2.0 Rappels sur les grammaires syntagmatiques

Grammaire formelle Pour définir une grammaire formelle, il faut définir
– un alphabet *terminal* (X),
– un alphabet *non terminal* disjoint (V), comprennant l'*axiome* $S \in V$,
– un ensemble de couples de $(X \cup V)^* V (X \cup V)^* \times (X \cup V)^*$, les *règles de production*.

Principe de réécriture Les grammaires formelles constituent un cas particulier de *système de réécriture* (ensemble de règles de réécriture).

Une *règle de réécriture* de la forme $r \to r'$ s'applique à l'objet syntaxique t si celui-ci contient une instance du membre gauche r, c'est-à-dire un sous-objet que l'on peut identifier à r. L'objet t se réécrit alors en un nouvel objet t', obtenu en remplaçant l'instance de r par l'instance du membre droit r' correspondante. Notation : $t \to t'$.

Dérivation Pour les grammaires formelles, on définit ainsi la réécriture (ou dérivation) immédiate (pour une règle $r : A \longrightarrow u$) : $f \xrightarrow{r} g$ ssi $\exists v, w$ t.q. $f = vAw$ et $g = vuw$

La *dérivation* est définie comme la fermeture de la dérivation immédiate :
$f \xrightarrow{\mathcal{G}*} g$ ssi $f = g$, ou
$\exists f_1 = f, f_2, ..., f_n = g,\ \exists r_i \in P$ t.q. $f_{i-1} \xrightarrow{r_i} f_i$

Proto-mot, langage engendré Un mot sur $(X \cup V)^*$ obtenu au cours d'une dérivation et contenant au moins un symbole non terminal est appelé *proto-mot*. L'ensemble des mots (sur X^*) engendrés à partir de l'axiome constitue le *langage engendré* par la grammaire.

Dérivations multiples En général, la suite de réécritures immédiates qui conduit de l'axiome à un mot donné u n'est pas unique. On distingue parmi les suites possibles la *dérivation gauche*, définie de telle sorte que le non-terminal réécrit à chaque étape est le premier non-terminal du proto-mot. Une grammaire telle que pour un mot donné il existe plusieurs dérivations gauches distinctes est une grammaire *ambigüe*.

Pouvoir expressif On peut distinguer des sous-classes de grammaires formelles, sur la base de la forme des règles de production. Ces classes de grammaires définissent à leur tour des classes de langages formels. Chomsky a proposé une hiérarchie de classes de grammaires/langages, dont les classes sont incluses les unes dans les autres.

Langages	type	règles	machine
récursivement énumérables	0	*pas de contraintes*	Machine de Turing
contextuels	1	$\alpha A \beta \to \alpha \gamma \beta$, avec $A \in V, \alpha, \beta, \gamma \in (X \cup V)^*, \gamma \neq \varepsilon$	Automates à pile linéairement bornée
algébriques	2	$A \to \gamma$, avec $A \in V, \gamma \in (X \cup V)^*$	Automates à pile
rationnels	3	$A \to Ba$, $A \to a$, avec $A, B \in V, a \in X$	Automates à nombre fini d'états

Arbre de dérivation Pour les **grammaires algébriques**, on peut représenter l'ensemble des dérivations équivalentes par un *arbre de dérivation*.

$\mathcal{G}_0 =$

S	$\to$	TZ	S	S
T	$\to$	$0T1C$	TZ	TZ
	$\mid$	ε	0T1CZ	0T1CZ
$C1$	$\to$	$1C$	00T1C1CZ	0T1Z2
CZ	$\to$	$Z2$	00ε1C1CZ	00T1C1Z2
$1Z$	$\to$	1	0011CCZ	00ε1C1Z2
			0011CZ2	0011CZ2
			001 1Z22	001 1Z22
			001122	001122

$L_{G_0} = 0^i 1^i 2^i$

$$\mathcal{G}_1 = \left\langle \{jean, dort\}, \{Np, SN, SV, V, S\}, S, \left\{ \begin{array}{l} S \to SN\ SV \\ SN \to Np \\ SV \to V \\ Np \to jean \\ V \to dort \end{array} \right\} \right\rangle\}$$

$\mathcal{G}_2 = \langle \{(,)\}, \{S\}, S, \{S \longrightarrow \varepsilon \mid (S)S\} \rangle$

$G_3 = E \to E + T \mid T, T \to T \times F \mid F, F \to (E) \mid a$

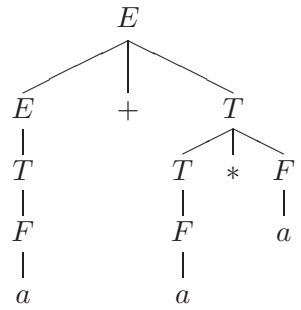
Voici un exemple de dérivation du mot $a + a * a$ par la grammaire $\mathcal{G}_2$ (8 étapes) :

$E \Rightarrow E{+}T \Rightarrow E{+}T{*}F \Rightarrow T{+}T{*}F \Rightarrow T{+}F{*}F \Rightarrow T{+}a{*}F \Rightarrow F{+}a{*}F \Rightarrow a{+}a{*}F \Rightarrow a{+}a{*}a$

Dérivation gauche :

$E \Rightarrow E + T \Rightarrow T + T \Rightarrow F + T \Rightarrow a + T \Rightarrow a + T * F \Rightarrow a + F * F \Rightarrow a + a * F \Rightarrow a + a * a$

Factorisation par un arbre de dérivation :



$G'_2 = \langle \{E\}, \{a, +, \times\}, E, \{E \to E + E \mid E \times E \mid a\} \rangle$

Deux arbres $\neq$ pour $a + a * a$: