6.1 Stratégies d'analyse

Comme les automates, les grammaires définissent un langage. On peut donc, comme avec les automates, chercher pour un mot donné s'il appartient au langage engendré par une grammaire donnée.

Mais la grammaire nous donne quelque chose de plus : une décomposition, une analyse, du mot donné, en fait, un **arbre syntaxique**. Il est inutile d'expliquer à un linguiste l'intérêt de ce résultat. Il faut noter qu'alors se pose la question, non seulement de trouver une grammaire qui engendre (*i.e.* reconnaît) le langage voulu, mais aussi de trouver la grammaire qui propose la « meilleure » analyse, *i.e.* le « meilleur » arbre syntaxique. On peut en effet trouver plusieurs grammaires qui reconnaissent le même langage, ce qui les distingue alors, c'est l'arbre syntaxique associé à chaque mot. Noter que la question de la meilleure grammaire reste ouverte (plus économique ? plus « naturelle » ?...)

En tout cas, l'analyse syntaxique, étant donnés une grammaire et un mot, va donc non seulement déterminer si le mot est engendré, mais encore fournir l'arbre syntaxique correspondant.

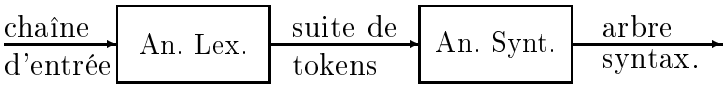
Le problème principal est que, contrairement au automates qui peuvent être déterminisés, toute grammaire, même *context-free*, ne peut pas être « déterminisée ».

6.2 Analyse lexicale

Prenons le cas d'une grammaire du langage naturel (sn, sv).

La grammaire étant un dispositif puissant, mais aussi lourd (non déterministe), on préfère en général confier la tâche de la reconnaissance des symboles qui constituent un mot (nom, déterminant) à un dispositif plus simple (automate, par exemple), plutôt que d'insérer cela dans les règles.

Alors, la grammaire a pour terminaux des **tokens**, dont la décomposition est faite par un module indépendant, d'analyse lexicale.



Dorénavant, on ne s'intéressera pas beaucoup à la partie lexicale, et on fera volontiers l'hypothèse que la grammaire travaille avec des tokens comme terminaux.

Encore une remarque : l'arbre syntaxique n'est que rarement le but ultime d'une analyse : on ajoute le plus souvent une étape dite « sémantique » qui interprète le mot reconnu, en exploitant l'arbre syntaxique.