



Traduction des langages



M1 Informatique – Développement Logiciel
Semestre 7

Cours donné par Christine MAUREL
Rédigé par Antoine de ROQUEMAUREL

2014

Avant-propos

- 7 séances de cours, 7 séances de TD \Rightarrow Rapide
- MCC : $1CT = 100\%$ ^a
- Plus de cours/TD \Rightarrow Cours magistraux.
- Moyenne S7 doit être ≥ 10 + note UE ≥ 6

a. 22 Novembre

Table des matières

1	Introduction	4
1.1	Intéprétation ou Compilation	4
2	Analyse lexicale	6
2.1	Token	6
2.2	Identificateurs	6
2.3	Fonctionnement	7
3	Analyse syntaxique	10
4	Génération de code	11
A	Rappels	12
A.1	Grammaire	12
A.2	Reconnaître un langage avec un automate à pile	13
B	Table des figures	15

1

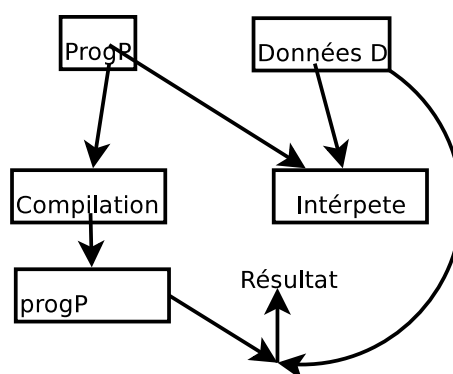
Introduction

La traduction des langages peut être assimilée à de la « compilation ». C'est à dire comprendre pourquoi un programme dans un langage de programmation est compris la machine où que les erreurs sont détectés.

1.1 Intérprétation ou Compilation

Une interprétation utilise un interpréteur et calcul lors de l'exécution du programme.

Une compilation utilise un compilateur et traduit le programme. Aucune execution n'est nécessaire.



	Avantages	Inconvénients
Interpréteur	<ul style="list-style-type: none">— Convivial— Mise au point rapide	<ul style="list-style-type: none">— Moins efficace
Compilateur	<ul style="list-style-type: none">— Efficacité— Optimisation possible	<ul style="list-style-type: none">— Plus lourd

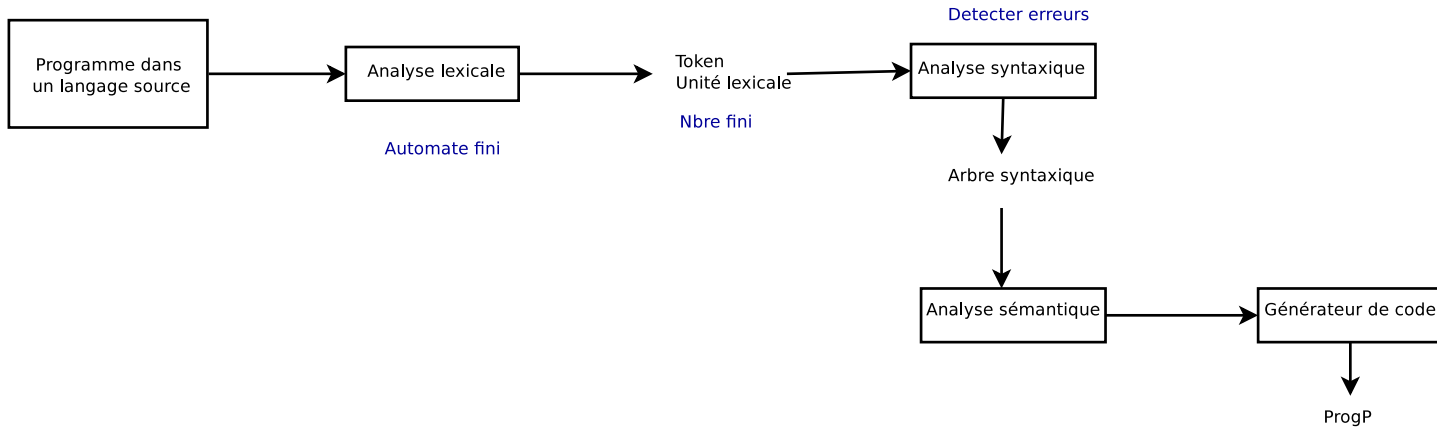


FIGURE 1.1 – Phases de compilation

2

Analyse lexicale

Un analyseur lexical doit découper un texte source en *tokens*, l'analyseur lexicale peut aussi être appelé scanner. L'analyseur lexical ne fonctionne pas tout seul, il est en général guidé par un analyseur syntaxique.

