Folha Prática 7

Uma **gramática independente de contexto (GIC)** é um quarteto $\mathcal{G} = (V, \Sigma, P, S)$, onde V e Σ são conjuntos de símbolos, tais que $V \cap \Sigma = \emptyset$, sendo ambos finitos e não vazios, $S \in V$, e P é um conjunto finito de regras da forma $X \to w$ com $X \in V$ e $w \in (V \cup \Sigma)^\star$. Formalmente, P é definido como uma relação binária de V em $(V \cup \Sigma)^\star$, constituída pelos pares (X, w) referidos. É usada a seguinte terminologia:

- V é o conjunto das **variáveis** ou símbolos **não terminais**;
- S é o **símbolo inicial**;
- Σ é o **alfabeto** ou conjunto dos símbolos **terminais**;
- P é o conjunto de **produções** ou **regras** (ou regras de produção) e escreve-se $X \to w$ se $(X, w) \in P$. Podemos referir as regras que definem X como X**-produções**, para $X \in V$.

Por vezes, usamos $X \to w_1 \mid w_2 \mid \ldots \mid w_k$ para representar um conjunto de X-produções de forma abreviada. Neste caso, teriamos as regras $X \to w_i$, com $1 \leq i \leq k$.

Sendo x e y sequências de $(V \cup \Sigma)^\star$ tais que x tem pelo menos um símbolo não terminal, diz-se que y se pode derivar de x por aplicação de uma regra $X \to w$ de $\mathcal{G}$ se e só se $x = x_1 X x_2$ e $y = x_1 w x_2$, com $X \in V$ e $x_1, w, x_2 \in (V \cup \Sigma)^\star$. Numa derivação, a aplicação da regra $X \to w$ para substituição de X em $x_1 X x_2$, substitui essa ocorrência de X por w independentemente do contexto em que X está. É por essa razão que a gramática se diz independente de contexto.

Escrevemos:

- $x \Rightarrow_{\mathcal{G}} y$ se se derivar y de x por aplicação de alguma regra de $\mathcal{G}$ (*derivação num passo*);
- $x \Rightarrow_{\mathcal{G}}^{n} y$ se se derivar y de x por aplicação de n regras de $\mathcal{G}$, não necessariamente distintas (*derivação em n passos*);
- $x \Rightarrow_{\mathcal{G}}^{\star} y$ se se derivar y de x por aplicação de um número finito de regras de $\mathcal{G}$, possivelmente zero (*derivação em zero ou mais passos*).

A derivação num passo $\Rightarrow_{\mathcal{G}}$ pode ser definida como uma relação binária em $(V \cup \Sigma)^\star$, sendo $\Rightarrow_{\mathcal{G}}^{\star}$ o seu fecho reflexivo e transitivo, e $\Rightarrow_{\mathcal{G}}^{n}$ a relação obtida por composição de $\Rightarrow_{\mathcal{G}}$ com si própria, n vezes. Mais formalmente, $(\Rightarrow_{\mathcal{G}}^{1}) = (\Rightarrow_{\mathcal{G}})$ e $(\Rightarrow_{\mathcal{G}}^{n}) = (\Rightarrow_{\mathcal{G}}^{n-1} \Rightarrow_{\mathcal{G}}) = (\Rightarrow_{\mathcal{G}} \Rightarrow_{\mathcal{G}}^{n-1})$, para $n \geq 2$.

Se estivermos a considerar apenas uma gramática, podemos omitir $\mathcal{G}$ em $\Rightarrow_{\mathcal{G}}$, $\Rightarrow_{\mathcal{G}}^{n}$, e $\Rightarrow_{\mathcal{G}}^{\star}$.

Num passo de derivação, podemos substituir uma qualquer variável (desde que tal nos permita obter a palavra pretendida, no fim da derivação). Uma **derivação pela esquerda** é uma derivação em que a variável que se substitui em cada passo é sempre a que estiver mais à esquerda. Uma **derivação pela direita** é uma derivação em que a variável que se substitui em cada passo é sempre a que estiver mais à direita. Numa derivação podemos optar por não seguir nenhum desses dois critérios.

