Aula 2 – Alfabeto – Palavras - Gramática

Uma Linguagem Formal, ao contrário de uma Linguagem Natural, é tal que:
- Tem uma sintaxe bem definida, de forma que, dada uma sentença, seja sempre possível saber se ela pertence ou não à linguagem;
- Tem uma semântica precisa, de modo que não contenha sentenças sem significado ou ambíguas.

O Dicionário Aurélio define linguagem como: *"O uso da palavra articulada ou escrita como meio de expressão e comunicação entre pessoas".*

Linguagem é um dos conceitos fundamentais da computação. Entretanto, para definir linguagem é necessário antes estudar e conhecer os conceitos de *alfabeto* e de *palavra* ou *cadeia de caracteres.*

Toda linguagem tem um alfabeto associado.

1 Alfabeto

Um alfabeto é um conjunto finito de símbolos ou caracteres. Portanto:
- Um conjunto infinito não é um alfabeto;
- O conjunto vazio é um alfabeto.

Exemplos de Alfabeto:

- Os seguintes conjuntos são exemplos de alfabetos:

{a,b,c} ⦸ (conjunto vazio)

- Os seguintes conjuntos não são exemplos de alfabetos:

N (conjunto dos números naturais) {a,b,AA,ab,ba,bb,aaa,...}

Exemplos de Alfabeto: Linguagens de Programação

O Alfabeto de uma linguagem de programação é o conjunto de todos os símbolos usados na construção de programas, denotado por **Σ,** incluindo:

- Letras;
- Dígitos;
- Caracteres especiais como ">", "/", etc.;
- Espaço ou "branco".

2 Palavra, Cadeia de Caracteres ou Sentença

Uma Palavra, Cadeia de Caracteres ou Sentença sobre um alfabeto é uma sequência finita se símbolos (do alfabeto) justapostos denotada por **W**.

Portanto, uma cadeia sem símbolos é uma palavra válida, e o símbolo:

ε *denota a cadeia vazia ou palavra vazia*.

Se Σ representa um alfabeto, então:

Σ^* denota o conjunto de todas as palavras possíveis sobre Σ

Σ^+ denota Σ^* - {ε} -> portanto sem a palavra vazia

Tamanho ou Comprimento de uma Palavra

O Tamanho ou Comprimento de uma palavra **w**, representado por **|w|**, é o número de símbolos que compõe a palavra. Portanto, para um dado alfabeto Σ, comprimento é uma função com domínio em Σ^* e condomínio em **N**.

Exemplo de comprimento de uma Palavra:

|abcde| = 5 | ε |= 0

2.1 Prefixo, Sufixo, subpalavra

Um Prefixo de uma palavra é qualquer sequência inicial de símbolos da palavra.

Um Sufixo de uma palavra é qualquer sequência final de símbolos da palavra.

Uma Subpalavra de uma palavra é qualquer sequência de símbolos contígua da palavra.

Exemplos:

a) **abcb** é uma palavra sobre o alfabeto $\Sigma = \{a,b,c\}$;
b) Se $\Sigma = \{a,b\}$, então $\Sigma_+ = \{a,b,aa,bb,ab,ba,aaa,...\}$ e $\Sigma_* = \{ \varepsilon, a,b,aa,bb,ab,ba,...\}$;
c) |abcb| = 4 e |ε| = 0;

2.2 Linguagem Formal

Uma Linguagem Formal, ou simplesmente *Linguagem L*, é um conjunto de palavras sobre um alfabeto Σ, ou seja: $L \subseteq \Sigma^*$

Exemplo: Suponha o alfabeto $\Sigma = \{a,b\}$. Então:

a) O conjunto vazio e o conjunto formado pela palavra vazia são linguagens sobre o alfabeto Σ (obviamente $\emptyset \neq \{\varepsilon\}$);

b) Os conjuntos Σ_* e Σ_+ são linguagens sobre um alfabeto Σ qualquer. Vale que: $\Sigma_* \neq \Sigma_+$

c) O conjunto de palíndromos (palavras que tem a mesma leitura da esquerda para a direita e vice-versa) sobre Σ é um exemplo de linguagem infinita. Assim, ε, a, b, aa, bb, aaa, aba, ... são palavras desta linguagem.

2.3 Concatenação

A concatenação é uma operação binária, definida sobre uma linguagem, a qual associa a cada par de palavras uma palavra formada pela justaposição da primeira com a segunda. Uma concatenação é denotada pela justaposição dos símbolos que representam as palavras componentes.

A operação de concatenação satisfaz às seguintes propriedades (suponha v,w,t palavras):

a) Associatividade: v(w.t) = (v.w).t;

b) Elemento Neutro à Esquerda e à Direita: ε.w = w = w.ε

Uma operação de concatenação definida sobre uma Linguagem **L** não é, necessariamente, fechada sobre **L**, ou seja, a concatenação de duas palavras de **L** não é, necessariamente, uma palavra de **L**.

