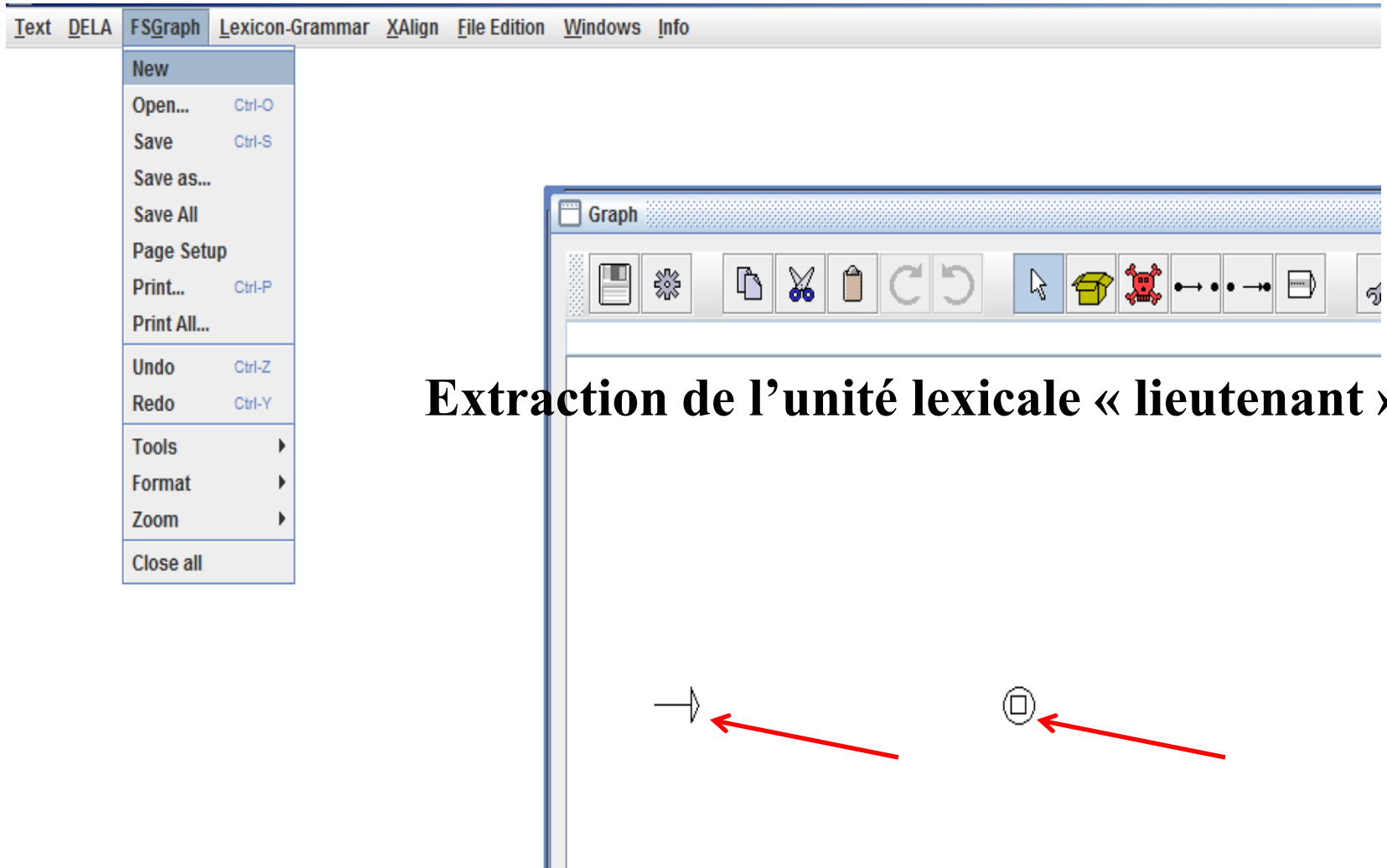


Extraction d'information

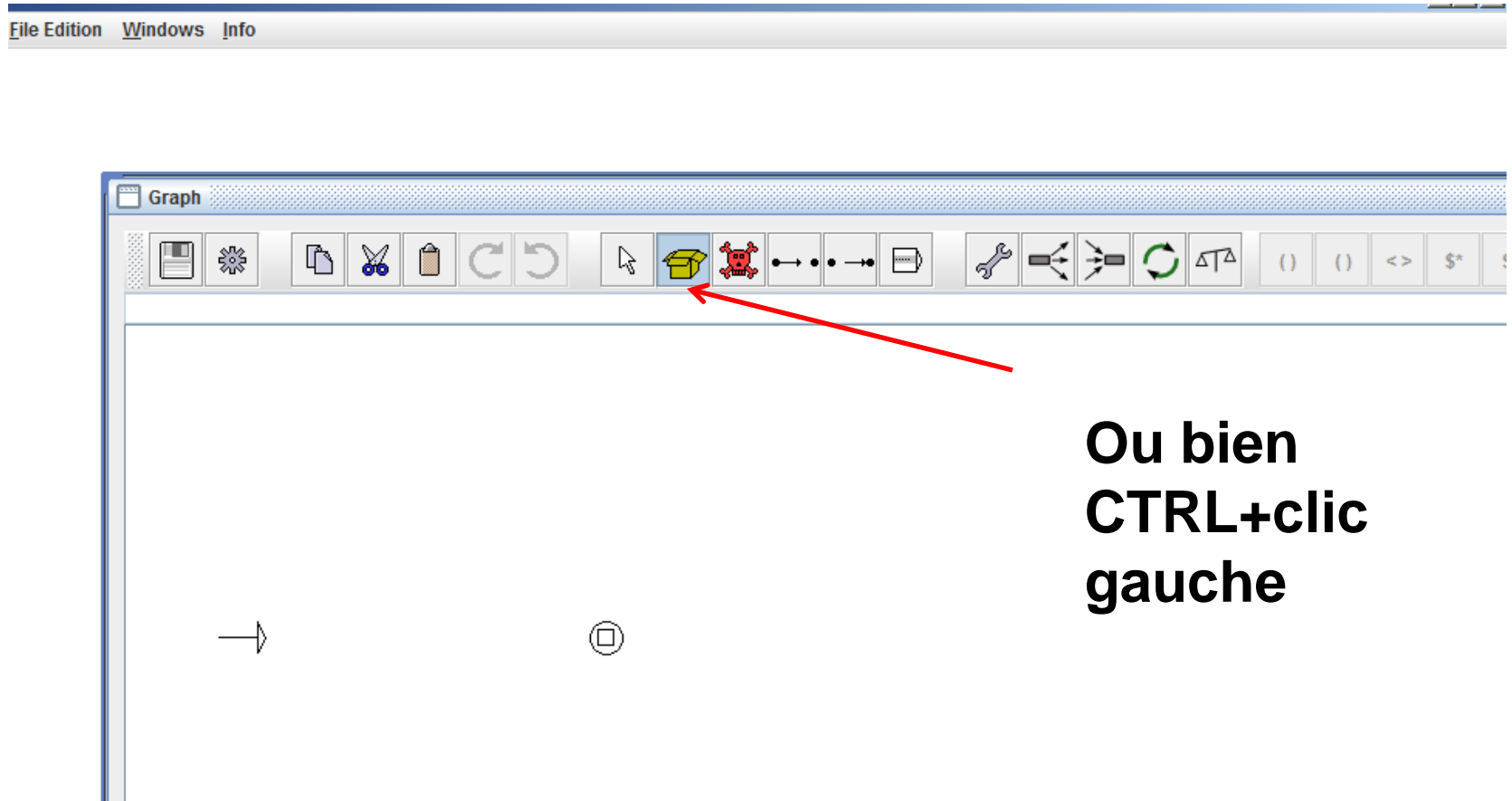
Cours 5

Nassim ZELLAL

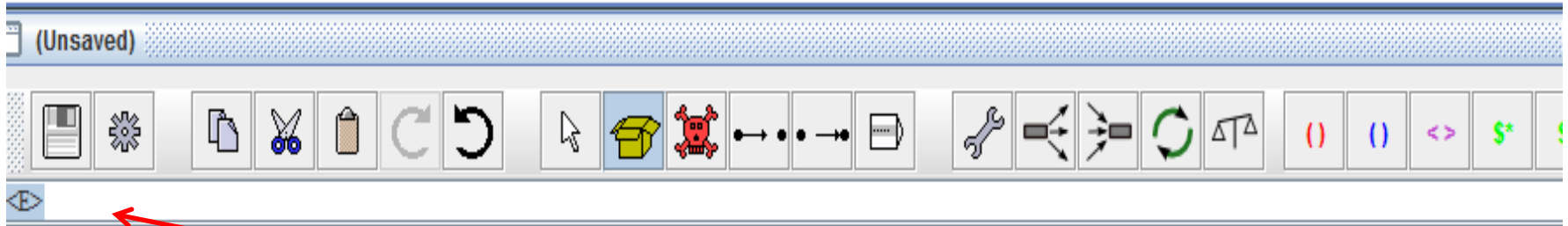
Création d'un graphe (grammaire) d'extraction UNITEX



