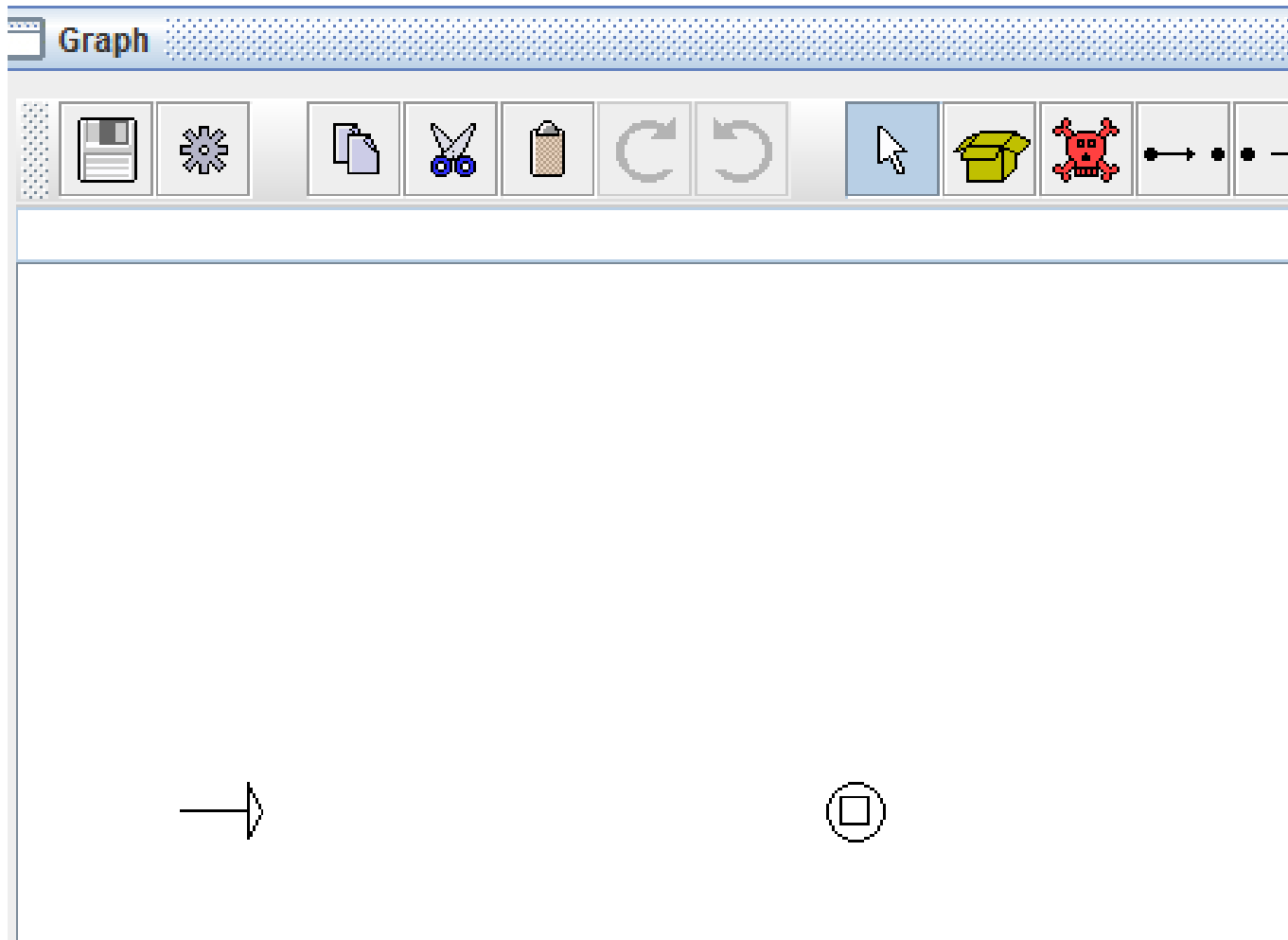


Extraction d'information

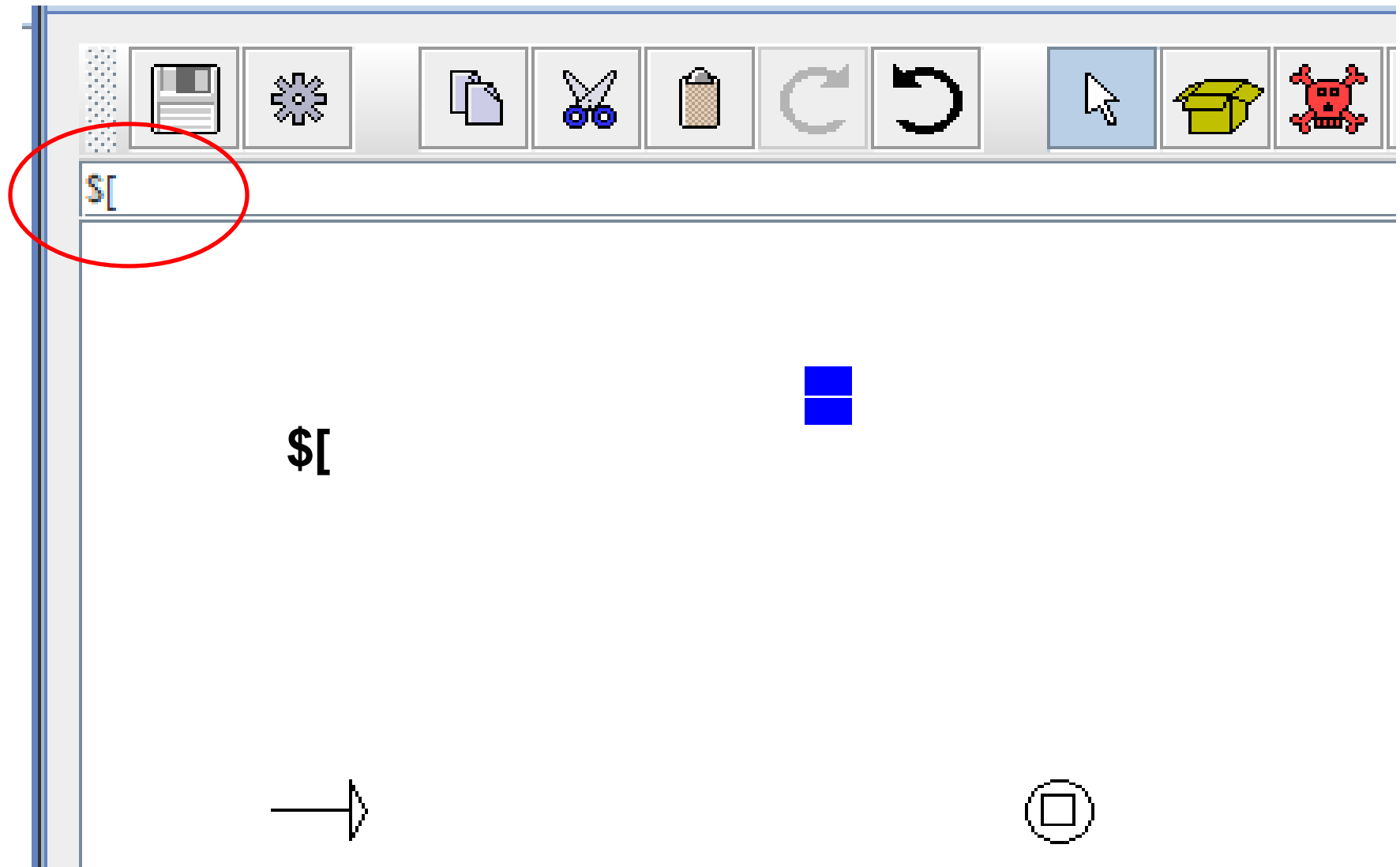
Cours 7

Nassim ZELLAL

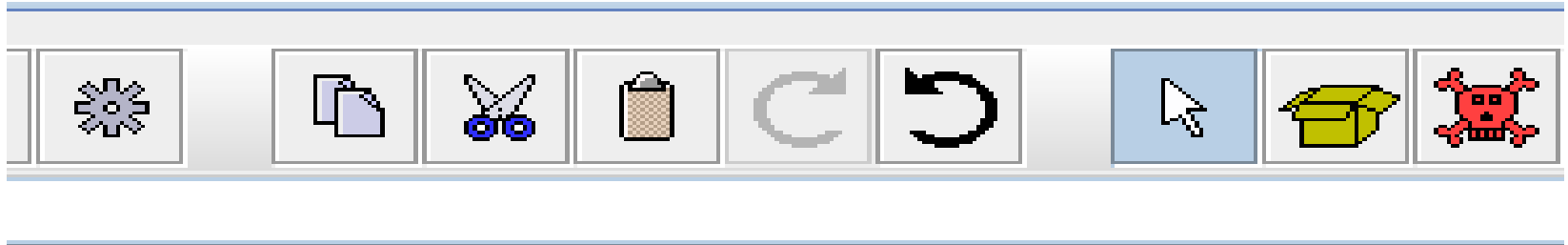
Création d'un contexte à droite



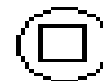
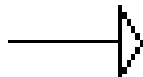
Création d'un contexte à droite



