

# Compiladores

## Introdução à Compilação

**Prof. Edson Alves**

Faculdade UnB Gama

# Sumário

1. Introdução
2. Análise do programa fonte
3. Fases de um compilador

# Histórico

- ▶ Os primeiros compiladores surgiram na década de 50

# Histórico

- ▶ Os primeiros compiladores surgiram na década de 50
- ▶ Não há registros preciso de qual foi o primeiro compilador

# Histórico

- ▶ Os primeiros compiladores surgiram na década de 50
- ▶ Não há registros preciso de qual foi o primeiro compilador
- ▶ Os primeiros compiladores lidavam com a tradução de fórmulas aritméticas (FORTRAN – *Formula Translator*)

# Histórico

- ▶ Os primeiros compiladores surgiram na década de 50
- ▶ Não há registros preciso de qual foi o primeiro compilador
- ▶ Os primeiros compiladores lidavam com a tradução de fórmulas aritméticas (FORTRAN – *Formula Translator*)
- ▶ Os compiladores eram considerados programas difíceis de se escrever

# Histórico

- ▶ Os primeiros compiladores surgiram na década de 50
- ▶ Não há registros preciso de qual foi o primeiro compilador
- ▶ Os primeiros compiladores lidavam com a tradução de fórmulas aritméticas (FORTRAN – *Formula Translator*)
- ▶ Os compiladores eram considerados programas difíceis de se escrever
- ▶ O primeiro compilador Fortran levou 18 homens-ano para ser escrito (1 homen-ano  $\approx$  2.080 horas)

# Histórico

- ▶ Os primeiros compiladores surgiram na década de 50
- ▶ Não há registros preciso de qual foi o primeiro compilador
- ▶ Os primeiros compiladores lidavam com a tradução de fórmulas aritméticas (FORTRAN – *Formula Translator*)
- ▶ Os compiladores eram considerados programas difíceis de se escrever
- ▶ O primeiro compilador Fortran levou 18 homens-ano para ser escrito (1 homen-ano  $\approx$  2.080 horas)
- ▶ Embora continue não sendo uma tarefa não trivial, a escrita de compiladores se beneficiou dos avanços da área desde então



# Definição

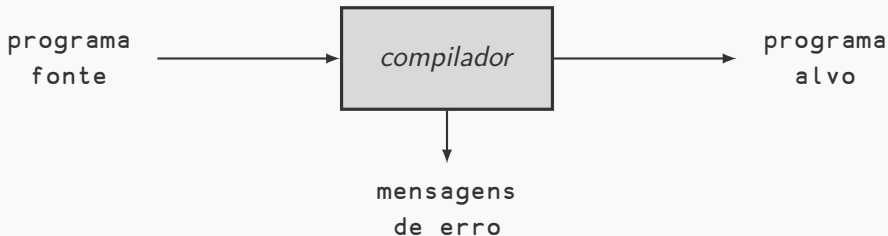
## Definição de compilador (informal)

Um compilador é um programa que lê um programa escrito em uma linguagem (linguagem fonte) e o traduz para uma outra linguagem (linguagem alvo).

# Definição

## Definição de compilador (informal)

Um compilador é um programa que lê um programa escrito em uma linguagem (linguagem fonte) e o traduz para uma outra linguagem (linguagem alvo).



# Características dos compiladores

- ▶ O processo de compilação deve identificar e relatar possíveis erros no programa fonte

# Características dos compiladores

- ▶ O processo de compilação deve identificar e relatar possíveis erros no programa fonte
- ▶ Em geral, as linguagens fonte são linguagens de programação tradicionais (C/C++, Java, Python, etc)

# Características dos compiladores

- ▶ O processo de compilação deve identificar e relatar possíveis erros no programa fonte
- ▶ Em geral, as linguagens fonte são linguagens de programação tradicionais (C/C++, Java, Python, etc)
- ▶ As linguagens alvo podem ser tanto linguagens tradicionais quanto linguagens de máquina