A **linguagem gerada pela gramática** $\mathcal{G}$ é o conjunto das palavras de $\Sigma^\star$ que se podem *derivar num número finito de passos* a partir do símbolo inicial de $\mathcal{G}$. Denota-se por $\mathcal{L}(\mathcal{G})$, e sendo S o símbolo inicial, tem-se:

$$\mathcal{L}(\mathcal{G}) = \{x \mid x \in \Sigma^\star, S \Rightarrow_{\mathcal{G}}^{\star} x\}.$$

Uma **linguagem é independente de contexto (LIC)** é, por definição, qualquer linguagem que possa ser gerada por uma gramática independente de contexto.

A derivação de uma palavra pode ser representada esquematicamente por uma **árvore de derivação** (ou **árvore sintática**) que é uma árvore orientada, com raíz S, em que os descendentes diretos de um nó interno X correspondem ao lado direito da regra aplicada para substituição desse X na derivação. Assim, se a regra aplicada para substituir esse X for $X \to \beta_1\beta_2\ldots\beta_m$, com $\beta_i \in V \cup \Sigma \cup \{\varepsilon\}$, para $1 \leq i \leq m$, o nó X será ligado aos m novos nós, criados para β_1, β_2, $\ldots$, β_m, ficando os ramos ordenados. O ramo de X para β_1 será o filho de X mais à esquerda e o ramo de X para β_m será o filho de X mais à direita. Cada folha da árvore contém um símbolo terminal ou ε. A palavra que deu origem a essa árvore de derivação é dada pela concatenação das folhas da árvore.

Uma **gramática independente de contexto é ambígua** se existir alguma palavra $x \in \mathcal{L}(\mathcal{G})$ que admita mais do que uma derivação pela esquerda (ou, equivalentemente, que admita mais do que uma derivação pela direita) em $\mathcal{G}$. É conhecido que cada derivação pela esquerda (ou pela direita) determina uma e uma só árvore de derivação, e vice-versa. Assim, uma gramática $\mathcal{G}$ é ambígua se existir alguma palavra $x \in \mathcal{L}(\mathcal{G})$ que admita duas ou mais árvores de derivação em $\mathcal{G}$.

Uma **linguagem independente de contexto é (inerentemente) ambígua** se todas as gramáticas independentes de contexto que a geram são ambíguas. Existem linguagens independentes de contexto que são ambíguas mas está fora do âmbito da disciplina provar que uma linguagem é ambígua. No entanto, sempre que for simples perceber que uma linguagem não é ambígua, faremos um esforço por a definir por uma gramática não ambígua.

Exemplo

Considere a gramática $\mathcal{G} = (\{E, N\}, \{\texttt{0}, \texttt{1}, +\}, P, E)$, onde P é constituído pelas regras

$$E \;\to\; E{+}E \mid \texttt{0} \mid \texttt{1}N \qquad\qquad N \;\to\; \texttt{1}N \mid \texttt{0}N \mid \varepsilon$$

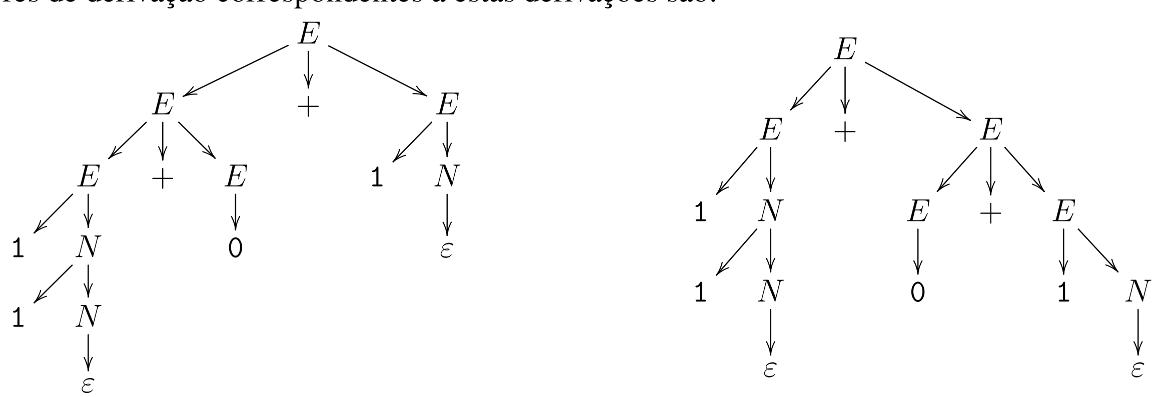
A palavra $\texttt{11}{+}\texttt{0}{+}\texttt{1}$ pertence a $\mathcal{L}(\mathcal{G})$, como mostra a seguinte derivação pela esquerda:

$$E \Rightarrow_{\mathcal{G}} E{+}E \Rightarrow_{\mathcal{G}} E{+}E{+}E \Rightarrow_{\mathcal{G}} \texttt{1}N{+}E{+}E \Rightarrow_{\mathcal{G}} \texttt{11}N{+}E{+}E \Rightarrow_{\mathcal{G}} \texttt{11}\varepsilon{+}E{+}E \Rightarrow_{\mathcal{G}} \texttt{11}{+}\texttt{0}{+}E \Rightarrow_{\mathcal{G}} \texttt{11}{+}\texttt{0}{+}\texttt{1}N \Rightarrow_{\mathcal{G}} \texttt{11}{+}\texttt{0}{+}\texttt{1}\varepsilon$$

A gramática é ambígua. Por exemplo, para a mesma palavra, existe uma outra derivação pela esquerda:

$$E \Rightarrow_{\mathcal{G}} E{+}E \Rightarrow_{\mathcal{G}} \texttt{1}N{+}E \Rightarrow_{\mathcal{G}} \texttt{11}N{+}E \Rightarrow_{\mathcal{G}} \texttt{11}\varepsilon{+}E \Rightarrow_{\mathcal{G}} \texttt{11}{+}E{+}E \Rightarrow_{\mathcal{G}} \texttt{11}{+}\texttt{0}{+}E \Rightarrow_{\mathcal{G}} \texttt{11}{+}\texttt{0}{+}\texttt{1}N \Rightarrow_{\mathcal{G}} \texttt{11}{+}\texttt{0}{+}\texttt{1}\varepsilon$$

As árvores de derivação correspondentes a estas derivações são:



As palavras de $\mathcal{L}(\mathcal{G})$ representam números em binário sem $\texttt{0}$'s não significativos ou somas de números desse tipo, sendo $\mathcal{L}(\mathcal{G}) = (\{\texttt{0}\} \cup \{\texttt{1}\}\{\texttt{0}, \texttt{1}\}^\star)(\{+\texttt{0}\} \cup \{+\texttt{1}\}\{\texttt{0}, \texttt{1}\}^\star)^\star$. A ambiguidade de $\mathcal{G}$ resulta da regra $E \to E{+}E$. Se a substituirmos convenientemente, podemos obter uma gramática não ambígua equivalente (i.e., que gera a mesma linguagem), o que mostra que $\mathcal{L}(\mathcal{G})$ não é ambígua embora $\mathcal{G}$ seja. Por exemplo, a gramática $\mathcal{G}_1 = (\{E, N\}, \{\texttt{0}, \texttt{1}, +\}, P_1, E)$, onde P_1 é constituído pelas regras:

$$E \;\to\; \texttt{1}N{+}E \mid \texttt{0}{+}E \mid \texttt{0} \mid \texttt{1}N \qquad\qquad N \;\to\; \texttt{1}N \mid \texttt{0}N \mid \varepsilon$$

Exercícios

1. Defina uma gramática independente de contexto que gere a linguagem indicada em cada alínea (pode ser uma gramática ambígua). Para as sete primeiras, o alfabeto é $\{\mathtt{0}, \mathtt{1}\}$. Para as restantes é $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. Justifique sucintamente a correção da gramática, indicando o que gera cada variável e/ou a forma geral das sequências de símbolos terminais e/ou não terminais que se podem derivar em k passos, com $k \geq 1$.