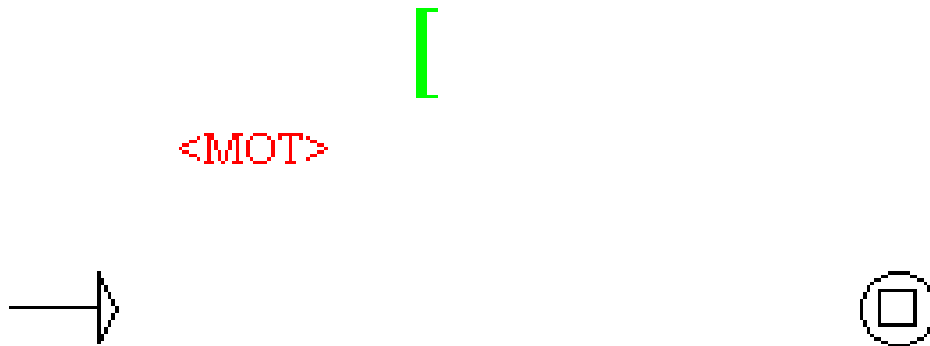
Création d'un contexte à droite



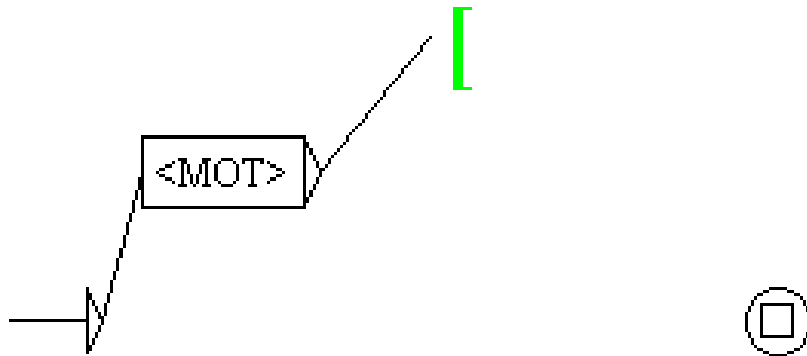
[



Création d'un contexte à droite

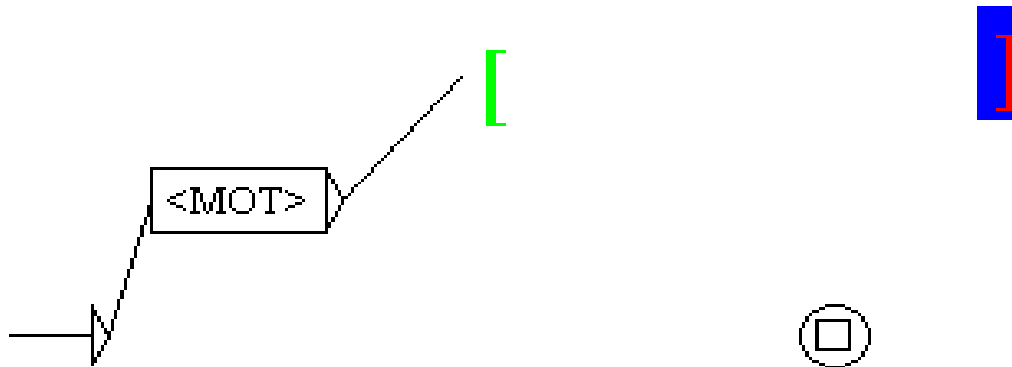


Création d'un contexte à droite

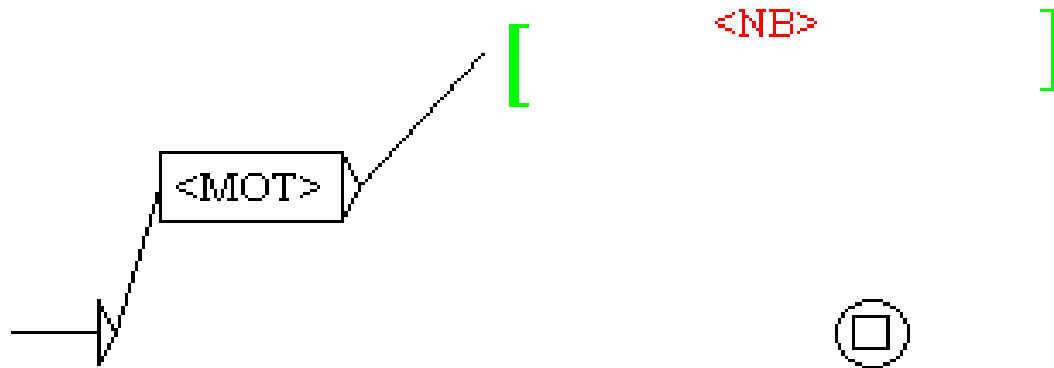


Création d'un contexte à droite

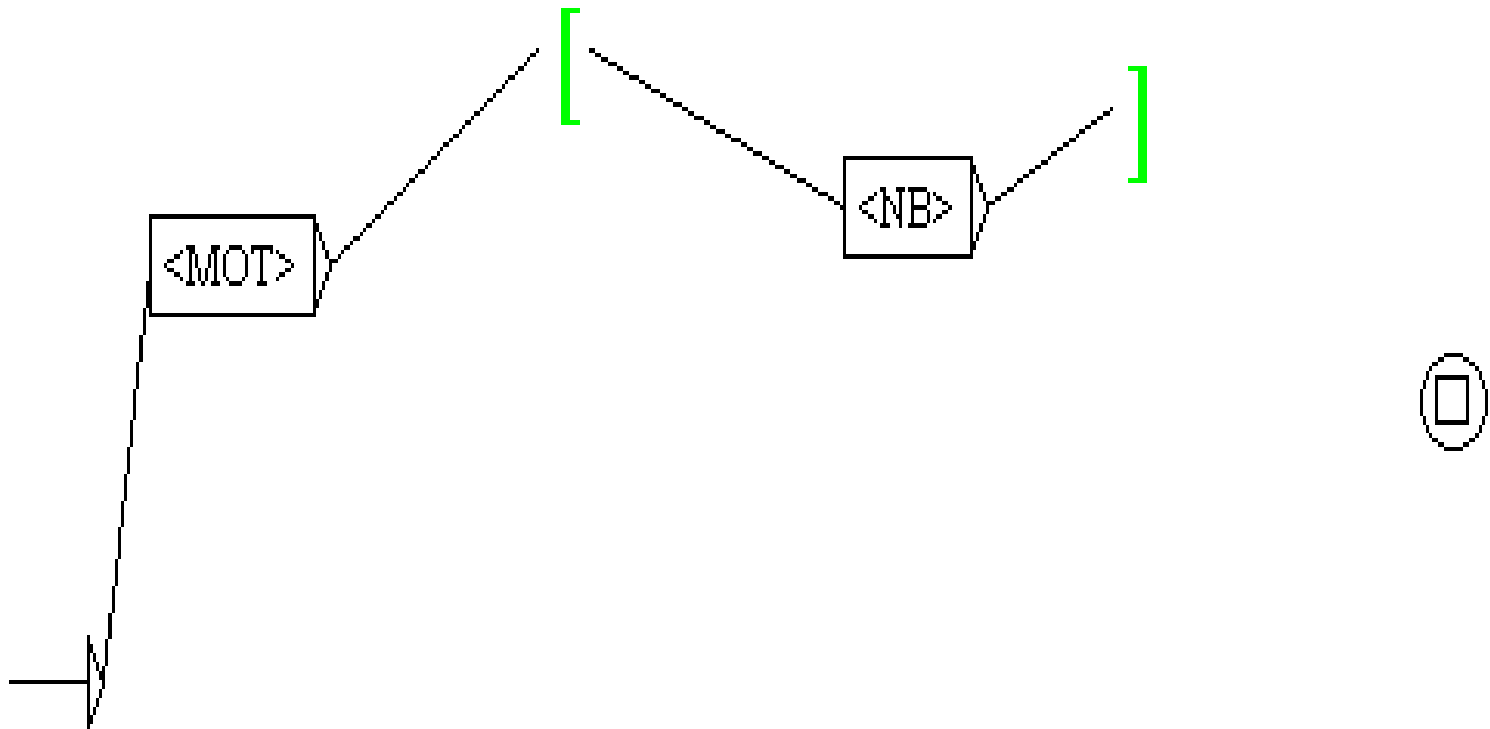
S]



