

Teoria da Computação

Aula Teórica 18:

Linguagens Independentes de Contexto

António Ravara

Departamento de Informática

20 de Maio de 2019

Linguagens Independentes de Contexto

- ▶ São uma classe de linguagens que inclui as regulares.
- ▶ São definíveis recursivamente:
usam-se as gramáticas independentes de contexto.
- ▶ Foram propostas por N. Chomsky no final dos anos 50 para definir formalmente as línguas naturais.
- ▶ Têm muitas aplicações na Informática:
 - ▶ para definir compiladores, pois permitem descrever concisa e rigorosamente a sintaxe das linguagens de programação; um compilador recebe uma gramática e devolve um analisador sintático (um “parser”) para a linguagem;
 - ▶ para descrever formatos de documentos, como o XML;
 - ▶ têm um modelo computacional equivalente – os autómatos de pilha.

Exemplo: A linguagem dos palíndromos

Mostra-se para o alfabeto $\{0, 1\}$, usando o Lema da Bombagem.

$\mathcal{L} \stackrel{\text{def}}{=} \{w \in \{0, 1\}^* \mid w = \text{Rev}(w)\}$ não é regular

1. P_1 fixa $p > 0$ e P_2 escolhe $0^p 10^p$ com $2p + 1$ símbolos.
2. Como $y \neq \epsilon$ e $|xy| \leq p$, P_1 define $x = 0^n, y = 0^m$, com $n + m \leq p$ e $m > 0$.
3. P_2 escolhe $i = 0$ e mostra que a linguagem não é regular, pois $xy^0z = 0^n 10^p \notin \mathcal{L}$
(tem n 0s à esquerda e p 0s à direita, com $n < p$ pois $m > 0$).

$\mathcal{L} \stackrel{\text{def}}{=} \{w \in \{0, 1\}^* \mid w = \text{Rev}(w)\}$ é definível indutivamente

- ▶ Casos base: $\epsilon, 0$ e 1 são palíndromos.
- ▶ Passo: se w é um palíndromo,
 - ▶ $0w0$ é um palíndromo;
 - ▶ $1w1$ é um palíndromo.

Gramáticas Independentes de Contexto: um exemplo

- ▶ É conveniente apresentar os casos de uma definição indutiva como regras.
- ▶ Usam-se variáveis para representar a recursão.

Gramática para $\mathcal{L} \stackrel{\text{def}}{=} \{w \in \{0, 1\}^* \mid w = \text{Rev}(w)\}$

1. $P \rightarrow \epsilon$
2. $P \rightarrow 0$
3. $P \rightarrow 1$
4. $P \rightarrow 0P0$
5. $P \rightarrow 1P1$

De forma compacta, escreve-se:

$$P \rightarrow \epsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1$$

Gramáticas Independentes de Contexto: outro exemplo

- ▶ Viu-se que não é regular a linguagem $\mathcal{L} \stackrel{\text{def}}{=} \{0^n 1^n \mid n \in \mathbb{N}_0\}$.
- ▶ Mostra-se agora que é independente de contexto.

Definição Recursiva de $\mathcal{L} \stackrel{\text{def}}{=} \{0^n 1^n \mid n \in \mathbb{N}_0\}$

- ▶ $\epsilon \in \mathcal{L}$
- ▶ Se $w \in \mathcal{L}$ então $0w1 \in \mathcal{L}$

Note-se que:

- ▶ $\epsilon = 0^0 1^0$
- ▶ tomando $w = 0^n 1^n$, tem-se $0w1 = 00^n 1^n 1 = 0^{n+1} 1^{n+1}$

Gramática para $\mathcal{L} \stackrel{\text{def}}{=} \{0^n 1^n \mid n \in \mathbb{N}_0\}$

$$P \rightarrow \epsilon \mid 0P1$$

Gramáticas Independentes de Contexto: definição

- ▶ Chama-se ao alfabeto o conjunto (finito) dos símbolos *terminais* (vão formar as palavras).
- ▶ O conjunto (finito) das variáveis diz-se o dos símbolos *não terminais*. Cada variável está associada a uma linguagem.
- ▶ A variável associada à linguagem que se pretende definir diz-se o símbolo *inicial*; as restantes são auxiliares.
- ▶ O conjunto (finito) das regras da definição recursiva da linguagem diz-se o das *produções*.

