

LF – Théorie des langages formels

Sylvain Brandel

2023 – 2024

sylvain.brandel@univ-lyon1.fr



CM 3

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**

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

- **Génération**

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

Grammaires

- 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 $A \rightarrow w$

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

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 $A \rightarrow w \in R$

- Exemple

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 $E \rightarrow baE$

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

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

La suite $w_0 \Rightarrow_G w_1 \Rightarrow_G \dots \Rightarrow_G w_n$ 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

$G = (V, \Sigma, R, \text{Phrase})$ avec :

- $V = \{ \text{Phrase, Sujet, Verbe, Complément, Article, Nom, Adjectif} \}$
- $\Sigma = \{ \text{il, elle, nous, je, mange, voit, suis, le, la, les, frites, montagne, bleue, dorées, déprimé, " " } \}$
- $R = \{ \text{Phrase} \rightarrow \text{Sujet Verbe Complément,}$
 $\text{Sujet} \rightarrow \text{il | elle | nous | je,}$
 $\text{Verbe} \rightarrow \text{mange | voit | suis,}$
 $\text{Complément} \rightarrow \text{Article Nom Adjectif,}$
 $\text{Article} \rightarrow \text{le | la | les,}$
 $\text{Nom} \rightarrow \text{frites | montagne,}$
 $\text{Adjectif} \rightarrow \text{bleue | dorées | déprimé } \}$

Grammaires algébriques

- Exemple

$G = (V, \Sigma, R, \text{Paragraphe})$ avec :

- $V = \{ \text{Paragraphe}, \text{Phrase}, \text{Sujet}, \text{Verbe}, \text{Complément}, \text{Article}, \text{Nom}, \text{Adjectif} \}$
- $\Sigma = \{ \text{il}, \text{elle}, \text{nous}, \text{je}, \text{mange}, \text{voit}, \text{suis}, \text{le}, \text{la}, \text{les}, \text{frites}, \text{montagne}, \text{bleue}, \text{dorées}, \text{déprimé}, \text{.} \}$
- $R = \{ \text{Paragraphe} \rightarrow \text{Phrase Paragraphe} \mid \text{Phrase}, \text{Phrase} \rightarrow \text{Sujet Verbe Complément.}, \text{Sujet} \rightarrow \text{il} \mid \text{elle} \mid \text{nous} \mid \text{je}, \text{Verbe} \rightarrow \text{mange} \mid \text{voit} \mid \text{suis}, \text{Complément} \rightarrow \text{Article Nom Adjectif}, \text{Article} \rightarrow \text{le} \mid \text{la} \mid \text{les}, \text{Nom} \rightarrow \text{frites} \mid \text{montagne}, \text{Adjectif} \rightarrow \text{bleue} \mid \text{dorées} \mid \text{déprimé} \}$

Grammaires algébriques

- Exemple

$G = (V, \Sigma, R, \text{Chapitre})$ avec :

- $V = \{ \text{Chapitre, Paragraphe, Phrase, Sujet, Verbe, Complément, Article, Nom, Adjectif} \}$
- $\Sigma = \{ \text{il, elle, nous, je, mange, voit, suis, le, la, les, frites, montagne, bleue, dorées, déprimé, .} \}$
- $R = \{ \text{Chapitre} \rightarrow \text{Paragraphe Chapitre} \mid \text{Paragraphe, Paragraphe} \rightarrow \text{Phrase Paragraphe} \mid \text{Phrase, Phrase} \rightarrow \text{Sujet Verbe Complément.}, \text{Sujet} \rightarrow \text{il} \mid \text{elle} \mid \text{nous} \mid \text{je, Verbe} \rightarrow \text{mange} \mid \text{voit} \mid \text{suis, Complément} \rightarrow \text{Article Nom Adjectif, Article} \rightarrow \text{le} \mid \text{la} \mid \text{les, Nom} \rightarrow \text{frites} \mid \text{montagne, Adjectif} \rightarrow \text{bleue} \mid \text{dorées} \mid \text{déprimé} \}$

Grammaires algébriques

- Exemple

$G = (V, \Sigma, R, \text{Livre})$ avec :