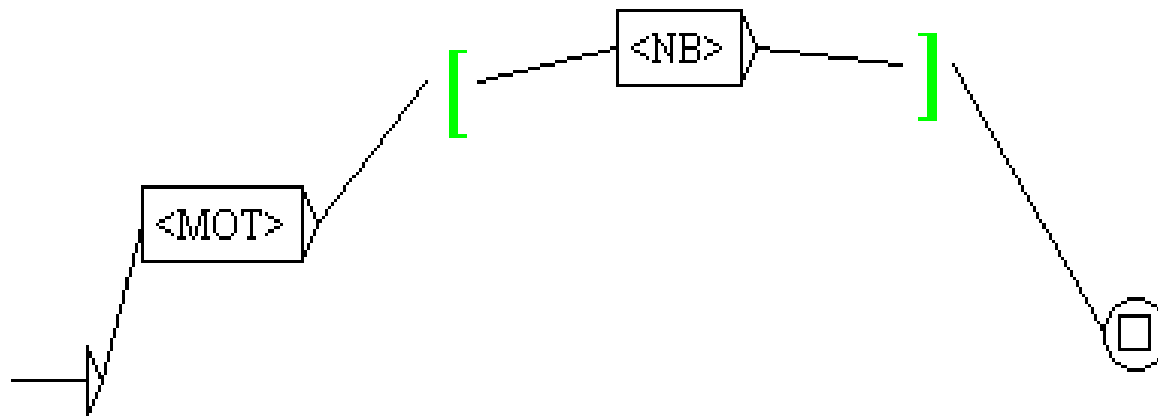
Création d'un contexte à droite



Création d'un contexte à droite



Création d'un contexte à droite

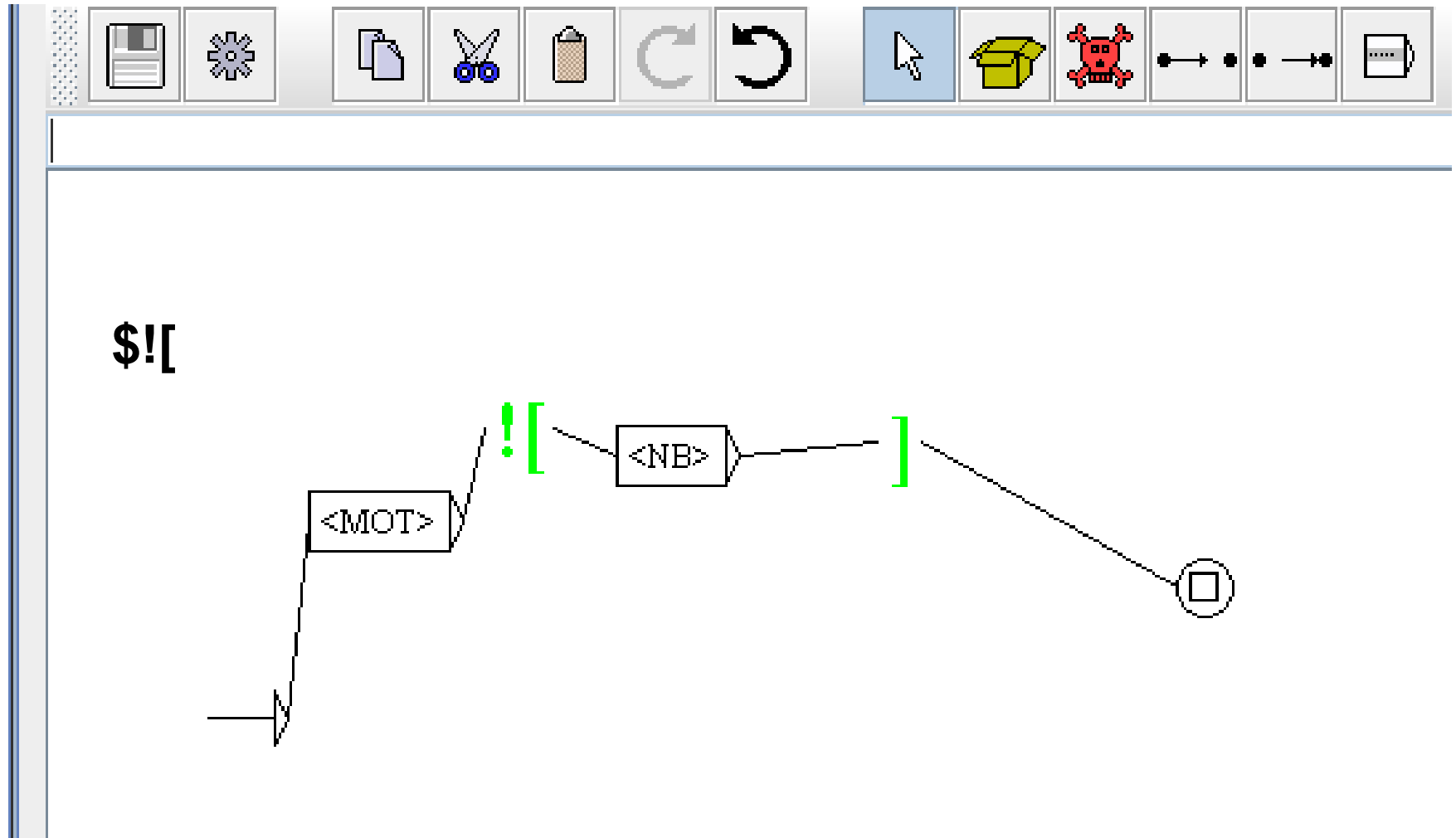


Résultat de l'application du graphe contextuel

104 matches

sous-lieutenants à 7 000 F, brigadiers à 60 000, généraux à 100 000. [Le
Compagnie des Indes : sous-lieutenants à 7 000 F, brigadiers à 60 000, c
7 000 F, brigadiers à 60 000, généraux à 100 000. [Le traitement des fo
AÎTRE, L'AUTRE COMME DOMESTIQUE {S}En l'année 1872, la maison portant le
cembre.{S} Du 12, sept heures du matin, au 21, huit heures quarante-cinq
ISTOIRE MORMONE {S}Pendant la nuit du 5 au 6 décembre, le train courut a
s les neuf jours comptés du 12 décembre au 21, pouvait franchir les trois
vait pas de temps à perdre.{S} On était au 30 octobre, et le lendemain m
leas Fogg portait l'arrivée du paquebot au 5.{S} Or, il n'arrivait que le
Kong le 25 à midi.{S} Nous ne sommes qu'au 22, et nous arriverons à temp
i indiquait _ depuis le 2 octobre jusqu'au 21 décembre _ le mois, le qua
les obstacles ! {S}Il arriva aussi que, ce 23 novembre, Passepartout épre
Il en était encore à la cité légendaire de 1849, à la ville des bandits,
t le traité de Nanking, après la guerre de 1842, assura la possession à
0 000 F, et le gouverneur général, plus de 600 000 F. {S}(Note de l'aut
olide.{S} Leur pesanteur spécifique est de 1 170, celle de l'eau distille
jours, répondit Mr. Fogg, le samedi 21 décembre 1872, à huit heures qua
épendance indoue est encore absolue. {S}Depuis 1756 _ époque à laquelle :

Contexte négatif à droite



Résultat de l'application du contexte négatif

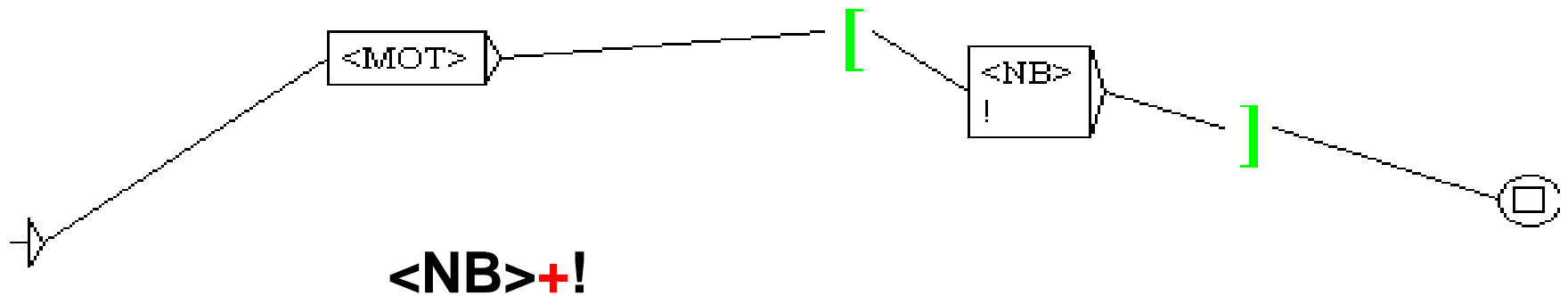
Concordance: C:\Users\Public\Documents\Unitex-GramLab\Unitex\French\Corpus\80jours_snt\concord.html

200 matches

cu mille ans sans vieillir. {S}Anglais, à coup sûr, Phileas Fogg n'était de Londres, bien qu'il semblât prendre à tâche de ne rien faire qui pût ne l'avait jamais vu ni à la Bourse, ni à la Banque, ni dans aucun des comptoirs de la Cité. On ne l'avait jamais vu ni à la Bourse, ni à la Banque, ni dans aucun des comptoirs de la Cité. hable quant aux pieds _ , mais un Byron à moustaches et à favoris, un Byron impassible, qui aurait vécu mille ans sans vieillir. {S}On disait qu'il ressemblait à Byron _ automobile par la tête, car il était irréprochable