Création d'une boîte



Création d'une boîte



Supprimer le mot vide <E>

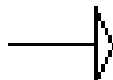


Création d'une boîte

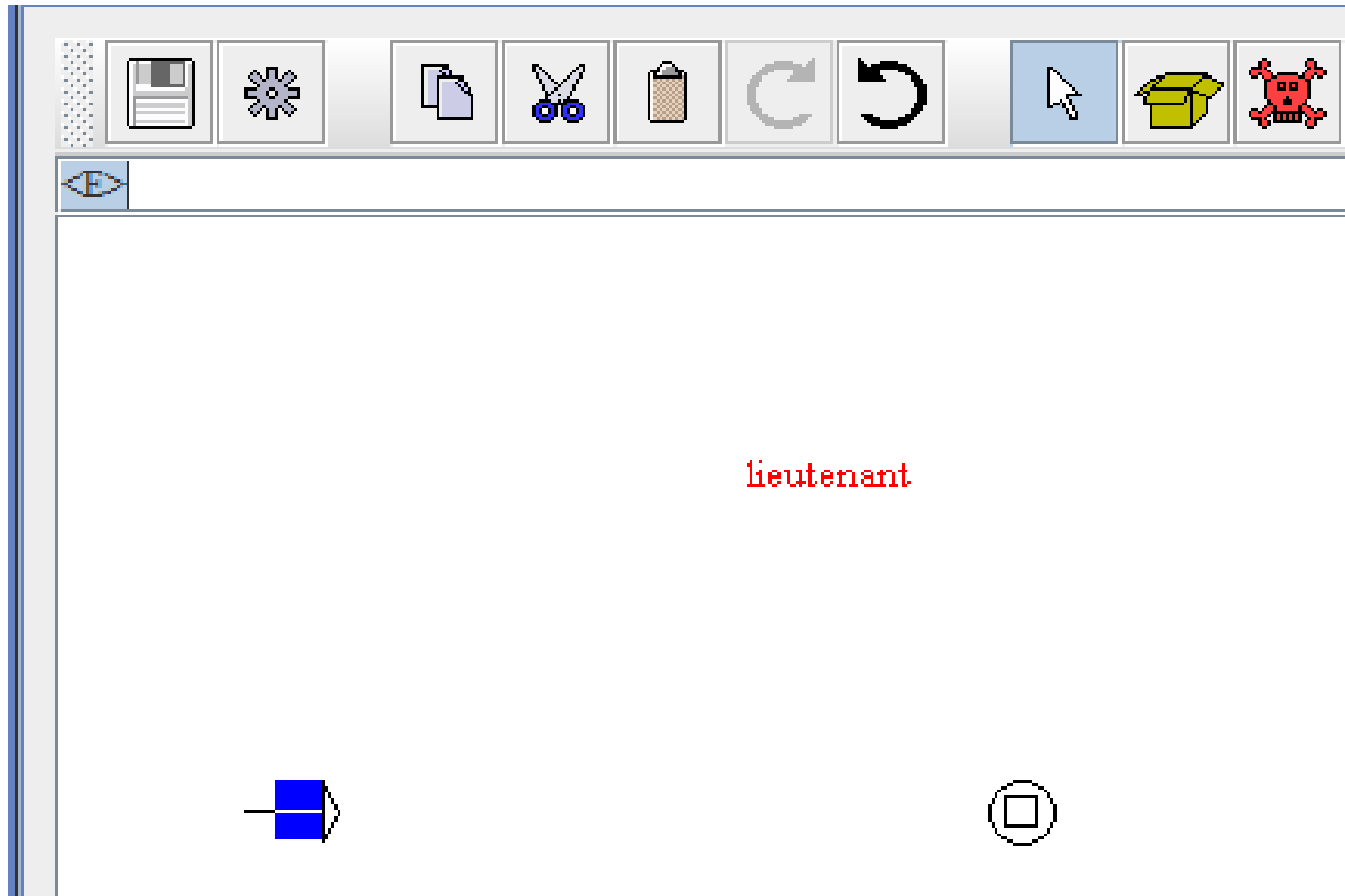


lieutenant

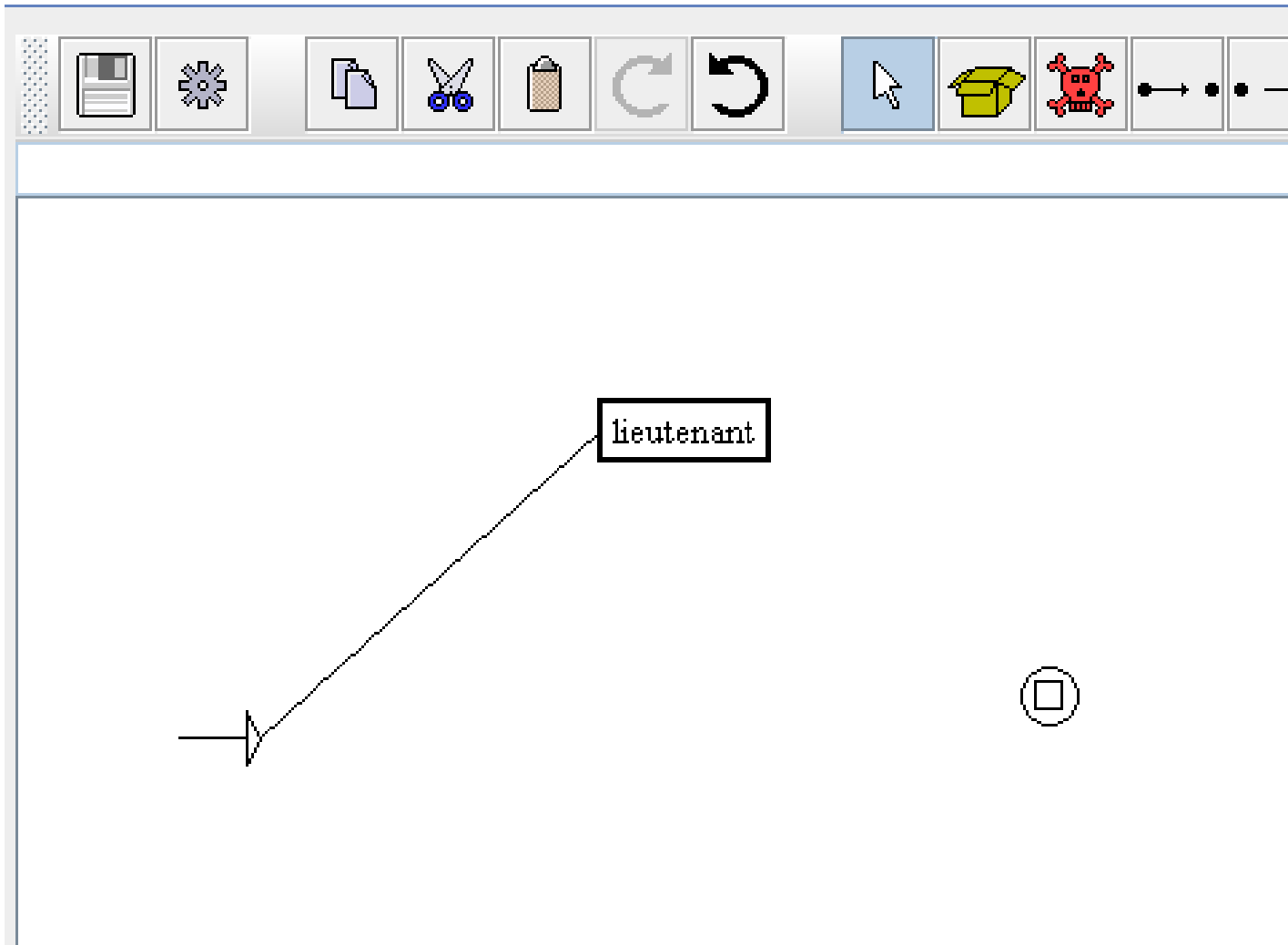
**Saisir le mot « lieutenant »
puis appuyez sur Entrée**



