

Université Bordeaux 1, Master Informatique première année
Introduction au Traitement Automatique des Langues (C. Retoré)
Examen du 24 mai 2004, 14h-15h30

Le sujet est sans doute un peu long. On essaiera de traiter au moins une partie de chacun des trois exercices.

1. On considère la grammaire suivante, dont les Non-Terminaux sont S, SN, B, VI et les terminaux les mots.

S → SN B que S | SN VI
B → croit | pense
VI → dort | court | mange | rit
SN → Amélie | Benoît | Céline | Didier

- (a) Quel est le type de cette grammaire?
- (b) Donner deux exemples de phrases reconnues dont un contenant trois verbes.
- (c) Existe-t-il une grammaire régulière équivalente? Si oui, donner une expression régulière décrivant ce langage.
- (d) Mettre cette grammaire en forme normale de Chomsky et analyser votre phrase à trois verbes avec l'algorithme de Cocke Younger Kasami.
- (e) On rajoute la règle

SN → SN et SN

et on modifie les règles de B et VI en conséquence

B → croit | pense | croient | pensent
VI → dort | court | mange | rit | dorment | courent | mangent | rient

Ecrire une grammaire de clauses définies (DCG) qui soit équivalente à la grammaire ainsi modifiée, et qui de plus vérifie que les verbes (de type B ou VI) s'accordent en nombre avec leur sujet.

-
2. En italien, la grande majorité des noms se finit par l'une des voyelles *o*, *a*, *e*, *i*. Le pluriel des noms en *o*, *e*, *i* est en *i*, celui des noms en *a* est en *e*. Si la fin du nom est *co*, *ca*, *ga* ou *go*, au pluriel un *h* est intercalé entre le *c* ou le *g* et le *i* ou le *e* du pluriel. Exemples:

singulier → pluriel
libro → *libri*
casa → *case*
ponte → *ponti*
banco → *banchi*
barca → *barche*
albergo → *alberghi*
alga → *alghe*

- (a) Ecrire un transducteur qui étant donné une entrée de la forme *nom_au_singulier+p#* donne le pluriel du nom. On pourra considérer une étape intermédiaire et écrire deux transducteurs, auquel cas on utilisera le symbole « ^ » pour séparer les morphèmes.
- (b) Si on a le temps on raffiner ce(s) transducteur(s) pour prendre en compte les noms en *cia* et *gia* qui ont leur pluriel en *ce* ou *ge* si la lettre avant est une consonne et en *cie* ou *gie* si la lettre avant le *a* est une voyelle. Exemples:
- | | | |
|-----------------|---|----------------|
| singulier | → | pluriel |
| <i>doccia</i> | → | <i>docce</i> |
| <i>frangia</i> | → | <i>frange</i> |
| <i>camicia</i> | → | <i>camicie</i> |
| <i>ciliegia</i> | → | <i>cilegie</i> |

3. On considère le modèle de Markov caché (HMM) suivant:

ETATS: INIT DET NOM ADJ VERBE PRO FIN

PROBABILITÉS DE PASSER D'UN ÉTAT À L'AUTRE (L'état inscrit dans la première colonne est celui de départ, et celui en haut de la colonne sur la première ligne celui d'arrivée.)

→	INIT	DET	NOM	ADJ	PRO	VERBE	FIN
INIT	0	1	0	0	0	0	0
DET	0	0	0,5	0,5	0	0	0
NOM	0	0	0	0	0,4	0,4	0,2
ADJ	0	0	0,8	0	0	0	0,2
PRO	0	0	0	0	0	1	0
VERBE	0	0,5	0	0	0	0	0,5

PROBABILITÉ D'ÉMETTRE UN MOT LORSQU'ON EST DANS UN ÉTAT DONNÉ. Les états INIT et FIN n'émettent rien et les mots omis d'un état sont émis avec une probabilité 0.

DET	NOM	ADJ	PRO	VERBE
la:1	petite:0,2	petite:1	la:0,6	brise:0,5
	glace:0,4		se:0,4	glace:0,5
	brise:0,4			

- (a) Dessiner le HMM.
- (b) Y a-t-il plusieurs chemins produisant la chaîne $A=La\ petite\ brise\ la\ glace?$ A quelle(s) analyse(s) correspondent-ils?
- (c) Trouver la probabilité d'observer la chaîne A.
- (d) Quel le chemin le plus probable de la chaîne A? En déduire l'étiquetage grammatical de A par ce HMM.
- (e) Trouver les chemins de la chaîne $B=La\ brise\ la\ glace$ et donner le plus probable.
- (f) Modifier les probabilités du HMM de sorte que soient améliorées les probabilités des deux exemples A et B le long de leur chemin le plus probable.
- (g) Laquelle des deux possibilités de A va être renforcée par la modification des probabilités? (sans calcul)