- $V = \{ \text{Livre, Chapitre, Paragraphe, Phrase, Sujet, Verbe, Complément, Article, Nom, Adjectif} \}$
- $\Sigma = \{ \text{il, elle, nous, je, mange, voit, suis, le, la, les, frites, montagne, bleue, dorées, déprimé, .} \}$
- $R = \{ \text{Livre} \rightarrow \text{Chapitre Livre} \mid \text{Chapitre, Chapitre} \rightarrow \text{Paragraphe Chapitre} \mid \text{Paragraphe, Paragraphe} \rightarrow \text{Phrase Paragraphe} \mid \text{Phrase, Phrase} \rightarrow \text{Sujet Verbe Complément.}, \text{Sujet} \rightarrow \text{il} \mid \text{elle} \mid \text{nous} \mid \text{je, Verbe} \rightarrow \text{mange} \mid \text{voit} \mid \text{suis, Complément} \rightarrow \text{Article Nom Adjectif, Article} \rightarrow \text{le} \mid \text{la} \mid \text{les, Nom} \rightarrow \text{frites} \mid \text{montagne, Adjectif} \rightarrow \text{bleue} \mid \text{dorées} \mid \text{déprimé} \}$

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 :

$S \Rightarrow EE$
 $\Rightarrow EEEE$
 $\Rightarrow aEEE$
 $\Rightarrow abEEE$
 $\Rightarrow abaEE$
 $\Rightarrow ababEE$
 $\Rightarrow ababaE$
 $\Rightarrow ababaa$

(a)

$S \Rightarrow EE$
 $\Rightarrow aE$
 $\Rightarrow aEEE$
 $\Rightarrow abEEE$
 $\Rightarrow abaEE$
 $\Rightarrow ababEE$
 $\Rightarrow ababaE$
 $\Rightarrow ababaa$

(b)

$S \Rightarrow EE$
 $\Rightarrow Ea$
 $\Rightarrow EEEa$
 $\Rightarrow EEbEa$
 $\Rightarrow EEbaa$
 $\Rightarrow EbEbaa$
 $\Rightarrow Ebabaa$
 $\Rightarrow ababaa$

(c)

$S \Rightarrow EE$
 $\Rightarrow aE$
 $\Rightarrow aEEE$
 $\Rightarrow aEEa$
 $\Rightarrow abEEa$
 $\Rightarrow abEbEa$
 $\Rightarrow ababEa$
 $\Rightarrow ababaa$

(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 algébriques

- Exemple

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

- $V = \{ E, T, F \}$

- $\Sigma = \{ x, y, +, \times, (,) \}$

- $R = \{ E \rightarrow E + T \mid T, T \rightarrow T \times F \mid F, F \rightarrow (E) \mid x \mid y \}$

- Chaîne $(x + y \times y) \times (x + y)$ avec la dérivation la plus-à-gauche :

$E \Rightarrow T$	$\Rightarrow (x + y \times F) \times F$
$\Rightarrow T \times F$	$\Rightarrow (x + y \times y) \times F$
$\Rightarrow F \times F$	$\Rightarrow (x + y \times y) \times (E)$
$\Rightarrow (E) \times F$	$\Rightarrow (x + y \times y) \times (E + T)$
$\Rightarrow (E + T) \times F$	$\Rightarrow (x + y \times y) \times (T + T)$
$\Rightarrow (T + T) \times F$	$\Rightarrow (x + y \times y) \times (F + T)$
$\Rightarrow (F + T) \times F$	$\Rightarrow (x + y \times y) \times (x + T)$
$\Rightarrow (x + T) \times F$	$\Rightarrow (x + y \times y) \times (x + F)$
$\Rightarrow (x + T \times F) \times F$	$\Rightarrow (x + y \times y) \times (x + y)$
$\Rightarrow (x + F \times F) \times F$	

Grammaires

- Grammaire ambiguë
 - Lorsqu'une grammaire peut produire plusieurs arbres distincts associés à un même mot, on 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

cf. preuves de *non-rationalité* (plus tard)

- 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
 - Automates à états finis

Langages réguliers
- Type 2
 - Grammaires algébriques
 - Automates à pile

Langages algébriques
- Type 1
 - Grammaires contextuelles
 - Machines de Turing à mémoire linéairement bornée
- Type 0
 - Grammaires générales
 - Machines de Turing

Langages récursivement énumérables
- 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.