quant aux pieds _ , mais un Byron à moustaches et à favoris, un Byron impassible, qui aurait vécu mille ans sans vieillir. {S}Anglais, à coup sûr, Phileas Fogg n'était de Londres, bien qu'il semblât prendre à tâche de ne rien faire qui pût attirer l'attention. {S}A l'un des plus grands orateurs de l'époque, PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT L'UN COMME MAÎTRE, L'AUTRE COMME DOMESTIQUE {S}En l'absence de tout autre maître, le domestique se faisait lui-même maître. {S}On disait qu'il ressemblait à Byron _ automobile par la tête, car il était irréprochable quant aux pieds _ , mais un Byron à moustaches et à favoris, un Byron impassible, qui aurait vécu mille ans sans vieillir. {S}Anglais, à coup sûr, Phileas Fogg n'était de Londres, bien qu'il semblât prendre à tâche de ne rien faire qui pût attirer l'attention. {S}A l'un des plus grands orateurs de l'époque, PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT L'UN COMME MAÎTRE, L'AUTRE COMME DOMESTIQUE {S}En l'absence de tout autre maître, le domestique se faisait lui-même maître.

L'union

(C:\Users\Public\Documents\Unitex-GramLab\Unitex\French\Graphs)

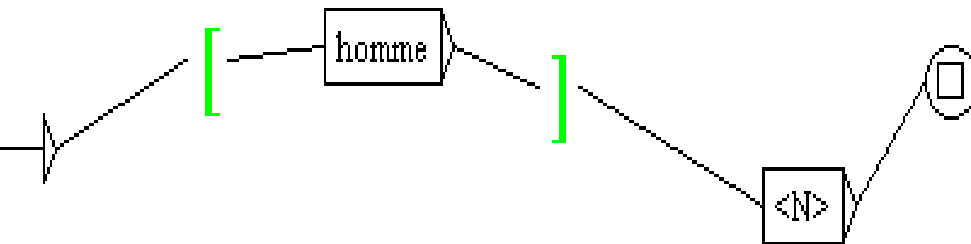


Résultat de l'union

lieutenants à 7 000 F, brigadiers à 60 000, généraux à 100 000. [Le
rie des Indes : sous-lieutenants à 7 000 F, brigadiers à 60 000, c
?, brigadiers à 60 000, généraux à 100 000. [Le traitement des for
-être la seule, et avec de tels abrutis !... " {S}En tout cas, Pa
sement : «Cela me va ! voilà mon affaire !{S} Nous nous entendrons
en Afrique ? _ En Afrique. _ En Afrique ! répéta Passepartout.{S
e assez vulgaire de son pays : " Ah ! bien se dit-il, elle est fo
commun.{S} Il y a autre chose. _ Ah ! c'est un original, ce Mr. Fo
adres un mandat d'arrestation. _ Ah ! cela, monsieur Fix, c'est vo
aux premières heures du jour. _ Ah ! les gueux ! s'écria Passepa
votre passeport est-il visé ? _ Ah ! c'est vous, monsieur, répon
ce gentleman, si casanier jusqu'alors ! {S}A huit heures, Passepa
s'écria Passepartout.{S} Prends, ami !{S} Kiouni est un brave et c
Kiouni est un brave et courageux animal ! " {S}Et, allant à la bêt
L'AUTRE COMME DOMESTIQUE {S}En l'année 1872, la maison portant le
se le tour du monde, même en dix ans ! et il paria cinq mille liv
25 à midi.{S} Nous ne sommes qu'au 22, et nous arriverons à temps

