Chapitre 1 : Introduction à la Théorie des Langages

Dr. Mandicou BA

mandicou.ba@esp.sn
http://www.mandicouba.net

Diplôme D'Ingénieur de Conception (DIC, 2^e année)

Option Informatique



Plan du Chapitre

- Définitions des concepts généraux
- 2 Les langages
- Reconnaisseurs de langages
- Grammaire de réécriture

Sommaire

- Définitions des concepts généraux
 - Symboles, alphabets et mots
- 2 Les langages
- Reconnaisseurs de langages
- Grammaire de réécriture

Définition 1 (Symboles)

 Les symboles sont des éléments indivisibles qui vont servir de briques de base pour la construction d'un mot

Exemple 1

Caractère de l'alphabet ou Chiffre de 0 à 9 est un symbole.

Définition 2 (Alphabet)

- Un alphabet est un ensemble fini et non vide de symboles.
- Conventionnellement, un alphabet est noté ∑

- 2 $\Sigma = \{a, b, \dots, z\}$: l'ensemble des caractères minuscules
- L'ensemble des caractères ASCII

Définition 3 (Mot)

- Un mot m sur Σ (encore appelé chaîne) est une suite de symboles, appartement à l'alphabet Σ, mis bout à bout
- 01101 est un mot de l'alphabet binaire $\Sigma = \{0, 1\}$

Longueur d'un mot

- La longueur d'un mot m, noté |m|, représente le nombre de symboles qui le constituent
- Le mot 01101 est de longueur 5.

Nombre de symboles dans un mot

- Le nombre de symbole s que possède un mot m est noté $|m|_s$.
- \bullet $|01101|_1 = 3.$

Mot vide

• Le mot vide d'un alphabet \sum , noté ε , est le mot de longueur zéro.

i^{ieme} symbole d'un mot m

- Soit m est un mot et i tel que $1 \le i \le |m|$.
- Le i^{ieme} symbole de m est le symbole situé à la position i et est noté m[i].

Puissance d'un alphabet

• $\forall \sum, \sum^{k}$ désigne l'ensemble de tous les mots de longueur k.

Exemple 3 (Soit l'alphabet binaire $\Sigma = \{0, 1\}$)

- **1** NB: quel que soit l'alphabet Σ , alors on obtient $\Sigma^0 = \{\varepsilon\}$

Ensemble de tous les mots d'un alphabet

• On note \sum^* l'ensemble de tous les mots d'un alphabet \sum .

Exemple 4 (pour l'alphabet $\Sigma = \{0, 1\}$)

- $\Sigma^* = \{0,1\}^* = \{\epsilon,0,00,10,110,000,\cdots\}$
- Autrement dit : $\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \cdots$

Ensemble de tous les mots non vides d'un alphabet

• On note Σ^+ l'ensemble de tous les mots non vides de l' alphabet Σ .

Conséquence 1

- $\Sigma^* = \Sigma^+ \cup \{\epsilon\}$

Concaténation de mots

Définition

• Soient α et β deux mots définis dans Σ . La concaténation de α et β , notée $\alpha \cdot \beta$ (ou simplement $\alpha\beta$), consiste à juxtaposer les symboles du mot β à la suite de ceux du mot α .

Exemple 5

- lacksquare Soient $lpha=a_1a_2\cdots a_i$ et $eta=b_1b_2\cdots b_j$ deux mots de Σ

Propriété 1 (Propriétés de la concaténation)

- La concaténation est associative : $(\alpha\beta)\gamma = \alpha(\beta\gamma)$
- ϵ est l'élément neutre pour la concaténation : $\epsilon\alpha=\alpha\epsilon=\alpha$
- L'itération de la concaténation d'un mot m donne les puissance de m :
 - $m^0 = \varepsilon$
 - $\forall n \in \mathbb{N}. m^{n+1} = mm^n$

Facteurs de mots

- Soient α , β et γ trois mots définis sur un alphabet Σ .
 - β est un facteur (ou sous-chaîne) du mot αβγ
 - ightharpoonup est un facteur gauche (ou préfixe) du mot αβ
 - \bullet β est un facteur droit (ou suffixe) du mot αβ
 - Si $\alpha \neq \beta$ et α est un préfixe de β , alors α est dit *préfixe propre* de β
 - Si $\alpha \neq \beta$ et α est un suffixe de β , alors α est dit suffixe propre de β
 - \bullet ϵ est suffixe, préfixe et facteur de tout mot m sur tout alphabet \sum

