

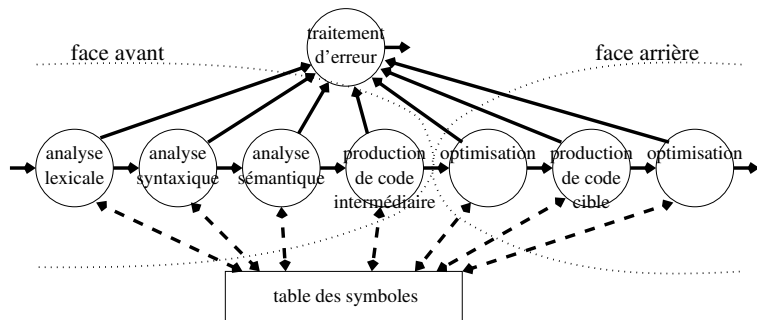
# Programmation Dirigée par la Syntaxe (PDS)

## CM5 - Génération de code

ISTIC, Université de Rennes 1  
Sebastien.Ferre@irisa.fr

PDS, M1 info

# Les phases de la compilation



# Plan

- 1 Introduction
- 2 Machine intermédiaire
  - Opérateurs arithmétiques
  - Opérateurs logiques
  - Opérateurs de comparaison
  - Opérateurs de contrôle
  - Opérateurs d'adressage
- 3 Langage source
- 4 Génération de code 3-adresses
  - Expressions arithmétiques
  - Expressions booléennes
  - Instructions
  - Appel et définitions de fonctions

# Plan

- 1 Introduction
- 2 Machine intermédiaire
  - Opérateurs arithmétiques
  - Opérateurs logiques
  - Opérateurs de comparaison
  - Opérateurs de contrôle
  - Opérateurs d'adressage
- 3 Langage source
- 4 Génération de code 3-adresses
  - Expressions arithmétiques
  - Expressions booléennes
  - Instructions
  - Appel et définitions de fonctions

# Phase : Génération de code cible

## Traduction de l'AST source vers le langage cible

- vient après les analyses sémantiques (**vérification de types**)
- 2 aspects (rappel) :
  - 1 **impératif** : préservation de la **sémantique**  
que ça fasse ce que l'on veut
  - 2 **secondaire** : code généré **efficace** (optimisation)  
que ça aille aussi vite que possible  
que ça consomme le moins de ressources possible

# Motivation du code intermédiaire

Les 2 aspects de la traduction motivent un découpage de la génération de code en 2 phases :

AST décoré  $\xrightarrow{1}$  code intermédiaire  $\xrightarrow{2}$  code cible

- ① se concentre sur la **traduction** proprement dite
  - **machine intermédiaire** offrant une **abstraction** des différentes machines
    - **machine = processeur + mémoire + pile**
  - la programmation de cette phase doit être assez **simple** pour garantir la préservation de la sémantique
    - **approche compositionnelle** (via **ASD** attribuée)
  - on accepte que le code produit soit **inefficace**
- ② spécialisation du code vers une **machine particulière**
  - traduction relativement simple : **mot à mot**

Optimisations possible à chaque niveau

# Motivation du code intermédiaire

Autre bénéfice du découpage en 2 phases :

- factorisation d'un compilateur à l'autre

(1).....

# Nature de la machine intermédiaire

soit **machine** au sens classique

- instructions, emplacements mémoires
- mais **abstrait** des détails et contraintes
- ex : **machine 3-adresses**

soit **langage** pour lequel il existe déjà un compilateur

- (2).....

- implique 2 phases d'analyse syntaxique



# Nature de la machine intermédiaire

soit **machine** au sens classique

- instructions, emplacements mémoires
- mais **abstrait** des détails et contraintes
- ex : **machine 3-adresses**

soit **langage** pour lequel il existe déjà un compilateur

- (2).....

