



# Plate-formes logicielles pour le TAL 1 : graphes et grammaires locales dans Unix

Karën Fort

karen.fort@sorbonne-universite.fr / <http://karenfort.org>

30 novembre 2018



# Quelques sources d'inspiration

- Manuel d'Unitex :  
<http://www-igm.univ-mlv.fr/~unitex/index.php?page=4>
- M. Constant (Université de Marne-la-Vallée / IGM), qui a patiemment répondu à toutes mes questions

- 1 Sources
- 2 Correction des exercices du cours précédent
- 3 Manipuler les transducteurs
- 4 Avant-goût : construire des dictionnaires
- 5 Annexes
- 6 Pour finir

## Grammaire des dates

Modifiez la grammaire des dates pour extraire des résultats intéressants sur le corpus du *Tour du Monde en 80 jours*

# Groupes nominaux

Construisez une grammaire :

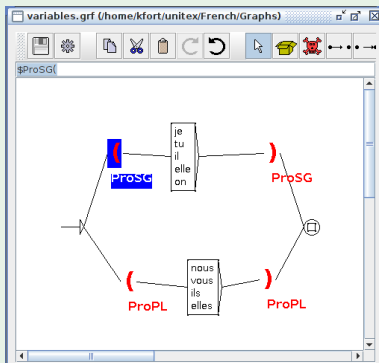
- reconnaissant des groupes nominaux simples, en tenant compte des accords en genre et en nombre
- insérez des sorties dans la grammaire afin qu'à partir du texte « après tout, son énorme gaffe n'est pas sérieuse. », on puisse obtenir la concordance suivante :  
*son énorme gaffe,.GN*

- 1 Sources
- 2 Correction des exercices du cours précédent
- 3 Manipuler les transducteurs**
  - Utiliser des variables
  - Opérations sur les variables
- 4 Avant-goût : construire des dictionnaires
- 5 Annexes
- 6 Pour finir

# Créer une variable (d'entrée)

## Création d'une variable

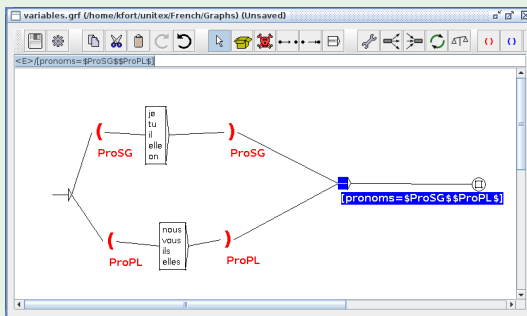
- créer un graphe qui reconnaît :
  - ▶ les pronoms personnels singuliers (je, tu, il, elle, on) ou
  - ▶ les pronoms personnels pluriels (nous, vous, ils, elles)
- insérer des états englobant définissant la variable **nomDeVariable** :  
**\$nomDeVariable(** et **\$nomDeVariable)**



# Utiliser les variables

## Fusionner des résultats

- ouvrir le graphe précédemment créé
- ajouter, juste avant l'état final, un état contenant  
`< E > / [nom de l'élément qu'on décrit = $nomDeVariable$]`





# Variables en mode *debug*

Concordance: /home/kfort/unitex/French/Corpus/80jours\_snt/concord.html

Tag	Output	Matched
<E>		
\$ProPL(		
nous	nous	
\$ProPL)		
<E>	[pronoms=nous]	

200 matches

ez de bonnes chaussures.{S} D'ailleurs, nous marcherons peu ou pas.{S} Allez. " {S} Passepartout, eux cents minutes.{S} Acceptez-vous ? \_ Nous acceptons, répondirent MM. Stuart, Fallentin, Sull : «Cela me va ! voilà mon affaire ! {S} Nous nous entendrons parfaitement, Mr. Fogg et moi ! {S} et je ne vous fais pas de reproche.{S} Nous partons dans dix minutes pour Douvres et Calais. " -t-il. \_ Oui, répondit Phileas Fogg.{S} Nous allons faire le tour du monde. " {S} Passepartout, paires de bas.{S} Autant pour vous.{S} Nous achèterons en route.{S} Vous descendrez mon ma , au contraire, dit Gauthier Ralph, que nous mettrons la main sur l'auteur du vol. {S} Des inspe ela me va ! voilà mon affaire ! {S} Nous nous entendrons parfaitement, Mr. Fogg et moi ! {S} Un h Et c'est ce qui, dans le cas dont nous nous occupons, rendra les recherches plus rapides. \_ Et s.{S} Et c'est ce qui, dans le cas dont nous nous occupons, rendra les recherches plus rapides. ns que le disait le *Morning Chronicle*, on avait lieu de supposer que l'auteur du vol ne faisa

Double-click to open the graph:



les variables sont globales

## Exercice : une autre utilisation des variables

### Inverser des motifs

- créer un graphe qui reconnaît les mois/années
- ajouter les variables \$mois et \$annee
- inverser les mois et les années

# Annotation vs variable

Annotation :

- ajoutée à la bande de sortie du FST
- ne peut pas être testée ou comparée

Variable :

- permet de stocker une chaîne de la bande d'entrée du FST
- peut être testée :
  - ▶ en insérant `$xxx.SET$` à la sortie d'une boîte
  - ▶ si une variable dénommée `xxx` a été définie, cette séquence est ignorée et la reconnaissance continue, sinon, elle s'arrête et le programme repart en arrière
- peut être comparée :
  - ▶ en insérant `$abc.EQUAL=xyz$` à la sortie d'une boîte
  - ▶ agit comme un interrupteur qui permet de bloquer l'exploration de grammaire si la valeur de la variable `abc` est différente de la valeur de la variable `xyz`

