Théorie des Langages - EISTI - ING 1

Yannick Le Nir

Ecole Internationale des Sciences du Traitement de l'Information

yannick.lenir@eisti.fr

La classification de Chomsky

Langages décidables et hiérarchie de classes

- Les langages de type 3 : rationnels ou réguliers
- Les langages de type 2 : algébriques ou hors contexte
- Les langages de type 1 : sensibles au contexte
- Les langages de type 0 : tous les autres décidables

Grammaire de type 3

Définition

Une grammaire $G =< T, N, S, P >$ est de type 3 si les règles de production sont de la forme :
$A \rightarrow a$ où $A \in N$ et $a \in T^*$ $A \rightarrow B\ a$ (ou $a\ B$) où $A, B \in N$ et $a \in T^*$

Langage associé

Un langage est de type 3 s'il peut être engendré par une grammaire de type 3.

Utilisation des grammaires de type 3

Domaines

- Occurrence de motifs dans une chaîne (Recherche d'informations)
- Expressions régulières (Shell, C, emacs)
- Séquence de l'ADN (Génôme)
- Apprentissage de grammaires pour l'IA

Exemple de grammaire

Affectation numérique

La grammaire `G = <T, N, S, P>` du cours précédent pour l'analyse lexicale de l'affectation numérique est de type 3 :

- $T = lettre \cup chiffre \cup \{+, -, *, /, =, .\}$
- $N = \{mot, nombre, operateur, nombre, identifiant\}$
- $S = mot$
- $P =$
$$\left\{ \begin{array}{l} mot \longrightarrow operateur, \\ operateur \longrightarrow +|-|*|/|=, \\ mot \longrightarrow nombre, \\ nombre \longrightarrow (chiffre^+)|(chiffre^+).(chiffre^+), \\ mot \longrightarrow identifiant, \\ identifiant \longrightarrow lettre\ |identifiant(lettre\ |chiffre) \end{array} \right\}$$

Grammaire de type 2

Définition

Une grammaire $G =< T, N, S, P >$ est de type 2 si les règles de production sont de la forme :
$A \rightarrow \alpha$ où $A \in \mathcal{N}$ et $\alpha \in (N \cup T)^*$

Langage associé

Un langage est de type 2 s'il peut être engendré par une grammaire de type 2.

Grammaires de type 2

Exemple de grammaire

Le fameux langage $a^n b^n$ peut être engendré par la grammaire hors contexte suivante :

- $T = \{a, b\}$
- $N = \{S\}$
- $S = S$
- $P = \left\{ \begin{array}{l} S \rightarrow a\ S\ b, \\ S \rightarrow a\ b \end{array} \right\}$

Domaines d'application

- Langages de programmation
- La plupart des constructions des langues naturelles

Autres grammaires

Grammaires de type 1

Une grammaire $G =< T, N, S, P >$ est de type 1 si les règles de production sont de la forme :
$u\ A\ v \rightarrow u\ w\ v$ où $A \in N$, $u, v \in T^*$ et $w \in (N \cup T)^*$

Grammaires de type 0

Une grammaire $G =< T, N, S, P >$ est de type 0 si les règles de production sont quelconques.

Théorèmes

- Une grammaire n'engendre qu'un seul langage. La réciproque est fausse.
- Les différentes familles de langages sont incluses les unes dans les autres : $L_3 \subset L_2 \subset L_1 \subset L_0$

Forme normale de Chomsky

Définition

Une grammaire de type 2 (sans ε) est dite sous forme normale de Chomsky si et seulement si toutes les règles sont de la forme :

$$A \rightarrow a \text{ (où } a \in T)$$
$$A \rightarrow BC \text{ (où } B, C \in N)$$

Théorème

Tout langage hors-contexte sans ε peut être engendré par une grammaire en forme normale de Chomsky.

