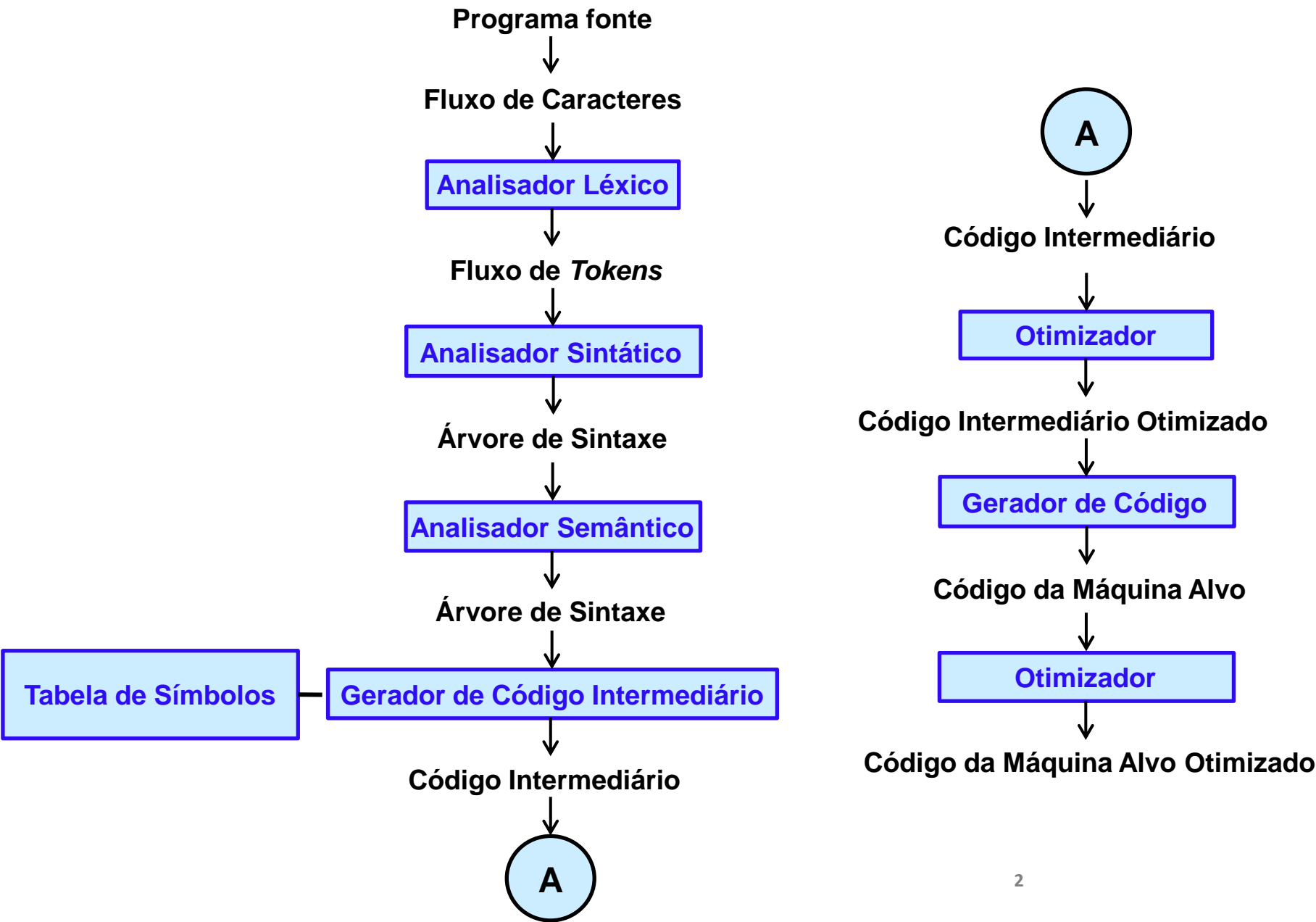