## Características dos compiladores

- ▶ O processo de compilação deve identificar e relatar possíveis erros no programa fonte
- ▶ Em geral, as linguagens fonte são linguagens de programação tradicionais (C/C++, Java, Python, etc)
- ▶ As linguagens alvo podem ser tanto linguagens tradicionais quanto linguagens de máquina
- ▶ Os compiladores podem ser classificados de diversas formas, dependendo de seu objetivo ou como foi construído (de uma passagem, múltiplas passagens, depuradores, etc)

## Criação do programa executável

- ▶ Além do compilador, outros programas podem ser usados na criação do programa executável

## Criação do programa executável

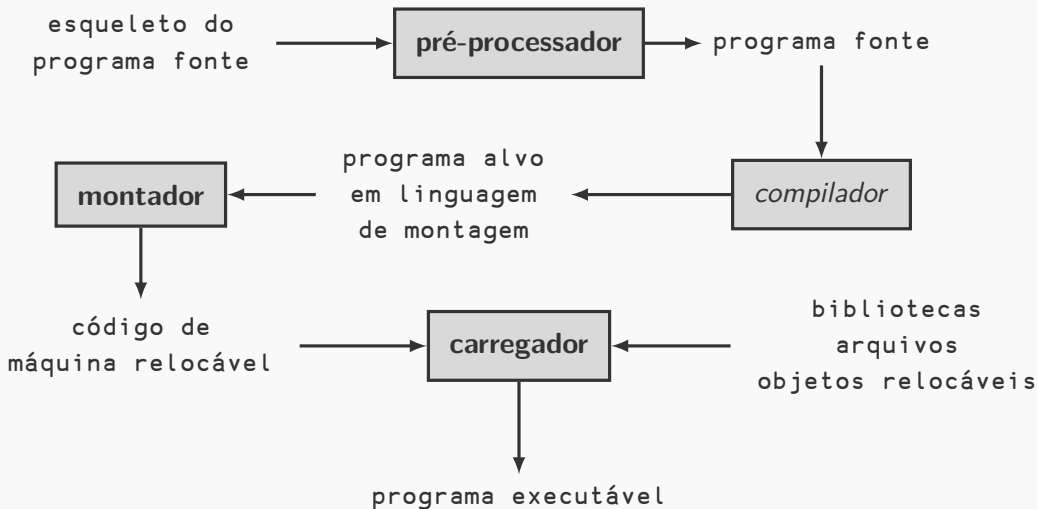
- ▶ Além do compilador, outros programas podem ser usados na criação do programa executável
- ▶ Antes de ser passado para o compilador, o programa alvo pode ser pré-processado (por exemplo, o pré-processador da linguagem C processa as diretivas, como `#include` e `#define`)



## Criação do programa executável

- ▶ Além do compilador, outros programas podem ser usados na criação do programa executável
- ▶ Antes de ser passado para o compilador, o programa alvo pode ser pré-processado (por exemplo, o pré-processador da linguagem C processa as diretivas, como `#include` e `#define`)
- ▶ Após a compilação, o programa alvo pode demandar processamento adicional para a construção do executável (novamente no caso da linguagem C, temos o montador e o *linkeditor*)

## Exemplo de fluxo de geração de um programa executável



# Análise e síntese

- ▶ A compilação é composta por duas partes: análise e síntese

# Análise e síntese

- ▶ A compilação é composta por duas partes: análise e síntese
- ▶ A análise divide o programa fonte em partes constituintes e as organiza em uma representação intermediária

