4- Grammaire

- Comment définir rigoureusement un langage ?
  - En donnant des règles (syntaxiques) de « grammaire »
    - Qui explicitent comment construire les mots sur l'alphabet
  - Définition « constructive » du langage
  - Intuitivement, il faut pour cela 4 composants
    - Un alphabet
    - Les règles de construction, dites de « production »
    - Pour exprimer ces règles, un ensemble de symboles spécifiques
    - Parmi les symboles spécifiques, un symbole particulier à partir duquel toute construction commence
  - NB : On verra qu'il existe d'autres moyens pour définir les langages

Grammaires : définition

- Une <u>grammaire</u> G est un quadruplet <X, N, P, S> où :
  - X est un ensemble fini de <u>terminaux</u> (l'alphabet)
  - N est un ensemble fini de <u>non terminaux</u> (avec $N \cap X = \emptyset$)
  - P est un ensemble fini de <u>règles de production</u> (règles de réécriture)
    - P = {$\alpha \rightarrow \beta$ où

      α est une séquence non vide de symboles de l'ensemble $N \cup X$
      β est une séquence non vide de symboles de l'ensemble $N \cup X$
      ou $\beta = \lambda$}
    - La signification intuitive de $\alpha \rightarrow \beta$ est que la partie gauche α de la règle peut être remplacée par la partie droite β
    - En pratique, nous manipulerons des grammaires t.q. $\alpha \in N$
      - C.-à-d. uniquement 1 non terminal en partie gauche des règles
    - Les règles peuvent être récursives
  - $S \in N$ appelé <u>axiome</u>

Grammaires : définition

- Dérivation
  - Soient
    - $G = <N, X, P, S>$
    - $w_1 \in (N \cup X)^* \setminus \{\lambda\}$
    - $w_2 \in (N \cup X)^*$
  - On dit que w1 se dérive en une étape en w2 (ou que w_2 dérive directement de w_1), noté $\mathbf{w_1 \Rightarrow w_2}$,
    si $\exists\, x,y \in (N \cup X)^*$ tels que $w_1 = x\,u\,y$ et $w_2 = x\,v\,y$
    et si la règle $u \rightarrow v \in P$
  - On dit que w1 se dérive (en un nombre quelconque d'étapes) en w2, noté $\mathbf{w_1 \Rightarrow^* w_2}$,
    si $w_1 = w_2$
    ou si $\exists\, w_3 \in (N \cup X)^*$ t.q. $w_1 \Rightarrow w_3$ et $w_3 \Rightarrow^* w_2$

Grammaires : définition

- Langage « engendré » par une grammaire
  - Soit $G = <N, X, P, S>$
  - Le langage engendré par la grammaire G, noté L(G), est l'ensemble des mots $\{w \in X^* \,/\, S \Rightarrow^* w\}$

Grammaires : définition

- Arbres de dérivation
  - Soit G = <N, X, P, S> et w ∈ L(G)
  - Un arbre de dérivation du mot w est une représentation (structurelle, non séquentielle) des dérivations de w à partir de S
    - Aux feuilles d'un arbre de dérivation, il n'y a que des terminaux ! (on dit que l'arbre est clos)
  - Rq : L(G) est l'ensemble des mots de X* qui admettent un arbre de dérivation

Grammaires : définition

- Ambiguïté
  - Une grammaire est <u>ambiguë</u> si ∃ w ∈ L(G) qui admet au moins deux arbres de dérivations différents
  - Rq : En pratique, les grammaires ambiguës posent problème puisqu'un mot peut être « compris » de plusieurs façons différentes
    - 2 x 3 + 4

Égalité sur les langages

- Soient
  - Une grammaire G
    - G engendre un langage noté L(G)
  - Un langage L
- Comment montrer que L(G) = L ?
  - Par « double inclusion » c.-à-d.
    - $L \subseteq L(G)$
      - Par ex., par récurrence sur la longueur (ou …) des mots
    - $L(G) \subseteq L$
      - Par ex., par récurrence sur le nombre d'utilisation d'une ou des règles

Hiérarchie de Chomsky

- Classification des grammaires donc des langages
  - En fonction de leur complexité
- Grammaire « context-free »
  - Les règles sont de la forme
    $A \rightarrow \beta$ avec $A \in N$
    (et $\beta \in (N \cup X)^*$)
- Grammaire « linéaire à droite »
  - Les règles sont de la forme
    $A \rightarrow a\,B \mid \lambda$ avec $A,B \in N$
    et $a \in X$