# Test

cf. Manuel d'Unitex p. 148

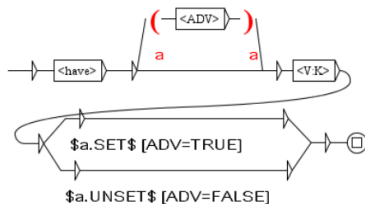
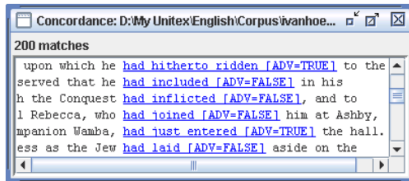


FIGURE 6.50 – Test d'une variable



- 1 Sources
- 2 Correction des exercices du cours précédent
- 3 Manipuler les transducteurs
- 4 Avant-goût : construire des dictionnaires**
- 5 Annexes
- 6 Pour finir

# Construction des dictionnaires

- ❶ construction d'un dictionnaire de formes canoniques (ou formes de base)
- ❷ construction de modules de flexion automatique (transducteurs)
- ❸ à chaque forme de base, on associe une classe flexionnelle (un ensemble de règles)

# Étape 1 : créer le fichier DELAS (formes non fléchies)

Menu File Edition / New File

Ajouter (1 par ligne) des mots (unités lexicales simples) qui sont toujours au masculin :

- ballon
- livre
- (votre exemple)

Quelle flexion ? On va la créer : N1000

Ce qui donne :

ballon,N1000

livre,N1000

**RETOUR À LA LIGNE**

Enregistrer le fichier **sous Dela** avec une extension .dic

## Étape 2 : créer le graphe de flexion

Menu FS Graph / New

Créer un graphe permettant :

- d'ajouter un s au (masculin) pluriel : s/ :mp
- de ne rien ajouter au (masculin) singulier : <E>/ :ms

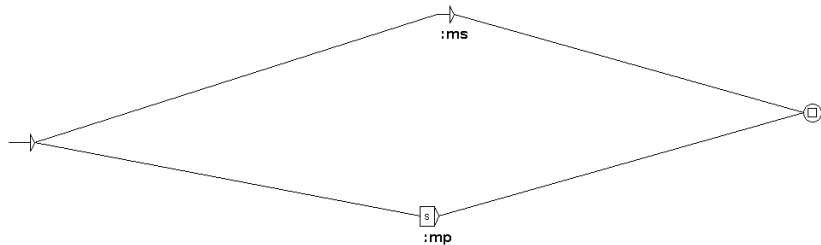
L'enregistrer **sous Inflection** avec le nom N1000 (.grf). Le compiler (Unitex va créer un .fst2).

! pas d'espace



## Étape 2 : résultat

ballon

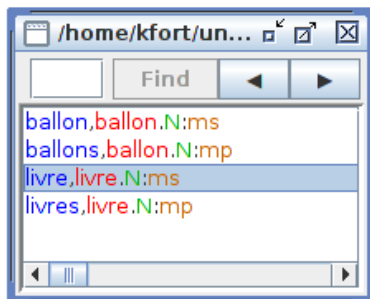


## Étape 3 : créer le dictionnaire fléchi

Menu DELA / Open

- sélectionner le fichier dictionnaire précédemment créé
- DELA / Inflect...
- Inflect Dictionary

## Étape 3 : résultat



## Mode console : UnitexTool

- permet d'exécuter les programmes externes d'Unitex
- permet d'enchaîner les commandes

Exemple : faire un locate et lancer la concordance

```
UnitexTool {
  Locate "-tD:\My Unitex\English\Corpus\ivanhoe.snt"
"D:\My Unitex\English\regexp.fst2"
"-aD:\My Unitex\English\Alphabet.txt" -L -I -n200
"--morpho=D:\Unitex2.0\English\Dela\del-a-en-public.bin" -b -Y
}
{
Concord "D:\My Unitex\English\Corpus\ivanhoe_snt\concord.ind"
"-fCourier new" -s12 -l40 -r55 --CL --html
"-aD:\My Unitex\English\Alphabet_sort.txt"
}
```

# Symboles spéciaux

Caractère	Signification	Codage
"	les guillemets délimitent des séquences qui ne doivent ni être interprétées par Unitex, ni subir de variantes de casse	\ "
+	+ sépare les différentes lignes boîtes	" + "
:	: sert à introduire à appel à un sous-graphe	" : " or \ :
/	/ indique le début de la sortie d'une boîte	\ /
<	< indique le début d'un motif ou d'un méta	" < " or \ <
>	> indique la fin d'un motif ou d'un méta	" > " or \ >
#	# sert à interdire la présence de l'espace	" # "
\	\ sert à déspecialiser la plupart des caractères spéciaux	\ \

- 1 Sources
- 2 Correction des exercices du cours précédent
- 3 Manipuler les transducteurs
- 4 Avant-goût : construire des dictionnaires
- 5 Annexes
- 6 **Pour finir**
  - CQFR : Ce Qu'il Faut Retenir
  - TD



Savoir :

- utiliser des variables
- construire un dictionnaire fléchi de mots simples

# Ajouter des entrées au dictionnaire

## Créer les entrées fléchies des mots suivants

- twittos
- followeur
- dégagisme
- frotteur
- grossophobie