

Université Bordeaux 1 — Master Informatique
UE INF 316
Traitement Automatique des Langues
Examen bis¹ du lundi 22 mai, 9h – 12h
Sujet de 4 pages
documents autorisés / calculatrice bienvenue

Exercice A Grammaires, automates et expressions régulières

On considère la grammaire suivante, G dont les terminaux sont *anne*, *bernard*, *regarde*, *qui* et dont les non-terminaux sont S , SN , SV , NP , VS , et on note $L(G)$ le langage qu'elle engendre.

$$\begin{aligned} S &\rightarrow NP VT NP REL \\ REL &\rightarrow que VT NP REL \\ REL &\rightarrow \varepsilon \\ NP &\rightarrow anatole \\ NP &\rightarrow brigitte \\ VT &\rightarrow regarde \end{aligned}$$

- (A.i) Donner l'arbre de dérivation d'une phrase de $L(G)$ ayant neuf mots.
- (A.ii) Quelle est le type de G (régulière, hors-contexte, contextuelle, générale)?
- (A.iii) Donner un automate qui engendre $L(G)$.
- (A.iv) Quelle est le type le plus spécifique d'une grammaire engendrant $L(G)$ parmi ceux-ci: régulière, hors-contexte, contextuelle, générale?
- (A.v) Donner une grammaire équivalente du type le plus petit possible et sans production vide.
- (A.vi) Les relatives introduites par "que" ont ici un sujet situé après le verbe. Ecrire une grammaire G' qui produise les même phrase que G à ceci près que les relatives introduites par "que" auront toujours un sujet situé à sa place habituelle, c'est-à-dire avant le verbe.
- (A.vii) Quel est le type de G' ?
- (A.viii) Quel est le type le plus spécifique d'une grammaire produisant $L(G')$? Justifiez votre réponse.

1. En remplacement de celui du 9 mai pour ceux ayant eu un examen en même temps.

Exercice B Transducteurs

On simplifie (de beaucoup) la formation du pluriel des noms français ainsi: on forme le pluriel d'un nom en ajoutant un "s" sauf si le mot se finit par "al" auquel cas "al" devient "aux".

(B.i) Ecrire un transducteur qui, étant donné, un nom français au singulier (par exemple *cheval+S* ou *chiens+S*) rende ce nom au pluriel (*chevaux+P* ou *chiens*) et qui ne modifie pas les mots que ne se finissent pas par +S.

Exercice C Modèles de Markov Cachés

Le tableau ci-dessous donne les catégories syntaxiques des mots *le*, *la*, *sceau*, *marque*, *sublime*.

	article	adjectif	nom	pronom	verbe
la	x			x	
le	x			x	
sceau			x		x
sublime		x	x		x
marque			x		x

On suppose que

- une phrase ne peut commencer que par un article,
- une phrase ne peut pas finir par un pronom, ni par un article,
- les seules successions possibles de catégories sont les suivantes:
 - article → adjectif
 - article → nom
 - adjectif → nom
 - adjectif → verbe
 - adjectif → pronom
 - nom → adjectif
 - nom → verbe
 - nom → pronom
 - pronom → verbe
 - verbe → article

(C.i) Construire un Modèle de Markov Caché (HMM) avec état initial et état final en supposant qu'il y a équi-probabilité entre les transitions possibles et que la probabilité d'émettre chaque mot, pour une catégorie donnée, est la même.

(C.ii) Pour la phrase suivante, trouver tous les étiquetages syntaxiques différents et grammaticalement corrects (il y en a au moins deux) :

(1) le sceau sublime la marque

(C.iii) Calculer la probabilité de l'étiquetage où "la" est un article (algorithme forward).

(C.iv) Quel est l'étiquetage donné par le Modèle de Markov Caché? (chemin de Viterbi)

(C.v) Réestimer les paramètres par la méthode de Viterbi à partir des exemples suivants:

- (2) Le sceau sublime la marque. (déjà traité)
- (3) Le sublime la marque.
- (4) La marque sublime le sceau.