

## Analyser le français, en toute modestie !

Eric de la Clergerie et André Bittar

4 Décembre 2008

### Introduction

Le but de ce TD est de compléter la grammaire jouet DCG des TD 2 et TD 3, en assumant acquis le passage vers les structures de traits (si nécessaire partir de *french4.dcg* fournie avec ce TD) et en se focalisant maintenant sur :

- la gestion de phrases enchassées (argument phrastique, relatives)
- la gestion des extractions de constituants avec les BMG (pour les relatives et interrogatives), sans oublier les contraintes d’îlots pour bloquer des déchargements aux mauvaises places.
- la construction de formes sémantiques :
  - La sémantique d’une phrase avec sujet sera supposée être de la forme  $E^{\text{Formula}}$  avec  $E$  correspondant à une variable d’évènement.
  - La sémantique d’une phrase sans sujet sera supposée être de la forme  $E^{\text{X}}^{\text{Formula}}$  avec  $E$  correspondant à une variable d’évènement et  $\text{X}$  à une variable d’entité (le sujet implicite).

En final, on doit pouvoir analyser une phrase comme « *Paul mange le gâteau que Claude demande à Marie de couper* » et obtenir :

```
./french8 — Paul mange le gâteau que Claude demande à Marie de couper
Answer:
L = [ Paul , mange , le , gâteau , que , Claude , demande , à , Marie , de , couper ]
Tree = s(gn(lexical{lemma=> Paul , form=> Paul}), gv(lexical{lemma=> manger
, form=> mange}, gn(gn(lexical{lemma=> le , form=> le}, lexical{lemma=>
gâteau , form=> gâteau}), srel(lexical{lemma=> que , form=> que}, s(gn(
lexical{lemma=> Claude , form=> Claude}), gv(lexical{lemma=> demander , form
=> demande}, s(gv(lexical{lemma=> couper , form=> couper}, trace))), gp(
lexical{lemma=> à , form=> à}, gn(lexical{lemma=> Marie , form=> Marie}))))))
Sem = ((_A : manger(Paul , _B) & (def(_B) & gâteau(_B)) & (_C : demander(
Claude , _D, Marie) & (_D : couper(Marie , _B) & true) & true) &
true)
XSem = (_A : manger(Paul , _B) & def(_B) & gâteau(_B) & _C : demander(
Claude , _D, Marie) & _D : couper(Marie , _B))
```

Le matériel pour ce TD peut être récupéré sous forme d’archive .tgz à <http://alpage.inria.fr/~clerger/PrologTAL08/td3.tgz>. Il comprend :

- une ébauche de grammaire DCG pour le français *french4.dcg*, s’appuyant sur des structures de traits et construisant les arbres d’analyse.
- une mini-segmenteur *lex2db* permettant de construire un treilli de mots à partir d’une chaîne.
- un jeu de quelques phrases *sentences.txt*
- un fichier *Makefile*

L’ensemble du jeu de phrases peut être testé en boucle avec la commande suivante (pour une version récente de DyALog) :

```
cat sentences.txt | ./lex2db | ./french1 -loop -utime
```

Ne pas oublier de consulter également le matériel des TD 2 et TD 3.

**NOTE :** Lors de la compilation avec `dyacc`, l'option `-warning` peut être ajoutée pour obtenir des informations sur des prédicats ou non-terminaux utilisés mais non définis, ce qui traduit en général une typo ou une erreur d'arité sur le nombre d'arguments.