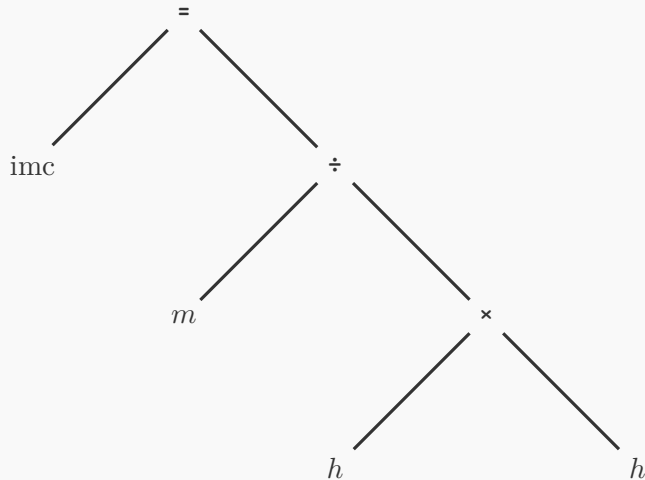
# Análise e síntese

- ▶ A compilação é composta por duas partes: análise e síntese
- ▶ A análise divide o programa fonte em partes constituintes e as organiza em uma representação intermediária
- ▶ Em geral, a representação intermediária consiste em uma árvore sintática, onde cada nó representa uma operação e cada filho representa um operando

# Análise e síntese

- ▶ A compilação é composta por duas partes: análise e síntese
- ▶ A análise divide o programa fonte em partes constituintes e as organiza em uma representação intermediária
- ▶ Em geral, a representação intermediária consiste em uma árvore sintática, onde cada nó representa uma operação e cada filho representa um operando
- ▶ A síntese constrói o programa alvo a partir desta representação intermediária

## Exemplo de árvore sintática



**Figura:** Árvore sintática da fórmula  $imc = m/h^2$ .

# Análise do programa fonte

A análise é composta por três fases:



# Análise do programa fonte

A análise é composta por três fases:

1. *análise linear*: o fluxo de caracteres que compõem o programa alvo é lido, da esquerda para direita, e agrupado em *tokens* (sequência de caracteres com significado coletivo)

# Análise do programa fonte

A análise é composta por três fases:

1. *análise linear*: o fluxo de caracteres que compõem o programa alvo é lido, da esquerda para direita, e agrupado em *tokens* (sequência de caracteres com significado coletivo)
2. *análise hierárquica*: os *tokens* são ordenados hierarquicamente em coleções aninhadas com significado coletivo

# Análise do programa fonte

A análise é composta por três fases:

1. *análise linear*: o fluxo de caracteres que compõem o programa alvo é lido, da esquerda para direita, e agrupado em *tokens* (sequência de caracteres com significado coletivo)
2. *análise hierárquica*: os *tokens* são ordenados hierarquicamente em coleções aninhadas com significado coletivo
3. *análise semântica*: verificação que garante que os componentes do programa se combinam de forma significativa

# Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento.

## Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

## Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

1. o identificador `C`

## Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

1. o identificador `C`
2. o símbolo de atribuição `=`

## Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

1. o identificador `C`
2. o símbolo de atribuição `=`
3. a constante em ponto flutuante `1.8`



## Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

1. o identificador `C`
2. o símbolo de atribuição `=`
3. a constante em ponto flutuante `1.8`
4. o símbolo de multiplicação `*`

## Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

1. o identificador C
2. o símbolo de atribuição =
3. a constante em ponto flutuante 1.8
4. o símbolo de multiplicação \*
5. o identificador F

## Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

1. o identificador `C`
2. o símbolo de atribuição `=`
3. a constante em ponto flutuante `1.8`
4. o símbolo de multiplicação `*`
5. o identificador `F`
6. o símbolo de adição `+`

# Análise léxica

Em um compilador, a análise linear também é denominada análise léxica ou esquadrinhamento. Por exemplo, no enunciado

$$C = 1.8 * F + 32$$

a análise léxica identificaria os seguintes *tokens*:

1. o identificador `C`
2. o símbolo de atribuição `=`
3. a constante em ponto flutuante `1.8`
4. o símbolo de multiplicação `*`
5. o identificador `F`
6. o símbolo de adição `+`
7. a constante inteira `32`

# Análise sintática

A análise hierárquica também é denominada análise sintática ou gramatical. Ela agrupa os *tokens* hierarquicamente, em geral em uma árvore gramatical.