Sommaire

- Définitions des concepts généraux
- 2 Les langages
- Reconnaisseurs de langages
- Grammaire de réécriture

Définition d'un langage

Définition 4 (Langage \mathcal{L} sur un alphabet Σ)

Un ensemble de mots construits sur l'alphabet ∑

Conséquence 2

- ullet Si Σ est un alphabet et $\mathcal{L} \subseteq \Sigma^*$, alors \mathcal{L} est un langage sur Σ Autrement dit, tout langage \mathcal{L} défini sur Σ est une partie de Σ^*
- L'ensemble de tous les langages qui peuvent être définis sur Σ^* est dit l'ensemble des parties de Σ^* et est noté $\mathcal{P}(\Sigma^*)$

Quelques langages particuliers

- ightharpoonup est un langage pour tout alphabet.
- \bullet 0, le langage vide, est un langage pour tout alphabet
- lacksquare $\{\epsilon\}$, le langage constitué du mot vide, est un langage pour tout alphabet
- NB : $\emptyset \neq \{\epsilon\}$

Opérations sur les langages

Soient \mathcal{L}_1 et \mathcal{L}_1 deux langages définis sur un alphabet \sum

- **●** Union de deux langages : $\mathcal{L}_1 \cup \mathcal{L}_2 \equiv \{x | x \in \mathcal{L}_1 \lor x \in \mathcal{L}_2\}$
- **●** Intersection de deux langages : $\mathcal{L}_1 \cap \mathcal{L}_2 \equiv \{x | x \in \mathcal{L}_1 \land x \in \mathcal{L}_2\}$
- **☞** Différence de deux langages : $\mathcal{L}_1 \mathcal{L}_2 \equiv \{x | x \in \mathcal{L}_1 \land x \notin \mathcal{L}_2\}$
- Complément d'un langage : $\overline{L} \equiv \{x \in \Sigma^* | x \notin L\}$
- Concaténation de deux langages : $L_1 L_2 \equiv \{xy | x \in L_1 \land y \in L_2\}$
- Auto concaténation d'un langage : $\overbrace{\mathcal{L}\cdots\mathcal{L}}\equiv\mathcal{L}^n$
- Fermeture de Kleene : $\mathcal{L}^* \equiv \bigcup_{k>0} \mathcal{L}^k$

Exemple 6 (Soit $\Sigma = \{0,1\}$ l'alphabet binaire)

- ightharpoonup Le langage de tous les mots constitués de n 0 suivit de n 1, avec n \geq 0
 - $\mathcal{L}_1 = \{\epsilon, 01, 001, 00011, 00111, \cdots\}$
- - $\mathcal{L}_2 = \{\epsilon, 01, 0011, 000111, 1001, \cdots\}$
- Le langage de tous les mots binaires qui sont des nombres premiers
 - $\mathcal{L}_3 = \{\epsilon, 01, 11, 101, 1011, \cdots\}$

Exemple 7 (Le langage français)

- $lue{}$ Une phrase du français peut être représentée par un mot $m \in \sum *$
- L est l'ensemble des phrases syntaxiquement correctes du français
 - **O** Ceci est une phrase correcte $\in \mathcal{L}$
 - ② Ceci est une correcte phrase $\notin \mathcal{L}$
 - lacksquare Le cœur a ses raisons que la raison ne connaît point $\in \mathcal{L}$

Exemple 8 (Le langage arithmétique)

- $\Sigma = \{0 \cdots 9, +, \times, /, -, (,)\}$
- Une expression arithmétique peut être représenté e par un mot $m \in \sum *$
- - **1** (1+3) × 45 ∈ \mathcal{L}
 - **②** /43 ∉ *L*
 - **③** (1+234 ∉ *L*

Exemple 9 (Un exemple de langage logique)

- $\Sigma = \{a, \dots, z, \vee, \wedge, \neg, \rightarrow, \longleftrightarrow\}$
- Une formule de la logique des propositions peut être représentée par un mot $m \in \sum *$
- L est l'ensemble des formules de la logique des propositions bien formées
 - $a \lor b \land c \in \mathcal{L}$
 - ② ∨bc ∉ L

Exemple 10 (Le langage de programmation C)

