

# Linguistique de corpus 2

## Outils et méthodes de traitement de corpus 2

Patrick Paroubek

LIMSI-CNRS  
Dépt. CHM - Groupe LIR  
Bât. 508 Université Paris XI, 91403 Orsay Cedex  
[pap@limsi.fr](mailto:pap@limsi.fr)

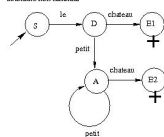
mercredi 08 Février 2017 / L3 - Cours S2-2

# Automates d'états finis déterministes, pratique UNITEX

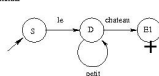
Éditer le graphe d'un automate d'état fini déterministe avec le menu : FS\_GRAPH+new Ici un graphe qui reconnaît une séquence arbitraire de “le” traduit en sortie en “the”. Rappel, le texte à entrer pour annoter le nœud est : le/the

le château  
le petit château  
le petit petit château  
le petit petit petit château  
le petit petit petit...petit château

automate non minimal

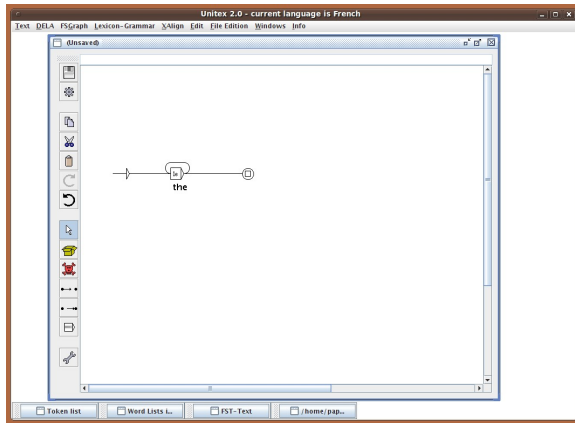


automate minimal



Pour appliquer le graphe comme patron de recherche, activer l'item de menu : Text/locate pattern/Graph+Set+Search

# Automates d'états finis déterministes, pratique UNITEX



# Automates d'états finis déterministes, pratique UNITEX

Pour créer un nœud qui génère une sortie, il faut entrer comme texte du nœud le symbole d'entrée, suivi du caractère “/”, suivi du mot de sortie (ce dernier apparaît alors sous la boîte représentant le nœud dans l'éditeur de graphe.

L'application d'un automate sur un texte (menu fichier/locate\_pattern/graph), offre 3 options :

- ❶ ignorer les symboles de sortie
- ❷ ajouter les symboles de sortie au texte
- ❸ remplacer les séquences de symboles d'entrée par les symboles de sortie.

# Automates d'états finis déterministes, pratique UNITEX

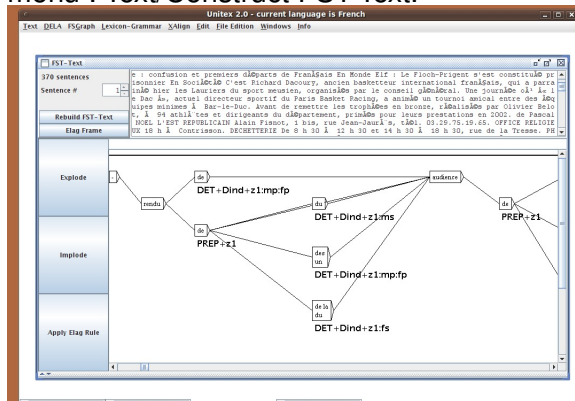
Charger le texte du roman de *Jules Verne tour du monde en 80 jours*, puis éditer un automate d'états finis déterministe qui reconnaisse les expressions temporelles, c'est-à-dire qui font référence à un point ou bien un interval dans le temps, e.g. lundi, le 23 mars, le 3 avril 2011, etc.

L'édition de graphe est expliquée à partir de la page 67 du manuel UNITEX.

Le manuel en français (pdf) d'UNITEX est disponible à l'URL : <http://igm.univ-mlv.fr/unitex/manuel.html>

# Automates d'états finis déterministes, pratique UNITEX

Unitex permet de transformer automatiquement un texte en automates d'états finis déterministes (un par phrase) et de le visualiser avec l'éditeur d'automates, en activant l'item de menu : Text/Construct FST-Text.



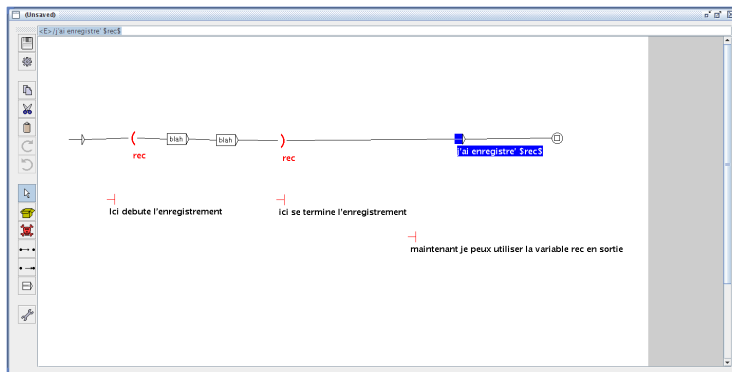
# Unitex

## Variables et transducteurs en Unitex

### En Unitex

- une variable “**FOO**” se déclare en créant 2 boîtes-états dont le texte est respectivement : \$FOO( et \$FOO).
- à partir de ce moment là, le texte rencontré de n'importe quel parcours du graphe d'automate entre les 2 boîtes est stocké dans la variable FOO.
- Pour restituer son contenu, on utilise \$FOO\$ dans le texte de sortie d'une boîte.

# Exemple de variable Unitex





# Appel de sous-graphe

Il est possible de diviser un automate d'états finis déterministe en plusieurs automates modulaires s'appelant les uns les autres afin de ne pas surcharger l'écran et de rationaliser la gestion des phénomènes linguistiques gérés par les automates. Cela se fait en créant plusieurs automates et en donnant comme texte d'un nœud de l'automate **appelant le nom d'un fichier contenant l'automate appelé précédé du caractères " : " (deux-points)**.

Notez que l'on peut soit donner le chemin complet depuis la racine, un chemin relatif et si le fichier de l'automate se trouve dans le repertoire par défaut pour les graphes d'automates d'unitex, il est possible de donner uniquement le nom du fichier sans son suffixe.

# Appel de sous-automate