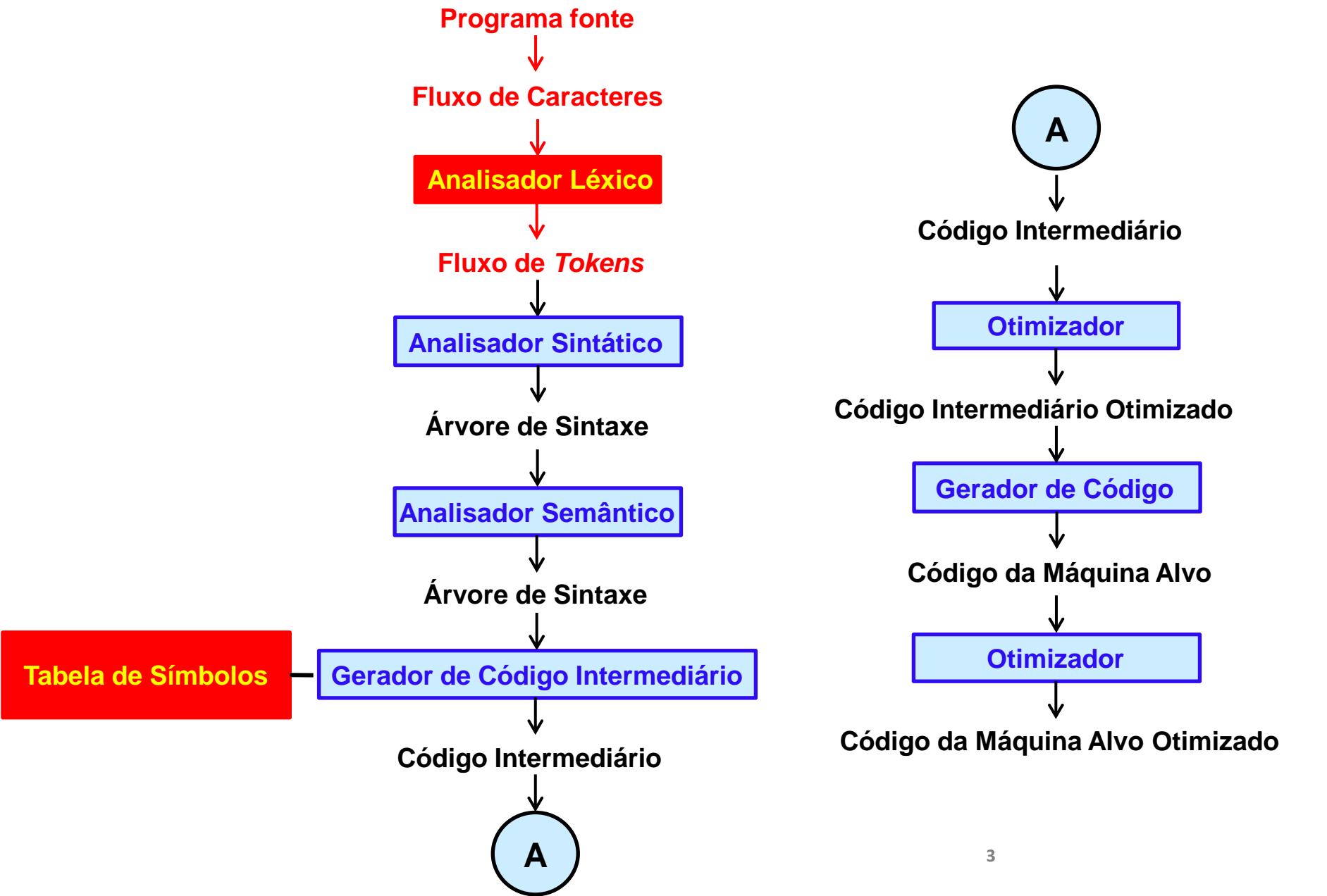