## Análise sintática

A análise hierárquica também é denominada análise sintática ou gramatical. Ela agrupa os *tokens* hierarquicamente, em geral em uma árvore gramatical.

A estrutura hierárquica pode ser definida por meio de regras recursivas.

## Análise sintática

A análise hierárquica também é denominada análise sintática ou gramatical. Ela agrupa os *tokens* hierarquicamente, em geral em uma árvore gramatical.

A estrutura hierárquica pode ser definida por meio de regras recursivas. Por exemplo, considere as seguintes regras:

# Análise sintática

A análise hierárquica também é denominada análise sintática ou gramatical. Ela agrupa os *tokens* hierarquicamente, em geral em uma árvore gramatical.

A estrutura hierárquica pode ser definida por meio de regras recursivas. Por exemplo, considere as seguintes regras:

1. qualquer identificador é uma expressão



# Análise sintática

A análise hierárquica também é denominada análise sintática ou gramatical. Ela agrupa os *tokens* hierarquicamente, em geral em uma árvore gramatical.

A estrutura hierárquica pode ser definida por meio de regras recursivas. Por exemplo, considere as seguintes regras:

1. qualquer identificador é uma expressão
2. qualquer número é uma expressão

## Análise sintática

A análise hierárquica também é denominada análise sintática ou gramatical. Ela agrupa os *tokens* hierarquicamente, em geral em uma árvore gramatical.

A estrutura hierárquica pode ser definida por meio de regras recursivas. Por exemplo, considere as seguintes regras:

1. qualquer identificador é uma expressão
2. qualquer número é uma expressão
3. se  $E_1$  e  $E_2$  são expressões, também são expressões  $E_1 + E_2$  e  $E_1 * E_2$

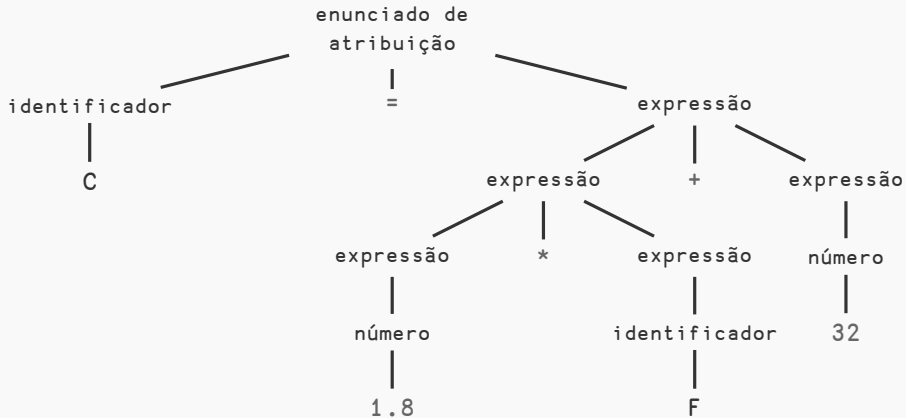
# Análise sintática

A análise hierárquica também é denominada análise sintática ou gramatical. Ela agrupa os *tokens* hierarquicamente, em geral em uma árvore gramatical.