Contexte à gauche

On extrait des noms qui
sont « homme »



Il est possible aussi de
mettre la négation sur
un contexte à gauche :
\$![...]

Honnête homme, Phileas
honnête homme. - Monsi
honnête homme, et que
honnête homme ! - Oui
honnête homme ? - Non
honnête homme ! " Pas
honnête homme du monde
honnête homme ! - Pos
honnête homme, il éta
un brave homme ! " Le
jeuna en homme pour qu
cte, mon homme a déjà
ête mon homme ou il

Contexte gauche avec le symbole *

\$*

homme

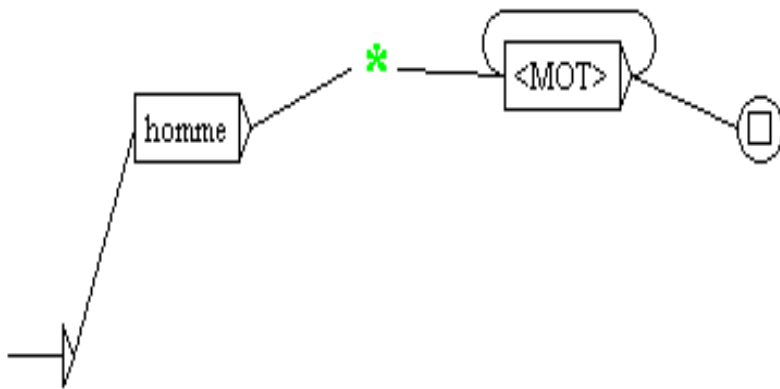
*

<MOT>



rated London News. S'il avait été homme à s'étonner
ys indou, si Phileas Fogg eût été homme à les demand
ant, en frais de toute sorte, mon homme a déjà laiss
nné que jamais, ne semblait point homme à se jeter d
onde. Passepartout, n'étant point homme à laisser la
arfaitemment, Mr. Fogg et moi ! Un homme casanier et
blement ces mers. John Bunsby, un homme de quarante-
cela. - Tiens ! Mais vous êtes un homme de coeur ! d
ain à la station d'Elko, était un homme de haute tai
se présenta aussitôt. C'était un homme de cinquante

Boucler sur le motif simple <MOT>

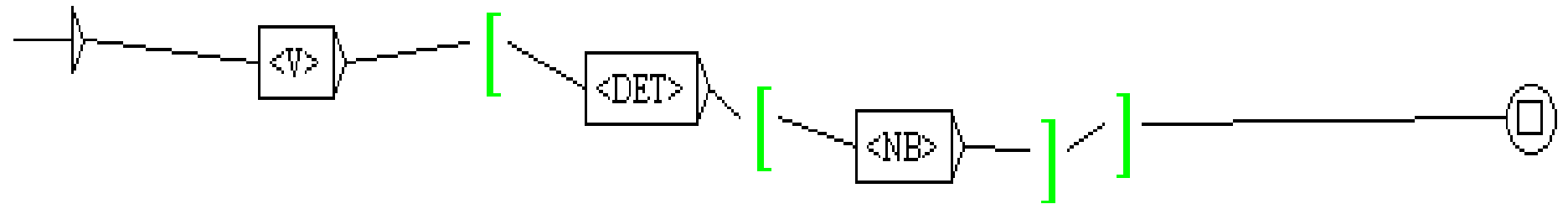


**Dans Locate Pattern,
Sélectionner « Longest
matches » (séquences
les plus longues)**

... cet homme à se jeter dans ce lac. Cette prér
t point homme à se jeter dans ce lac. Cette prér
moi ! Un homme casanier et régulier ! Une véritab
cher à l'homme de barre ! On n'eût pas manoeuvré
était un homme de cinquante ans, une sorte le lou
êtes un homme de coeur ! dit Sir Francis Cromart
ser ", l'homme de confiance de la Compagnie, l'éq
était un homme de haute taille, très brun, moust
asby, un homme de quarante-cinq ans environ, vige
trait l'homme dont le signalement avait été fou
honnête homme du monde ! - Qu'en savez-vous ? ré
ect d'un homme du monde. " Le capitaine ? demanda
montré l'homme en question qui se promenait deva
... - Cet homme est votre domestique ? ajouta le p

Que va extraire ce graphe ?

Contextes imbriqués



Contextes imbriqués

Concordance: C:\Users\Public\Documents\Unitex-GramLab\Unitex\French\Corpus\80jours_snt\concord.html