Uma produção consiste:

- ▶ numa *cabeça* (a variável à esquerda); seguida de
- ▶ o símbolo $\rightarrow$; seguido de
- ▶ uma sequência (eventualmente vazia) de símbolos terminais e não terminais.

Gramáticas Independentes de Contexto: formalmente

Gramática Independente de Contexto, ou CFG

É um quádruplo $\langle V, T, P, S \rangle$, sendo

- ▶ V um conjunto finito não vazio (de variáveis ou símbolos não terminais).
- ▶ T um conjunto finito (de símbolos terminais – o alfabeto), disjunto de V .
- ▶ P um conjunto finito (de regras), sendo $P \subseteq V \times (V \cup T)^*$.
- ▶ $S \in V$ o símbolo inicial.

Exemplos

- ▶ $\langle \{P\}, \{0, 1\}, \{(P, \epsilon), (P, 0), (P, 1), (P, 0P0), (P, 1P1)\}, P \rangle$
define $\mathcal{L} \stackrel{\text{def}}{=} \{w \in \{0, 1\}^* \mid w = \text{Rev}(w)\}$
- ▶ $\langle \{P\}, \{0, 1\}, \{(P, \epsilon), (P, 0P1)\}, P \rangle$
define $\mathcal{L} \stackrel{\text{def}}{=} \{0^n 1^n \mid n \in \mathbb{N}_0\}$

Linguagem gerada por uma gramática

Derivação

- ▶ Para obter a linguagem definida por uma gramática, é preciso *expandir* a sua definição recursiva.
- ▶ Primeiro expande-se o símbolo inicial, usando uma das suas produções.
- ▶ Depois pega-se na palavra obtida e expande-se uma das variáveis usando uma das suas produções.
- ▶ O processo termina quando se tiver obtido uma palavra sem variáveis.
- ▶ Note-se que não há uma única forma de obter a linguagem: Uma variável pode ter várias produções e uma palavra que ocorre numa produção pode ter várias variáveis.

A linguagem é o conjunto de todas as palavras obtidas desta forma.

Linguagem gerada por uma gramática

Dada uma gramática $G = \langle V, T, P, S \rangle$, escreve-se $X \rightarrow_G w$ se $(X, w) \in P$.

Passo de derivação

Para $x, y \in (V \cup T)^*$ e $X \in V$, escreve-se $x X y \Rightarrow_G x w y$, se $X \rightarrow_G w$

A linguagem gerada por G é definida por

$$\mathcal{L}(G) \stackrel{\text{def}}{=} \{w \in T^* \mid S \Rightarrow_G^* w\}$$

sendo $\Rightarrow_G^*$ o fecho reflexivo e transitivo de $\Rightarrow_G$

Linguagem gerada por uma gramática

Exemplos de derivações

► $0110 \in \{w \in \{0,1\}^* \mid w = \text{Rev}(w)\}$

$$\begin{aligned} P &\Rightarrow 0 P 0 && \text{(Regra 4)} \\ &\Rightarrow 0 1 P 1 0 && \text{(Regra 5)} \\ &\Rightarrow 0 1 \epsilon 1 0 && \text{(Regra 1)} \\ &= 0 1 1 0 \end{aligned}$$

► $0011 \in \{0^n 1^n \mid n \in \mathbb{N}_0\}$

$$\begin{aligned} P &\Rightarrow 0 P 1 && \text{(Regra 2)} \\ &\Rightarrow 0 0 P 1 1 && \text{(Regra 2)} \\ &\Rightarrow 0 0 \epsilon 1 1 && \text{(Regra 1)} \\ &= 0 0 1 1 \end{aligned}$$

Linguagens Independentes de Contexto: um exemplo

$\mathcal{L} \stackrel{\text{def}}{=} \{(01)^n 2^n \mid n \in \mathbb{N}\}$ é gerada pela gramática $\langle V, T, P, S \rangle$

- ▶ $V = \{X, Y\}$
- ▶ $T = \{0, 1, 2\}$
- ▶ $S = X$
- ▶ $P = \{X \rightarrow 01Y2, Y \rightarrow \epsilon \mid 01Y2\}$

$010122 \in \mathcal{L}$

$$\begin{aligned} X &\Rightarrow 01Y2 && \text{(Regra 1)} \\ &\Rightarrow 0101Y22 && \text{(Regra 3)} \\ &\Rightarrow 0101\epsilon22 && \text{(Regra 2)} \\ &= 010122 \end{aligned}$$

Uma gramática equivalente (que gera a mesma linguagem) é $\langle \{X\}, \{0, 1, 2\}, \{X \rightarrow 012 \mid 01X2\}, X \rangle$.