- implique 2 phases d'analyse syntaxique

# Plan

## 1 Introduction

## 2 Machine intermédiaire

- Opérateurs arithmétiques
- Opérateurs logiques
- Opérateurs de comparaison
- Opérateurs de contrôle
- Opérateurs d'adressage

## 3 Langage source

## 4 Génération de code 3-adresses

- Expressions arithmétiques
- Expressions booléennes
- Instructions
- Appel et définitions de fonctions

# Machine intermédiaire

- Machine 3-adresses

- exécutant du code 3-adresses
- code 3-adresses = liste d'instructions 3-adresses

- instruction 3-adresses :  $(op, x, y, z)$

- $op$  : opération élémentaire, type d'instruction
- $x, y, z$  : emplacement mémoires ou registres

indifférenciés à ce stade

- en général :  $x$  = résultat,  $y, z$  = opérandes
- en fait : *au plus* 3-adresses

toutes les opérations n'utilisent pas les 3 adresses

# Opérateurs arithmétiques

## Opérandes et résultats de type entier

nom	instruction	notation	arité
addition	$(+, x, y, z)$	$x = y + z$	binaire
soustraction	$(-, x, y, z)$	$x = y - z$	binaire
multiplication	$(*, x, y, z)$	$x = y * z$	binaire
division	$(/, x, y, z)$	$x = y / z$	binaire
modulo	$(\%, x, y, z)$	$x = y \% z$	binaire
- unaire	$(-UNAIRE, x, y, \_)$	$x = -y$	unaire
constante ( $N \in \mathbb{N}$ )	$(CONST\ N, x, \_, \_)$	$x = N$	0-aire

# Opérateurs logiques

## Opérandes et résultats de type booléen

nom	instruction	notation	arité
et	(AND, $x$ , $y$ , $z$ )	$x = y \text{ and } z$	binaire
ou	(OR, $x$ , $y$ , $z$ )	$x = y \text{ or } z$	binaire
non	(NOT, $x$ , $y$ , $\_$ )	$x = \text{not } y$	unaire
vrai	(CONST true, $x$ , $\_$ , $\_$ )	$x = \text{true}$	0-aire
faux	(CONST false, $x$ , $\_$ , $\_$ )	$x = \text{false}$	0-aire

# Opérateurs de comparaison

Opérandes de type entier et résultat de type booléen

nom	instruction	notation	arité
eq	(EQ, $x, y, z$ )	$x = y == z$	binaire
neq	(NEQ, $x, y, z$ )	$x = y != z$	binaire
leq	(LEQ, $x, y, z$ )	$x = y <= z$	binaire
lt	(LT, $x, y, z$ )	$x = y < z$	binaire
geq	(GEQ, $x, y, z$ )	$x = y >= z$	binaire
gt	(GT, $x, y, z$ )	$x = y > z$	binaire

# Opérateurs de contrôle

nom	instruction	notation	arité
saut inconditionnel	(GOTO $L, \_, \_, \_$ )	goto $L$	0-aire
saut si zero	(IFZ $L, x, \_, \_$ )	ifz $x$ goto $L$	unaire
saut si non-zero	(IFNZ $L, x, \_, \_$ )	ifnz $x$ goto $L$	unaire
entrée fonc/proc	(BEGIN $L, \_, \_, \_$ )	begin $L$	0-aire
retour fonc	(RETURN, $x, \_, \_$ )	return $x$	unaire
retour proc	(RETURN, $\_, \_, \_$ )	return	0-aire
passage argument	(ARG, $x, \_, \_$ )	arg $x$	unaire
appel fonc	(CALL $L, x, \_, \_$ )	$x = \text{call } L$	0-aire
appel proc	(CALL $L, \_, \_, \_$ )	call $L$	0-aire

où  $L$  est une **étiquette de code**

- position dans le code
- constante car connue à la compilation

# Opérateurs d'affectation et d'adressage

nom	instruction	notation	arité
copie	$(=, x, y, \_)$	$x = y$	unaire
déréf. droit	$(*D, x, y, \_)$	$x = *y$	unaire
	si $y$ contient une adresse		
déréf. gauche	$(*G, x, y, \_)$	$*x = y$	binaire
	si $x$ contient une adresse		
“adresse de”	$(\&, x, y, \_)$	$x = \&y$	unaire
	si $y$ est un emplacement mémoire		

Diagrammes mémoire de ces opérateurs : (3).....



# Plan

- 1 Introduction
- 2 Machine intermédiaire
  - Opérateurs arithmétiques
  - Opérateurs logiques
  - Opérateurs de comparaison
  - Opérateurs de contrôle
  - Opérateurs d'adressage
- 3 Langage source**
- 4 Génération de code 3-adresses
  - Expressions arithmétiques
  - Expressions booléennes
  - Instructions
  - Appel et définitions de fonctions

# Langage source

Dans le cadre de ce cours, on considère le langage **BABIL**

- langage **impératif** tel que C ou PASCAL
- contenant les principales constructions
  - expressions arithmétiques et booléennes
  - structures de contrôle : conditionnelles et boucles
  - définitions et appels de fonctions et procédures
- ordre d'évaluation des expressions non fixé (comme en C, Java)
  - sauf pour les expressions booléennes (comme en C, Java)

# Syntaxe abstraite BABIL : types de base et types énumérés

- types de base : *string*, *int*, *bool*
- types synonymes et énumérés

*id* ::= *string*

*op* ::= + | - | \* | /

*rop* ::= == | =< | >= | < | > | !=

*bop* ::= **And** | **Or**

# Syntaxe abstraite BABIL : expressions arith/bool

*expr(ession)* ::= **Const**(*int*)  
                  | **Var**(*id*)  
                  | **Binop**(*op*, *expr*, *expr*)  
                  | **Minus**(*expr*)  
                  | **Call**(*id*, *arg*\*)  
  
*arg(ument)* ::= *expr*  
*cond(ition)* ::= **Bool**(*bool*)  
                  | **Comp**(*rop*, *expr*, *expr*)  
                  | **Logop**(*bop*, *cond*, *cond*)  
                  | **Not**(*cond*)

# Syntaxe abstraite BABIL : programmes, fonctions, instructions

<i>prog(ram)</i>	::=	<b>Prog</b> ( <i>function*</i> )
<i>func(tion)</i>	::=	<b>Func</b> ( <i>id</i> , <i>param*</i> , <i>stat</i> )
<i>param(eter)</i>	::=	<i>id</i>
<i>stat(ement)</i>	::=	<b>Assign</b> ( <i>id</i> , <i>expr</i> )
		<b>IfThenElse</b> ( <i>cond</i> , <i>stat</i> , <i>stat?</i> )
		<b>While</b> ( <i>cond</i> , <i>stat</i> )
		<b>SCall</b> ( <i>id</i> , <i>arg*</i> )
		<b>Return</b> ( <i>expr?</i> )
		<b>Bloc</b> ( <i>stat*</i> )

# Plan

- 1 Introduction
- 2 Machine intermédiaire
  - Opérateurs arithmétiques
  - Opérateurs logiques
  - Opérateurs de comparaison
  - Opérateurs de contrôle
  - Opérateurs d'adressage
- 3 Langage source
- 4 Génération de code 3-adresses**
  - Expressions arithmétiques
  - Expressions booléennes
  - Instructions
  - Appel et définitions de fonctions

# Génération de code 3-adresses

- code 3-adresses = séquence d'instructions  
→ impose de **linéariser** le code

- exemple :  $x + 2 * z$

- 3 opérations :  $+$ ,  $*$ ,  $2$
- donc minimum 3 instructions :  $+$ ,  $*$ ,  $CONST\ 2$
- ordre :  $CONST\ 2 \rightarrow * \rightarrow +$

- code :
 

1	$(CONST\ 2, t1, \_, \_)$	$t1 = 2$
2	$(*, t2, t1, z)$	$t2 = t1 * z$
3	$(+, t3, x, t2)$	$t3 = x + t2$

- $x, y$  : variables du programme source

a priori en mémoire

- $t1, t2, t3$  : variables intermédiaires  
→ inventées (allouées) par le compilateur

a priori en registre

- $t3$  contient le résultat de l'expression

# Génération de code 3-adresses

- code 3-adresses = séquence d'instructions  
→ impose de **linéariser** le code

- exemple :  $x + 2 * z$

- 3 opérations :  $+$ ,  $*$ ,  $2$
- donc minimum 3 instructions :  $+$ ,  $*$ ,  $CONST\ 2$
- ordre :  $CONST\ 2 \rightarrow * \rightarrow +$

- code :
 