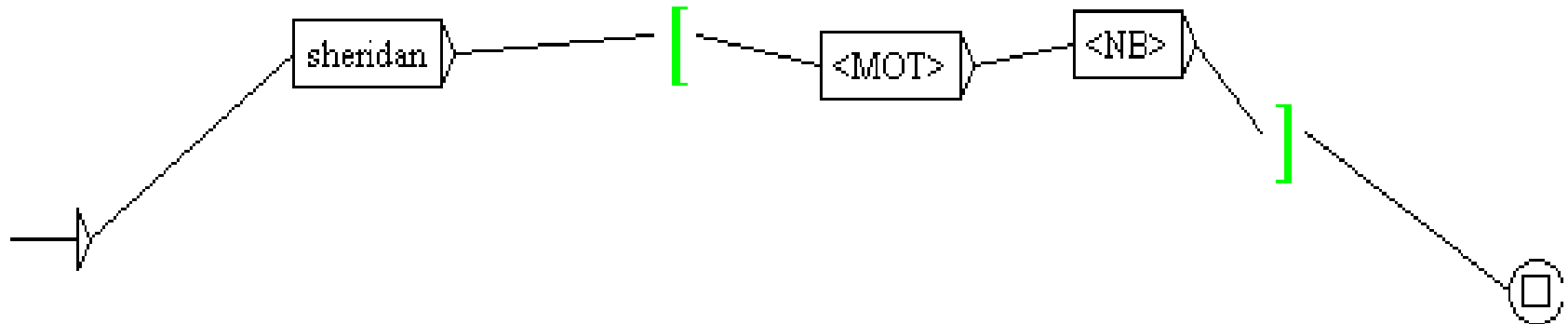
14 matches

er que le 22 octobre à Bombay. Or, il y arrivait le 20. C'était donc, de
chine, l'Henrietta, dans les neuf jours comptés du 12 décembre au 21, pour
chaudières, le Carnatic fût parti à la date du 5 novembre, et les voyageurs
iron dix-sept mille cinq cents, à cette date du 23 novembre. Mais maintenant
le 15 octobre seulement au matin, y entraît le 14 au soir. C'était une
tière solide. Leur pesanteur spécifique est de 1 170, celle de l'eau distillée
permet pas de reposer. Le lendemain, c'était le 12 décembre. Du 12, sept
UN COURS D'HISTOIRE MORMONE Pendant la nuît du 5 au 6 décembre, le train
rendre à Hong-Kong le bateau qui devait partir le 6 novembre pour Yokohama
it à faiblir. En effet, un long article parut le 7 octobre dans le Bulletin
s, 300 000 F, et le gouverneur général, plus de 600 000 F. (Note de l'auteur)
Fogg était donc autorisé à croire que, rendu le 2 décembre à San Francisco
rendu le 2 décembre à San Francisco, il serait le 11 à New York et le 20
le départ du Carnatic, et il faut que je sois le 14, au plus tard, à Yokohama

Exercice 1



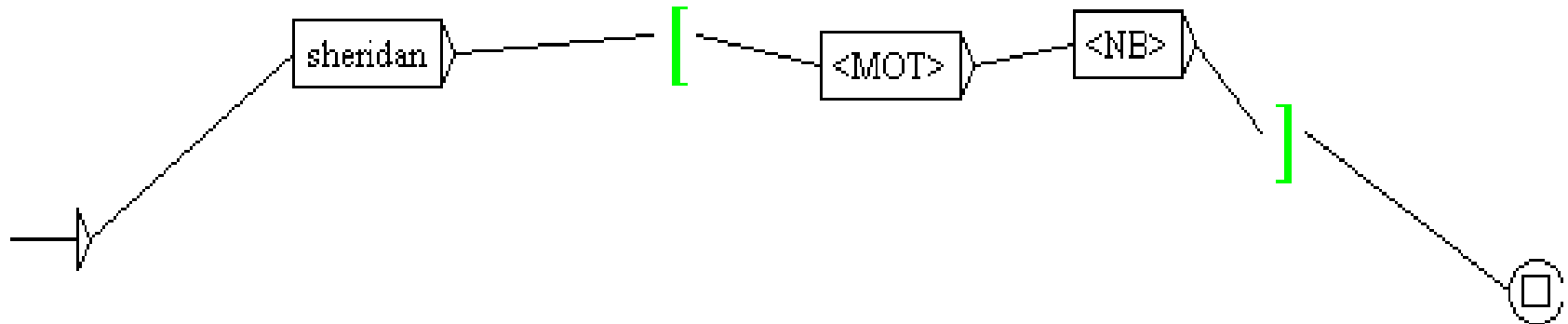
Ce graphe doit être modifié afin qu'il extraie :
Sheridan mourut en 1814



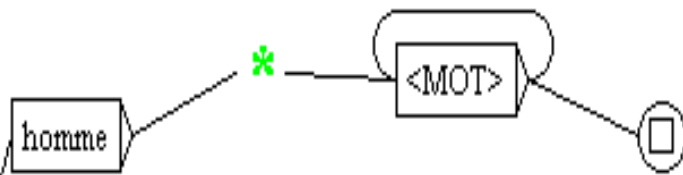
Exercice 2



Ce graphe doit être modifié afin qu'il extraie :
« Sheridan » suivi par « mourut en 1814 »



Exercice 3



Modifier ce graphe pour qu'il reconnaisse les phrases sans s'arrêter aux signes (symboles) de ponctuation forte en fin de phrase.

Ne pas inclure le délimiteur de phrases {S}

le notre homme a pris passage à bord du Mongolia
rait été homme à s'étonner de quelque chose, c'e
t point homme à se jeter dans ce lac. Cette pren
moi ! Un homme casanier et régulier ! Une véritab
cher à l'homme de barre ! On n'eût pas manoeuvré
était un homme de cinquante ans, une sorte le lou
êtes un homme de coeur ! dit Sir Francis Cromart
ser ", l'homme de confiance de la Compagnie, l'éq
était un homme de haute taille, très brun, moust
sby, un homme de quarante-cinq ans environ, vigo
trait l'homme dont le signalement avait été four
honnête homme du monde ! - Qu'en savez-vous ? re
ect d'un homme du monde. " Le capitaine ? demanda
montré l'homme en question qui se promenait devant
... - Cet homme est votre domestique ? ajouta le p

Exercice 4

- Construire un graphe d'extraction pour extraire un verbe, qui n'est pas suivi par une préposition, sauf si cette dernière est suivie par un nombre.
- Remarque : L'étiquette de préposition est : <PREP>

Exercice 5

- Extraire via un graphe tous les "noms" qui ne sont pas le token "homme" au singulier ou au pluriel, sauf si ce token est suivi par le caractère ?
-

Exercice 6

- Extraire via un graphe les "verbes", qui sont précédés du token "homme" au singulier. Ces verbes ne doivent pas être ambigus avec des adjectifs, comme « instruit », qui peut être: <A> (Adjectif) ou <V> (Verbe).
- Cependant, la forme fléchie "est" du verbe "être", qui peut être un adjectif, comme dans "le versant est de l'île", ne doit pas être exclue.

