Langages Formels

Anne Grazon – L3 Info Rennes 1

2015 – 2016, S1

1 Monoïdes libres, langages

1.1 Monoïdes

La théorie des langages est née dans les années 60 de la volonté des linguistes de formaliser la notion de grammaire (des langages naturels). Parmi eux, Noam Chomsky a défini quatre types de grammaires associées à quatre types de langages (type 0, 1, 2 et 3). Dans ce cours, nous étudierons les langages algébriques (type 2) et rationnels (type 3).

def 1.1 Quelques définitions :

1. Alphabet — Ensemble Σ fini non vide de symboles, appelés lettres.
2. Mot sur Σ — Suite finie de lettres. On définit sa longueur $|u| = n$.
3. Le mot vide est noté ϵ ou 1_Σ, $|1_\Sigma| = 0$.
4. Σ^* désigne l'ensemble de tous les mots sur Σ.
5. La loi de composition interne sur Σ^* notée $\cdot$ est la concaténation, qui est associative et admet 1_Σ comme élément neutre.
6. Et ça, c'est un monoïde.

def 1.2 Un monoïde est dit libre lorsque la décomposition d'un élément quelconque en "éléments de base" suivant sa loi de composition, est unique. Σ^* est alors le monoïde libre engendré par Σ.
Remarque 1 — On voit immédiatement que deux mots sont égaux si et seulement si ils sont de même longueur, et ont leur lettres égales deux à deux. Cette propriété caractérise les monoïdes libres.

def 1.3 v est un facteur de $u \Leftrightarrow \exists \alpha, \beta \in \Sigma^*, u = \alpha v \beta$; c'est un facteur droit (resp. gauche) si $\beta = 1_\Sigma$ (resp. si $\alpha = 1_\Sigma$). C'est un facteur propre si $v \neq u$ et $v \neq 1_\Sigma$.

def 1.4 Pour $x \in \Sigma$, $|\cdot|_x$ est le nombre d'occurrences de x dans un mot ; on définit de même le nombre d'occurrences d'un mot dans un autre.

1.2 Langages

Un langage est un ensemble quelconque de mots ($L \subseteq \Sigma^*$, $L \in \mathcal{P}(\Sigma^*)$). L'union, l'intersection et le complémentaire sont définis intuitivement sur les langages.

def 1.5 Les autres opérations usuelles sont le produit $L \cdot M = \{uv | u \in L \text{ et } v \in M\}$, l'étoile de Kleene $L^* = \cup_{n \geqslant 0} L^n$ et l'étoile propre $L^+ = L^* \setminus L^0$.

2 Grammaires algébriques

def 2.1 Une grammaire est un quadruplet $G = (\Sigma, V, S, P)$ où Σ est l'alphabet terminal, V disjoint de Σ l'alphabet des non-terminaux, $S \in V$ l'axiome de G et $P \subsetneq V \times (X \cup V)^*$ l'ensemble des règles de production.

Exemple

$G_1 = (\Sigma, V, S, P)$ avec $\Sigma = \{a, b\}$, $V = \{S, T\}$ et les règles de production $P \to aSb + aT$, $T \to b$.

def 2.2 Une grammaire est dite linéaire (resp. à droite et à gauche) si $P \subset V \times (\Sigma^* \times V \times \Sigma^* \cup \Sigma^*)$ (resp. $\Sigma^* \times V$ and so on).

EXEMPLE

G_1 est linéaire.

La dérivation consiste à engendrer un mot à partir d'un autre en suivant une règle de production. Elle est notée $\rightarrow$, sa fermeture réflexive $\rightarrow^*$ et une dérivation à l'ordre n, $\rightarrow^n$.

def 2.3 Le langage engendré par une grammaire est $L(G) = \{f \in \Sigma^* | S \rightarrow^* f\}$ (le langage élargi accepte aussi V^*). Réciproquement, un langage est dit algébrique s'il existe une grammaire G telle que $L = L(G)$.

EXEMPLE

$L(G_1) = \{a^n b^n | n \geqslant 1\}$