- ullet set l'ensemble des identifiants, des constantes, des opérateurs et des mots clefs du langage C
- Un programme en langage C peut être représenté par un mot $m \in \sum *$
- L est l'ensemble des programmes syntaxiquement corrects du langage de programmation C
 - main(void) {printf("Hello word");} $\in \mathcal{L}$
 - ② main(void) {printf("Hello word)} ∉ L

Problème de décision en théorie des langages

Problème de décision

- \blacksquare Entrée : $x \in S$
- Question : x satisfait-il à la propriété \mathcal{P} de \mathcal{S} ?
- \longrightarrow S étant un ensemble quelconque, dont les éléments sont appelés instances ou entrée du problème.
- ${}^{\blacksquare}$ ${}^{\blacksquare}$ est une propriété de ${}^{\square}$
- Représentons $x \in S$ par un mot m
- Appelons \(\mathcal{L} \) l'ensemble des mots qui représentent les instances qui vérifient \(\mathcal{P} \).

Problème de décision en théorie des langages

- **Entrée** : un mot m
- \blacksquare Question : m appartient-il à \mathcal{L} ?

Description d'un langage

Comment décrire un langage

- Description littérale
 - Ensemble de tous les mots construits l'alphabet $\Sigma = \{a, b\}$, de longueur paire
- Énumération
 - $\mathcal{L} = \{\varepsilon, aa, bb, ab, ba, aaaa, aaab, aaba, \dots\}$
- Expressions régulières
 - Formule permettant de dénoter de manière concise tous les mots du langage : ((a+b)(a+b))*.
- Reconnaisseur
 - Machine permettant de reconnaitre tous les mots du langage
- Grammaire de réécriture
 - Système permettant de générer tous les mots du langage

Sommaire

- Définitions des concepts généraux
- 2 Les langages
- Reconnaisseurs de langages
- Grammaire de réécriture

Composition d'un reconnaisseur

- Bande de lecture
- 2 Tête de lecture
- Mémoire
- Unité de contrôle

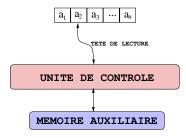


Figure: Illustration schématique d'un reconnaisseur

Les éléments d'un reconnaisseur

Bande de lecture

- Composée d'une succession de cases
- Chaque ne peut contenir qu'un seul symbole d'un alphabet d'entrée
- Les symboles du mot à reconnaître sont écris sur les cases de la bande de lecture

Tête de lecture

- Elle ne peut lire qu'une seule case à un instant donné
- La case sur laquelle se trouve la tête de lecture est appelée case courante
- La tête peut être déplacée par le reconnaisseurs pour se positionner sur la case située immédiatement à gauche ou à droite de la case courante
- Un reconnaisseur qui ne peut déplacer la tête de lecture que de la gauche vers la droite est appelé reconnaisseur à sens unique

Les éléments d'un reconnaisseur

Mémoire auxiliaire

- La mémoire peut prendre des formes différentes
- On suppose un alphabet de mémoire finie et que la mémoire contient uniquement des symboles de cet alphabet
- On suppose qu'à un instant donné, il est possible de déterminer avec exactitude le contenu et la structure de la mémoire
- On suppose également que la taille de la mémoire peut varier aléatoirement au cours du temps
- Deux fonctions principales :
 - fonction de correspondance
 - fonction de transition

Les éléments d'un reconnaisseur

Unité de contrôle

- Elle est le cœur du reconnaisseurs : représente le programme qui dicte au reconnaisseur son comportement
- Elle est définie par :
 - un ensemble d'états finis
 - associé à une fonction de transition qui décrit le passage d'un état vers un autre en fonction :
 - 1 du contenu de la case courante de la bande de lecture
 - et du contenu de la mémoire
- Elle décide :
 - 1 du sens de déplacement de la tête de lecture
 - 2 et des symboles à stocker dans la mémoire

Fonctionnement d'un reconnaisseur

Principe de fonctionnement

- Un reconnaisseur opère par des séquences de déplacement
- Au début de chaque déplacement, la case courante et l'état de la mémoire sont évalué par la fonction de transition
- Ainsi, le symbole courant, l'information issue de la mémoire et l'état de l'unité de contrôle détermine le sens de déplacement . Ce déplacement consiste :
 - Déplacement de la tête de lecture vers la case de gauche, ou vers la case de droite, ou immobilisation sur la case courante
 - Stockage des informations sur la mémoire
 - Changement de l'état de l'unité de contrôle