Linguagens Regulares vs. Linguagens Independ. de Contexto

Há linguagens independ. de contexto que não são regulares

Como se provou, as linguagens usadas nos exemplos anteriores:

- ▶ não são regulares;
- ▶ podem-se definir indutivamente (e com uma gramática), logo são independentes de contexto.

Todas as linguagens regulares são independentes de contexto

Facilmente se define uma função para converter DFAs em CFGs:

- ▶ o conjunto dos estados é o conjunto das variáveis, sendo o estado inicial a variável inicial;
- ▶ o alfabeto é o conjunto dos símbolos terminais;
- ▶ cada transição passa a ser uma produção;
- ▶ para cada estado final junta-se uma produção em que a palavra à direita é ϵ .

Derivações *leftmost* e *rightmost*

- ▶ Para simplificar a derivação de palavras, pode-se fixar uma estratégia de substituição de variáveis.
- ▶ Uma derivação diz-se *leftmost*, se se substitui sempre a variável mais à esquerda por uma das palavras duma sua produção. Denota-se por $\Rightarrow_{lm}$.
- ▶ Uma derivação diz-se *rightmost*, se se substitui sempre a variável mais à direita por uma das palavras duma sua produção. Denota-se por $\Rightarrow_{rm}$.

Derivação canónica

Numa CFG, qualquer palavra da linguagem tem uma derivação *leftmost* (e uma *rightmost*).

Outro exemplo de gramática: parênteses bem balanceados

$$G \stackrel{\text{def}}{=} \langle \{B\}, \{ (,) \}, P, B \rangle$$

P é dado por

$$B \rightarrow \epsilon \mid (B) \mid BB$$

Exemplos de derivações

$$\begin{aligned} B &\Rightarrow (B) && \text{(Regra 2)} \\ &\Rightarrow (BB) && \text{(Regra 3)} \\ &\Rightarrow ((B)B) && \text{(Regra 2)} \\ &\Rightarrow ((\epsilon)B) && \text{(Regra 1)} \\ &= (()B) \\ &\Rightarrow (() (B)) && \text{(Regra 2)} \\ &\Rightarrow (() (\epsilon)) && \text{(Regra 1)} \\ &= (()()) \end{aligned}$$

$$\begin{aligned} B &\Rightarrow BB && \text{(Regra 3)} \\ &\Rightarrow (B)B && \text{(Regra 2)} \\ &\Rightarrow (B)(B) && \text{(Regra 2)} \\ &\Rightarrow (\epsilon)(B) && \text{(Regra 1)} \\ &= ()(B) \\ &\Rightarrow ()((B)) && \text{(Regra 2)} \\ &\Rightarrow ()((\epsilon)) && \text{(Regra 1)} \\ &= ()(()) \end{aligned}$$

Exemplos de derivações numa gramática

Considere-se de novo a gramática

$$G \stackrel{\text{def}}{=} \langle \{B\}, \{ (,) \}, \{ B \rightarrow \epsilon \mid (B) \mid BB \}, B \rangle$$

Exemplos de derivações

$$\begin{aligned} B &\Rightarrow BB \\ &\Rightarrow (B)B \\ &\Rightarrow (B)(B) \\ &\Rightarrow (B)(\epsilon) \\ &= (B)() \\ &\Rightarrow (\epsilon)() \\ &= ()() \end{aligned}$$

$$\begin{aligned} B &\Rightarrow_{lm} BB \\ &\Rightarrow_{lm} (B)B \\ &\Rightarrow_{lm} (\epsilon)B \\ &= ()B \\ &\Rightarrow_{lm} ()(B) \\ &\Rightarrow_{lm} ()(\epsilon) \\ &= ()() \end{aligned}$$

$$\begin{aligned} B &\Rightarrow_{rm} BB \\ &\Rightarrow_{rm} B(B) \\ &\Rightarrow_{rm} B(\epsilon) \\ &= B() \\ &\Rightarrow_{rm} (B)() \\ &\Rightarrow_{rm} (\epsilon)() \\ &= ()() \end{aligned}$$