Langages et Grammaires

Jean-Paul ARCANGELI
Jean-Baptiste RACLET
Jean-Baptiste.Raclet@irit.fr

UNIVERSITE TOULOUSE III PAUL SABATIER
Faculté Sciences et Ingénierie
D[i] Département Informatique
IRIT Institut de Recherche en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J

LANGAGES ET AUTOMATES
UPS – Licence 3 Informatique – S5
2020-2021

Plan du chapitre

1. Introduction
2. Alphabet et mots
3. Langage
4. Grammaire
5. Classification des langages

1- Introduction

- Langages… formels
- Alphabet, mot, langage, grammaire…
  - Vous connaissez ces termes
  - Nous allons les (re)définir !

2- Alphabet et mots : définitions

- Un <u>alphabet</u> est un ensemble fini non vide de <u>symboles</u>
  - Notation habituelle de l'alphabet : X
  - On emploie parfois le terme « lettre » à la place de « symbole »

2- Alphabet et mots : définitions

- Un mot construit sur l'alphabet X est une séquence finie de symboles de X
  - Séquence = suite
    - Attention à l'ordre des symboles !
  - La séquence peut être vide !
  - Le mot vide, noté λ, est le mot ne contenant aucun symbole
  - Pas de mot composé d'une infinité de symboles

2- Alphabet et mots : définitions

- La longueur d'un mot w, notée $|w|$, est le nombre de symboles qui constituent w
  - Donc $|\lambda| = 0$

- Soit un mot w sur l'alphabet X et a un symbole de X. On note $|w|_a$ est le nombre d'occurrences du symbole a dans w
  - Pour construire un mot, un symbole peut être utilisé un nombre quelconque de fois
  - $|w| = \Sigma_{x \in X} |w|_x$ (bien sûr !)

2- *Alphabet et mots : définitions*

- On note X^n l'ensemble des mots de longueur n sur X
  - $X^0 = \{\lambda\}$, $X^1 = X$…
- On note X^* l'ensemble de tous les mots sur l'alphabet X
  - Y compris λ
  - $\boxed{\mathbf{X^* = \cup_{n \in N} X^n}}$
    - $X^* = X^0 \cup X^1 \cup X^2 \cup X^3$ …
  - X^* est un ensemble infini !
    - On peut construire une infinité de mots sur un alphabet X
- Attention !!!!
  - λ est un mot, $\lambda \in X^*$
  - <u>λ n'est pas un symbole</u> de l'alphabet : $\lambda \notin X$!!!

Alphabet et mots : définitions

- La <u>concaténation</u> est une opération qui permet de construire un mot à partir de deux mots donnés
  - Notation : $\bullet$
  - $\forall\, w_1,w_2 \in X^*$, $w_1 \bullet w_2 \in X^*$ tel que
    - Si $|w_1| = 0$ (c.-à-d. $w_1 = \lambda$), $w_1 \bullet w_2 = w_2$
    - Si $|w_1| \neq 0$ ($\exists\, x \in X$, $\exists\, w_1' \in X^*$ et $w_1 = x \bullet w_1'$),
      $w_1 \bullet w_2 = (x \bullet w_1') \bullet w_2 = x \bullet (w_1' \bullet w_2)$
  - Théorème : $|w_1 \bullet w_2| = |w_1| + |w_2|$
  - L'opération $\bullet$ est <u>associative</u> et admet λ comme <u>élément neutre</u>
    - $\forall\, u,v,w \in X^*$, $(u \bullet v) \bullet w = u \bullet (v \bullet w)$
    - $\forall\, w \in X^*$, $\lambda \bullet w = w \bullet \lambda = w$
    - *On dit que (X^*, $\bullet$) est un « monoïde »*
    - <u>L'opération $\bullet$ n'est pas commutative</u> !!!
      - $w_1 \bullet w_2 \neq w_2 \bullet w_1$ dans le cas général

3- Langage : définitions

- Un <u>langage</u> sur un alphabet X est un <u>ensemble de mots</u> construits à partir de symboles de X
  - Pour tout langage L sur l'alphabet X, $L \subseteq X^*$ ($L \in \mathcal{P}(X^*)$) : un langage sur X est une partie de X*
  - L'ensemble vide $\varnothing$ est un langage
  - $X^0 = \{\lambda\}$, X^1, X^2, X^3 … sont des langages
    - Attention : $\{\lambda\} \neq \varnothing$
  - Toute partie de X* est un langage
  - X* est un langage

  - $\mathcal{P}(X^*)$ est l'ensemble de tous les langages sur l'alphabet X

3- Langage

- Il y a plusieurs façons de définir/décrire un langage
  - Textuellement « en français »
  - Dans le langage mathématique
    - Par énumération exhaustive des mots (pour un langage fini) c.-à-d. « en extension »
    - En spécifiant les propriétés caractéristiques c.-à-d. « en compréhension »
  - Mais aussi de manière complètement formelle au moyen de
    - Grammaires
    - Automates ou autres machines abstraites
    - Expressions régulières
      - Seulement pour une catégorie particulière de langages dits « réguliers »

3- Langages : définitions

- Les langages sont des ensembles, donc toutes les opérations sur les ensembles sont applicables
  - Inclusion, égalité
  - Différence, complémentaire, union, intersection
  - …

- *Rappels : Soient L et L' deux langages (ensembles) sur X*
  - *Différence : $L \setminus L' = \{w \in X^* / w \in L$ et $w \notin L'\}$*
  - *Complémentaire : $X^* \setminus L = \{w \in X^* / w \notin L\}$*
  - *Union : $L \cup L' = \{w \in X^* / w \in L$ ou $w \in L'\}$*
  - *Intersection $L \cap L' = \{w \in X^* / w \in L$ et $w \in L'\}$*

3- Langage : définitions

produit concaténation

- Le <u>produit</u> est une opération sur les langages qui permet de construire un langage à partir de deux langages donnés
  - Notation : $\bullet$
    - Identique à la concaténation, mais il n'y a pas de confusion possible
  - $\forall L_1, L_2 \subseteq X^*$, $L_1 \bullet L_2 = \{w \in X^* / \exists w_1 \in L_1, \exists w_2 \in L_2, w = w_1 \bullet w_2\}$
    - Le produit de deux langages est un langage (loi interne)
  - Le produit est <u>associatif</u> et admet un langage <u>élément neutre</u> $\{\lambda\}$
    - $\forall A,B,C \subseteq X^*$, $(A \bullet B) \bullet C = A \bullet (B \bullet C)$
    - $\forall L \subseteq X^*$, $\{\lambda\} \bullet L = L \bullet \{\lambda\} = L$
    - *D'où ($\mathcal{P}(X^*)$, $\bullet$) est un « monoïde »*
  - <u>Le produit n'est pas commutatif</u> !!! (puisque la concaténation ne l'est pas)
    - $L_1 \bullet L_2 \neq L_2 \bullet L_1$ dans le cas général

3- Langage : définitions

- Opération « puissance »
  - Soit $L \subseteq X^*$, $L^0 = \{\lambda\}$ et $L^n = L \bullet L^{n-1}$ pour tout $n>0$ (définition par récurrence)
- Remarque
  - On retrouve la notation X^n précédemment introduite

3- Langage : définitions

- Opération « étoile »
  - Soit $L \subseteq X^*$, $L^* = \cup_{n\in N} L^n = L^0 \cup L^1 \cup L^2 \cup \ldots = \{\lambda\} \cup L \cup (L\bullet L) \cup \ldots$
- Remarque
  - On retrouve la définition $X^* = \cup_{n\in N} X^n$

3- Propriétés sur les langages (théorèmes)

Soient L, L_1, L_2, L_i ($i \in I$) des langages sur l'alphabet X

- $L_1 \subseteq L_2 \Rightarrow L \bullet L_1 \subseteq L \bullet L_2$

Le produit de langages est distributif sur l'union

- $L \bullet (\cup_{i \in I} L_i) = \cup_{i \in I} (L \bullet L_i)$

3- Propriétés sur les langages (théorèmes)

Le produit de langages n'est pas distributif sur l'intersection

- $L \bullet (L_1 \cap L_2) \subseteq (L \bullet L_1) \cap (L \bullet L_2)$
- $L \bullet (L_1 \cap L_2) \neq (L \bullet L_1) \cap (L \bullet L_2)$
  - Pas d'égalité dans le cas général
- $\{x\} \bullet (L_1 \cap L_2) = \{x\} \bullet L_1 \cap \{x\} \bullet L_2$
  - Egalité dans ce cas ; ici, x est un mot composé d'1 symbole de X

L'étoile n'est pas distributive sur l'intersection

- $(L_1 \cap L_2)^* \subseteq L_1^* \cap L_2^*$
- $(L_1 \cap L_2)^* \neq L_1^* \cap L_2^*$
  - Pas d'égalité dans le cas général