Linguagens, Autômatos e Computabilidade - LAC

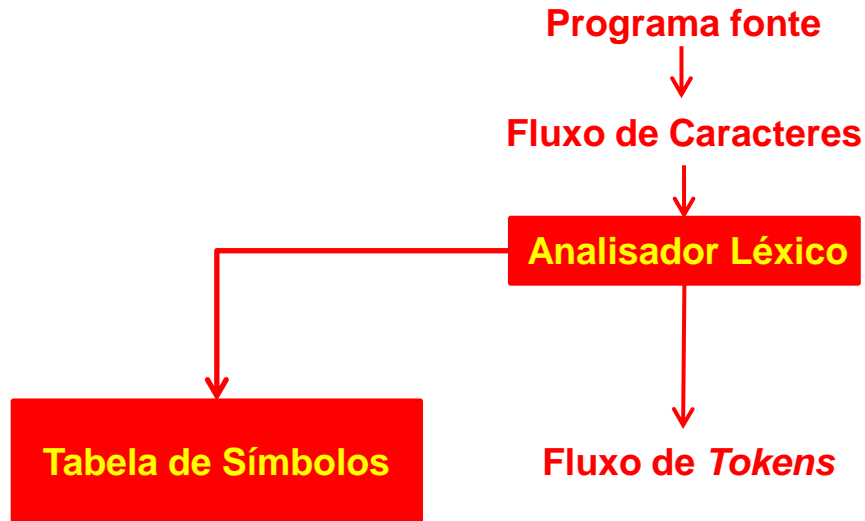
Prof. MSc. Fabio H. Pimentel
proffabio.pimentel@fiap.com.br

Análise Léxica - Parte 2





ANÁLISE LÉXICA (SÍMBOLOS)



Leitura do fluxo de caracteres do programa fonte, criação de **lexemas**, produção de sequencias de **tokens** para cada lexema e montagem da **Tabela de Símbolos**.

ANÁLISE LÉXICA

Demais atividades:

Remoção de:

- Comentários

- Espaços em branco

- Quebra de linha

- Tabulações

- E demais caracteres que não dizem respeito ao programa fonte em si

Correlacionar as mensagens de erros geradas pelo compilador com o programa fonte

Pode registrar o número de caracteres de quebra de linha para poder associar o número da linha a cada mensagem de erro.

ANÁLISE LÉXICA: TOKENS, PADRÕES e LEXEMAS

Tokens

É um par consistindo de um **nome** e um **valor atribuído**.

O **nome** do token é um **símbolo** que representa um **tipo de unidade léxica**.

Exemplo: uma palavra-chave, uma sequência de caracteres.

Os nomes dos tokens **são símbolos da entrada** que o analisador sintático processa.

Os nomes dos tokens **são colocados em negrito**.

Padrões

Descrição das formas que os lexemas de um token podem assumir.

Exemplos: Token **palavra-chave**, o padrão é uma sequência de caracteres esperada.

Token **identificador**, o padrão é a uma letra ou uma sequência de letras seguidas de letras ou de números.

ANÁLISE LÉXICA: TOKENS, PADRÕES e LEXEMAS

Lexemas

É a sequência de caracteres do programa fonte **que casa** com o padrão para algum token.

Exemplos:

TOKEN	PADRÃO	LEXEMA
id	letra seguida por letras ou/e dígitos	printf, placar, pi, D2
number	Qualquer constante numérica	3.14159, 0.6, 2e23
literal	qualquer caractere diferente de “, cercado por “ ”	“O total foi”, “Oi”
if	palavra-chave de caracteres i, f	if
else	palavra-chave de caracteres e, l, s, e	else
comparison	operadores relacionais < > <= >= == !=	<, >, <=, >=, ==, !=
end of statement	;	;

ANÁLISE LÉXICA: TOKENS, PADRÕES e LEXEMAS

Classes dos Tokens

1. Um token para cada palavra-chave.

O padrão é a própria palavra-chave

2. Tokens para operadores

Podem ser individualmente, ou em classes

3. Um token representando todos os identificadores

4. Um ou mais tokens representando constantes

Números e cadeias de literais

5. Tokens para cada símbolo de pontuação

Parênteses, vírgula, ponto e vírgula, etc.