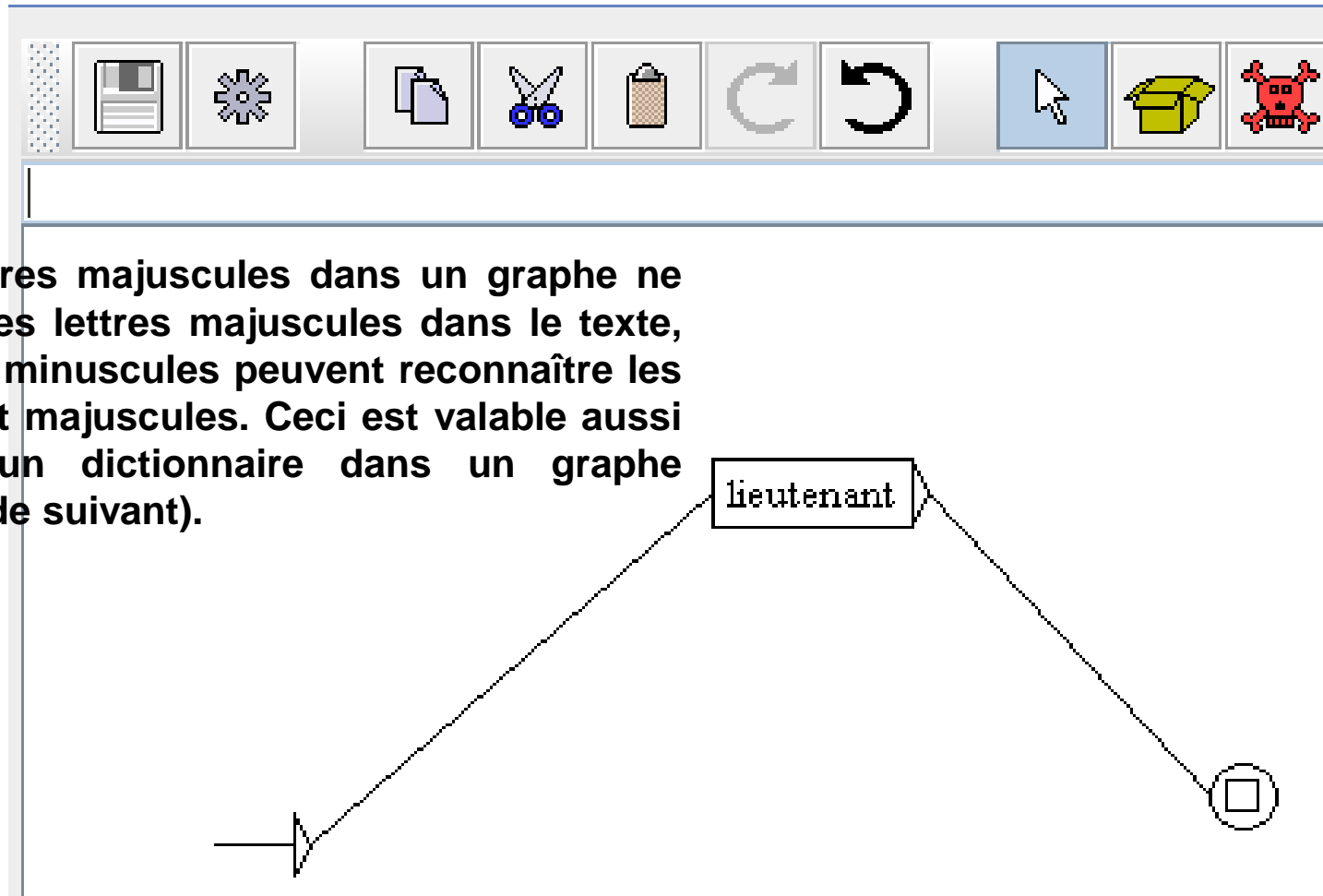
Création d'une boîte



Création d'une boîte

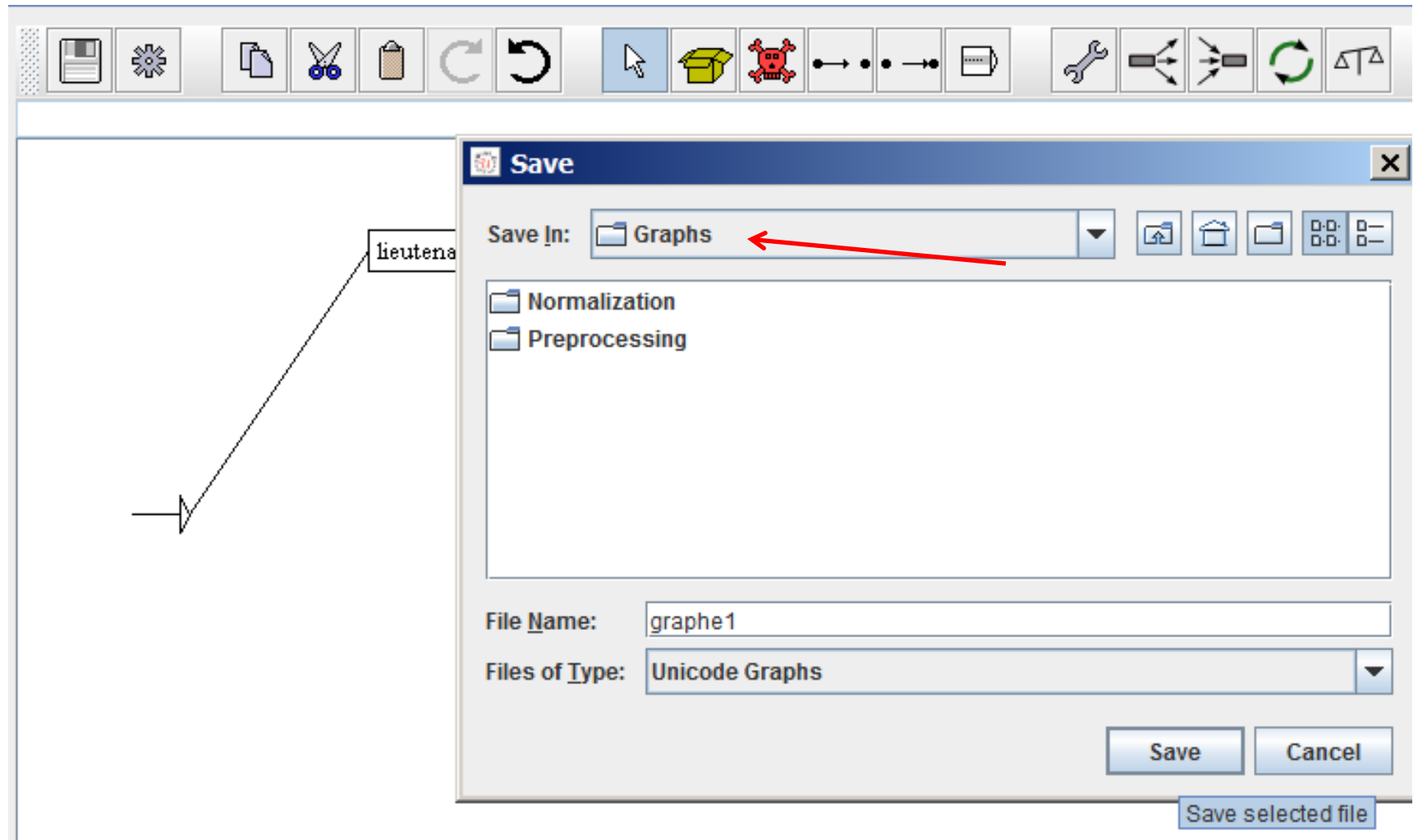


Création d'une boîte

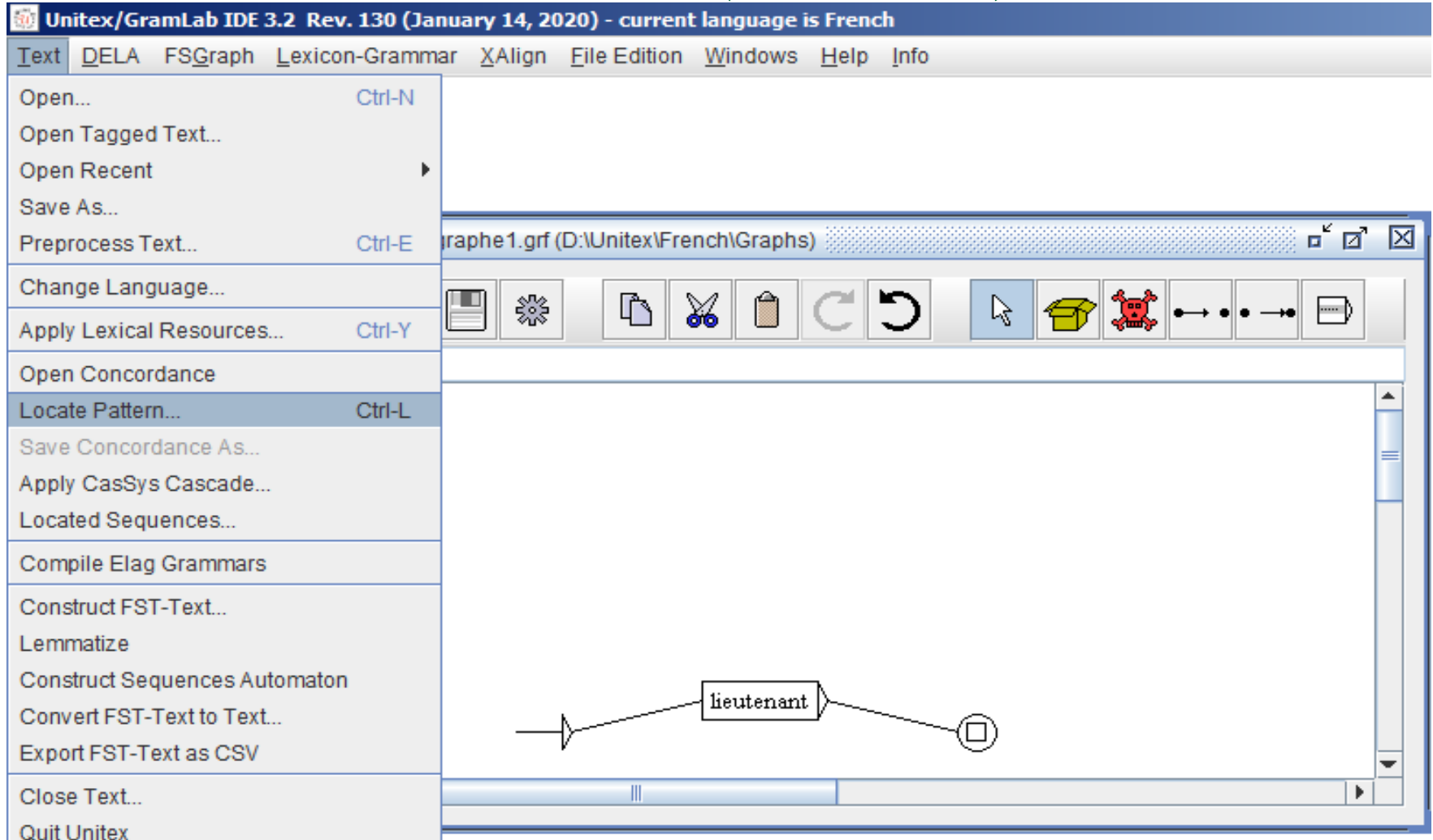


Remarque : Les lettres majuscules dans un graphe ne reconnaissent que les lettres majuscules dans le texte, alors que les lettres minuscules peuvent reconnaître les lettres minuscules et majuscules. Ceci est valable aussi lorsqu'on appelle un dictionnaire dans un graphe d'extraction (voir slide suivant).

