Plate-formes logicielles pour le TAL
Unitex, révisions++ par la pratique

Karën Fort

karen.fort@sorbonne-universite.fr / https://members.loria.fr/KFort/

LETTRES
SORBONNE
UNIVERSITÉ

Quelques sources d'inspiration

- Manuel d'Unitex : http://unitexgramlab.org/releases/3.1/man/Unitex-GramLab-3.1-usermanual-fr.pdf
- Denis Maurel : son tutoriel et ses conseils

Annoter les nombres écrits en lettres

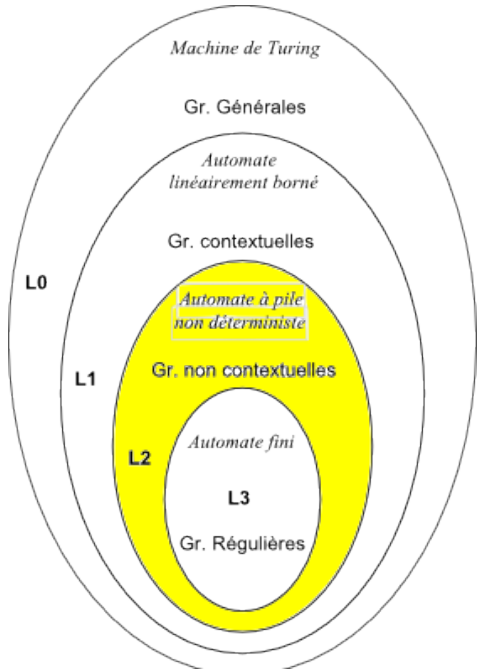
dans *Le Tour du Monde en 80 jours*

Exercice

Créez les graphes permettant d'annoter :

1. *deux* à *dix-neuf* (avec annotation $<$ *nombre* $>$)
2. *deux* à *quatre-vingt-dix-neuf* : utilisez le sous-graphe précédent
3. *deux* à *neuf cent quatre-vingt-dix-neuf*
4. *deux* à 999 999

Rappel : grammaires hors contexte ou algébriques



Grammaires hors contexte

Ce sont des grammaires contextuelles où le contexte est vide, ce qui signifie que les symboles non terminaux sont traités indépendamment de la place où ils apparaissent.
[Wikipédia, Grammaires formelles, consultée le 21/09/2014]

Rappel : grammaires manipulées par Unitex

Grammaires algébriques étendues

Définition

Les grammaires algébriques étendues sont des grammaires algébriques où les membres droits des règles ne sont plus des suites de symboles mais des expressions rationnelles. [Manuel d'Unitex, p. 94]

$S \rightarrow aS$ devient $S \rightarrow a*$
$S \rightarrow \varepsilon$

Les grammaires (ou graphes) d'Unitex intègrent également la notion de transduction (elles peuvent produire des sorties)

Conséquences ?

Exercice

Créez un graphe permettant de reconnaître :
Président, **sauf** dans le contexte de *de la République*

Conséquences ?

Exercice

Créez un graphe permettant de reconnaître :
Président, **sauf** dans le contexte de *de la République*

Grammaires hors contexte !

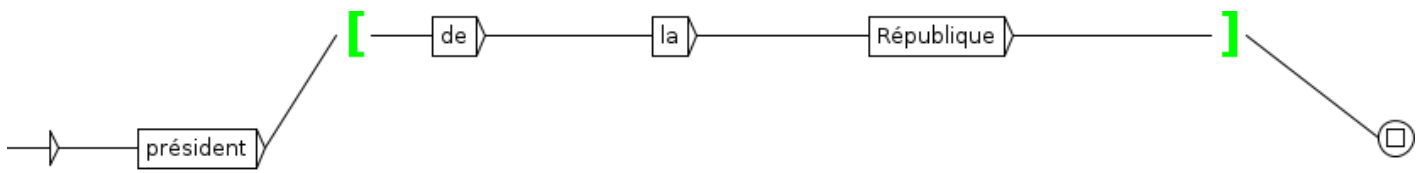
Solution ?

Grammaires contextuelles !

Prendre en compte le contexte (droit)

Contexte droit

- créer un graphe reconnaissant *président de la république*
- ajouter deux états délimitant le contexte
- y écrire : *$[* et *$]* respectivement
- enregistrer le graphe



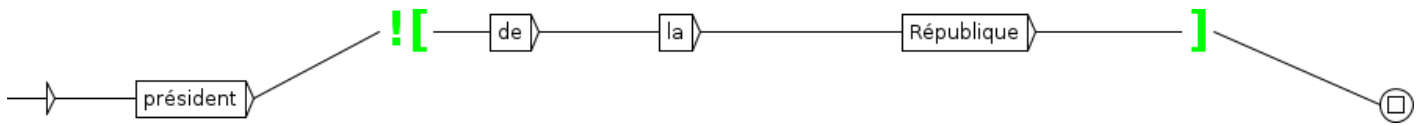
L'appliquer sur un texte que vous créerez, contenant *président* et *président de la République*.

