LIF15 – Théorie des langages formels

Sylvain Brandel

2016 – 2017

sylvain.brandel@univ-lyon1.fr

CM 8

GRAMMAIRES ALGÉBRIQUES

Grammaires algébriques

- <u>Exemple</u>
  - L = a(ab ∪ 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.

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

  - Génération :

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

Grammaires

- Une grammaire algébrique est composée :
  - d'un alphabet Σ de symboles terminaux à partir desquels sont construits les mots du langage (*a* et *b* dans l'exemple),
  - d'un ensemble de symboles non terminaux (S et E dans l'exemple) parmi lesquels on distingue un symbole particulier (souvent S pour *Start*),
  - d'un ensemble fini de règles ou production de la forme

    symbole non terminal → suite finie de symboles terminaux et / ou non terminaux.

Grammaires

- Définition

  Une grammaire algébrique est définie par un quadruplet

  $$G = (V, \Sigma, R, S) \text{ où :}$$

  – Σ est un ensemble fini de symboles terminaux appelé alphabet,

  – V est un ensemble fini de symboles non terminaux tels que

  $$V \cap \Sigma = \varnothing,$$

  – $S \in V$ est le symbole initial

  – R est un sous-ensemble fini de $V \times (V \cup \Sigma)^*$

  Les éléments de R sont appelés règles ou productions.

Grammaires

- <u>Définition</u>

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

  On dit que v <u>dérive directement de</u> 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 $A \rightarrow w \in R$

- <u>Exemple</u>

  En utilisant la grammaire G définie par les règles suivantes :

  $S \rightarrow aE$

  $E \rightarrow abE \mid baE \mid b$

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

- <u>Définition</u>

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

Grammaires

- <u>Définition</u>

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

  On dit que v <u>dérive de</u> 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 <u>dérivation</u> (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 <u>longueur</u> de la dérivation.

Grammaires

- <u>Définition</u>

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

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

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

- <u>Définition</u>
  - Un langage est dit <u>algébrique</u> s'il peut être engendré par une grammaire algébrique.

Grammaires

- <u>Exemple</u>

  G = (V, Σ, R, S) avec :
  - V = { S, E }
  - Σ = { a, b }
  - R = { S → EE, E → EEE | bE | Eb | a }

- La chaîne ababaa peut être dérivée de plusieurs façons :

S ⇒ EE	S ⇒ EE	S ⇒ EE	S ⇒ EE
⇒ EEEE	⇒ aE	⇒ Ea	⇒ aE
⇒ aEEE	⇒ aEEE	⇒ EEEa	⇒ aEEE
⇒ abEEE	⇒ abEEE	⇒ EEbEa	⇒ aEEa
⇒ abaEE	⇒ abaEE	⇒ EEbaa	⇒ abEEa
⇒ ababEE	⇒ ababEE	⇒ EbEbaa	⇒ abEbEa
⇒ ababaE	⇒ ababaE	⇒ Ebabaa	⇒ ababEa
⇒ ababaa	⇒ ababaa	⇒ ababaa	⇒ 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 (*left-most derivation*).
  - (c) : dérivation la-plus-à-droite (*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

- Lorsqu'une grammaire peut produire plusieurs arbres distincts associés à un même mot, ont dit que la grammaire est <u>ambiguë</u>.

- <u>Définition</u>

  Deux grammaires qui engendrent le même langage sont dites <u>équivalentes</u>.

Langages rationnels et langages algébriques

- Il existe des langages algébriques qui ne sont pas rationnels.

- <u>Définition</u>

  Une grammaire algébrique $G = (V, \Sigma, R, S)$ telle que

  $$R \subseteq V \times \Sigma^*(V \cup \{e\})$$

  est appelée <u>grammaire régulière</u> (ou encore <u>linéaire à droite</u>).

  (<u>rappel</u> : dans une grammaire algébrique (non régulière),

  $$R \subseteq V \times (V \cup \Sigma)^*.)$$

Langages rationnels et grammaires algébriques

- <u>Exemple</u>

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

  - $V = \{ S \}$
  - $\Sigma = \{ a, b \}$
  - $R = \{ S \rightarrow aaS \mid bbS \mid e \}$

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

Langages rationnels et grammaires algébriques

- <u>Théorème</u>

  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 :
  - $V = \{ S, X, Y \}$
  - $\Sigma = \{ a, b \}$
  - $R = \{ S \rightarrow aX \mid bY, X \rightarrow aS \mid a, Y \rightarrow bS \mid b \}$

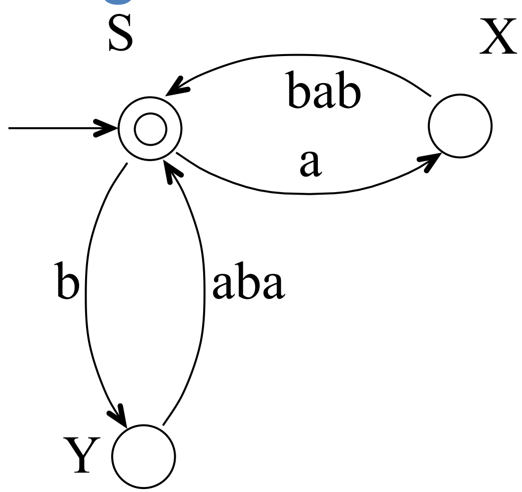
  Construisons l'automate M acceptant L(G), en utilisant la preuve du théorème précédent.

  Pour cela on associe à chaque non-terminal de G 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

- <u>Exemple</u> (M $\Rightarrow$ G)

  Soit l'automate :



Construisons une grammaire régulière G engendrant le même langage, en utilisant la preuve du théorème.

Pour cela on associe à chaque transition de M 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 $\rightarrow$ e.