1	$(CONST\ 2, t1, \_, \_)$	$t1 = 2$
2	$(*, t2, t1, z)$	$t2 = t1 * z$
3	$(+, t3, x, t2)$	$t3 = x + t2$

- $x, y$  : variables du programme source

a priori en mémoire

- $t1, t2, t3$  : variables intermédiaires  
→ inventées (allouées) par le compilateur

a priori en registre

- $t3$  contient le résultat de l'expression



# Génération de code 3-adresses

- code 3-adresses = séquence d'instructions  
→ impose de **linéariser** le code

- exemple :  $x + 2 * z$

- 3 opérations :  $+$ ,  $*$ ,  $2$
- donc minimum 3 instructions :  $+$ ,  $*$ ,  $CONST\ 2$
- ordre :  $CONST\ 2 \rightarrow * \rightarrow +$

- code : 

1	$(CONST\ 2, t1, \_, \_)$	$t1 = 2$
2	$(*, t2, t1, z)$	$t2 = t1 * z$
3	$(+, t3, x, t2)$	$t3 = x + t2$

- $x, y$  : variables du programme source

a priori en mémoire

- $t1, t2, t3$  : variables intermédiaires  
→ inventées (allouées) par le compilateur

a priori en registre

- $t3$  contient le résultat de l'expression

# Génération de code 3-adresses

Comme pour la vérification de types ou la production de l'AST

- le plus simple est de procéder de **façon compositionnelle**
  - définir une “valeur” (ici : code généré) pour chaque **construction** du langage, en isolation
  - **ASD attribuée** avec comme **attribut synthétisé** principal le **code 3-adresse généré**  $\tau.code$
- + **“prises”** : attributs hérités et synthétisés supplémentaires
  - permettant aux constructions englobantes d'assembler les sous-séquences de code
  - ex :  $t3$  comme **emplacement du résultat de l'expression**  $expr.place$
  - ex :  $t2$  comme **résultat de**  $2*z$ , **utilisé dans**  $x + 2*z$
  - ces prises diffèrent d'un type à l'autre

# Génération de code 3-adresses

Rappel : **impératif** = préserver la sémantique

- génération compositionnelle & systématique (ASD attribuée)
- pas d'optimisation
  - on ne réutilise pas les **variables intermédiaires**  
→ fonction `nouvar()`
  - idem pour les **étiquettes de code**  
→ fonction `nouvetiq()`

# Génération de code 3-adresses

## Remarque

La génération de code est le lieu de rencontre de **3 programmes** :

- 1 le programme **source** (types et variants ASD)
- 2 le programme **générateur** (code du compilateur)
- 3 le programme **généré** (code 3-adresses)

Ils sont entremêlés dans l'ASD attribuée :

- 1 source : variants ASD
- 2 générateur : actions/calculs associés aux variants
- 3 généré : codes 3-adresses manipulés comme valeurs par les actions/calculs

*C'est sans doute la principale difficulté de la compilation !*

# Génération de code pour les expressions arithmétiques

(4).....

# Grammaire attribuée pour les expressions arithmétiques

```

expr ::= Const(int)
        { expr.place := nouvar()
          { expr.code := (CONST int.vallex, expr.place, _, _)
        | Var(id)
          { expr.place := TS.place(id.vallex)
            { expr.code := vide
          | Binop(op, expr1, expr2)
            { expr.place := nouvar()
              { expr.code := expr1.code @@ expr2.code
                @@ (op.vallex, expr.place, expr1.place, expr2.place)
            | Minus(expr1)
              { expr.place := nouvar()
                { expr.code := expr1.code
                  @@ (-UNAIRE, expr.place, expr1.place, _)

```

# Exemple de génération pour les expressions arithmétiques

$E = (a + b) * (a + b)$

(5).....