Prendre en compte le contexte (droit)

en négatif

Contexte négatif

- modifier le graphe précédent
- dans l'état d'entrée du contexte, écrire : *$![*
- enregistrer le graphe

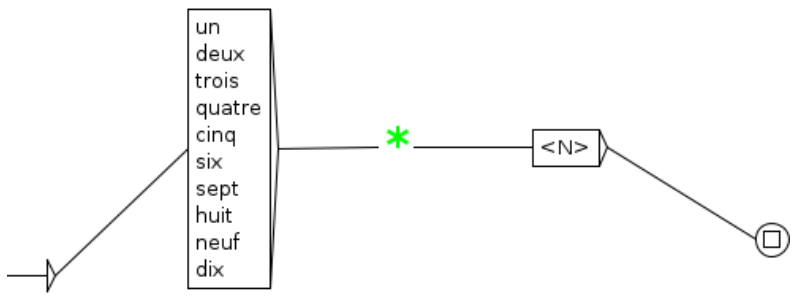


L'appliquer sur le texte créé précédemment.

Prendre en compte le contexte (gauche)

Contexte gauche

- créer un graphe reconnaissant *un* ou *deux* ou... suivi d'un nom
- enregistrer le graphe
- l'appliquer sur le Tour du monde en 80 jours
- modifier le graphe : dans l'état de sortie du contexte gauche, écrire *$\$*$*
- enregistrer le graphe
- l'appliquer sur le Tour du monde en 80 jours



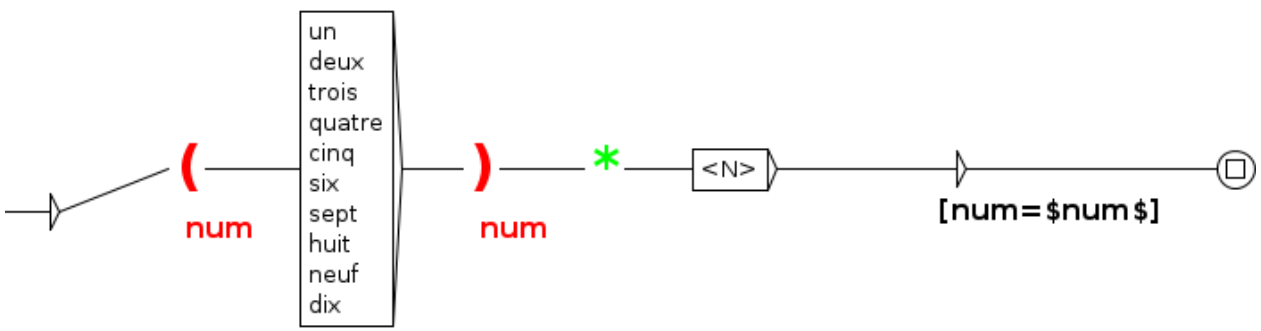
Différence ?

Prendre en compte le contexte (gauche)

et utiliser les variables

Contexte gauche + variable

- modifier le graphe précédent :
  - ajouter une variable `num`
  - fusionner les résultats (afficher l'instance)
- enregistrer le graphe
- l'appliquer sur le Tour du monde en 80 jours



Exercice

Créez un graphe permettant de reconnaître :
un mot constitué du préfixe *in* suivi d'un adjectif en *able*

Exercice

Créez un graphe permettant de reconnaître :
un mot constitué du préfixe *in* suivi d'un adjectif en *able*

Impossible : les graphes Unitex s'appliquent par défaut sur les tokens !

Solution ?

Mode morphologique

- créer un graphe reconnaissant un mot constitué du préfixe *in* suivi d'un adjectif en *able*
- ajouter deux états délimitant le mode morphologique
- y écrire : $\$<$ et $\$>$ respectivement
- enregistrer le graphe

L'appliquer sur un texte que vous créerez, contenant
Il est inacceptable qu'il soit acceptable d'écrire cela.

Graphe morphologique : quelques règles

- les filtres s'appliquent au caractère courant
- les espaces doivent être explicités
- pour utiliser les informations contenues dans un dictionnaire, celui-ci doit être préalablement déclaré comme dictionnaire du mode morphologique
- si vous atteignez la fin de la zone sans être à la fin du token, la reconnaissance échoue

Sources

Des nombres

Un pas en avant, un pas en arrière

Pour finir

CQFR : Ce Qu'il Faut Retenir

- créer des graphes
- ajouter des annotations
- appeler des sous-graphes (modulariser son code)
- utiliser les contextes et les variables
- utiliser le mode morphologique