

Linguistique de corpus

Désambiguïsation sémantique

Patrick Paroubek

LIMSI-CNRS
Dépt. CHM - Groupe LIR
Bât. 508 Université Paris XI, 91403 Orsay Cedex
pap@limsi.fr

mercredi 22 mars 2017 / Semestre 2 - Cours 7

Désambiguïisation sémantique avec les grammaires ELAG en UNITEX

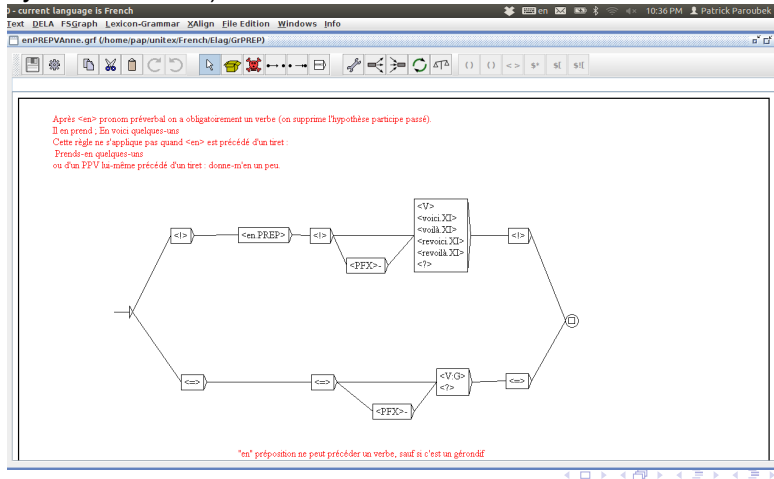
Les automates ELAG ont une syntaxe spéciale, ils sont constituées d'un ensemble de règles composées de deux parties :

- une partie condition (description de la situation ambiguë en contexte avec un patron linguistique)
- une partie conséquence (le choix effectué, c.a.d. le sous-patron linguistique choisi)

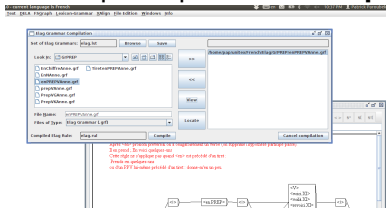
Les deux parties de règle (condition/conséquence) sont synchronisées grâce à des « marqueurs » de deux types (<! > pour la condition et <=> pour la conséquence).

Lorsque l'automate analyse un texte (nécessairement compilé sous-forme d'automate d'état fini avec le menu **File/Construct FST-Text**), si un chemin (avec ses annotations) de la partie « condition » est reconnu, il doit aussi être présent dans la partie « conséquence » pour être préservé, sinon il est éliminé de la sortie (constitué donc d'une version simplifiée du texte sous forme d'automate). On élimine ainsi les ambiguïtés.

Les deux parties d'une règle ELAG (avec ses marqueurs de synchronisation) :



Les règles d'une grammaire ELAG sont éditées à partir d'un menu spécifique : **File/Compile ELAG grammars**



Les étapes de l'utilisation d'une grammaire ELAG :

- 1 charger une texte dans Unitex et le normaliser
- 2 construire l'automate du texte **File/Construct FST-Text**
- 3 éditer une ou plusieurs règles ELAG avec le menu **File/Compile ELAG grammars**
- 4 Compiler la grammaire ELAG (ensemble de règles) ainsi modifié, le résultat est stocké dans un fichier **elag.rul**

Unitex

ELAG

Pour appliquer la grammaire ELAG compilée sur le texte chargé, il faut passer par le menu **Apply ELAG rule** de la fenêtre d'exploration du texte compilé sous forme d'automate