Suponha o $\Sigma = \{a,b\}$. Então, para as palavras v = baaaa e w = bb, vale que:

v.w = baaabb

v.ε = v = baaaa

2.4 Concatenação Sucessiva

A Concatenação Sucessiva de uma palavra (como ela mesma), representada na forma de expoente w^n, onde **w** é a palavra e **n** indica o número de concatenações sucessivas, é definida indutivamente a partir da concatenação binária, como segue:

$w^0 = \varepsilon$

$w^n = w.w^{n-1}$, para $n > 0$

Exemplo: Seja **w** uma palavra e **a** um símbolo, então:

$w^2 = ww$ $\qquad$ $a^5 = aaaaa$

$w^1 = w$ $\qquad$ $a^n = aaa...a$ (o símbolo **a** repetido **n** vezes)

3 Gramática

Uma Gramática de Chomsky, Gramática Irrestrita ou simplesmente Gramática, é definida como uma quádrupla ordenada **G = (V, T, P, S)**, onde:

V é o conjunto finito de símbolos variáveis ou não-terminais;

T é o conjunto finito de símbolos terminais disjuntos de **V**;

P é o conjunto finito de pares chamado de regras de produção, tal que a primeira componente é palavra de $(V \cup T)^+$ e a segunda componentes é palavra $(V \cup T)^*$;

S é o elemento de **V** denominado *variável inicial ou símbolo inicial.*

Uma regra de produção (α,β) é representada por $\alpha \rightarrow \beta$

Por simplicidade, um grupo de regras de produção da forma:

$$\alpha \rightarrow \beta_1, \alpha \rightarrow \beta_2, \ldots, \alpha \rightarrow \beta_n$$

(mesma componente no lado esquerdo) é usualmente abreviada como:

$$\alpha \rightarrow \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n.$$

A aplicação de uma regra de produção é denominada derivação de uma palavra. A aplicação sucessiva de regras de produção permite derivar as palavras da linguagem representada pela gramática.

3.1 Relação de Derivação

Seja **G =(V, T, P, S)** uma gramática. Uma Derivação é um par da relação denotada por $\Rightarrow$ com domínio em $(V \cup T)_+$ e contra-domínio em $(V \cup T)_*$. Um par (α, β) é representado: $\alpha \Rightarrow \beta$

A relação $\Rightarrow$ é indutivamente definida com segue:

- Para toda a produção da forma $S \rightarrow \beta$ (a primeira componente é o símbolo inicial de G), tem-se que: $S \Rightarrow \beta$;

- Para todo o par $\alpha \Rightarrow \beta$, onde $\beta = \beta_u\beta_v\beta_w$, se $\beta_v \rightarrow \beta_t$ é regra de P então: $\beta \Rightarrow \beta_u\beta_t\beta_w$.

Portanto, uma Derivação é a substituição de uma subpalavra de acordo com uma regra de produção. Quando for desejado explicitar a regra de produção **p** $\in$ **P** que define a derivação $\alpha \Rightarrow \beta$, a seguinte notação é usada: $\alpha \Rightarrow_p \beta$.

Uma Gramática é considerada um formalismo de geração, pois permite derivar ("gerar") todas as palavras da linguagem que representa.

Os sucessivos passos de derivação são definidos com segue:

$\Rightarrow$ ***** fecho transitivo e reflexivo da relação $\Rightarrow$, ou seja, zero ou mais passos de derivações sucessivas;

$\Rightarrow$ **+** fecho transitivo da relação $\Rightarrow$, ou seja, um ou mais passos de derivações sucessivas;

$\Rightarrow$ **i** exatos **i** passos de derivação sucessivas, onde **i** e um número natural.

3.2 Linguagem Gerada

Seja **G = (V,T,P,S)** uma gramática.

A Linguagem Gerada pela gramática G, denotada por **L(G)** ou **GERA(G)**, é composta por todas as palavras de símbolos terminais deriváveis a partir do símbolo inicial S, ou seja: **L(G) =** $\{w \in T^* \mid S \Rightarrow^+ w\}$ **a palavra "w" pertence a T (conjunto finito) com o vazio, tal que S (símbolo inicial) deriva em uma palavra com mais de uma derivação (símbolo do + antes da palavra).**

Exemplo: Gramática, Derivação, Linguagem Gerada. A gramática G = (V,T,P,N) onde:

V = {N,D}

T = {0,1,2,...,9}

P = {N $\rightarrow$ D | DN,

D $\rightarrow$ 0 | 1 | ... | 9}

Gera, sintaticamente, o conjunto dos números naturais. Note que se distingue o zero à esquerda. Por exemplo, distingue 123 de 0123.

Exemplo 1: Derivação do número 243:

N $\Rightarrow$	N $\rightarrow$ DN
DN $\Rightarrow$	D $\rightarrow$ 2
2N $\Rightarrow$	N $\rightarrow$ DN
2DN $\Rightarrow$	D $\rightarrow$ 4
24N $\Rightarrow$	N $\rightarrow$ D
24D $\Rightarrow$	D $\rightarrow$ 3
243	

Logo, pode-se indicar que:

S $\Rightarrow^+$243 (um ou mais passos de derivações sucessivas)

S $\Rightarrow^6$243 (exatos 6 passos de derivação sucessiva)

Observe que, no exemplo acima, a seguinte interpretação indutiva pode ser dada à gramática em questão:

- Base de Indução: todo dígito é um número natural

- Passo de Indução: se n é um número natural, então a concatenação de n com qualquer dígito também é um número natural.

Exemplo 2: A seguinte gramática:

$G = (\{S, X, Y, A, B, F\},\{a, b\}, P, S)$

Na qual:

$$
\begin{array}{ll}
P = \{S \rightarrow XY & \\
\quad X \rightarrow XaA \mid XbB \mid F & Bb \rightarrow bB \\
\quad Aa \rightarrow aA, & BY \rightarrow Yb, \\
\quad Ab \rightarrow bA & Fa \rightarrow aF \\
\quad AY \rightarrow Ya, & Fb \rightarrow bF, \\
\quad Ba \rightarrow aB, & FY \rightarrow \varepsilon\}
\end{array}
$$

Gera a linguagem cujas palavras são tais que a primeira metade é igual à segunda metade:

$$L(G) = \{ww \mid w \text{ é palavra de } \{a,b\}^*\}$$

Verificar se as palavras abaixo pertencem a Linguagem acima:

a) baba b) abab c) bbaabb d) aabaab