



Travaux Dirigés d'informatique linguistique n°8

Cours d'Introduction à l'informatique linguistique

—Licence Informatique 3ème année—

Analyse syntaxique par grammaires algébriques

Ce TP est **noté**. Il est consacré à l'analyse syntaxique par grammaires algébriques.

1 Instructions

Ce TP noté dure 2h. Vous devez répondre aux questions ci-dessous sous la forme de scripts python. Les réponses à des questions non implémentables devront être mises en commentaire dans les scripts. Ces scripts devront être envoyés à la fin de la séance à mconstan@univ-mlv.fr. Vous devrez envoyer un compte rendu de quelques pages sur ce TP, pour le jeudi 5 avril 14h, à la même adresse.

2 Ressources

Pour cette séance, nous vous fournissons un certain nombre de ressources se trouvant dans le répertoire suivant :

<http://igm.univ-mlv.fr/~mconstan/enseignement/l3/infolingu/data/>

- `td8-script-exemple.py`, un exemple de script d'analyse syntaxique d'une phrase anglaise par grammaire algébrique,
- `td8-text1.txt`, un texte,
- `td8-text2.txt`, un texte.

Vous pouvez également vous aider des documentations sur NLTK, situées à l'url suivante :

<http://nltk.sourceforge.net/docs.html>

3 Analyse syntaxique par grammaires algébriques

1. Adapter l'exemple `td8-script-exemple.py` au français pour la phrase : "Max aime les grosses pommes en chocolat". Que constatez-vous ? Expliquer.
2. Placer votre texte dans un fichier. Adapter votre programme précédent pour qu'il parse un fichier texte.

3. Modifier dans votre script le tokeniseur de `td8-script-exemple.py` pour qu'il sépare la ponctuation des mots dans le texte du fichier `td8-text1.txt`. Ne passer pas trop de temps sur cette question. Si vous n'y arrivez pas, mettez juste des espaces avant et après les ponctuations.
4. Écrire une grammaire la plus générale possible qui permet d'analyser le fichier texte `td8-text1.txt`. Dans un premier temps, vous préciserez les terminaux (tokens) en ne gardant que leur analyse dans le texte (même si ce token est potentiellement ambigu). Par exemple, pour le token `la`, on n'a qu'une seule production :

```
Det -> la
```

Dans un deuxième temps, vous préciserez les terminaux en tenant compte de l'ambiguïté potentielle. Pour chacune des ambiguïtés, vous donnerez un exemple d'emploi. Pour le token `la`, il existe trois analyses possibles :

```
Det -> la // "la carte de Luc est belle"
```

```
Pro -> la // "je la donne à Paul"
```

```
Nom -> la // ambiguïté un peu trop artificielle "le la sonne faux sur ton piano."
```

Que constatez vous ?

4 Analyse syntaxique par grammaires lexicalisées

1. Construire une grammaire lexicalisée pour chacun des verbes locatifs (`plonger`, `aller`, `venir`) de telle manière que les phrases du fichier `td8-text2.txt` soient analysées.
2. Essayer de représenter d'autres variantes dans votre grammaire.