Théorie des Langages

Yannick Le Nir Gaspard Férey

CY tech

ING1

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

Chaque ensemble est strictement inclus dans ceux de numéro inférieur.

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 :

- soit $A \to a$ où $A \in N$ et $a \in T \cup \{\varepsilon\}$
- soit $A \to B\ a$ (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 :

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

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 $A \in N$ et $\alpha \in (N \cup T)^*$

Le membre gauche est donc **exactement un** non terminal alors que le membre droit est quelconque.

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 $G =< T, N, S, P >$ avec :

- $T = \{a, b\}$
- $N = \{S\}$
- $S = S$
- $P = \{\ S \rightarrow a\ S\ b \mid a\ b\ \}$

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 :
$u\ A\ v \to u\ w\ v$ 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 :

- soit $A \to a$ (où $a \in T$)
- soit $A \to BC$ (où $B, C \in N$)

Théorème

Tout langage hors-contexte (de type 2) 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 $X \to x$
2. Toute règle $X \to YZW$ est remplacée par $X \to YV$ et $V \to ZW$
3. remplacer les règles $X \to Y$ par $X \to WZ$ si $Y \to WZ$

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

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

Analyse CKY

Algorithme CKY (Cocke, Kasami et Younger)

- Découvert indépendamment par Cocke, Kasami et Younger.
- Algorithme efficace pour reconnaître l'appartenance à un langage hors-contexte produit par une grammaire sous forme normale de Chomsky.

Algorithme CKY

Problème et algorithme

- Données :
  - G une grammaire sous forme normale de Chomsky
  - w : mot de longueur n
- Question : w appartient-il à $L(G)$?
- Idée :
  - Pour chaque sous-mot m de w ($m = w[i,j]$), on peut construire successivement

$$M_{i,j} = \{X \in N \ / \ \text{tels que } X \rightarrow^* m\}$$

  - Il suffit alors de vérifier si le symbole axiome S fait partie des non-terminaux obtenus pour le mot entier

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 $Y \to X_p X_{p+l}$.

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	∅	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	$\emptyset$			
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

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

Analyse CKY

Déroulement du CKY

4				
3	$\emptyset$	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

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

Analyse CKY

Déroulement du CKY

4	$\emptyset$			
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	$\emptyset$	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

Algorithme

Données : $G =< T, N, P, S >$ et w un mot de T^*

$n \leftarrow |w|$

$V \leftarrow \text{matrix}(n, n)$

// Sous-mots d'une lettre

pour $i = 1$ **à** n **faire**

- $V[i, 1] = \{A \mid A \to a \in P \text{ et } w[i] = a\}$

// Sous-mots de plusieurs lettres

pour $j = 2$ **à** n **faire**

- **pour** $i = 1$ **à** $n - j + 1$ **faire**
  - $V[i, j] \leftarrow \emptyset$
  - **pour** $k = 1$ **à** $j - 1$ **faire**
    - $V[i, j] = V[i, j] \bigcup \{A \mid A \to BC \in P,\ B \in V[i, k] \text{ et } C \in V[i + k, j - k]\}$

w reconnu par G ssi $S \in V[1, n]$.

Analyse CKY

Complexité

- n^2 coupures 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.

Analyse CKY

JFLAP

L'algorithme est illustré dans JFLAP 7.0 et (en plus détaillé) dans JFLAP 8.0 .