ANÁLISE LÉXICA: TOKENS, PADRÕES e LEXEMAS

Atributos dos Tokens

Se mais de um lexema casar com um padrão, o analisador léxico precisa fornecer informações adicionais sobre qual foi o lexema em particular casado.

Exemplo:

Padrão de constante numérica, Token number

Os lexemas 0, 1, 46, 29, 73, etc podem casar

?? É importante para o Gerador de Código saber exatamente qual foi

Por isso o analisador léxico monta, além do nome do token, um **valor de atributo**.

O valor de atributo descreve o lexema.

Exemplo: Token id

Atributo: apontador para sua entrada na tabela de símbolos, na qual é registrado seu lexema, tipo, localização no programa fonte.

ANÁLISE LÉXICA: TOKENS, PADRÕES e LEXEMAS

Atributos dos Tokens

Exemplo em Fortran: $E = M * C ** 2$

< id, apontador para entrada da tabela de símbolos de E >

< assign_op >

< id, apontador para entrada da tabela de símbolos de M >

< mult_op >

< id, apontador para entrada da tabela de símbolos de C >

< exp_op >

< number, apontador para entrada da tabela de símbolos de 2 >

ANÁLISE LÉXICA

Erros Léxicos

É difícil a um analisador léxico saber que existem erros no programa fonte.

Exemplo: fi (a)

fi está errado o correto é if

É difícil ele decidir se fi é a palavra-chave if escrita ao contrário, ou se é um identificador de uma função que não foi declarada.

Ele simplesmente escolhe o que lhe é mais conveniente, neste caso indicar à próxima etapa que se trata de um token id, e deixar que o erro seja detectado em uma fase futura do compilador.

Mas caso o erro não coincida com nenhum dos padrões possíveis, o analisador léxico é incapaz de prosseguir.

Neste caso ele pode iniciar o “**Modo de Pânico**”...

ANÁLISE LÉXICA

Erros Léxicos

Modo de Pânico

Possibilidades

1. Remover um caractere da entrada restante

Ele remove o caractere seguinte da entrada restante, avalia o que restou e verifica se casa com algum token.

2. Inserir um caractere que falta na entrada restante

Ele insere um caractere para casar com algum token.

3. Substituir um caractere por outro

Ele substitui um caractere por outro que permita casar com algum token.

4. Transpor 2 caracteres adjacentes

Pula 2 caracteres vizinhos e verifica se o restante casa com algum token.

ANÁLISE LÉXICA

Exercícios

1. Divida o seguinte programa em C++:

```
float limitedSquare(x) float x {  
    /* retorna x ao quadrado, mas nunca mais do que 100 */  
    return (x<=-10.0 || x>=10.0)?100:x*x;  
}
```

Em lexemas apropriados, fornecendo informações de token e padrão de cada um.

Quais lexemas devem ter valores léxicos associados?

Quais deverão ser esses valores?

ANÁLISE LÉXICA

Exercícios

2. Divida o seguinte programa em C:

```
for(a=0;a<=b;a++){  
    r=a*a;  
    printf("%dx%d=%d\n",a,a,r);  
}
```

Em lexemas apropriados, fornecendo informações de token e padrão de cada um.

Quais lexemas devem ter valores léxicos associados?

Quais deverão ser esses valores?

ANÁLISE LÉXICA

Exercícios

3. Divida o seguinte programa em C:

```
do{  
    r=r*a;  
    cont++;  
}while(cont<=b);
```

Em lexemas apropriados, fornecendo informações de token e padrão de cada um.

Quais lexemas devem ter valores léxicos associados?

Quais deverão ser esses valores?

Continua...