Fonctionnement d'un reconnaisseur

Configurations d'un reconnaisseur

- Le comportement d'un reconnaisseur peut être, conventionnellement, décrit en termes de configuration
- Une configuration est une « cartographie » du reconnaisseur décrivant :
 - L'état de l'unité de contrôle
 - 2 Le contenu de la bande de lecture
 - La position de la tête de lecture
 - Le contenu de la mémoire

Unité de Contrôle de Déterministe vs Unité de Contrôle de Non Déterministe

- L'unité de contrôle est dit déterministe si et seulement si à chaque configuration, un et un seul déplacement est possible.
- L'unité de contrôle est dit non déterministe si à chaque configuration, plusieurs déplacements sont possibles.

Configuration initiale et configuration finale

Configuration initiale

- La configuration initiale d'un reconnaisseur est une configuration où :
 - L'unité de contrôle se trouve dans son état initial
 - La tête de lecture est positionnée sur la case la plus à gauche de la bande de lecture
 - La mémoire a le contenu initial spécifié

Configuration finale

- La configuration finale d'un reconnaisseur est une configuration où :
 - L'unité de contrôle se trouve un état d'acception spécifié
 - La tête de lecture est positionnée sur la case la plus à droite de la bande de lecture
 - La mémoire satisfait les conditions de la configuration finale

Condition d'acceptation d'un reconnaisseur

Acceptation

Un reconnaisseur accepte un mot m si et seulement si, en partant d'une configuration initiale où m est placé sur la bande de lecture, il effectue une séquence de déplacements successifs qui termine sur une configuration finale.

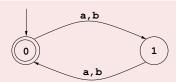
Cas d'un reconnaisseur non déterministe

 Si au moins une des séquences de déplacements possibles conduit à une configuration finale, alors la mot passé en entrée va être accepté

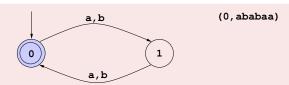
Langage défini par un reconnaisseur

 Un langage défini par un reconnaisseur représente l'ensemble des mots acceptés par ce reconnaisseur

- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- \blacksquare Est-ce que le mot ababaa $\in \mathcal{L}$?

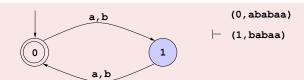


- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- \blacksquare Est-ce que le mot ababaa $\in \mathcal{L}$?



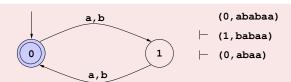


- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- \blacksquare Est-ce que le mot ababaa $\in \mathcal{L}$?



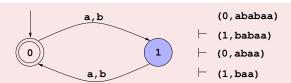


- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- Est-ce que le mot ababaa $\in \mathcal{L}$?



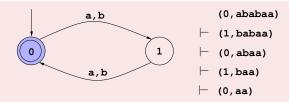


- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- Est-ce que le mot ababaa $\in \mathcal{L}$?



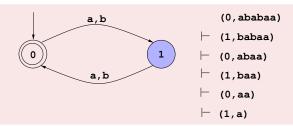


- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- \blacksquare Est-ce que le mot ababaa $\in \mathcal{L}$?



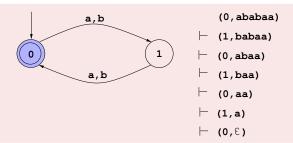


- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- \blacksquare Est-ce que le mot ababaa $\in \mathcal{L}$?





- Soit le langage ⊥ défini par un reconnaisseur ci-dessous
- \blacksquare Est-ce que le mot ababaa $\in \mathcal{L}$?





Sommaire

- Définitions des concepts généraux
- 2 Les langages
- Reconnaisseurs de langages
- Grammaire de réécriture

Description générale d'une grammaire

- Les grammaires sont, probablement, la plus importante classe de générateur de langage
- Une grammaire est un système mathématique permettant de générer tous les mots d'un langage
- Pour un langage L, une grammaire se fonde sur deux ensembles disjoints de symboles
 - $\textbf{ \mathbb{N} L'ensemble des symboles non-terminaux, noté \mathcal{N} et appelé alphabet non-terminal }$
 - Les non-terminaux représentent les symboles élémentaires du langage défini par la grammaire
 - ${f 2}$ L'ensemble des symboles terminaux, noté ${f \Sigma}$ et appelé alphabet terminal
 - Les terminaux représentent des mots constitués de non-terminaux et utilisés dans la génération des mots du langages