Transformation en forme normale de Chomsky

Algorithme

1. remplacer tous les terminaux x en partie droite des règles par des non-terminaux X en ajoutant les règles $X \rightarrow x$
2. Toute règle $X \rightarrow YZW$ est remplacée par $X \rightarrow YV$ et $V \rightarrow ZW$
3. remplacer les règles $X \rightarrow Y$ par $X \rightarrow WZ$ si $Y \rightarrow WZ$

Exemple

Le langage $a^n b^n$ peut être engendré par la grammaire sous forme normale de Chomsky
suivante :$< \{a, b\}, \{A, B, S\}, S, P >$ avec

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, S \rightarrow A\ X, X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Forme normale de Chomsky

- La Forme Normale de Chomsky permet l'obtention d'un algorithme efficace pour reconnaître l'appartenance à un langage hors-contexte.
- Cet algorithme est du à Cocke, Kasami et Younger (CKY).

Idée de l'algorithme

- G : grammaire sous forme normale de Chomsky
- w : mot de longueur n
- Question : $w \in L(G)$?
- Chercher pour chaque sous-mot m de w, $X \in N$ tels que $X \rightarrow^* m$
- A la sortie, vérifier si le symbole axiome S fait partie des non-terminaux obtenus.

Analyse CKY

Construction

- Tableau triangulaire : colonnes numérotées par $i = 1, 2, ..., n$ (les positions de début de mot dans w) et lignes par $j = 1, 2, ..., n$ (les longueurs possibles).
- Remplissage de la ligne 1 puis 2 ...
- La $p$ième case de la ligne k correspond aux non-terminaux qui peuvent engendrer $w[p, k]$.
- On examine donc toutes les coupures du mot $w[p, k]$ en deux sous-mots : $w[p, l]$ et $w[p + l, k - l]$.
- Pour tous les Y possibles et tous les l compris entre 1 et k, on regarde toutes les expressions X_p X_{p+l} en recherchant s'il existe Y dans la grammaire tel que $Y \rightarrow X_p X_{p+1}$.

Analyse CKY

Déroulement du CKY

4				
3				
2				
1				
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3				
2				
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3				
2	$\emptyset$			
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3				
2	$\emptyset$	S		
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3				
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \to A\ B, \\ S \to A\ X, \\ X \to S\ B, \\ A \to a, \\ B \to b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3				
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3	$\emptyset$			
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3	$\emptyset$			
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3	$\emptyset$	X		
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3	∅	X		
2	∅	S	∅	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4				
3	$\emptyset$	X		
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4	$\emptyset$			
3	$\emptyset$	X		
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4	$\emptyset$			
3	$\emptyset$	X		
2	$\emptyset$	S	$\emptyset$	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \to A\ B, \\ S \to A\ X, \\ X \to S\ B, \\ A \to a, \\ B \to b \end{array} \right\}$$

Analyse CKY

Déroulement du CKY

4	S			
3	∅	X		
2	∅	S	∅	
1	A	A	B	B
input	a	a	b	b

$$P = \left\{ \begin{array}{l} S \rightarrow A\ B, \\ S \rightarrow A\ X, \\ X \rightarrow S\ B, \\ A \rightarrow a, \\ B \rightarrow b \end{array} \right\}$$

Analyse CKY

Algorithme

```
Pour i = 1 à n faire :
  V[i, 1] <-- {A; A -> a est une règle
               et le ième symbole de w est a}
Pour j = 2 à n faire :
  pour i = 1 à n - j + 1 faire
   V[i, j] <-- Ø ;
   pour k = 1 à j - 1 faire
    V[i, j] <--- V[i, j] union
    {A ; A -> BC est une règle, B appartient à V[i, k]
                        et C appartient à V[i+k, j-k]}
```

Analyse CKY

Complexité

- n^2 fois une action consistant à couper en deux un mot d'au plus n lettres
- n coupures à effectuer à chaque fois
- Complexité de l'algorithme de l'ordre de n^3 par rapport à la longueur n du mot d'entrée.
- Les langages hors-contexte peuvent être reconnus en un temps polynomial.