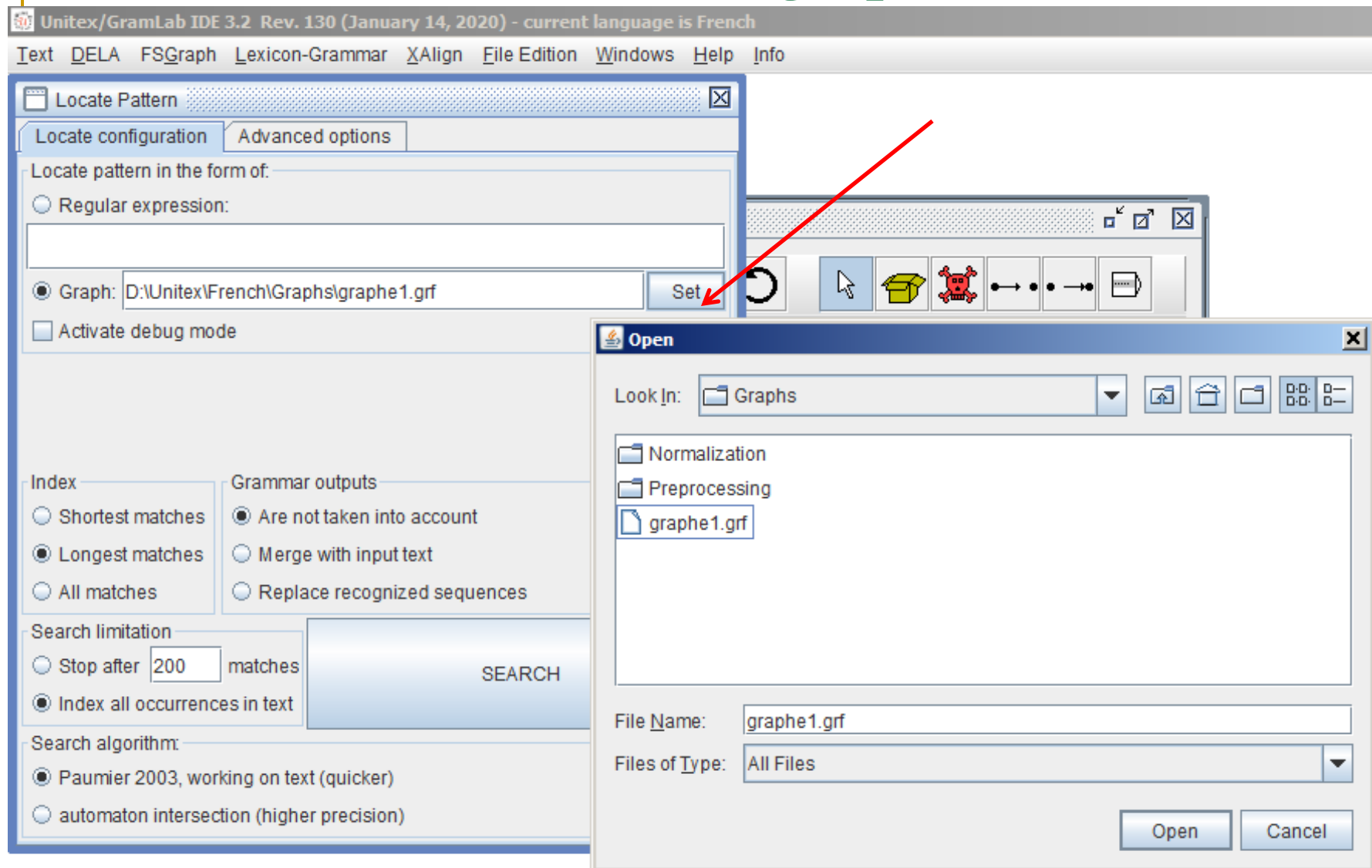
Enregistrement du graphe



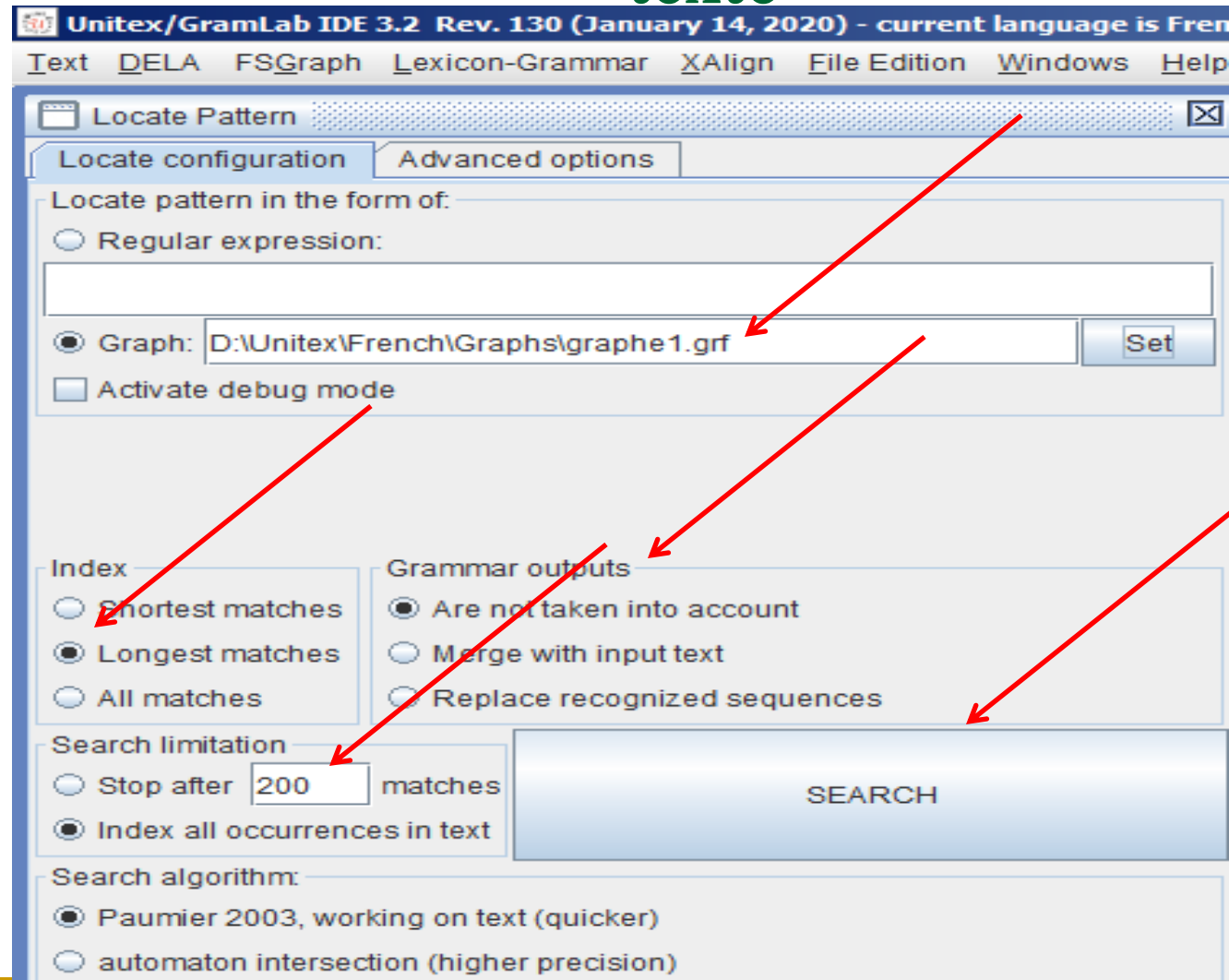
Utilisation du graphe pour extraire l'information souhaitée (CTRL+L)



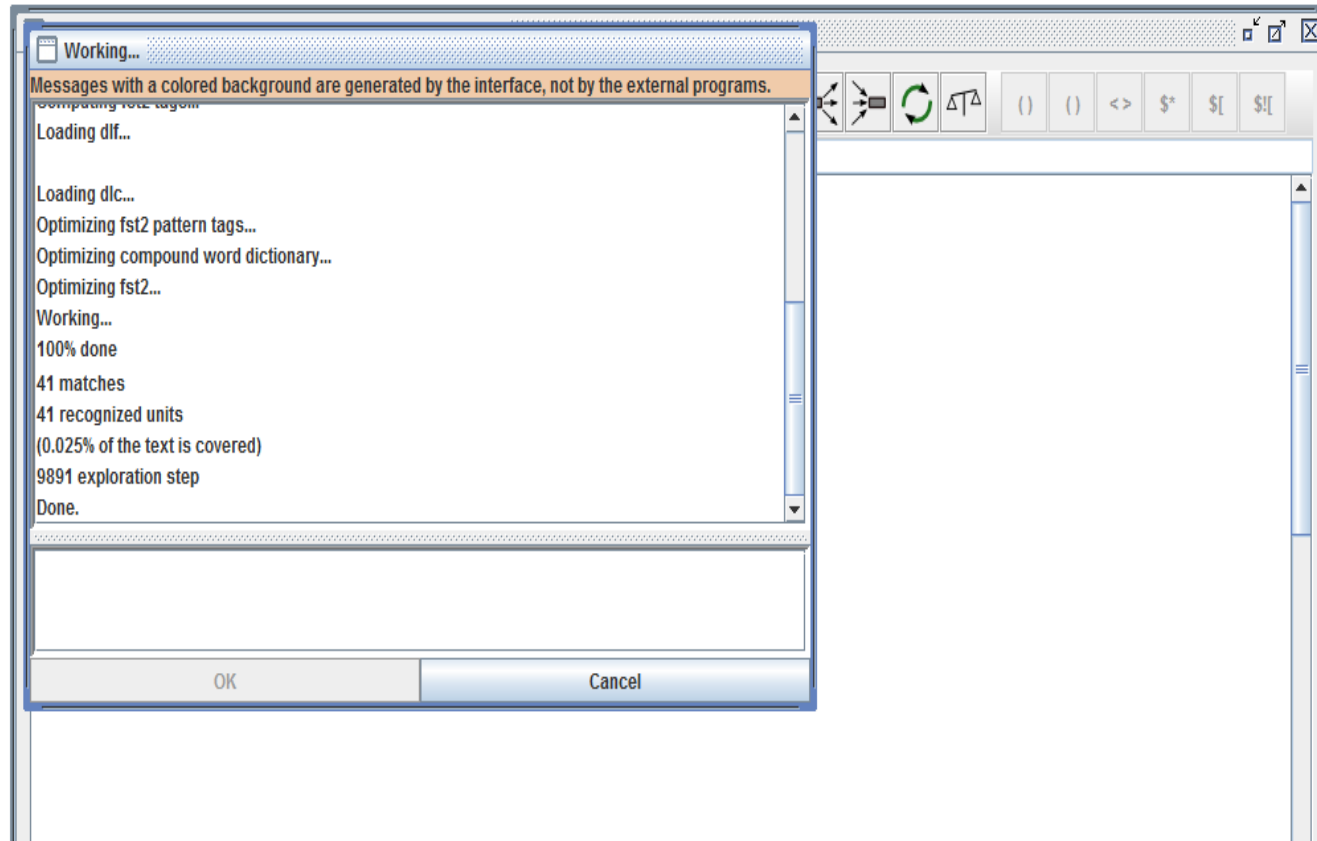
Sélection du graphe



Paramétrage de l'application du graphe sur le texte



Application du graphe



Application du graphe

3652 sentence delimiters, 165243 (9453 diff) tokens, 71860 (9423) simple forms, 438 (10) digits
67454 occurrences (12770 DLF entries) simple words, 1878 occurrences (2031 DLC entries) compound words, 4221 occurrences (47

Chapitre I

{S}DANS LEQUEL PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT L'UN COMME MAÎTRE, L'AUTRE COMME DOMESTIQUE

{S}En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens _ maison dans laquelle Sheridan mourut en 1814 _ , était habitée par Phileas Fogg, esq., l'un des membres les plus singuliers et les plus remarqués du Reform-Club 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 qui honorent l'Angleterre, succédait donc ce Phileas Fogg, personnage énigmatique, dont on ne savait rien, sinon que c'était l'un des plus beaux gentlemen de la haute société anglaise.

{S}On disait qu'il ressemblait à Byron _ automobile par la tête quant aux pieds _ , mais un Byron à moustaches et à favoris, vécu mille ans sans vieillir.

{S}Anglais, à coup sûr, Phileas Fogg n'était peut-être pas Lo vu ni à la Bourse, ni à la Banque, ni dans aucun des comptoirs ni les docks de Londres n'avaient jamais reçu un navire ayant Ce gentleman ne figurait dans aucun comité d'administration. {S} retenti dans un collège d'avocats, ni au Temple, ni à Lincoln's-Inn, ni à Gray's-Inn. {S} Jamais il ne plaida ni à la Cour du chancelier, ni au Banc de la Reine, ni à l'Échiquier, ni en Cour ecclésiastique. {S} Il n'était ni industriel, ni négociant, ni marchand, ni agriculteur. {S} Il ne faisait partie ni de l'Institution royale de la Grande-Bretagne, ni de l'Institution de Londres, ni de l'Institution des Artisans, ni de l'Institution Russell, ni de l'Institution littéraire de l'[Ouest], ni de l'Institution du Droit, ni de cette Institution des Arts et des Sciences réunis, qui est placée sous le patronage direct de Sa Gracieuse Majesté. {S} Il n'appartenait enfin à aucune des nombreuses sociétés qui pullulent dans la capitale de l'Angleterre, depuis la Société de l'Armonica jusqu'à la Société entomologique, fondée principalement dans le but de détruire les insectes nuisibles. {S} Phileas Fogg était membre du Reform-Club, et voilà tout.

Result Info

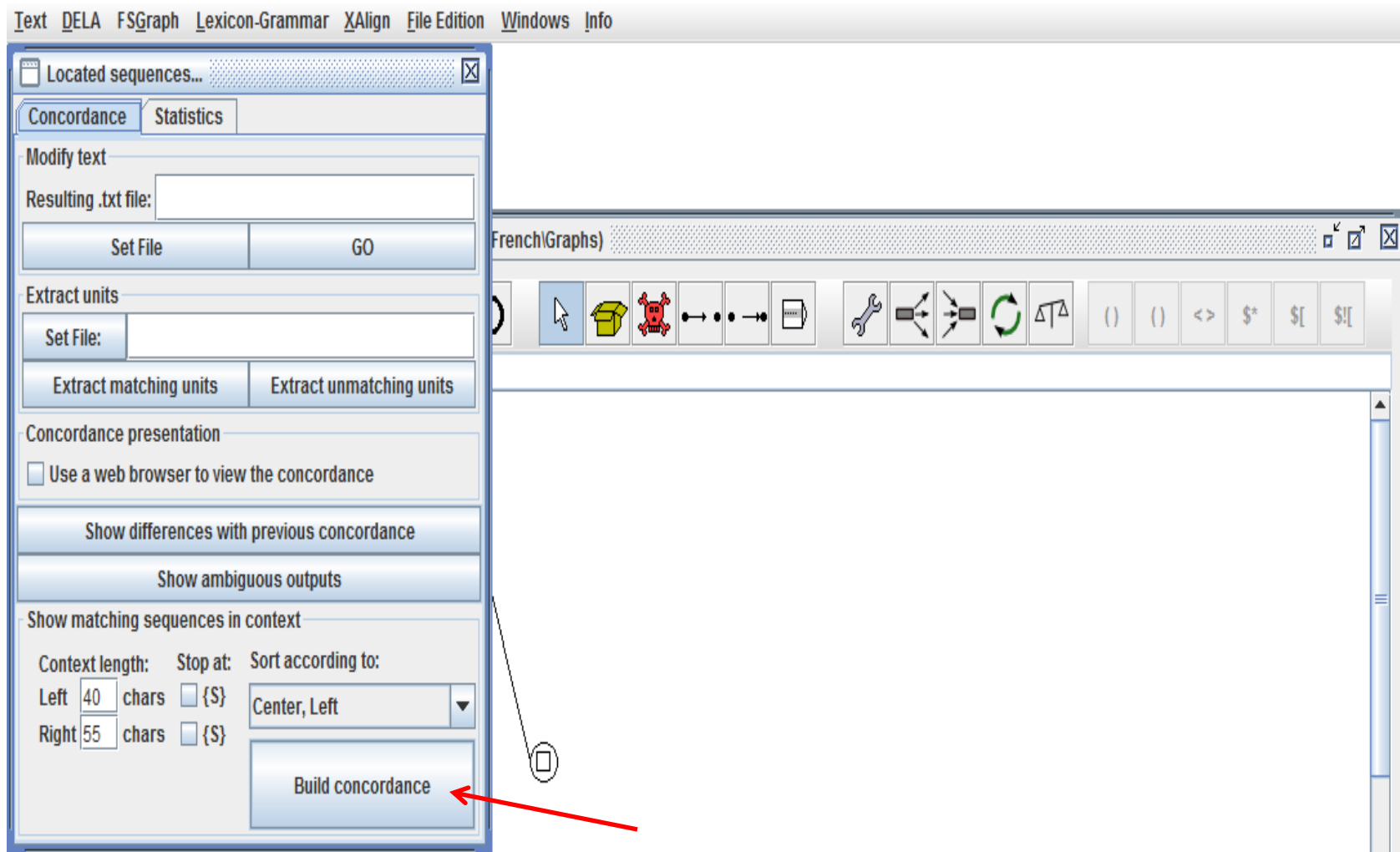
1 match

1 recognized units

(0.000% of the text is covered)