A estrutura hierárquica pode ser definida por meio de regras recursivas. Por exemplo, considere as seguintes regras:

1. qualquer identificador é uma expressão
2. qualquer número é uma expressão
3. se  $E_1$  e  $E_2$  são expressões, também são expressões  $E_1 + E_2$  e  $E_1 * E_2$
4. se  $I$  é um identificador e  $E$  uma expressão, então  $I = E$  é um enunciado

## Exemplo de árvore gramatical



**Figura:** Árvore gramatical do enunciado  $C = 1.8 * F + 32$

# Análise semântica

- ▶ A análise semântica verifica potenciais erros semânticos no programa alvo

# Análise semântica

- ▶ A análise semântica verifica potenciais erros semânticos no programa alvo
- ▶ Ela usa a árvore da análise sintática para identificar operadores e operandos das expressões e enuncados

# Análise semântica

- ▶ A análise semântica verifica potenciais erros semânticos no programa alvo
- ▶ Ela usa a árvore da análise sintática para identificar operadores e operandos das expressões e enuncados
- ▶ Ela também faz a verificação de tipos

# Análise semântica

- ▶ A análise semântica verifica potenciais erros semânticos no programa alvo
- ▶ Ela usa a árvore da análise sintática para identificar operadores e operandos das expressões e enuncados
- ▶ Ela também faz a verificação de tipos
- ▶ Caso os tipos dos operandos não sejam compatíveis com os tipos esperados pelos operadores, esta análise ou retorna um erro ou adicionar uma promoção (ou conversão) de tipo, a depender da linguagem alvo



# Análise semântica

- ▶ A análise semântica verifica potenciais erros semânticos no programa alvo
- ▶ Ela usa a árvore da análise sintática para identificar operadores e operandos das expressões e enunciados
- ▶ Ela também faz a verificação de tipos
- ▶ Caso os tipos dos operandos não sejam compatíveis com os tipos esperados pelos operadores, esta análise ou retorna um erro ou adicionar uma promoção (ou conversão) de tipo, a depender da linguagem alvo
- ▶ Por exemplo, na expressão à direita do enunciado  $C = 1.8 * F + 32$ , o operando à esquerda da soma tem tipo ponto flutuante e o da direita tipo inteiro: o valor 32 deve ser promovido para ponto flutuante ou deve ser sinalizado um erro de tipo

# As fases de um compilador

---

- ▶ Conceitualmente, o compilador opera em fases

# As fases de um compilador

---

- ▶ Conceitualmente, o compilador opera em fases
- ▶ Cada fase manipula o programa fonte e entrega o resultado para a próxima fase

# As fases de um compilador

- ▶ Conceitualmente, o compilador opera em fases
- ▶ Cada fase manipula o programa fonte e entrega o resultado para a próxima fase
- ▶ Na prática, algumas fases podem ser agrupadas, e a representação intermediária entre elas pode ser não se construída explicitamente