# Génération de code pour les expressions booléennes

- On peut faire comme pour les expressions arithmétiques
  - attributs *cond.code* et *cond.place*
- mais *cond* utilisé uniquement comme condition de branchement
  - le résultat ne sert qu'à décider du branchement
- de plus, dans  $(c_1 \text{ and } c_2) \text{ or } c_3$ 
  - si  $c_1 = \text{false}$ , inutile d'évaluer  $c_2$
  - si  $c_1 = \text{true}$  et  $c_2 = \text{true}$ , inutile d'évaluer  $c_3$



# Code court-circuit pour les expressions booléennes

On va fait du **code court-circuit**

- branchement vers une étiquette *cond.siVrai* si *cond* = *true* et vers *cond.siFaux* si *cond* = *faux*
  - plus de résultat explicite  
→ l'attribut *cond.place* n'est plus défini
  - *cond.siVrai* et *cond.siFaux* sont des attributs **hérités** qui informant *cond* de "où brancher" en fonction de la valeur de *cond*
- en n'évaluant que ce qui est nécessaire (aspect "**court-circuit**")

# Génération de code court-circuit pour les expressions booléennes

(6).....

# Grammaire attribuée pour les expressions booléennes

```

cond ::= Bool(bool)
      {
        cond.code := si bool.vallex
                     alors (GOTO cond.siVrai, __, __, __)
                     sinon (GOTO cond.siFaux, __, __, __)
      }
      | Comp(rop, expr1, expr2)
      {
        resultat   := nouvar()
        cond.code := expr1.code @@ expr2.code
                     @@ (rop.vallex, resultat, expr1.place, expr2.place)
                     @@ (IFNZ cond.siVrai, resultat, __, __)
                     @@ (GOTO cond.siFaux, __, __, __)
      }
      | Not(cond1)
      {
        cond1.siVrai   := cond.siFaux
        cond1.siFaux   := cond.siVrai
        cond.code      := cond1.code
      }
  
```

# Grammaire attribuée pour les expressions booléennes

**Logop(And,  $cond_1$ ,  $cond_2$ )**

{	$cond_1.siFaux$	$:=$	$cond.siFaux$
	$cond_2.siFaux$	$:=$	$cond.siFaux$
	$cond_1.siVrai$	$:=$	$nouveliq()$
	$cond_2.siVrai$	$:=$	$cond.siVrai$
	$cond.code$	$:=$	$cond_1.code$
			@@ $cond_2.code$

**Logop(Or,  $cond_1$ ,  $cond_2$ )**

{	$cond_1.siFaux$	$:=$	$nouveliq()$
	$cond_2.siFaux$	$:=$	$cond.siFaux$
	$cond_1.siVrai$	$:=$	$cond.siVrai$
	$cond_2.siVrai$	$:=$	$cond.siVrai$
	$cond.code$	$:=$	$cond_1.code$
			@@ $cond_2.code$

# Exemple de génération pour les expressions booléennes

$B = (a = d \text{ and } b > e) \text{ or } c < f$   
(7).....

# Génération de code pour les instructions

(8).....

# Grammaire attribuée pour les instructions

```

stat ::= Assign(id, expr)
      {
        stat.code := expr.code
        @@ (=, TS.place(id.vallex), expr.place, _)
      }
| IfThenElse(cond, stat1, ε)
  {
    cond.siVrai := alors := nouveliq()
    cond.siFaux := fin := nouveliq()
    stat.code := cond.code
    @@ (LABEL alors, __, __, __) @@ stat1.code
    @@ (LABEL fin, __, __, __)
  }
| IfThenElse(cond, stat1, stat2)
  {
    cond.siVrai := alors := nouveliq()
    cond.siFaux := sinon := nouveliq()
    fin := nouveliq()
    stat.code := cond.code
    @@ (LABEL alors, __, __, __) @@ stat1.code
    @@ (GOTO fin, __, __, __)
    @@ (LABEL sinon, __, __, __) @@ stat2.code
    @@ (LABEL fin, __, __, __)
  }

```

| **While**(*cond*, *stat*<sub>1</sub>)

{	<i>boucle</i>	:=	<i>nouvetiq</i> ()
	<i>cond.siVrai</i>	:=	<i>corps</i> := <i>nouvetiq</i> ()
	<i>cond.siFaux</i>	:=	<i>sortie</i> := <i>nouvetiq</i> ()
	<i>stat.code</i>	:=	(LABEL <i>boucle</i> , __, __, __) @@ <i>cond.code</i>
		@@	(LABEL <i>corps</i> , __, __, __) @@ <i>stat</i> <sub>1</sub> .code
		@@	(GOTO <i>boucle</i> , __, __, __)
		@@	(LABEL <i>sortie</i> , __, __, __)

| **Bloc**(*stat*\*)