OK

Construction de la concordance contenant le résultat



Concordance contenant les occurrences trouvées

1 match

gouverneurs à Madras, à Bombay, au Bengale, et un lieutenant-gouverneur à Agra. {S}Mais l'I

Concordance contenant les occurrences trouvées

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

1 match

gouverneurs à Madras, à Bombay, au Bengale, et un lieutenant-gouverneur à Agra. {S}Mais l'Inde ang

80jours.snt (C:\Users\Public\Documents\Unitex-GramLab\Unitex\French\Corpus)

3652 sentence delimiters, 165243 (9453 diff) tokens, 71860 (9423) simple forms, 438 (10) digits
67454 occurrences (12770 DLF entries) simple words, 1878 occurrences (2031 DLC entries) compound words, 4221 occurrences (473 ...

{S}Personne n'ignore que l'Inde _ ce grand triangle renversé dont la base est au nord et la pointe au sud _ comprend une superficie de quatorze cent mille milles carrés, sur laquelle est inégalement répandue une population de cent quatre-vingts millions d'habitants.{S} Le gouvernement britannique exerce une domination réelle sur une certaine partie de cet immense pays.{S} Il entretient un gouverneur général à Calcutta, des gouverneurs à Madras, à Bombay, au Bengale, et un lieutenant-gouverneur à Agra.

{S}Mais l'Inde anglaise proprement dite ne compte qu'une superficie de sept cent mille milles carrés et une population de cent à cent dix millions d'habitants.{S} C'est assez dire qu'une notable partie du territoire échappe encore à l'autorité de la reine ;{S} et, en effet, chez certains rajahs de l'intérieur, farouches et terribles, l'indépendance indoue est encore absolue.

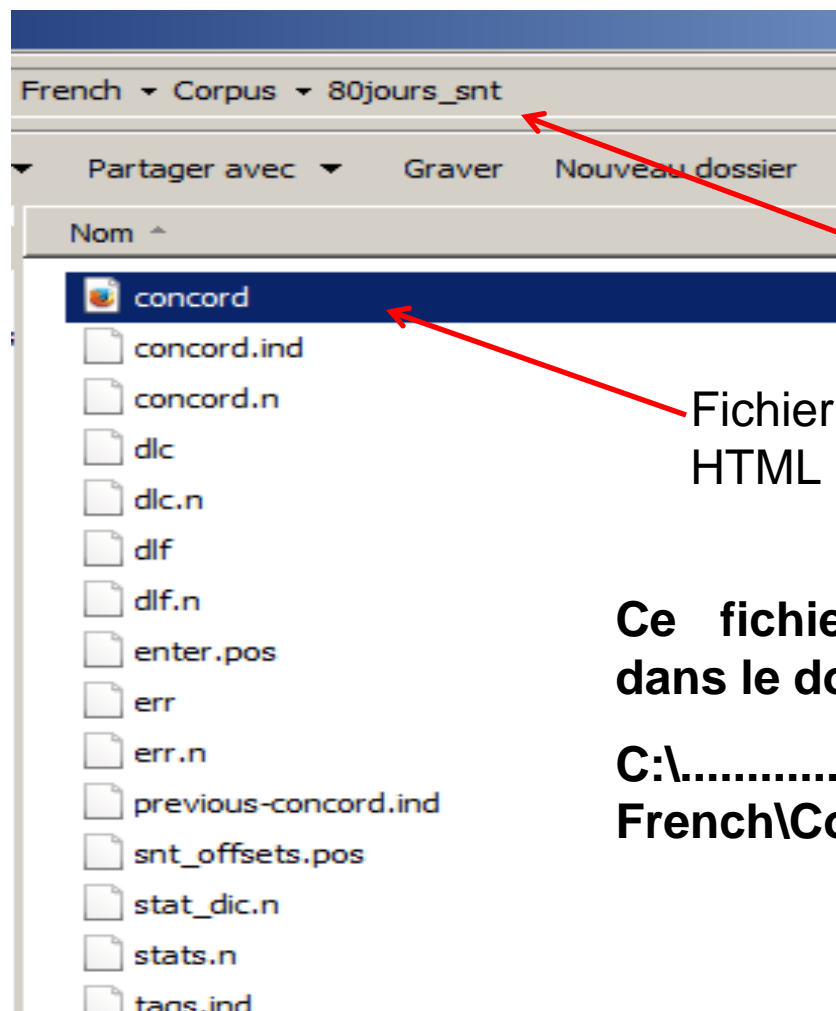
{S}Depuis 1756 _ époque à laquelle fut fondé le premier établissement anglais sur l'emplacement aujourd'hui occupé par la ville de Madras _ jusqu'à cette année dans laquelle éclata la grande insurrection des cipayes, la célèbre Compagnie des Indes fut toute-puissante.{S} Elle s'annexait peu à peu les diverses provinces, achetées aux rajahs au prix de rentes qu'elle payait peu ou point ;{S} elle nommait son gouverneur général et tous ses employés civils ou militaires ;{S} mais maintenant elle n'existe plus, et les possessions anglaises de l'Inde relèvent directement de la couronne.

{S}Aussi l'aspect, les moeurs, les divisions ethnographiques de la péninsule tendent à se modifier chaque jour.{S} Autrefois, on y voyageait par tous les antiques moyens de transport, à pied, à cheval, en charrette, en brouette, en palanquin, à dos d'homme, en coach, etc.

{S}Maintenant, des steamboats parcourent à grande vitesse l'Indus, le Gange, et un chemin de fer, qui traverse l'Inde dans toute sa largeur en se ramifiant sur son parcours, met Bombay à trois jours seulement de Calcutta.

{S}Le tracé de ce chemin de fer ne suit pas la ligne droite à travers l'Inde.{S} La distance à vol d'oiseau n'est que de mille à onze cents milles, et des trains, animés d'une vitesse moyenne seulement, n'emploieraient pas trois jours à la franchir ;{S} mais cette distance est accrue d'un tiers, au moins, par la corde que décrit le railway en s'élevant jusqu'à

Fichier « concord.html »



Fichier
HTML

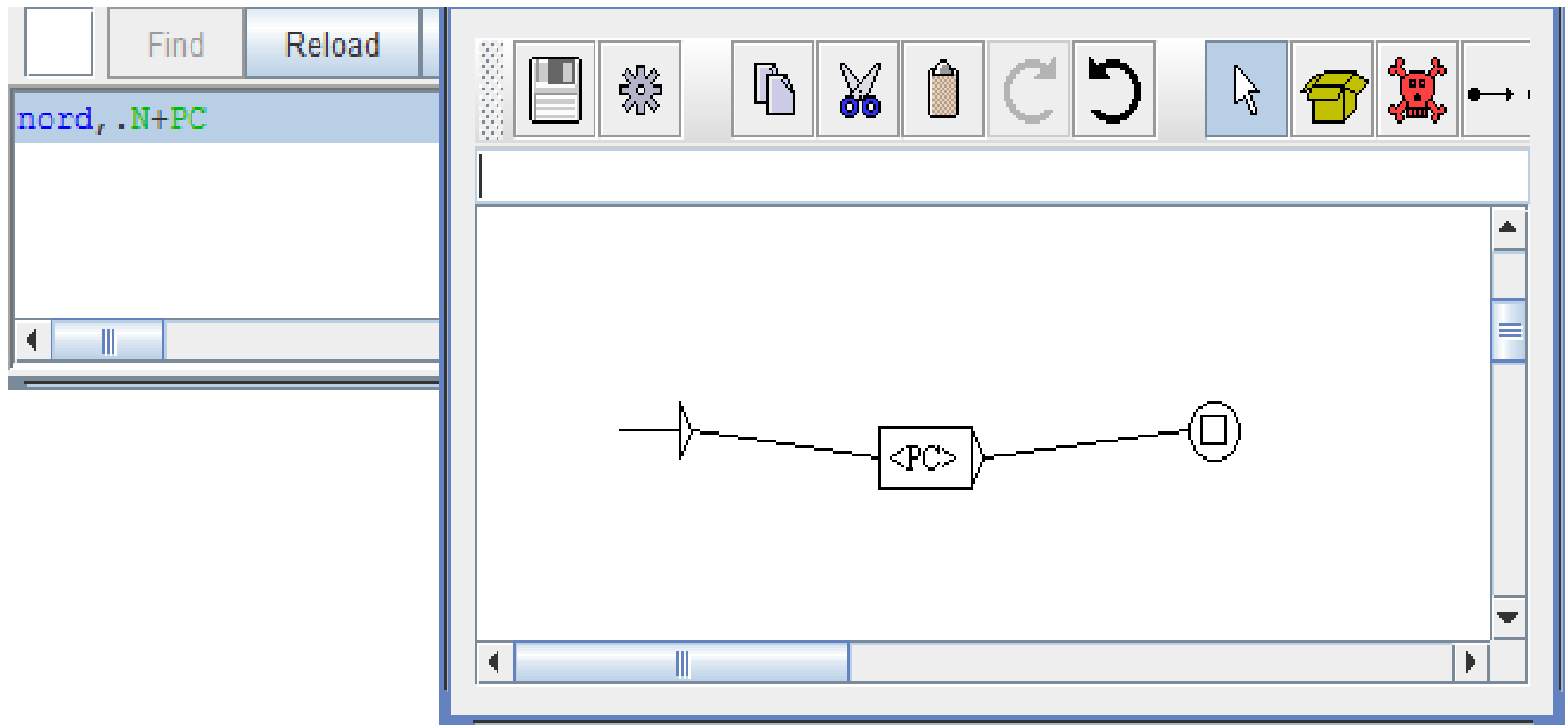
**Ce fichier HTML se trouve
dans le dossier :**