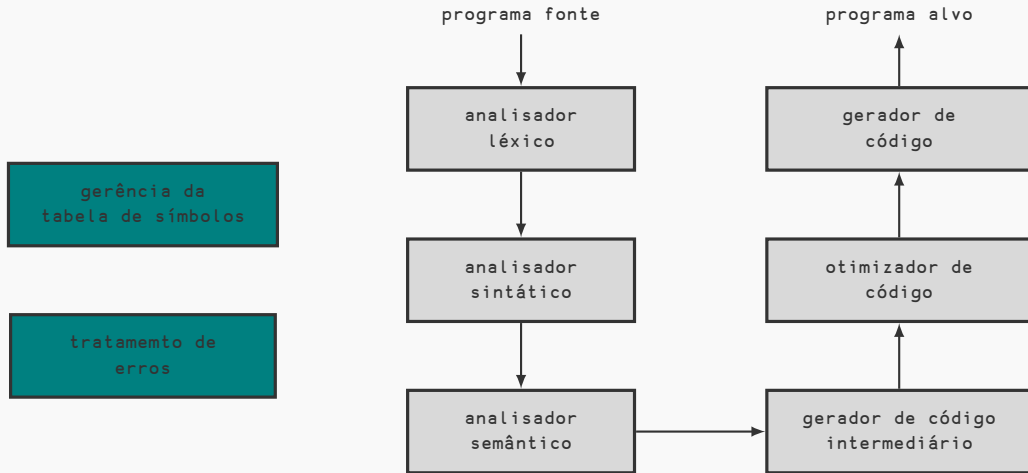
# As fases de um compilador

- ▶ Conceitualmente, o compilador opera em fases
- ▶ Cada fase manipula o programa fonte e entrega o resultado para a próxima fase
- ▶ Na prática, algumas fases podem ser agrupadas, e a representação intermediária entre elas pode ser não se construída explicitamente
- ▶ As primeiras fases estão relacionadas à análise do programa fonte, as últimas estão relacionadas à síntese (construção do programa alvo)

# As fases de um compilador

- ▶ Conceitualmente, o compilador opera em fases
- ▶ Cada fase manipula o programa fonte e entrega o resultado para a próxima fase
- ▶ Na prática, algumas fases podem ser agrupadas, e a representação intermediária entre elas pode ser não se construída explicitamente
- ▶ As primeiras fases estão relacionadas à análise do programa fonte, as últimas estão relacionadas à síntese (construção do programa alvo)
- ▶ Duas atividades interagem com todas as fases: a gerência da tabela de símbolos e o tratamento de erros

# Representação típica das fases de um compilador



## Gerenciamento da tabela de símbolos

- ▶ Esta atividade registra os identificadores do programa alvo e identifica seus diversos atributos



# Gerenciamento da tabela de símbolos

- ▶ Esta atividade registra os identificadores do programa alvo e identifica seus diversos atributos
- ▶ Exemplos de possíveis atributos de um identificador: nome, tipo, memória e escopo

## Gerenciamento da tabela de símbolos

- ▶ Esta atividade registra os identificadores do programa alvo e identifica seus diversos atributos
- ▶ Exemplos de possíveis atributos de um identificador: nome, tipo, memória e escopo
- ▶ Caso o identificador se refira a um procedimento, dentre seus atributos devem constar a quantidade de seus parâmetros e respectivos tipos, modo de passagem (cópia ou referência) e o tipo do retorno, se houver

## Gerenciamento da tabela de símbolos

- ▶ Esta atividade registra os identificadores do programa alvo e identifica seus diversos atributos
- ▶ Exemplos de possíveis atributos de um identificador: nome, tipo, memória e escopo
- ▶ Caso o identificador se refira a um procedimento, dentre seus atributos devem constar a quantidade de seus parâmetros e respectivos tipos, modo de passagem (cópia ou referência) e o tipo do retorno, se houver
- ▶ Os identificadores e seus respectivos atributos são armazenados em uma estrutura denominada tabela de símbolos

# Tratamento de erros

- ▶ A cada fase da compilação podem acontecer um ou mais erros

# Tratamento de erros

- ▶ A cada fase da compilação podem acontecer um ou mais erros
- ▶ Após a identificação do erro, o compilador deve tratá-lo de alguma maneira e, se possível, continuar o processo em busca de outros erros

## Tratamento de erros

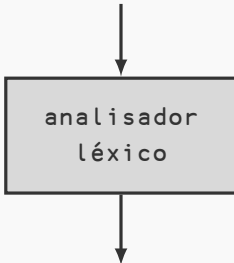
- ▶ A cada fase da compilação podem acontecer um ou mais erros
- ▶ Após a identificação do erro, o compilador deve tratá-lo de alguma maneira e, se possível, continuar o processo em busca de outros erros
- ▶ Abortar a compilação logo no primeiro erro pode diminuir a utilidade do compilador (por exemplo, o prosseguimento da compilação após um erro léxico pode ajudar na geração de uma sugestão de correção para o erro)

## Tratamento de erros

- ▶ A cada fase da compilação podem acontecer um ou mais erros
- ▶ Após a identificação do erro, o compilador deve tratá-lo de alguma maneira e, se possível, continuar o processo em busca de outros erros
- ▶ Abortar a compilação logo no primeiro erro pode diminuir a utilidade do compilador (por exemplo, o prosseguimento da compilação após um erro léxico pode ajudar na geração de uma sugestão de correção para o erro)
- ▶ As análises sintática e semântica podem identificar uma parcela considerável dos erros no programa fonte

