LIF15 – Théorie des langages formels

Sylvain Brandel

2019 – 2020

sylvain.brandel@univ-lyon1.fr

CM 8

GRAMMAIRES

Grammaires

- Exemple
  - $L = a(ab \cup ba)^*b$.
  - Mots générés : un *a* suivi de *ab* ou *ba* un certain nombre de fois suivi d'un *b*
  - Un mot de L peut naturellement être décomposé en un début, un milieu et une fin

$$S \to aE$$
$$E \to abE$$
$$E \to baE$$
$$E \to b$$

  - Génération

$$S \to aE \to aabE \to aabbaE \to aabbab$$
$$S \to aE \to abaE \to ababaE \to ababaabE \to ababaabb$$

Grammaires

- Grammaire algébrique, ou hors-contexte (ang. *Context-free*)
  - Un ensemble de symboles terminaux à partir desquels sont construits les mots du langage (*a* et *b* dans l'exemple)
  - Un ensemble de symboles non terminaux (S et E dans l'exemple) parmi lesquels on distingue un symbole particulier (souvent S pour *Start*)
  - Un ensemble fini de règles ou production de la forme

    symbole non terminal $\rightarrow$ suite finie de symboles terminaux et / ou non terminaux

- Grammaire contextuelle (ang. *Context-sensitive*)
  - Dans les règles, le symbole non terminal est entouré de deux mots appelés le contexte

- Grammaire générale
  - Pas de restriction sur les règles

Grammaires

- Grammaire algébrique : quadruplet $G = (V, \Sigma, R, S)$ où
  - Σ est un ensemble fini de symboles terminaux appelé alphabet
  - V est un ensemble fini de symboles non terminaux tels que $V \cap \Sigma = \emptyset$
  - $S \in V$ est le symbole initial
  - $R \subset V \times (V \cup \Sigma)^*$ de la forme $A \to w$

Les éléments de R sont appelés règles ou productions

- Grammaire contextuelle
  - $R \subset (V \cup \Sigma)^* V (V \cup \Sigma)^* \times (V \cup \Sigma)^*$ de la forme $uAv \to uwv$

- Grammaire générale
  - $R \subset (V \cup \Sigma)^* V (V \cup \Sigma)^* \times (V \cup \Sigma)^*$ de la forme $z \to w$

Grammaires algébriques

- Dérivation directe

  Soient u et v $\in (V \cup \Sigma)^*$

  On dit que v dérive directement de u, et on note $u \Rightarrow_G v$,

  ssi $\exists$ x, y, w $\in (V \cup \Sigma)^*$, $\exists A \in V$ tels que

  $u = xAy$ et $v = xwy$ et $A \to w \in R$

- Exemple

  En utilisant la grammaire G définie par les règles suivantes :

  $S \to aE$

  $E \to abE \mid baE \mid b$

  on obtient $aabE \Rightarrow_G aabbaE$ par application de la règle $E \to baE$

- Dérivation

  La relation $\Rightarrow_G^*$ est la fermeture réflexive transitive de la relation $\Rightarrow_G$

Grammaires algébriques

- Dérivation

Soient u et v $\in (V \cup \Sigma)^*$

On dit que v dérive de u, et on note $u \Rightarrow_G^* v$

ssi $\exists\, w_0, \ldots, w_n \in (V \cup \Sigma)^*$ tels que

$u = w_0$ et $v = w_n$ et $w_i \Rightarrow_G w_{i+1}\ \forall\, i < n$

La suite $w_0 \Rightarrow_G w_1 \Rightarrow_G \ldots \Rightarrow_G w_n$ est appelée une dérivation

(lorsqu'il n'y a pas d'ambiguïté, on peut omettre la référence à la grammaire G dans les symboles $\Rightarrow_G$ et $\Rightarrow_G^*$)

La valeur de n ($n \geq 0$) est la longueur de la dérivation

Grammaires algébriques

- Langage engendré

  Soit $G = (V, \Sigma, R, S)$ une grammaire algébrique

  Le langage engendré par G, noté L(G), est :

  $$L(G) = \{ w \in \Sigma^* \mid S \Rightarrow_G^* w \}$$

- Langage algébrique
  - Un langage est dit algébrique s'il peut être engendré par une grammaire algébrique

Grammaires algébriques

- Exemple

  $G = (V, \Sigma, R, S)$ avec :
  - $V = \{ S, E \}$
  - $\Sigma = \{ a, b \}$
  - $R = \{ S \rightarrow EE,\ E \rightarrow EEE \mid bE \mid Eb \mid a \}$

- La chaîne ababaa peut être dérivée de plusieurs façons :

$S \Rightarrow EE$	$S \Rightarrow EE$	$S \Rightarrow EE$	$S \Rightarrow EE$
$\Rightarrow EEEE$	$\Rightarrow aE$	$\Rightarrow Ea$	$\Rightarrow aE$
$\Rightarrow aEEE$	$\Rightarrow aEEE$	$\Rightarrow EEEa$	$\Rightarrow aEEE$
$\Rightarrow abEEE$	$\Rightarrow abEEE$	$\Rightarrow EEbEa$	$\Rightarrow aEEa$
$\Rightarrow abaEE$	$\Rightarrow abaEE$	$\Rightarrow EEbaa$	$\Rightarrow abEEa$
$\Rightarrow ababEE$	$\Rightarrow ababEE$	$\Rightarrow EbEbaa$	$\Rightarrow abEbEa$
$\Rightarrow ababaE$	$\Rightarrow ababaE$	$\Rightarrow Ebabaa$	$\Rightarrow ababEa$
$\Rightarrow ababaa$	$\Rightarrow ababaa$	$\Rightarrow ababaa$	$\Rightarrow ababaa$
(a)	(b)	(c)	(d)