**C:\.....\Unitex\
French\Corpus\80jours_snt**

Code source du fichier « concord.html »

```
<html lang=en>
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
  <title>1 match</title>
</head>
<body>
<table border="0" cellpadding="0" width="100%" style="font-family: 'Courier new'; font-size: 12">
<tr><td nowrap>s à Madras, à Bombay, au Bengale, et un<nbsp><a href="79036 79046 607 0">lieutenant</a>-gouverneur à Agra. {S}Mais l'Inde anglaise p<nbsp></td></tr>
</table></body>
</html>
```

Appel d'un dictionnaire DELAF dans un graphe



Appel d'un dictionnaire DELAF dans un graphe

The image displays a software interface with three main components:

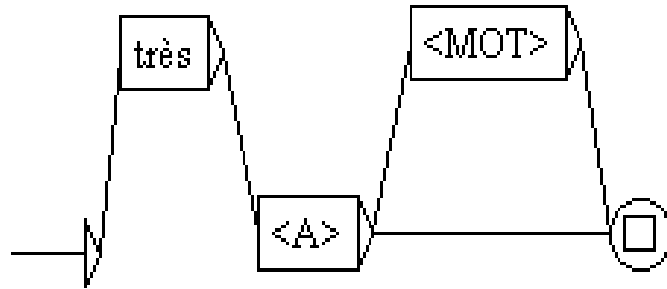
- Top Left Window:** A small window titled "D:\Unitex\French\Delaf\pc.dic". It contains a search bar with the text "nord, .N+PC" and buttons for "Find", "Reload", and navigation arrows.
- Top Right Window:** A window titled "D:\Unitex\French\Graphs". It features a toolbar with icons for file operations (copy, paste, delete, undo, redo) and a large empty area for graph visualization.
- Bottom Window:** A window titled "Concordance: D:\Unitex\French\Corpus\80jours_snt\concord.html". It displays a list of 26 matches for the word "nord". The text is as follows:

26 matches

joignait campaient a quelques milles au nord de la ville. Le brigadier ge
fornie, la ligne courait directement au nord-est, en longeant American-r
tant plus que les courants y portent au nord. - Pilote, répondit Phileas
haïne des Ghâtes-Occidentales, court au nord-est jusqu'à Burhampour, sil
d triangle renversé dont la base est au nord et la pointe au sud - compre
u territoire remontait, du sud-ouest au nord-ouest, par Grand-Island, Col
r jusqu'à l'embouchure de la branche du nord, suit la branche du sud, tra
evint très fort. Fixé dans la partie du nord-ouest, il contraria la march
huit heures quarante, entre la gare du Nord et la gare de Lyon, à trave
te et des voyageurs entre l'Amérique du Nord, la Chine, le Japon et les
voir un coup de vent. - Viendra-t-il du nord ou du sud ? demanda simpleme
evada, suivant toujours la direction du nord-est. A midi, il quittait Re
e North-Platte-river. Tout l'horizon du nord et de l'est était couvert p

A red oval highlights the word "nord" in the concordance results, which is linked to the dictionary window above it.

Index - Shortest/Longest/All



1 Très grand. Très grand homme.

Index - Shortest

2 matches

Très grand. Très grand homme.
Très grand. Très grand homme.

Index - Longest

2 matches

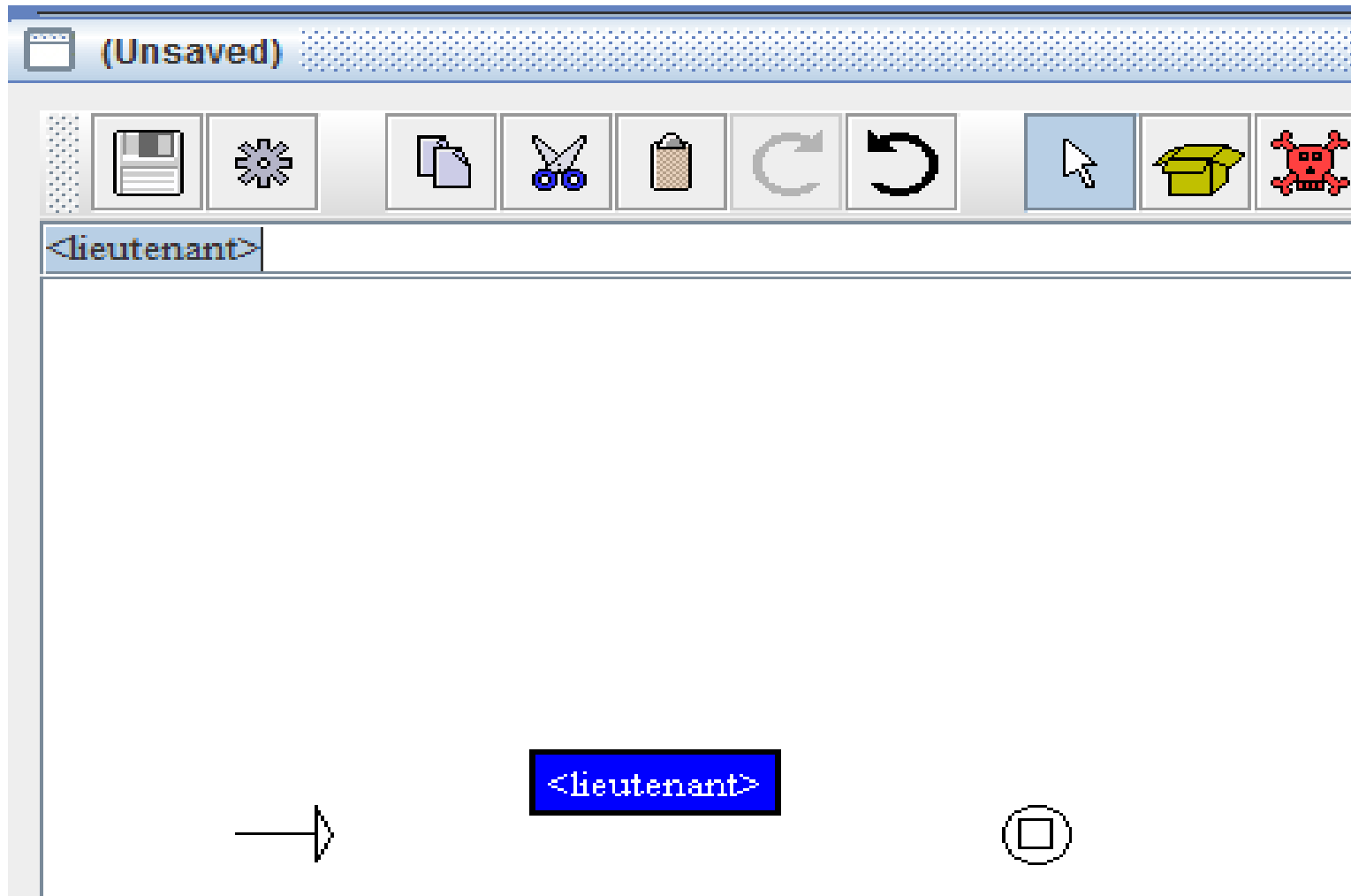
Très grand. Très grand homme.
Très grand. Très grand homme.

Index - All

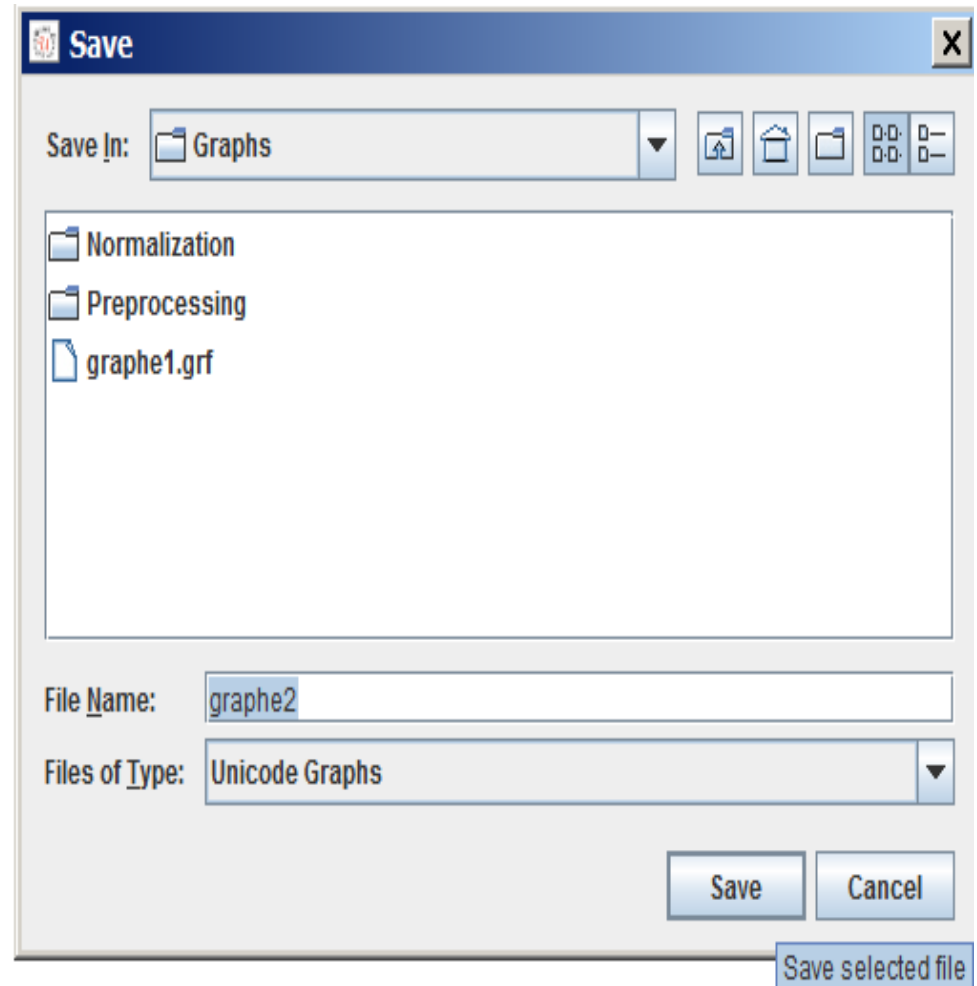
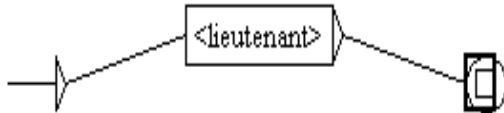
3 matches

[Très grand](#). Très grand homme.
Très grand. [Très grand](#) homme.
Très grand. [Très grand homme](#).