{ *stat.code* := @@(*stat* \* .code)



# Exemple de génération pour les instructions

```
S = i = 1; f = 1; while i < n and f < 100 do i = i + 1; f =  
f * i done  
(9).....
```

# Appels de fonctions/procédures

- 2 constructions
  - $expr ::= \mathbf{Call}(id, arg*)$  : appel de fonction (expression)
  - $stat ::= \mathbf{SCall}(id, arg*)$  : appel de procédure (instruction)
- il existe 2 sémantiques du **passage de paramètres**
  - **par nom** : nom d'une mémoire, emplacement
    - permet à la fonction de modifier le contenu de cette mémoire
    - ex : **paramètre var de PASCAL**
    - ex : **références sur objets en JAVA**
  - **par valeur** : contenu d'une mémoire
    - la valeur du paramètre est recopiée dans une zone mémoire réservée aux arguments **sur la pile ou dans des registres**
    - ne permet pas de modifier l'argument **mais l'argument peut être un pointeur...**
    - c'est la sémantique de C... et de BABIL
- c'est l'appelant qui a la charge d'installer les paramètres avant d'appeler la fonction
  - opérateur ( $ARG, x, \_, \_$ )

# Appels de fonctions/procédures

- 2 constructions
  - $expr ::= \mathbf{Call}(id, arg*)$  : appel de fonction (expression)
  - $stat ::= \mathbf{SCall}(id, arg*)$  : appel de procédure (instruction)
- il existe 2 sémantiques du **passage de paramètres**
  - **par nom** : nom d'une mémoire, emplacement
    - permet à la fonction de modifier le contenu de cette mémoire
    - ex : **paramètre var de PASCAL**
    - ex : **références sur objets en JAVA**
  - **par valeur** : contenu d'une mémoire
    - la valeur du paramètre est recopiée dans une zone mémoire réservée aux arguments **sur la pile ou dans des registres**
    - ne permet pas de modifier l'argument **mais l'argument peut être un pointeur...**
    - c'est la sémantique de C... et de BABIL
- c'est l'appelant qui a la charge d'installer les paramètres avant d'appeler la fonction
  - opérateur ( $ARG, x, \_, \_$ )

# Génération de code pour les appels de fonctions

(10).....

# Grammaire attribuée pour les appels de fonctions

$$\begin{aligned}
 \text{expr} &::= \mathbf{Call}(id, arg*) \\
 &\quad \left\{ \begin{array}{ll} \text{expr.place} &:= \text{nouvar}() \\ \text{expr.code} &:= @@(arg*.code) \\ &@@ (\text{CALL } TS.label(id.vallex), \text{expr.place}, \_, \_) \end{array} \right. \\
 \\
 \text{stat} &::= \mathbf{SCall}(id, arg*) \\
 &\quad \left\{ \begin{array}{ll} \text{stat.code} &:= @@(arg*.code) \\ &@@ (\text{CALL } TS.label(id.vallex), \_, \_, \_) \end{array} \right. \\
 \\
 \text{arg} &::= \text{expr} \\
 &\quad \left\{ \begin{array}{ll} \text{arg.code} &:= \text{expr.code} @@ (\text{ARG}, \text{expr.place}, \_, \_) \end{array} \right.
 \end{aligned}$$

# Génération de code pour les définitions de fonctions

(11).....

# Grammaire attribuée pour les définitions de fonctions

```
stat ::= Return( $\epsilon$ )  
      { stat.code := (RETURN, __, __, __)  
      | Return(expr)  
      { stat.code := expr.code  
      { @@ (RETURN, expr.place, __, __)  
  
func ::= Func(id, param*, stat)  
      { func.code := (BEGINFUNC TS.label(id.vallex), __, __, __)  
      { @@ stat.code  
  
prog ::= Prog(func*)  
      { prog.code := @@(func * .code)
```

# Exemple de génération pour les fonctions

**Exemple :** `define fact(n) if n==0 then return 1 else  
return n * fact(n-1) end  
(12).....`