Description générale d'une grammaire

Règles de production

- Le cœur d'une grammaire est un ensemble fini, noté \mathcal{P} , de règles de production (encore appelé *production*)
- Les règles de production décrivent comment une phrase du langage définie par une grammaire va être générée
- Une règle de production (α, β) (ou règle de réécriture) un élément de l'ensemble
 - $(\mathcal{N} \cup \Sigma)^* \mathcal{N} (\mathcal{N} \cup \Sigma)^* \times (\mathcal{N} \cup \Sigma)^*$
 - $(\mathcal{N} \cap \Sigma)^* \mathcal{N} (\mathcal{N} \cup \Sigma)^*$: désigne n'importe quelle chaîne contenant au moins un non-terminal
 - $(\mathcal{N} \cup \Sigma)^*$: désigne n'importe quelle chaîne

Convention

• Si (α,β) est une règle de production, alors la notation suivante est utilisée : $\alpha \to \beta$

Description formelle d'une grammaire

Définition 5

- Une grammaire de réécriture est un 4-uplet $\langle \mathcal{N}, \Sigma, \mathcal{P}, \mathcal{S} \rangle$ où :

 - $ightharpoonup \Sigma$ est un ensemble de symboles terminaux, appelé alphabet terminal, tel que $\mathcal{N} \cap \Sigma = \emptyset$
 - \mathcal{P} est un sous ensemble fini de :

$$(\mathcal{N} \cup \Sigma)^* \mathcal{N} (\mathcal{N} \cup \Sigma)^* \times (\mathcal{N} \cup \Sigma)^*$$

Un élément de (α,β) $\mathcal P$, aussi noté $\alpha\to\beta$, est une règle de production

ullet S est un élément de $\mathcal N$ appelé l'axiome de la grammaire

Exemple 12

- Soit le langage ⊥ défini par la grammaire suivante
- $G = \langle \{P, I\}, \{a, b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- ightharpoonup Le mot ababaa $\in \mathcal{L}$?

Exemple 12

- Soit le langage ⊥ défini par la grammaire suivante
- $G = \langle \{P, I\}, \{a, b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- Le mot ababaa $\in \mathcal{L}$?

Exemple 12

- Soit le langage ⊥ défini par la grammaire suivante
- $G = \langle \{P,I\}, \{a,b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- Le mot ababaa $\in \mathcal{L}$?

Exemple 12

- Soit le langage ⊥ défini par la grammaire suivante
- $G = \langle \{P,I\}, \{a,b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- \blacksquare Le mot ababaa $\in \mathcal{L}$?

$$P \Rightarrow al$$
 $\Rightarrow abP$
 $\Rightarrow abal$
 $\Rightarrow ababP$
 $\Rightarrow ababal$
 $\Rightarrow ababaaP$
 $\Rightarrow ababaaP$
 $\Rightarrow ababaaP$

Exemple 12

- Soit le langage ⊥ défini par la grammaire suivante
- $G = \langle \{P, I\}, \{a, b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- \blacksquare Le mot ababaa $\in \mathcal{L}$?

Exemple 12

- Soit le langage L défini par la grammaire suivante
- $G = \langle \{P, I\}, \{a, b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- Le mot ababaa $\in \mathcal{L}$?

Exemple 12

- Soit le langage L défini par la grammaire suivante
- $G = \langle \{P, I\}, \{a, b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- Le mot ababaa $\in \mathcal{L}$?

$$P \Rightarrow al$$

$$\Rightarrow$$
 abP

$$\Rightarrow$$
 abal

Exemple 12

- Soit le langage L défini par la grammaire suivante
- $G = \langle \{P, I\}, \{a, b\}, \{P \rightarrow \varepsilon, P \rightarrow aI, P \rightarrow bI, I \rightarrow aP, I \rightarrow bP\}, P \rangle$
- Le mot ababaa $\in \mathcal{L}$?

$$P \Rightarrow al$$

$$\Rightarrow$$
 abP

$$\Rightarrow$$
 abal

Chapitre 1 : Introduction à la Théorie des Langages

Dr. Mandicou BA

mandicou.ba@esp.sn http://www.mandicouba.net

Diplôme D'Ingénieur de Conception (DIC, 2^e année) Option Informatique

