préposition
CONJ	conjonction
PRO	pronom

o Codes relatifs aux temps des verbes

PR	present
INF	infinitif
I	imparfait
F	futur
C	conditionnel
S	présent du subjonctif
IS	imparfait du subjonctif
PP	participe passé

o Nombre

m	masculin
s	singulier

¹ Tiré du manuel de Monsieur iMarcel Rux, de l'Université de Marseille (France)

```

# NooJ V1
# Dictionary properties' definition
# Language is: fr
# Special Characters: '=' '+' '#' ''
# List categories and properties associated with features
#   Example: N_Number = m + f;
# Special KEYWORD: INFLECTION lists all inflectional features (used by variables
#   $xF)
# Example: INFLECTION = m + f + Present + Futur;

```

```

INFLECTION = m + f + s + p +
  1 + 2 + 3 +
  Inf + G + PP +
  P + Ip + S + C +
  I + PS + IS +
  F;

```

```

A_Genre = m + f;
A_Nombre = s + p;

```

```

DET_Genre = m + f;
DET_Nombre = s + p;

```

```

N_Genre = m + f;
N_Nombre = s + p;
N_Sem = Anl + Hum + HumColl + Conc + Alim + Anim + Abst + Pol + PR;
N_Struct = AAN + AN + E01 + NA + NAA + NAN + NCN + NDN + NN + NPN +
NPNPN + NPV + NX + NX4 + PN + VN + VV + VX + XA + XN + XV + XX;

```

```

PRO_Genre = m + f;
PRO_Nombre = s + p;
PRO_Pers = 1 + 2 + 3;

```

```

V_Pers = 1 + 2 + 3;
V_Temps = INF + G + PP +
  PR + IP + S + C +
  I + PS + SI +
  F;

```

```

V_Synt = tr + intr + pro + aux;
V_Genre = m + f;
V_Nombre = s + p;

```

```

X_Genre = m + f;
X_Nombre = s + p;

```

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

TRAVAUX APPLIQUÉS EN SYNTAXE

Quel genre de questions se poser en syntaxe et les résoudre avec à NooJ?

Lest trois démarches illustrées vous proposent une marche à suivre pour accomplir un travail pratique soumis dans le cadre du cours de syntaxe à l'automne 2005. Vous avez le choix de travailler sur le corpus littéraire du module français, sur le corpus acadien du module ac, ou sur un texte de votre choix.

Question 1 : la sous-catégorisation des adjectifs (<A>).

- a) quel genre de complément peut avoir un adjectif?
- b) quels adjectifs prennent des compléments dans le texte à l'étude?
- c) lesquels n'ont pas de complément
 - est-ce que ce sont des compléments sélectionnés?
 - est-ce que ce sont des compléments non-sélectionnés?
- i. écrire une expression rationnelle pour trouver 100 occurrences.
- ii. dans la concordance, éliminer sélectionner les séquences qui ne sont pas des adjectifs (le texte n'a pas été étiqueté quant aux catégories).

-clic sur la 1^{re} séquence, CTRL+clic sur les autres séquences pour conserver toutes les séquences à éliminer.

CTRL+clic DRT >>> 'filter out *selected lines*'

→ le seq. « non adjectif » disparaissent, mais restent en mémoire. Le nombre de seq. 'filtrées' est déduit du total d'occurrences (voir dans la marge inférieure droite).

-dans concordance,

CTRL+clic DRT >>> 'export concordance'

→ vous choisissez l'endroit pour sauvegarder le fichier (la clé, la disquette, ou P)

- iii. dans concordance, cliquez sur les adjectifs qui ont un complément sélectionné, puis sauvegarder les séquences
 - CTRL+clic DRT >>> 'filter out unselected lines'
 - les séquences qui restent sont celles que vous avez identifiées comme ayant la structure A → Adj SP sélectionné
 - CTRL+clic DRT >>> 'export concordance'
 - nommez le fichier par un nom descriptif : ex. ADJ_SP
 - ouvrez le fichier (P (sur le bureau), et trouvez le fichier .txt

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

-ouvrir, OK pour les options proposées par Word
-choisir le format 'paysage' dans la mise en page et vous aurez un tableau clair. SEULES LES OCCURRENCES QUE VOUS AVIEZ SÉLECTIONNÉES SONT LÀ.

- iv. dans concordance, CTRL+clic DRT >>> reset
➔ toutes les concordances trouvées (y compris les 'non adjectifs) reviennent
-trouvez les adjectifs qui n'ont pas de SP sélectionné. S'il n'y en a pas, refaire la requête en élevant la limite. NOOJ NE RISQUE PAS DE TOMBER EN PANNE, SI VOUS N'ESSAYEZ PAS DE COLORIER LE TEXTE.
Vous pouvez en tout moment visionner une occurrence particulière dans le texte, sans crainte de panne, quelque soit le nombre d'occurrences. C'EST LE COLORIAGE SIMULTANÉ DE TOUTES LES SÉQUENCES QUI OCCASIONNE DES PANNES (jusqu'à preuve du contraire...)

Autres applications possibles

a) On pourrait appliquer la démarche aux formes verbales pour comparer par exemple, les usages des verbes acadiens et des verbes français (par ex., le verbe *appartenir* est-il le seul verbe acadien qui n'a pas le même cadre de sous-catégorisation qu'en français et une grille thématique différente (les rôles sémantiques sont inversés) (voir les contraintes sémantiques dans votre manuel de syntaxe.)

fr. : appartenir, V/ _____ SP
(la maison appartient à mon père)

ac. : appartenir, V/ _____ SN
(mon père appartient cette maison-là)

Appartenir (thème, bénéficiaire)

Appartenir (bénéficiaire, thème)

REM : appartenir est un verbe d'état plutôt qu'un verbe d'action. C'est pourquoi la personne qui est propriétaire n'est pas considérée comme un agent, mais comme un 'bénéficiaire'.

