



Session : 1

Durée de l'épreuve : **2 heures**

Date : **17 mai 2016 16:00**

Documents autorisés : **tous**

Master mention informatique 2^e année : Technologies de la langue **HMIN206**

Sujet : **2 pages**

Matériel utilisé : **aucun**

Chacune des trois parties 1 2 3 sera notée sur 7 (soit environ 40mn par partie)
On rendra une copie par partie

1 Grammaires Forme Normale de Chomsky (type 2) et grammaires de type 0 (V. Prince)

1.1

Ecrivez la **plus petite** grammaire en forme normale de Chomsky qui soit capable de générer le mot "x a b a b a x".

Vous appelez votre axiome S (S appartient au vocabulaire non terminal). $\{x, a, b\}$ est une partie de votre vocabulaire terminal.

Vous montrez que votre grammaire génère bien le mot en créant l'arbre dont les feuilles correspondent aux symboles du mot, et la racine est S.

Vous montrez que c'est la plus petite grammaire possible.

1.2

Proposez une grammaire (quelconque) d'analyse de la phrase suivante :

Le petit chat gris mange sa pâtée dans une assiette rouge.

Votre grammaire devra tenir compte des rôles que joue chaque ensemble de mots par rapport aux autres, c'est-à-dire qu'est-ce qui est sujet, qu'est-ce qui est complément, et de quoi.

Votre résultat doit être une structure parenthésée correspondant à une arborescence.

Par exemple :

" Je prends le premier train pour Toulouse."

pourrait donner lieu à une structure du type :

PH (PRNSUJ (Je) V (prendre) GNPREPOBJ (GNADJ (le premier train) PREP (pour) NOMP (Toulouse))))

PH : phrase (axiome) ; PRNSUJ : pronom sujet ; GNPREPOBJ : groupe nominal prépositionnel objet ; GNADJ : groupe nominal adjectival.

Cette structure est celle d'une grammaire non FN Chomsky, qui mêle les dépendances (sujet, objet), les constituants (GN...) et les catégories grammaticales (V, NOMP, ...).

Une autre solution pourrait être :

PH (PRNSUJ (Je) GV (prendre GNADJOBJ (le premier train)) GNPREP(pour Toulouse))))

qui montre un autre attachement. Pour proposer votre grammaire, il faut :

- donner un jeu d'étiquettes grammaticales (ex : DET, V, NOM, ADJ, etc...)
- donner un jeu d'étiquettes syntaxiques
- écrire les règles de grammaire correspondantes (format libre)
- écrire la structure parenthésée correspondante
- expliquer pourquoi vous avez choisi telle structure plutôt que telle autre.

suite du sujet au verso ...

2 HMM et DCG (Ch. Retoré)

2.1 HMM

On souhaite définir un Modèle de Markov caché (HMM) avec les états "déterminant, nom, adjectif, verbe, pronom" pour étiqueter des phrases construites sur le vocabulaire figurant dans les phrases suivantes, l'étiquetage étant déterminé par le chemin de Viterbi. [Il est assez ridicule de faire des statistiques et des probabilités sur si peu d'exemples, mais cet exercice a pour but de vérifier si vous avez compris tout en évitant de longs calculs.]

1. Ecrire le HMM avec les probabilités induites par les deux exemples suivants :

- (1) Un regard trouble le charme.
det nom adj pro vb
- (2) Son charme trouble le regard.
det nom vb det nom

2. Quel étiquetage votre HMM attribue-t-il à la phrase suivante ?

- (3) Son regard trouble le charme.

On justifiera sa réponse de préférence en utilisant l'algorithme de Viterbi vu en cours. Une justification directe et correcte est aussi admise, mais elle ne donnera que la moitié des points dévolus à cette question.

2.2 DCG

On considère le langage engendré par l'expression régulière suivante.

Je connais ((des | un) (amis | ami) qui (connaissent | connaît)) Donald.*

1. Donner trois exemples de phrases produites, de longueurs différentes.
2. Donner une grammaire de clauses définies (DCG) qui produise ces phrases en les accordants correctement en nombre. On pourra commencer par écrire une grammaire hors-contexte (algébrique) qui reconnaît ces phrases sans tenir compte de l'accord.

3 Morphologie, sémantique lexicale (M. Lafourcade)

Votre réponse à cette partie de devra pas dépasser 2 pages.

3.1

Expliquer l'intérêt de disposer d'un outil pouvant effectuer des dérivations morphologiques

3.2

Décrivez une stratégie (voire un algorithme) permettant de déduire des informations sémantiques à partir d'un terme et de sa forme. Vous pouvez supposer que vous disposez d'une base de connaissance.