Linguagens, Reconhecedores e Gramáticas

Já vimos que

Linguagem é um conjunto de *cadeias* de símbolos sobre um alfabeto/vocabulário, V. É um subconjunto específico de V*. Estas cadeias são denominadas *sentenças* da linguagem, e são formadas pela justaposição de elementos individuais, os símbolos da linguagem.

Como representar uma linguagem?

Pode-se representar uma linguagem:

(1) como um conjunto finito ou infinito de cadeias

Ex: Linguagem dos números pares; $L=\{ 0^n 1^n \mid n \geq 1\}$

(2) por meio de um Reconhecedor de Linguagem (Autômato)

(3) Por meio de um Gerador de Linguagem (Gramática)

A Gramática é o formalismo gerativo de linguagens, enquanto que os autômatos (a MT é um autômato) são reconhecedores de linguagens.

Um Autômato para uma linguagem L, ou um Reconhecedor de L, é uma Máquina de Estados que recebe como entrada uma cadeia e, após uma sequência de mudanças de estado, fornece como resposta SIM, se a cadeia pertence à linguagem (i.e. se é sentença de L), ou, eventualmente, NÃO, se não pertence.

A Máquina de Turing (MT) é um Autômato e as Linguagens para as quais é possível definir uma MT que as reconheça coincidem com as funções computáveis (Tese de Turing-Church).

Falaremos mais adiante sobre as MT e outros autômatos.

Gramáticas

Exemplo 1:

Vn

Vt

Axioma

Vt = Vocab. terminal

Vn = Vocab. não-terminal

G1 = ({S,A,B}, {a,b}, P ,S)

P = { 1. S -> aB
2. S -> bA
3. A -> a
4. A -> aS
5. A -> bAA
6. B -> b
7. B -> bS
8. B -> aBB }

P: Regras de Produção

Como G1 gera uma linguagem? Por um processo de substituição ou derivação de símbolos. As cadeias $\in$ Vt* geradas formam a linguagem L(G).

$S \rightarrow^{1} aB \rightarrow^{8} aaBB \rightarrow^{7} aabB \rightarrow^{7} aabb$

$S \rightarrow^{2} bA \rightarrow^{4} baS \rightarrow^{1} baaB \rightarrow^{7} baab$ etc.

Definição:

Formalmente, as gramáticas são caracterizadas como quádruplas ordenadas

G = (Vn, Vt, P, S)

onde:

1. Vn representa o vocabulário não terminal da gramática. Este vocabulário corresponde ao conjunto de todos os símbolos dos quais a gramática se vale para definir as leis de formação das sentenças da linguagem.

2. Vt é o vocabulário terminal, contendo os símbolos que constituem as sentenças da linguagem. Dá-se o nome de terminais aos elementos de Vt.

3. P representa o conjunto de todas as leis de formação utilizadas pela gramática para definir a linguagem.

Para tanto, cada construção parcial, representada por um não-terminal, é definida como um conjunto de regras de formação relativas à definicão do não-terminal a ela referente. A cada uma destas regras de formação que compõem o conjunto P dá-se o nome de produção da gramática.

Assumimos Vn $\cap$ Vt = $\emptyset$. Convencionamos que Vn U Vt = V
Cada produção P tem a forma:

$\alpha \rightarrow \beta$ $\alpha \in V^+$;**(qualquer cadeia não nula de V) e**
$\beta \in V^*$ **(qualquer cadeia de V, incluindo a nula)**

4. S ε Vn denota a principal categoria gramatical de G; é dito o símbolo inicial ou o axioma da gramática. Indica onde se inicia o processo de geração de sentenças.

Notação/Convenções

- Letras do alfabeto latino maiúsculas {A,B,..Z}: símbolos de Vn (ou variáveis)
- Letras do começo do alfabeto latino minúsculas {a,b,c,...}: símbolos de Vt (ou terminais)
- Letras do fim do alfabeto latino minúsculas {t,u,v,x,z}: cadeias de terminais
- Letras gregas minúsculas {α,β,γ,**δ,ε,** ..., ω}: cadeias de terminais e não terminais - cadeias de $(Vn \cup Vt)^*$.
- Regras de produção com mesmo lado esquerdo são simplificadas com a notação | **(ou)**

G1 = ({S,A,B}, {a,b}, P,S)

P = {S -> aB | bA

A -> a | aS | bAA

B -> b | bS | aBB }

Definida uma gramática G, qual é a linguagem gerada por ela?