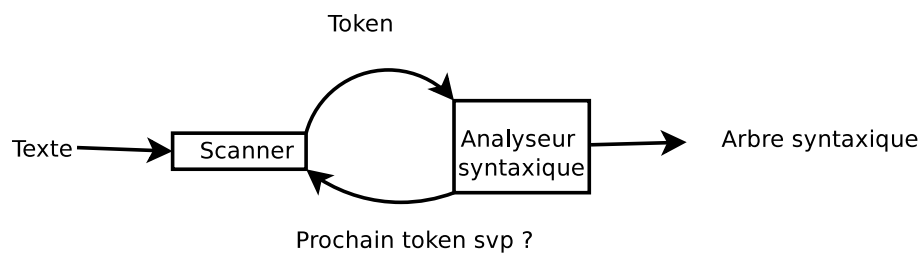


FIGURE 2.1 – Diagramme de traduction

2.1 Token

- Identificateur
- Mots clés
- Constantes numériques
- Opérateurs arithmétiques
- Opérateurs de comparaison
- Séparateur
- Commentaires
- Séparateurs

2.2 Identificateurs

- Commence par une lettre
- Suivi d'une suite éventuellement vide de lettres, de chiffres (et de caractères spéciaux)

Nombres entiers signés 2014, +2014, -2014

Alphabet $X = \{a, \dots, z, \dots, 0, \dots, 9, (+), (-)\}$

Automate fini qui reconnaît identificateurs et nombres

$$L_0 = l(L + c)^* + cc^* + (+)cc^* + (-)cc^*$$

$$L_0 =$$

2.3 Fonctionnement

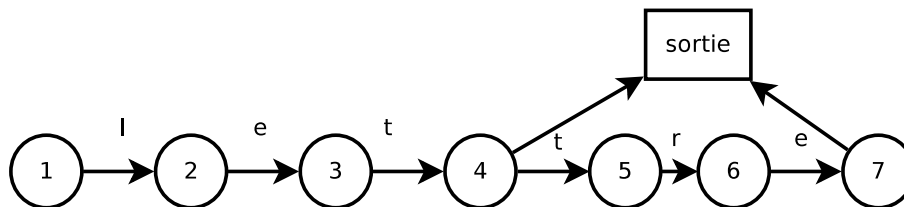
L'analyseur lexical lit le texte caractère par caractère, découpe et reconnaît des tokens (lexèmes) exprimé avec une regex.

Il existe des outils générateurs de scanner.

2.3.1 Les problèmes posés par l'analyse lexicale

2.3.1.1 Reconnaissance

Si on a `let` en mot clé et on a un identificateur `lettre`. On donne la priorité à l'unité syntaxique la plus longue, cf figure 2.3.1.1



2.3.1.2 Sur langage

Éventualité de faire un automate plus petit qui reconnaît L' tel que $L \subseteq L'^1$ avec des actions sémantiques plus « importantes ».

2.3.1.3 Le recul

Si on a `<`, on sait que l'unité syntaxique c'est `<`. Mais si on tombe sur `<=` on a une autre unité syntaxique ; `<=`.

Ce n'est pas réellement un problème car on a vu qu'on privilégie l'Unité Syntaxique(US) la plus longue. Si on tombe sur `< 1`, il faut remettre 1 dans le flot d'entrée

1. C'est-à-dire un sur langage de L

2.3.2 La table des symboles

La table des symboles, appelé TDS, est un endroit où ranger les *tokens* rencontrés avec toutes les informations associées. Cela permet de calcul le « *hashcode* » pour la gestion des synonymes.

3

Analyse syntaxique

Objectif Voir si les *tokens* trouvés par l'AL forment une « phrase correcte » ou non par rapport à la grammaire du langage.

L'analyseur syntaxique est une grammaire reconnue par un automate à pile.

4

Génération de code

A

Rappels

A.1 Grammaire

Une grammaire est fait pour raconter de quelle manières les mots du langages sont construit. On part de l'axiome S et on applique les règles de productions, ou réécriture.