## Exemplo da parte da análise do programa fonte

`celsius = 1.8 * fahrenheit + 32`

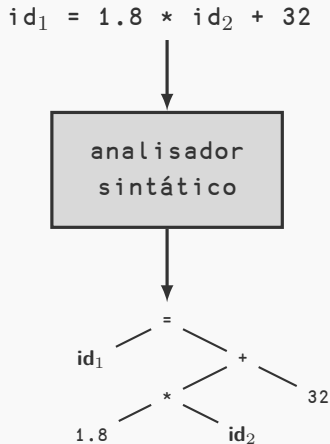


`id1 = 1.8 * id2 + 32`

**Figura:** Análise léxica do enunciado `celsius = 1.8 * fahrenheit + 32`

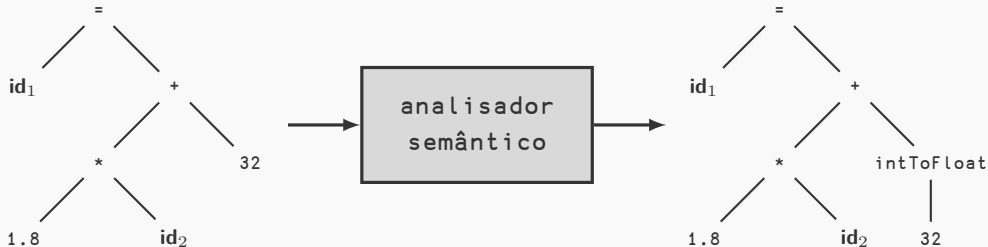


## Exemplo da parte da análise do programa fonte



**Figura:** Análise sintática

## Exemplo da parte da análise do programa fonte



**Figura:** Análise semântica

## Geração de código intermediário

- ▶ A árvore resultante da análise semântica é transformada pelo compilador em um código intermediário

## Geração de código intermediário

- ▶ A árvore resultante da análise semântica é transformada pelo compilador em um código intermediário
- ▶ Esta representação pode ser entendida como um código para uma máquina abstrata

## Geração de código intermediário

- ▶ A árvore resultante da análise semântica é transformada pelo compilador em um código intermediário
- ▶ Esta representação pode ser entendida como um código para uma máquina abstrata
- ▶ O código intermediário deve ter duas qualidades fundamentais:

## Geração de código intermediário

- ▶ A árvore resultante da análise semântica é transformada pelo compilador em um código intermediário
- ▶ Esta representação pode ser entendida como um código para uma máquina abstrata
- ▶ O código intermediário deve ter duas qualidades fundamentais:
  1. deve ser fácil de gerar

## Geração de código intermediário

- ▶ A árvore resultante da análise semântica é transformada pelo compilador em um código intermediário
- ▶ Esta representação pode ser entendida como um código para uma máquina abstrata
- ▶ O código intermediário deve ter duas qualidades fundamentais:
  1. deve ser fácil de gerar
  2. deve ser fácil de traduzir para o programa alvo

## Geração de código intermediário

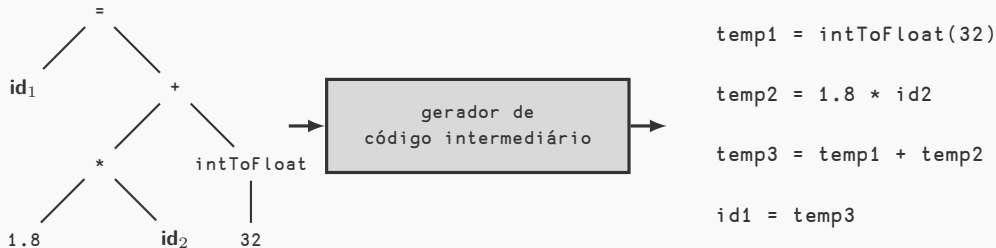
- ▶ A árvore resultante da análise semântica é transformada pelo compilador em um código intermediário
- ▶ Esta representação pode ser entendida como um código para uma máquina abstrata
- ▶ O código intermediário deve ter duas qualidades fundamentais:
  1. deve ser fácil de gerar
  2. deve ser fácil de traduzir para o programa alvo
- ▶ Uma representação possível é o código de três endereços