Retour au lemme <lieutenant>



Retour au lemme <lieutenant>



Résultat de l'application du lemme <lieutenant>

3652 sentence delimiters, 165243 (9453 diff) tokens, 71860 (9423) simple forms, 438 (10) digits
67454 occurrences (12770 DLF entries) simple words, 1878 occurrences (2031 DLC entries) compound words, 4221 occur

Chapitre I

{S}DANS LEQUEL PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT L'UN COMME MAÎTRE
L'AUTRE COMME DOMESTIQUE

{S}En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens _
dans laquelle Sheridan mourut en 1814 _ , était habitée _
membres les plus singuliers et les plus remarquables du R _
semblât prendre à tâche de ne rien faire qui pût attir

{S}A l'un des plus grands orateurs qui honorent l'Angl
personnage énigmatique, dont on ne savait rien, sinon
l'un des plus beaux gentlemen de la haute société angl
{S}On disait qu'il ressemblait à Byron _ automobile pa
quant aux pieds _ , mais un Byron à moustaches et à fa
vécu mille ans sans vieillir.

{S}Anglais, à coup sûr, Phileas Fogg n'était peut-être

Result Info

3 matches

3 recognized units

(0.001% of the text is covered)

OK

Résultat de l'application du lemme <lieutenant>

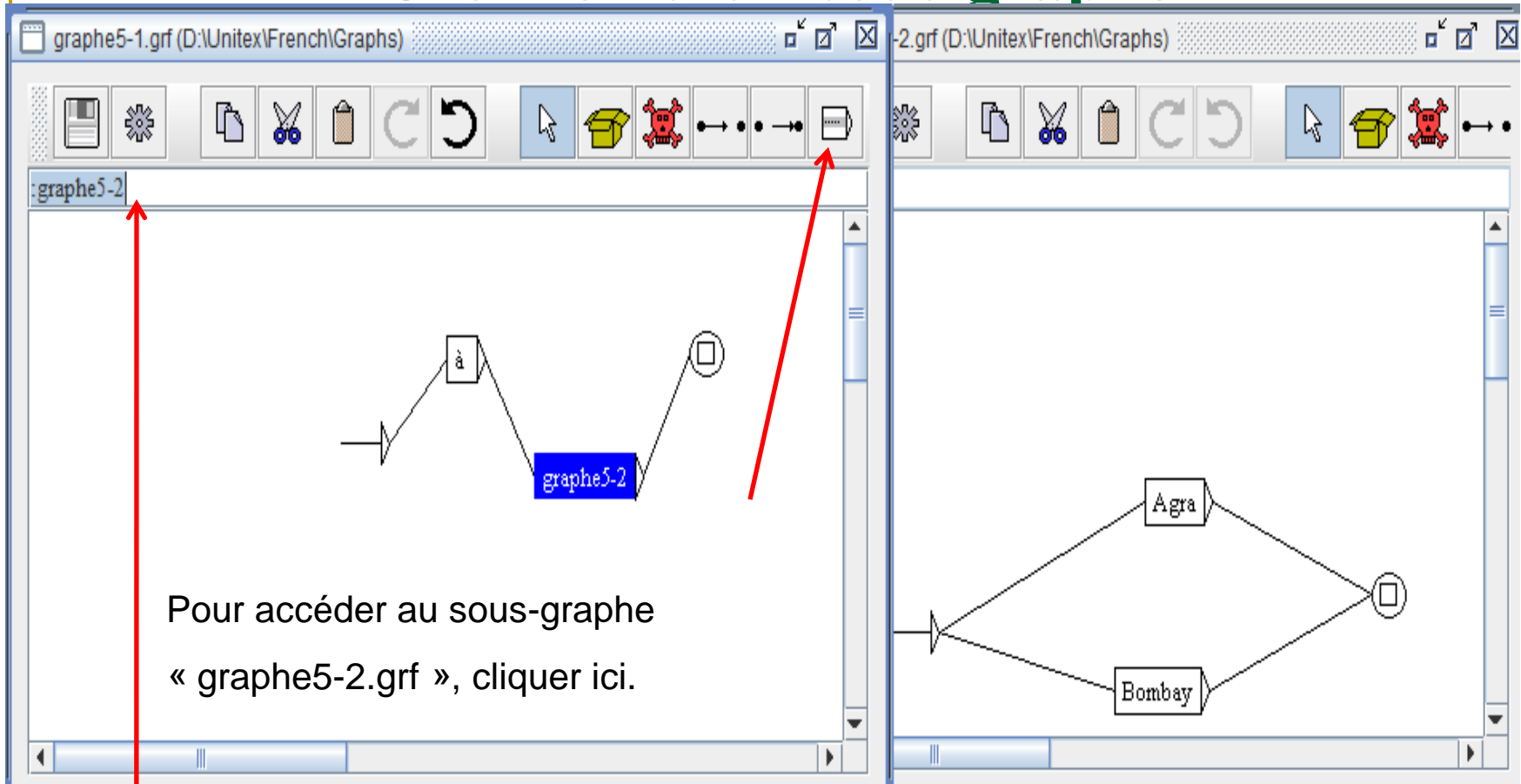
3 matches

s à Madras, à Bombay, au Bengale, et un lieutenant-gouverneur à Agra. {S}Mais l'Inde
pas, et d'un geste, appelant un de ses lieutenants, il lui donnait l'ordre de pousser
e l'ancienne Compagnie des Indes : sous-lieutenants à 7 000 F, brigadiers à 60 000,

Code source du fichier « concord.html »

```
<html lang=en>
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
  <title>3 matches</title>
</head>
<body>
<table border="0" cellpadding="0" width="100%" style="font-family: 'Courier new'; font-size: 12">
<tr><td nowrap>s à Madras, à Bombay, au Bengale, et un< a href="79035 79045 607 1">lieutenant</a>-gouverneur à Agra. {S}Mais l'Inde anglaise p
</td></tr>
<tr><td nowrap>&nbsp;pas, et d'un geste, appelant un de ses< a href="361211 361222 3000 2">lieutenants</a>, il lui donnait l'ordre de pousser une reco
</td></tr>
<tr><td nowrap>e l'ancienne Compagnie des Indes : sous-< a href="67898 67909 508 0">lieutenants</a>&nbsp;à 7 000 F, brigadiers à 60 000, généraux à&nbsp;&nbsp;
</td></tr>
</table></body>
</html>
```

Création d'un sous-graphe



Pour créer un sous-graphe, saisir son nom précédé de deux points ➔ **:graphe5-2**

Résultat de l'extraction

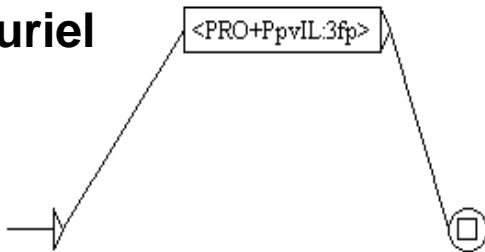
Concordance: D:\Unitex\French\Corpus\80jours_snt\concord.html

31 matches

au Bengale, et un lieutenant-gouverneur à Agra. Mais l'Inde anglaise prop
e le pavé de la pagode de Malebar-Hill, à Bombay, dans la journée du 20
aji, mais de la pagode de Malebar-Hill, à Bombay. - Et comme pièce de cor
eprendre leur victime, fût-ce à Madras, à Bombay, à Calcutta. Et Sir Fra
l à Calcutta, des gouverneurs à Madras, à Bombay, au Bengale, et un lieu
lez-vous donc ? - Mais, ainsi que vous, à Bombay. - C'est au mieux ! Est-
de Londres ? - Oui. - Et vous allez ? - A Bombay. - Bien, monsieur. Vous
! Non. Toute cette gymnastique cessera à Bombay, n'en doutez pas. - Et
it de Passepartout. Jusqu'à son arrivée à Bombay, il avait cru et pu croi
e à comprendre, - car à Calcutta, comme à Bombay, comme à Suez, le manda
lia ne devait arriver que le 22 octobre à Bombay. Or, il y arrivait le 20
s de Passepartout, qui devait le croire à Bombay. Mais il fut amené à re
rsée du Mongolia. Quoi ! je vous laisse à Bombay, et je vous retrouve sur
que je retarde son départ ! J'ai échoué à Bombay, j'ai échoué à Calcutta
'embarque sur le Mongolia, qui débarque à Bombay, où il dit devoir séjou
passagers du Mongolia avaient débarqué à Bombay, et le train de Calcutta
t régulièrement les voyages de Brindisi à Bombay par le canal de Suez. C
és, en effet, de Londres à Aden, d'Aden à Bombay, de Calcutta à Singapour
nvoyez sans retard mandat d'arrestation à Bombay (Inde anglaise). Fix, de
e de m'adresser un mandat d'arrestation à Bombay, m'embarquer sur le Mon
si, il y a quelques années, je résidais à Bombay, quand une jeune veuve
écanicien du Mongolia, si nous arrivons à Bombay avec une belle avance !
tout le temps que celui-ci demeurait à Bombay. Il ne doutait pas que

Résultat de l'application du masque lexical complexe <PRO+PpvIL:3fp> sous forme de graphe