Une grammaire $G = \langle N, X, P, S \rangle$ avec :

N L'ensemble des non terminaux. Majuscules. $\{A, S, B\}$

X Alphabet, ensemble des terminaux. Minuscules $\{a, b, c\}$

P Règles de productions. À gauche on a un seul non terminal $P\{A \rightarrow \alpha, A \in N, \alpha \in (N \cup X)^*\}$

S $S \in N$ Axiome

$$\begin{aligned} G_1 &= \langle N, X, P, S \rangle \\ N &= \{S\} \\ X &= \{a, b\} \\ P &= \{S \rightarrow abS, S \rightarrow a\} \end{aligned}$$

$$\begin{aligned} G_2 &= \langle N, X, P, S \rangle \\ N &= \{S, A\} \\ X &= \{a, b, c\} \\ P &= \{S \rightarrow aAc, \\ &\quad A \rightarrow bbA, \\ &\quad A \rightarrow b\} \end{aligned}$$

$\omega \in X^*$ est un mot engendré par G , $\Rightarrow \omega \in L(G)$, avec $L(G)$ qui est un langage engendré par G .

Avec $G_1 : S \Rightarrow abS \Rightarrow ababS \Rightarrow ababa \in L(G_1)$

A.1.1 Théorèmes

A.1.1.1 Théorème d'Arden

P règle de production G , système d'équation de langages, c'est à dire résoudre $L(G)$.

$X = r_1X + r_2 \Rightarrow X = r_1r_2^*$ est solution. Si $\lambda \notin r_1$ alors la solution est unique.

$$\begin{aligned} G_1 &\rightsquigarrow S = \underbrace{ab}_{r_1}S + \underbrace{a}_{r_2} \\ G_2 &\rightsquigarrow \begin{cases} S = aAc \\ A = \underbrace{bb}_{r_1}A + \underbrace{b}_{r_2} \end{cases} \end{aligned}$$

A.1.1.2 Théorème d'Arden bis

$$X = Xr_1 + r_2 \Rightarrow X = r_2r_1^*$$

A.1.1.3 Théorème A^nB^n

$$X = YXZ + T \Rightarrow X = Y^nTZ^n$$

A.2 Reconnaître un langage avec un automate à pile

Un automate à pile est défini par $\langle Q, X, q_0, \Gamma, Z_0, F \rangle$

Q Ensemble d'état

X Alphabet

$q_0 \in Q$ Etat initial

Γ Alphabet de pile

Z_0 Fond de pile $Z_0 \in \Gamma$

F Ensemble d'états finaux $F \subseteq Q$

σ Fonction de transition

$$S \rightarrow aAc \quad (\text{A.1})$$

$$A \rightarrow bbA \quad (\text{A.2})$$

$$A \rightarrow b \quad (\text{A.3})$$

$$N = S, \quad (\text{A.4})$$

$$X = a, b, c \quad (\text{A.5})$$

$$\sigma(q, \lambda, S) = (q, aAc) \quad (\text{A.6})$$

$$\sigma(q, \lambda, A) = (q, bbA) \quad (\text{A.7})$$

$$\sigma(q, \lambda, A) = (q, b) \quad (\text{A.8})$$

$$\sigma(q, a, a) = (q, \lambda) \quad (\text{A.9})$$

$$\sigma(q, b, b) = (q, \lambda) \quad (\text{A.10})$$

$$\sigma(q, c, c) = (q, \lambda) \quad (\text{A.11})$$

Analyse de la séquence abbbbbc

Ruban	Pile	Règle
$\lambda abbbbbc$	S	A.6
$abbbbbc$	aAc	A.9
$\lambda bbbbbc$	Ac	A.7
$\lambda bbbbbc$	$bbAc$	A.10
$bbbbc$	bAc	A.10
$bbbc$	Ac	A.7
$bbbc$	$bbAc$	A.10
bbc	bAc	A.10
λbc	Ac	A.8
bc	bc	A.10
c	c	A.11
mot lu	pile vide	

L'analyse doit se faire en une seule lecture du ruban et de façon déterministe.

Pour cela, on regarde k symboles sur le ruban pour pouvoir décider de façon unique de la règle à appliquer. Cette analyse efficace est appelée analyse k -prédictive.

On peut écrire un algo déterministe pour la reconnaissance de $\omega \in L(G_2)$

```

Utiliser A.6;
Utiliser A.9 pour dépiler "a"
while il y a "bb" sur le ruban, do:
    Utiliser A.7;
    Utiliser deux fois A.10 pour dépiler les 2 "b"
end

```

B

Table des figures

1.1	Phases de compilation	5
2.1	Diagramme de traduction	6