Exercice 7

- Extraire via un graphe les "verbes" qui sont précédés du token "homme" au singulier ou au pluriel. Ces verbes ne doivent pas être ambigus avec des adjectifs, comme instruit qui peut être: <A> (adjectif) ou <V> (verbe).
- Sont également exclus les "verbes" être et avoir.
- Proposer deux solutions (deux graphes d'extraction).

Exercice 8

À partir du texte « 80jours.txt » encodé en UTF-16-LE BOM.

- Extraire via un graphe d'extraction toutes les entités nommées de type « Personne », qui sont exprimées sous la forme : Prénom Nom

Exercice 9

Construire le graphe permettant d'extraire le résultat ci-dessous.
Ce résultat doit contenir :

□ 17 matches ;

□ 21 unités reconnues ;

□ 0.013% du texte couvert.

Concordance: D:\Unitex\French\Corpus\80jours_snt\concord.html

17 matches

gnal du départ. Le pays prit bientôt un <EXTRACTION NOM="aspect" ADJECTIF="sauvage"/> très sauvage. Aux g
ance qui sépare Liverpool de Londres _, <EXTRACTION NOM="chose" ADJECTIF="faisable"/> très faisable, quan
l Proctor. Ce cerveau brûlé trouvait la <EXTRACTION NOM="chose" ADJECTIF="faisable"/> très faisable. Il r
puis ce point, son profil décrivait une <EXTRACTION NOM="courbe" ADJECTIF="allongée"/> très allongée, s'a
je faire, monsieur Fogg ? dit-elle. _ C' <EXTRACTION NOM="est" ADJECTIF="simple"/> très simple, répondit l
ne journée de promenade, il se sentit l' <EXTRACTION NOM="estomac" ADJECTIF="creux"/> très creux. Il avait
is il se dit que l'infortuné Fix devait <EXTRACTION NOM="être" ADJECTIF="désappointé"/> très désappointé,
lèrent. Le temps était fort mauvais, le <EXTRACTION NOM="froid" ADJECTIF="vif"/> très vif. Fix, assis sur
e anglaise. _ A moins que ce ne soit un <EXTRACTION NOM="homme" ADJECTIF="fort"/> très fort, répondit le
du châssis, sur l'avant, se dressait un <EXTRACTION NOM="mât" ADJECTIF="élevé"/> très élevé, sur lequel s
aution indispensable à prendre dans ces <EXTRACTION NOM="mers" ADJECTIF="fréquentées"/> très fréquentées
well. A ce point, il s'engagea dans les <EXTRACTION NOM="montagnes" ADJECTIF="ramifiées"/> très ramifiées
aient contre son étrave. Il éprouva des <EXTRACTION NOM="mouvements de tangage" ADJECTIF="violents"/> très
s. A travers ses longs cils brillait un <EXTRACTION NOM="oeil" ADJECTIF="vif"/> très vif, mais dont il sa
r Passepartout. Il savait qu'il n'était <EXTRACTION NOM="pas" ADJECTIF="difficile"/> très difficile de fa
s confortables tabagies où l'on fume un <EXTRACTION NOM="tabac" ADJECTIF="fin"/> très fin, et non l'opium
et souliers décrivaient dans l'air des <EXTRACTION NOM="trajectoires" ADJECTIF="tendues"/> très tendues,

Exercice 10

Construire un graphe et ses sous-graphes, en utilisant des contextes et des variables d'entrée, afin d'extraire les 19 occurrences suivantes :

avait procuré en la personne de Fix. Le dimanche 20 [20 est un jour] octobre, vers midi, on eut re, 8 heures 45 soir. " Arrivé à Paris, jeudi 3 [3 est un jour] octobre, 7 heures 20 matin. " Q bre, 7 heures 20 matin. " Quitté Paris, jeudi, 8 [8 est une heure] heures 40 matin. " Arrivé p les notes suivantes : " Quitté Londres, mercredi 2 [2 est un jour] octobre, 8 heures 45 soir. " nt, onze heures vingt-neuf du matin, ce mercredi 2 [2 est un jour] octobre 1872, vous êtes à mo ier de poche, puisque c'est aujourd'hui mercredi 2 [2 est un jour] octobre, je devrai être de r ts couverts et sans se compromettre. Le mercredi 30 [30 est un jour] octobre, dans l'après-midi retard. Il inscrivit donc, ce jour-là, mercredi 9 [9 est un jour] octobre, son arrivée à Suez, samedi, 5 heures soir. " Arrivé à Suez, mercredi 9 [9 est un jour] octobre, 11 heures matin. " he concernant le sieur Phileas Fogg. Le mercredi 9 [9 est un jour] octobre, on attendait pour o , dans ce salon même du Reform-Club, le samedi 21 [21 est un jour] décembre, à huit heures quar tre-vingts jours, répondit Mr. Fogg, le samedi 21 [21 est un jour] décembre 1872, à huit heures he suivant l'itinéraire convenu ? Et le samedi 21 [21 est un jour] décembre, à huit heures quar veille au soir, à cette date fatale du samedi 21 [21 est un jour] décembre, à huit heures quar 7 heures 20 matin. " Arrivé à Brindisi, samedi 5 [5 est un jour] octobre, 4 heures soir. " Emba eures soir. " Embarqué sur le Mongolia, samedi, 5 [5 est une heure] heures soir. " Arrivé à Su in. " Arrivé par le Mont-Cenis à Turin, vendredi 4 [4 est un jour] octobre, 6 heures 35 matin. ua à Londres, alors qu'il n'était qu'au vendredi, 20 [20 est un jour] décembre, soixante dix ne bre, 6 heures 35 matin. " Ouitté Turin, vendredi, 7 [7 est une heure] heures 20 matin. " Arriv