**Pronoms personnels préverbaux
sujets, 3^{ème} personne du féminin,
pluriel**

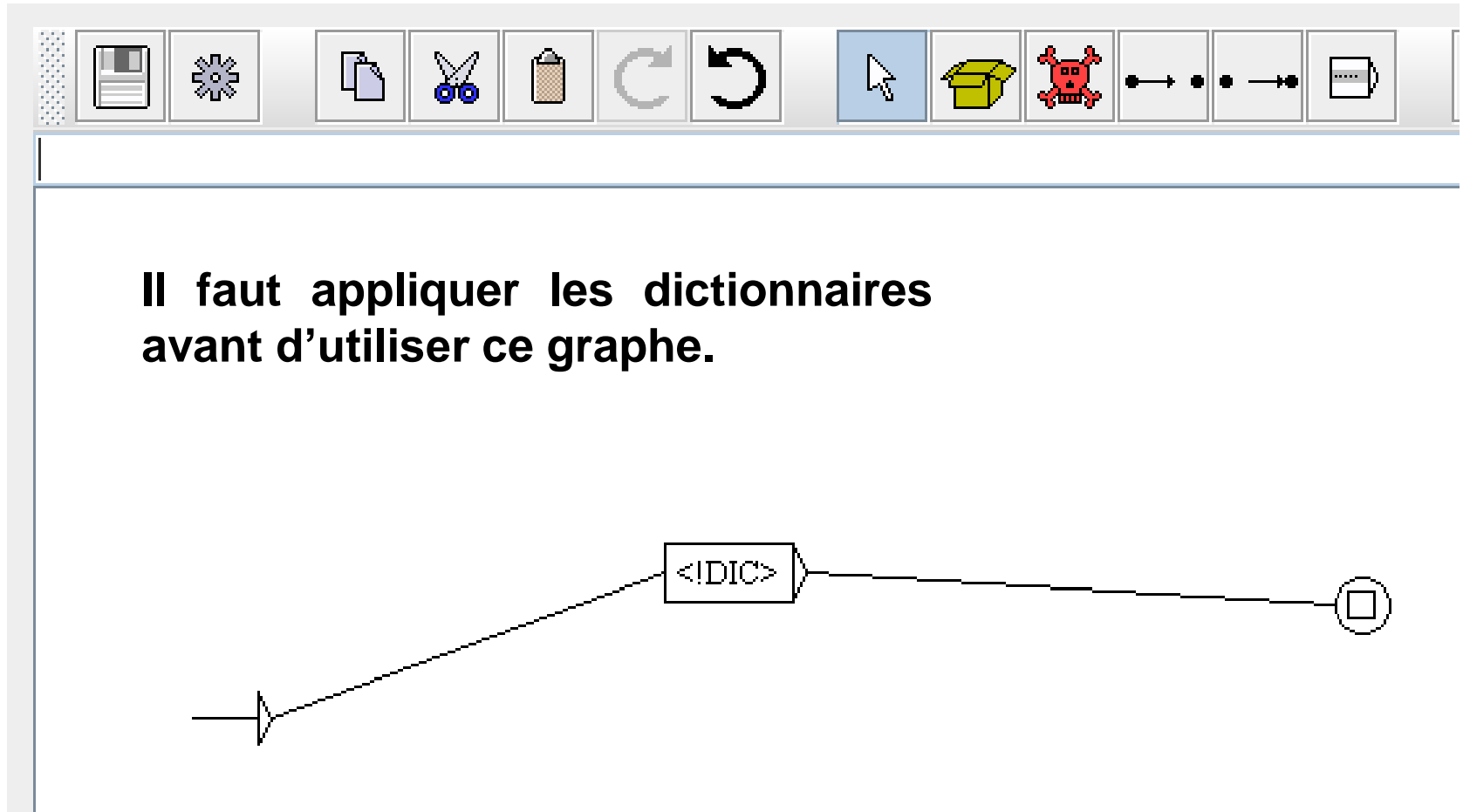


**On doit appliquer au
préalable le dictionnaire
système :**

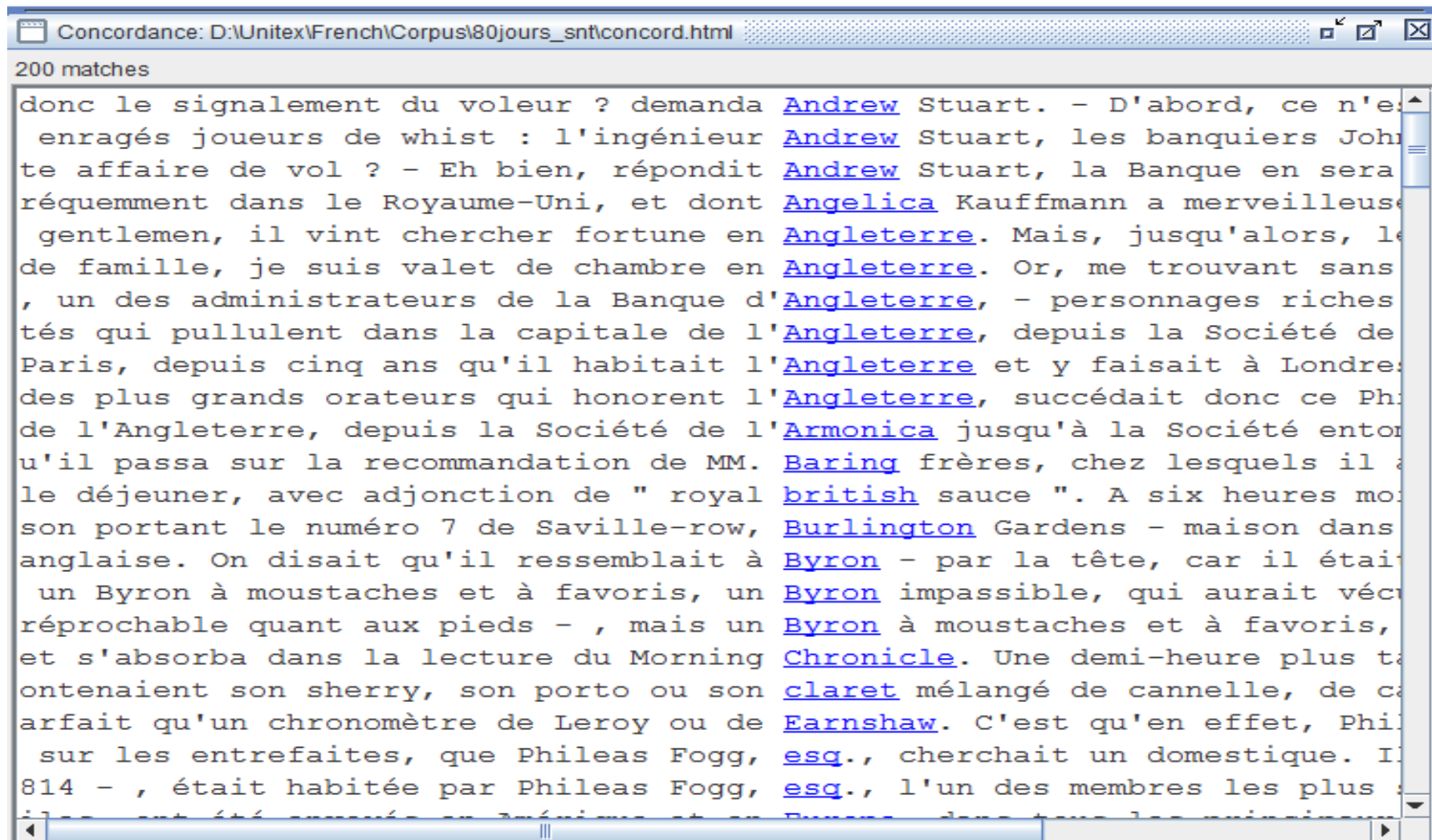
« motsGramf-.bin »

eur poche, et quand ils les retiraient, elles tournaient encore
es gravissaient des échelles de bambou, elles se dispersaient d
t surtout à cette époque de l'année.{S} Elles sont généralement
fées capricieuses venant de la côte.{S} Elles passaient, et la :
us d'un côté de la scène à l'autre ;{S} elles faisaient le tour
leurs leur interminable giration ;{S} elles couraient sur des
erses.{S} Les jongleurs jonglaient avec elles, et elles tournai
es, qu'il éteignit successivement quand elles passèrent devant
ers le désert et destiné à relier entre elles des villes et des
arties d'un bâtiment sont reliées entre elles. {S}Avec la nuit,
rain, les voitures communiquaient entre elles par des passerell
Passepourtout, et les perturbations, si elles existaient, eusse
s toujours prêts à partir.{S} De Dublin elles arrivent à Liverp
_ En effet, répondit le gentleman, mais elles passent. _ A bord
es détonations les accompagnèrent, mais elles ne venaient point
Les jongleurs jonglaient avec elles, et elles tournaient dans l
volants, avec des raquettes de bois, et elles tournaient toujou
ces rages secrètes, terribles parce qu'elles sont contenues, e

La négation d'une boîte



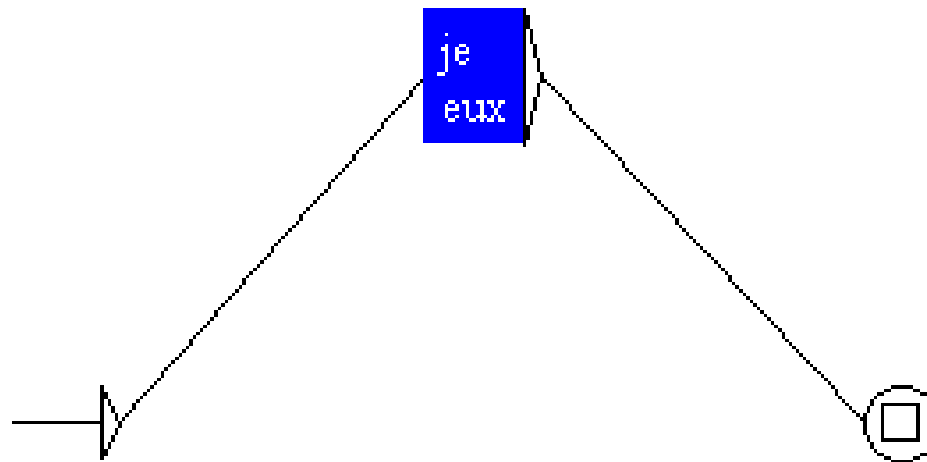
Affichage de tous les mots inconnus du texte



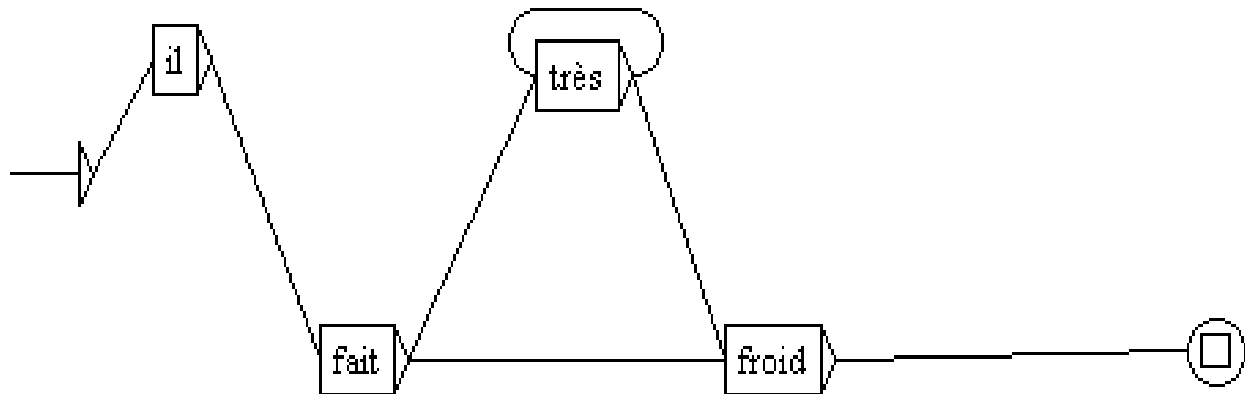
Union



Reconnaît « je » ou « eux »

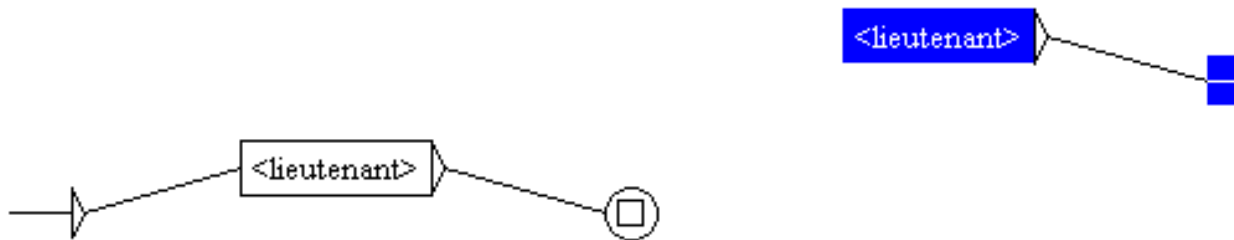


Boucles



Ce graphe reconnaît : il fait froid, il fait très froid, il fait très très froid, etc.

Copier une branche



Exercice

- Extraire à l'aide d'un graphe d'extraction les dates à partir du texte 80jours.txt
- Le résultat attendu est fourni dans le fichier concord.html
- Le résultat doit contenir :
 - 63 matches (occurrences trouvées);
 - 285 unités reconnues (unités couvertes par ces occurrences);
 - 0.176% (rapport entre le nombre d'unités couvertes et le nombre total d'unités du texte).