- À partir d'un corpus d'extraits d'attestations de phrases dans un corpus acadien, construites avec un verbe anglais, décrire le cadre de sous-catégorisation des verbes anglais, écrire le cadre de sous-catégorisation

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

Question 2 : les locutions prépositives

Les séquences PREP MOT PREP sont très ambiguës. Selon le sens, elles peuvent être des compléments sélectionnés, des compléments non sélectionnés, mais également des locutions prépositives.

Pour lever certaines ambiguïtés, on peut construire la liste des locutions prépositives et de les intégrer dans le dictionnaire de NooJ, sous forme de mots composés. Ces formes seront codées UAMB, cad, 'unambiguous'. Lorsque NooJ analyse un texte ('annotations') ces séquences ont priorités sur les formes simples, et réduisent, en principe, le risque de confondre les SP en séquence libre (les compléments) et les locutions prépositives.

Démarche suggérée

Composer une expression rationnelle (suites de mots ou symboles catégoriels) qui permette d'extraire des occurrences de la séquence ambiguë.

-Dans le texte, CTRL+Clic DRT >>> 'locate pattern' >>> Only 100 >>> écrire l'expression rationnelle >>> OK

→ Concordance :

- soit que vous avez une bonne formule, et NooJ a trouvé 100 occurrences
- soit vous avez une expression qui ne respecte pas la syntaxe de NooJ. Vérifier les symboles à la dernière page du guide de labo : il est facile d'oublier de fermer ou ouvrir les parenthèses, de confondre l'utilisation des MAJ et MIN...

-Dans la fenêtre concordance,

- i. se débarrasser des séquences inutiles : sélectionner les séquences non-conformes à ce que vous cherchez et filtrer les séquences sélectionnées

CTRL+clic DRT >>> 'filter out *selected lines*'

→ le séq. « non adjectif » disparaissent, mais restent en mémoire. Le nombre de séq. 'filtrées' est déduit du total d'occurrences (voir dans la marge inférieure droite).

- ii. sauvegarder les séquences restantes

-dans concordance,

CTRL+clic DRT >>> 'export concordance'

→ vous choisissez l'endroit pour sauvegarder le fichier (la clé, la disquette, ou P)

- iii. les importer dans un chiffrier (excel, par exemple)

Ouvrir excel, OK pour toutes les options.

Ajuster la largeur des marges pour y voir clair.

Éliminez les séquences qui ne conviennent toujours pas.

Dans la colonne des locutions, remplacer les prépositions élidées (d, a,) par de et à

Éliminez la colonne du contexte avant, et après la locution.

Écrivez PREP dans la colonne à droite de la 1^{re} locution, UNAMB dans la colonne suivante.

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.

POUR ACCÉLÉRER LE PROCESSUS : Copier le contenu de ces deux cellules (CTRL+ C) et le coller dans toutes les cellules jusqu'à la fin de la liste. (moyen rapide : Sélectionner les deux colonnes jusqu'au bout de la liste, puis faites CTRL+ V).

Dans une 3^e colonne, décrire la structure en suivant le modèle observé dans le 2^e laboratoire (le dictionnaire en tableau).

Sauvegarder fréquemment(CTRL+S).

Copier le fichier dans Word pour la suite.

Créer un dictionnaire : suivre le guide pour la création d'un nouveau dictionnaire / venir me consulter.

Suite à donner

Vous pouvez maintenant

- vous exercer sur un plus grand nombre d'occurrences,
- essayer de trouver les locutions avec déterminant,
- chercher des loc. prépositives dans un corpus acadien, pour voir si on y trouve les formes du français littéraire, s'il n'y aurait pas des formes non usitées en français littéraire.

REM : le français parlé comprend toujours moins de diversité de formes que le français écrit. Les formes PROTOTYPIQUE sont plus fréquentes que les formes plus complexes. 'Dans la mesure où' sera exprimé par 'si' dans le discours non surveillé.

Les locutions prépositives ont des séquences variées (de manière à, à la condition de, à partir de). Vous pouvez faire une requête par construction, ou, si vous avez la « bosse » mathématique et logique, essayer de formuler une expression rationnelle qui trouvera les différentes combinaisons. (Voir le manuel NooJ anglais pour plus de détails)

3^e Question de syntaxe : les déterminants

Les déterminants reconnus par NooJ, dans son dictionnaire actuel, ne couvre pas tous les types de déterminants. Mis à part l'article élidé, il n'extrait pas les déterminants qui ont une forme composée. (voir le paragraphe sur la définition des déterminants dans votre manuel). Beaucoup de, une foule de, etc.

En suivant essentiellement une démarche similaire à celle de la questions 2 sur les locutions prépositives, nous pourrions soit ajouter la liste des déterminants de forme composée dans le dictionnaire, et les étiquetter comme des formes UNAMB.

Suite possible

Voir si les locutions qui servent de déterminant sont utilisées en fr. acadien et si le fr. acadien en a qui ne sont pas ressorties du corpus littéraire que vous avez exploré.

Représenter la grammaire des déterminants sous forme de graphe pour extraire les occurrences dans d'autres corpus (voir le manuel INTEX et NooJ pour suivre le modèle)

En cas de panne mineure aller à (1) . En cas de panne majeure aller à (2) , puis rouvrir le fichier:

(1) Project >> close >> Project >> Run project >> La dame de trente ans.nop.

(2) Fermer NooJ >> DEMARRER >> NooJ >> Info >> Preferences >> General >> fr >> Lexical Analysis >> 'delaf.nod' et 'élision'.