Grammaires

- Types de dérivations
  - (a) et (b) : chaque étape de la dérivation consiste à transformer le non-terminal le plus à gauche. On appelle ce genre de dérivation une dérivation la-plus-à-gauche (ang. *left-most derivation*)
  - (c) : dérivation la-plus-à-droite (ang. *right-most derivation*) où le symbole transformé à chaque étape est le non-terminal le plus à droite
  - (d) : dérivation ni plus-à-droite, ni plus-à-gauche

- Une suite de dérivations peut être visualisée par un arbre de dérivation ou arbre syntaxique (ang. *parse tree*)
  - Un tel arbre indique quelles sont les règles appliquées aux non-terminaux
  - Il n'indique pas l'ordre d'application des règles
  - Les feuilles de l'arbre représentent la chaîne dérivée

Grammaires

- Grammaire ambiguë
  - Lorsqu'une grammaire peut produire plusieurs arbres distincts associés à un même mot, ont dit que la grammaire est ambiguë

- Grammaires équivalentes
  - Deux grammaires qui engendrent le même langage sont dites équivalentes

Langages rationnels et grammaires algébriques

- Il existe des langages algébriques qui ne sont pas rationnels

- Grammaire linéaire à droite
  - $G = (V, \Sigma, R, S)$ telle que $R \subseteq V \times \Sigma^*(V \cup \{\varepsilon\})$

  (rappel : dans une grammaire algébrique (non régulière),$R \subseteq V \times (V \cup \Sigma)^*$)

- Grammaire linéaire à gauche
  - $G = (V, \Sigma, R, S)$ telle que $R \subseteq V \times (V \cup \{\varepsilon\})\, \Sigma^*$

- Grammaire régulière
  - Grammaire linéaire à droite ou linéaire à gauche

Hiérarchie de Chomsky

- Type 3
  - Grammaires régulières — Langages réguliers
  - Automates à états finis
- Type 2
  - Grammaires algébriques — Langages algébriques
  - Automates à pile
- Type 1
  - Grammaires contextuelles
  - Machines de Turing à mémoire linéairement bornée
- Type 0
  - Grammaires générales — Langages récursivement énumérables
  - Machines de Turing
- Inclusion
  - $T3 \subset T2 \subset T1 \subset T0$

Langages rationnels et grammaires algébriques

- Exemple

  $G = (V, \Sigma, R, S)$ avec :

  – $V = \{ S \}$

  – $\Sigma = \{ a, b \}$

  – $R = \{ S \rightarrow aaS \mid bbS \mid \varepsilon \}$

  grammaire régulière : $L(G) = (aa \cup bb)^*$

Langages rationnels et grammaires algébriques

- Théorème

  Un langage est rationnel si et seulement si il est engendré par une grammaire régulière.

- Ce théorème exprime que tout langage rationnel est algébrique (et non l'inverse).

- On peut utiliser la preuve de ce théorème pour :
  - passer d'un automate (déterministe ou non) à une grammaire,
  - passer d'une grammaire à un automate.

Langages rationnels et grammaires algébriques

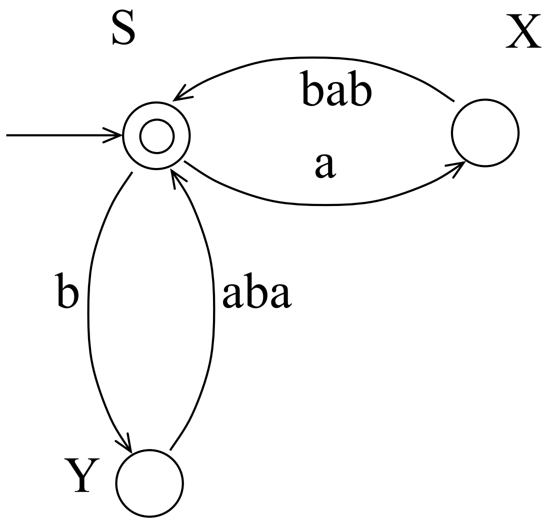
- Exemple ($G \Rightarrow M$)

  $G = (V, \Sigma, R, S)$ une grammaire régulière avec :
  - $V = \{ S, X, Y \}$
  - $\Sigma = \{ a, b \}$
  - $R = \{ S \rightarrow aX \mid bY,\ X \rightarrow aS \mid a,\ Y \rightarrow bS \mid b \}$

- Soit M tel que L(M) = L(G). Pour chaque non-terminal de G on crée un état dans M de la façon suivante :
  - Si $A \rightarrow wB \in R$ alors on crée dans M une transition de l'état A vers l'état B étiquetée par w
  - Si $A \rightarrow w \in R$ alors on crée dans M une transition de l'état A vers l'état F, où F est le seul état introduit dans M qui ne correspond à aucun non-terminal de G

Langages rationnels et grammaires algébriques

- Exemple (M $\Rightarrow$ G)
  Soit l'automate :



- Soit G une grammaire régulière telle que L(G) = L(M). Pour chaque transition de M on crée une règle dans G de la façon suivante :
  - Pour toute transition de l'état p vers l'état q étiquetée par w, on crée la règle correspondante dans G : P $\rightarrow$ wQ,
  - Pour tout état final f de M, on crée dans G une règle d'effacement : F $\rightarrow$ ε.