



Plate-formes logicielles pour le TAL

Unitex, révisions+ par la pratique

Karën Fort

karen.fort@sorbonne-universite.fr / <https://www.schplaf.org/kf/>

25 septembre 2020



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

Sources

Introduction

Présentation

Fonctionnalités

Installation

Avoir suivi d'un participe passé

Pour finir

Unitex

- ▶ <http://unitexgramlab.org/fr/>
- ▶ Université de Marne-la-Vallée
- ▶ licence LGPL (code) et LGPL-LR (ressources linguistiques)
- ▶ Java pour l'interface, C++ dessous (efficacité)
- ▶ Unicode, support de nombreuses langues
- ▶ projet GramLab (2010 à 2013) : surcouche d'Unitex
- ▶ utilisé par :
 - ▶ de très nombreuses universités
 - ▶ des entreprises du TAL (Kwaga, CEA, Sinequa, Systran, ...)

Caractéristiques linguistiques

- ▶ Unitex repose sur l'utilisation de données linguistiques
- ▶ dépendantes des langues
- ▶ types de données :
 - ▶ dictionnaires électroniques
 - ▶ grammaires locales
- ▶ voir :
<http://unitexgramlab.org/fr/language-resources>

Avantage

VOUS avez la main, sur tout.

Ressources

- ▶ site Internet officiel
<http://unitexgramlab.org/fr>
- ▶ site Internet de l'équipe TLN de l'Université de Tours
http://tln.li.univ-tours.fr/Tln_CasEN.html
- ▶ Unitex 3.1 Bêta - Manuel d'Utilisation (Paumier - Martineau 2006)
- ▶ ateliers réguliers
- ▶ liste de diffusion

À quoi ça sert ?

- ▶ recherche de motifs complexes dans des textes
 - ▶ concordance (visualisation des résultats en contexte)
 - ▶ annotation
 - ▶ analyse
- par la création de grammaires locales ou de transducteurs
- *via* une interface graphique

Installation d'Unitex-GramLab 3.1

- ▶ allez sur <http://unitexgramlab.org/fr>
- ▶ enregistrez sur votre disque (pas besoin de droits admin)
- ▶ créez-vous un répertoire perso (ailleurs)
- ▶ lancez Unitex beta
- ▶ Info / Preferences / Encoding : mettre UTF8

Sources

Introduction

Avoir suivi d'un participe passé

- Recherche de motifs

- Annotation

- Création de sous-graphes

Pour finir

Rechercher *avoir*, suivi d'un participe passé

dans *Le Tour du Monde en 80 jours*

Exercice

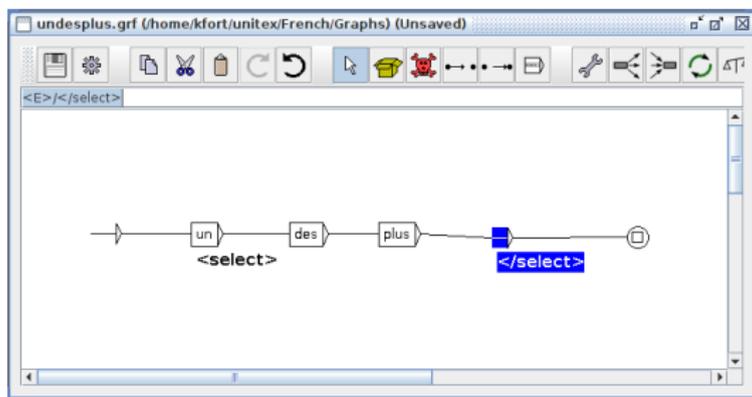
Créez les graphes permettant de rechercher

1. *avoir* suivi d'un participe passé : *a diminué*
2. avec prise en compte du passif : *a été anobli*
3. avec insertion possible d'un adverbe : *a enfin trouvé*

Rappel : création d'une annotation

Annoter une chaîne de caractères

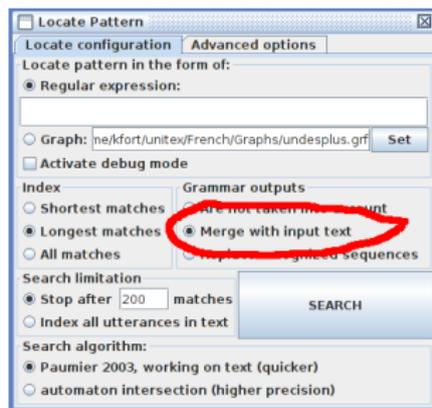
- ▶ écrire : *un*/*< select >*
- ▶ ajouter après l'expression l'état contenant :
< E >/< /select >
- ▶ enregistrer le graphe



Rappel : appliquer les annotations dans le texte

Annoter le texte

- ▶ ouvrir un texte (par exemple le *Tour du monde en 80 jours*)
- ▶ puis, menu Text / Locate Pattern
- ▶ sélectionner Grammar Outputs / Merge with input text



Annoter *avoir*, suivi d'un participe passé (suite)

dans *Le Tour du Monde en 80 jours*

Exercice

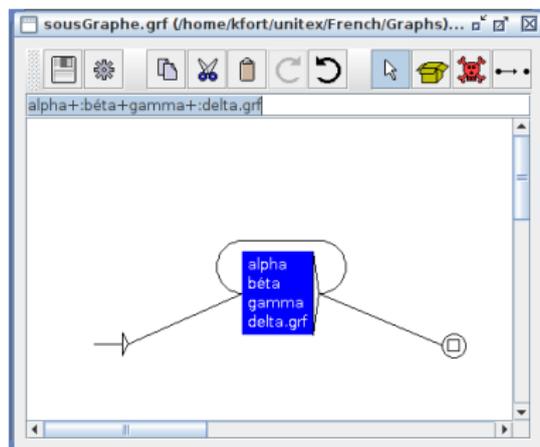
Créez les graphes permettant d'annoter *avoir* suivi d'un participe passé :

1. avec annotation des adverbes : < *adverbe* >< /*adverbe* >
2. avec insertion possible de plusieurs adverbes : *avait si bien marché*

Rappel : faire appel à un sous-graphe

Sous-graphe

- ▶ appeler un sous-graphe
alpha+ :beta+gamma+ :CheminVersFichier delta.grf



Annoter *avoir*, suivi d'un participe passé (suite)

dans *Le Tour du Monde en 80 jours*

Exercice

Créez les graphes permettant d'annoter *avoir* suivi d'un participe passé :

1. avec utilisation de sous-graphes : pour les adverbes

Annoter *avoir*, suivi d'un participe passé (suite 2)

dans *Le Tour du Monde en 80 jours*

Exercice

Créez les graphes permettant d'annoter *avoir* suivi d'un participe passé :

1. avec formes négatives : *qui n'a pas coûté*
 - ▶ sous-graphe Négations
 - ▶ sous-graphe Préverbaux
2. avec formes interrogatives : *avait-elle enfin sonné*
 - ▶ sous-graphe Postverbaux
3. avec formes interronégatives : *n'avait-il pas été entrepris*

Sources

Introduction

Avoir suivi d'un participe passé

Pour finir

CQFR : Ce Qu'il Faut Retenir



- ▶ créer des graphes
- ▶ ajouter des annotations
- ▶ appeler des sous-graphes (modulariser son code)