Sejam as relações

$\Rightarrow$ (deriva/gera diretamente) e

$\Rightarrow^*$ (deriva/gera)

definidas entre as cadeias de V* ($V = Vn \cup Vt$)

Def.1. Se $\alpha \to \beta$ é uma produção de P, e γ e δ são cadeias quaisquer de V*, então

$$\gamma\,\alpha\,\delta \Rightarrow \gamma\,\beta\,\delta$$

Ou: α deriva/gera β (β é derivado de α) por uma única produção, não importa o contexto (γ,δ) em que aparecem.

No Ex.1.:

S => aB , pois S -> ab;
aB => aaBB, pois B -> aBB;
aaBB=>aabB=>aabb, pois B ->b ou
S => aB =>aaBB=> aabB=> aabb

Def.2. Suponha que $\alpha_1\ \alpha_2\ \alpha_3\ \ldots\ \alpha_m$ são cadeias de V* e

$$\alpha_1 \Rightarrow \alpha_2, \quad \alpha_2 \Rightarrow \alpha_3, \quad \ldots, \quad \alpha_{m-1} \Rightarrow \alpha_m$$

Então dizemos que

$\boldsymbol{\alpha_1 \Rightarrow^{*} \alpha_m}$ ($\boldsymbol{\alpha_1}$ deriva/gera $\boldsymbol{\alpha_m}$ aplicando-se um número qualquer de produções de P)

Por convenção, $\boldsymbol{\alpha \Rightarrow^{*} \alpha}$ para toda cadeia α.

No Ex.1.: S =>* ab;
S =>* aaBB;
baS =>* baab;
aB =>* abbA

Def.3. Toda cadeia derivada/gerada do símbolo inicial S é chamada uma **forma sentencial**.

Ou seja, uma cadeia $\alpha \in V^*$ é uma forma sentencial se e só se $S \Rightarrow^* \alpha$

No Ex.1: aB, AB, S, ab são formas sentenciais.

Def.4. Uma forma sentencial, α, é uma **sentença** de G se for composta apenas de símbolos terminais.

Ou: $\alpha \in V^*$ é uma sentença de G se e só se
(a) $S \Rightarrow^* \alpha$ (for forma sentencial) e
(b) $\alpha \in Vt^*$ (só tem terminais).

Assim, as sentenças são as cadeias de terminais geradas pela gramática (por seu símbolo inicial).

Def.5. A Linguagem L gerada por uma gramática G é definida como o conjunto de sentenças de G. Ou seja,

L(G) = {x | x ϵ Vt^* e S $=>^*$ x} ou {x | x é sentença de G}

1. A cadeia consiste somente de terminais
2. A cadeia é derivada a partir do símbolo inicial da gramática

Def.6. Duas gramáticas G1 e G2 são equivalentes sse L(G1) = L(G2)

Exemplos de Gramáticas

G2 = ({S}, {0,1}, P1, S)
P: { 1. S -> 0S1
2. S -> 01 }

Qual é a linguagem gerada por G1? Aplicamos o processo de derivação para obter L(G1), que é o processo de obtenção de cadeias a partir de uma gramática.

G2

- A menor cadeia gerada é 01: $S \Rightarrow^2 01$

- Se aplicarmos n-1 vezes a produção 1, seguida da produção 2 teremos:

- $S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 0^3S1^3 \Rightarrow^*$
- $0^{n-1}S1^{n-1} \Rightarrow 0^n1^n$
- Portanto, $L(G2) = \{0^n1^n \mid n >= 1\}$

  ou $S \Rightarrow^* 0^n1^n$

Voltaremos às gramáticas mais tarde....