Cartouche du document

Année : ING 1 - Matière : Théorie des langages - Activité : Travail dirigé

Objectifs

- Notions de langage hors contexte ou algébrique
- Normalisation de Chomsky
- Algorithme CKY
- Application au traitement des langues

Définition :

Un langage hors-contexte est aussi appelé langage algébrique.

Une grammaire hors-contexte (ou algébrique) est un quadruplet T,N,S,P où :

- T : ensemble des éléments terminaux
- N : ensemble des éléments non terminaux
- S : élément non terminal initial (axiome)
- P : ensemble de règles de la forme :
  - X —> a où $a \in T$ et $X \in N$
  - X —> Y où $Y \in (N \cup T)^*$ et $X \in N$

Définition :

Une grammaire hors-contexte **qui ne produit pas** ε est dite sous forme normale de Chomsky si et seulement si toutes les règles sont de la forme :

A —> a où $a \in T$

A —> B C où $B, C \in N$

On rappelle que ε (le mot vide) peut être également noté λ (c'est le cas par défaut en JFLAP).

Sommaire des exercices

Corps des exercices

1 - Reconnaissance d'un mot par l'algorithme CKY

Énoncé :

Il s'agit dans l' exercice suivant de :

- Transformer une grammaire sous Forme Normale de Chomsky.
- Définir un algorithme qui teste l'appartenance d'un mot à cette grammaire normalisée.

Question 1)

Énoncé de la question

Soit la grammaire G basée sur l'alphabet {a, b}.

On a :

G = {

T = {a,b}

N = {S,A,B}

S = S

P = {

(1) S —> b A

(2) S —> a B

(3) A —> b A A

(4) A —> a S

(5) A —> a

(6) B —> a B B

(7) B —> b S

(8) B —> b

}

}

Mettre cette grammaire sous forme normalisée de Chomsky.

Procédure de transformation :

- 1. Remplacer tous les terminaux x en partie droite des règles par des non terminaux en ajoutant des règles de la forme X —> x
- 2. Transformer les parties droites des règles comme suit : X —> Y Z W par deux régles

  X —> Y V et V —> Z W
- 3. Transformer les parties droites des règles comme suit : X —> Y par X —> W Z si Y —> W Z

Question 2)

Énoncé de la question

Montrer que le mot **aabbab** appartient au langage engendré par la grammaire.

On utilisera l'algorithme de Coke, Younger et Kasami (CKY) qui permet de tester si un mot w est reconnu par une grammaire sous forme normale de Chomsky.

- On note n la longueur du mot w : n = |w|.
- On définit v une matrice de dimensions (n,n).

Le pseudo-code de l'algorithme CKY est le suivant :

Pour i = 1 à n
DEBUT
    v[i,1] = {A tel que A est le membre gauche d'une règle A —> a et a est le $i^{ème}$ symbole du mot w}
FIN
Pour j = 2 à n
DEBUT
    Pour i = 1 à n - j + 1
    DEBUT
        v[i,j] = ∅
        Pour k = 1 à j - 1
        DEBUT
            v[i,j] = v[i,j] ∪ {A tel que A est le membre gauche d'une règle A —> B C avec B ∈ v[i,k]
            et C ∈ v[i+k,j-k]}
        FIN
    FIN
FIN

Le mot w est reconnu par la grammaire <==> S ∈ v[1,n].

2 - Des propositions très relatives

Énoncé :

On s'intéresse aux constructions de phrases avec des propositions relatives.
On considère la grammaire hors contexte suivante :

G ={
T={que,qui,regarde,regardent,mange,mangent,dort,dorment,tombe,tombent,une,un,la,le,des,les,pommes, pomme,femme,femmes,Pierre,Marie}
N={s,sn,reln,rela,sv,proa,pron,vt,vi,det,n,np}
S=s
P={

(1) s —> sn sv

(2) sn —> det n reln | det n rela | np reln | np rela

(3) reln —> pron sv

(4) rela —> proa sn vt

(5) sn —> det n

(6) sn —> np

(7) sv —> vi | vt sn

(8) proa —> que

(9) pron —> qui

(10) vt —> regarde | regardent | mange | mangent

(11) vi —> dort | dorment | tombe | tombent

(12) det —> une | un | la | le | des | les

(13) n —> pommes | pomme | femme | femmes

(14) np —> Pierre | Marie

}

Légende :

- reln <==> proposition relative nominative
- rela <==> proposition relative accusative
- proa <==> pronom relatif accusatif
- pron <==> pronom relatif nominatif
- vi <==> verbe intransitif
- vt <==> verbe transitif
- det <==> déterminant
- n <==> nom commun
- np <==> nom propre

Question 1)

Énoncé de la question

Vérifier l'appartenance de la phrase **Une pomme que Pierre regarde tombe** au langage reconnu par la grammaire en utilisant un arbre syntaxique.

Question 2)

Énoncé de la question

Appliquer l'algorithme CKY pour vérifier :

- l'appartenance de la phrase **Une pomme que Pierre regarde tombe** au langage engendré par la grammaire
- le rejet de la phrase **Une pomme qui Pierre regarde tombe** du langage engendré par la grammaire

3 - JFLAP

Énoncé :

L'analyse CKY peut être effectuée par le logiciel JFLAP (http://jflap.org/jflaptmp).

Pour celà, on se place en mode **grammaire** (*Grammar*). Ensuite, on rentre les différentes règles de la grammaire non normalisée.

On peut alors vérifier qu'il s'agit bien d'une grammaire de type 2 (hors-contexte / *context-free*) grâce à l'item *Test for Grammar Type* de l'onglet *Test*.

On peut ensuite normaliser la grammaire grâce à l'item (*Transform Grammar*) de l'onglet *Convert*.
On peut ensuite réaliser l'analyse CKY grâce à la commande *CYK Parse* de l'onglet *Input*.