## Geração de código intermediário

- ▶ A árvore resultante da análise semântica é transformada pelo compilador em um código intermediário
- ▶ Esta representação pode ser entendida como um código para uma máquina abstrata
- ▶ O código intermediário deve ter duas qualidades fundamentais:
  1. deve ser fácil de gerar
  2. deve ser fácil de traduzir para o programa alvo
- ▶ Uma representação possível é o código de três endereços
- ▶ Além de computar expressões, esta representação também precisa tratar dos fluxos de controle e das chamadas de procedimentos

## Exemplo de geração de código intermediário



**Figura:** Representação por código de três endereços

# Otimização do código

- ▶ Esta fase procura formas de melhorar o código intermediário, com o intuito de melhorar a performance do código de máquina do programa alvo

# Otimização do código

- ▶ Esta fase procura formas de melhorar o código intermediário, com o intuito de melhorar a performance do código de máquina do programa alvo
- ▶ Algumas otimizações são triviais, outras demandam algoritmos sofisticados, impactando no tempo de compilação

# Otimização do código

- ▶ Esta fase procura formas de melhorar o código intermediário, com o intuito de melhorar a performance do código de máquina do programa alvo
- ▶ Algumas otimizações são triviais, outras demandam algoritmos sofisticados, impactando no tempo de compilação
- ▶ As otimizações não devem alterar a semântica do código intermediário

# Otimização do código

- ▶ Esta fase procura formas de melhorar o código intermediário, com o intuito de melhorar a performance do código de máquina do programa alvo
- ▶ Algumas otimizações são triviais, outras demandam algoritmos sofisticados, impactando no tempo de compilação
- ▶ As otimizações não devem alterar a semântica do código intermediário
- ▶ As otimizações podem melhorar, além do tempo de execução, o uso de memória do programa alvo

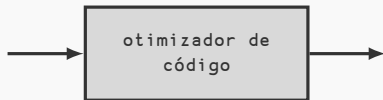
## Exemplo de otimização do código intermediário

```
temp1 = intToFloat(32)
```

```
temp2 = 1.8 * id2
```

```
temp3 = temp1 + temp2
```

```
id1 = temp3
```



```
temp1 = 1.8 * id2
```

```
id1 = temp1 + 32.0
```

**Figura:** Otimização

## Referências

1. APL Wiki. [Quad name](#), acesso em 23/09/2021.
2. **BROCKLEBANK**, Daniel. *APL – The Language*, John Hopkins APL Technical Digest, vol. 5, number 3, 1984.
3. Dyalog. [Dyalog APL Version 12.0](#), acesso em 23/09/2021.
4. Dyalog. [Try APL](#), acesso em 23/09/2021.
5. **IVERSON**, Kenneth E. *A Programming Language*, John Wiley and Sons, 1962.
6. Stack Overflow. [Add APL Keyboard Layout On Linux 20.04](#), acesso em 23/09/2021.
7. **ULMANN**, Bernd. *APL - One of the Greatest Programming Languages Ever*, Vintage Computer Festival Europe 2007.
8. Xah Lee. [Unicode APL Symbols](#), acesso em 23/09/2021.
9. Wikipédia. [APL \(programming language\)](#), acesso em 24/09/2021.
10. Wikipédia. [Array programming](#), acesso em 24/09/2021.