a) $\{\mathtt{0}^n \mid n \geq 0\}\{\mathtt{1}^{2n} \mid n \geq 1\}$

b) $\{\mathtt{0}^n\mathtt{1}^{2n} \mid n \leq 2\}$

c) $\{\mathtt{0}^n\mathtt{1}^{2n} \mid n \geq 2\}$

d) $\{\mathtt{0}^n\mathtt{1}^m \mid m \geq n \geq 0\}$

e) $\{\mathtt{0}^n\mathtt{1}^m \mid m \geq n \geq 0\}^\star$

f) $\mathcal{L}(\mathtt{00}^\star\mathtt{11} + (\mathtt{0} + \mathtt{101})^\star\mathtt{1})$

g) $\{wtw^R \mid w \in \{\mathtt{0}, \mathtt{1}\}^\star, t \in \{\mathtt{0}, \mathtt{1}, \varepsilon\}\}$

h) {palavras que terminam em `abc` ou têm exatamente dois `a`'s}

i) $\mathcal{L}(\mathtt{a}^\star\mathtt{abbb}^\star)$

j) $\{\mathtt{a}^n\mathtt{b}\mathtt{a}^n \mid n > 1\}$

k) $\{\mathtt{a}^i\mathtt{b}^{i+j}\mathtt{c}^j \mid i \geq 0, j \geq 0\}$

l) $\{\mathtt{a}^i\mathtt{b}^j\mathtt{a}^k\mathtt{c}^i \mid i, j, k \in \mathbb{N} \text{ e } k > 0 \text{ se } j > 0\}$

m) $(\{\mathtt{c}\}\{\mathtt{c}\}^\star\{\mathtt{a}^{2n}\mathtt{b}^n \mid n \geq 1\})^\star\{\mathtt{c}\}\{\mathtt{c}\}^\star$

n) {palavras cujo número de `a`'s é primo e não excede seis}

o) {palavras que não terminam em `c` se tiverem dois `b`'s consecutivos}

2. Seja L a linguagem das expressões regulares sobre $\Sigma = \{\mathtt{a}, \mathtt{b}\}$, a qual pode ser definida indutivamente, como uma linguagem de alfabeto $\{\mathtt{)}\,, \mathtt{(}, +, ^\star, \boxed{\varepsilon}, \emptyset\} \cup \Sigma$, por:

- $\boxed{\varepsilon} \in L$, $\mathtt{a} \in L$, $\mathtt{b} \in L$, $\emptyset \in L$
- $\mathtt{(}r^\star\mathtt{)} \in L$, $\mathtt{(}rs\mathtt{)} \in L$, e $\mathtt{(}r{+}s\mathtt{)} \in L$, quaisquer que sejam $r, s \in L$.

Para não confundir a palavra vazia com expressão ε, usámos $\boxed{\varepsilon}$ como símbolo em vez de ε. Determine uma gramática independente de contexto (não ambígua) que gere L. Apresente as árvores de derivação das palavras $(((\mathtt{a}{+}\mathtt{b})^\star) + (\emptyset\boxed{\varepsilon}))$ e $(((\mathtt{aa}) + (((\mathtt{a}{+}\mathtt{b})^\star) + (\emptyset\boxed{\varepsilon})))^\star)$.

3. Seja L a linguagem de alfabeto $\Sigma = \{\mathtt{p}, \wedge, \vee, \neg, \Rightarrow,), (\}$ assim definida indutivamente: (i) p pertence a L; (ii) se α e β pertencem a L, também as palavras $(\alpha\vee\beta)$, $(\alpha\wedge\beta)$, $(\alpha\Rightarrow\beta)$, e $(\neg\alpha)$ pertencem a L; (iii) as palavras de L são as que se obtêm por aplicação das regras (i) e (ii).

a) Verifique que $((\mathtt{p} \vee \mathtt{p}) \vee (\neg\mathtt{p}))$ e $((\mathtt{p} \vee \mathtt{p}) \Rightarrow (\neg(\mathtt{p} \wedge \mathtt{p})))$ pertencem a L.

b) Partindo da descrição dada, determine uma GIC G não ambígua que gere L.

c) Para a palavra $((\mathtt{p} \vee \mathtt{p}) \vee (\neg\mathtt{p}))$, apresente: derivação pela esquerda; derivação pela direita; derivação que não é nem pela esquerda nem pela direita; as árvores de derivação correspondentes.

d) Mostre que $((\mathtt{p} \vee \vee\mathtt{p})) \notin \mathcal{L}(G)$.

4. Mostre que a linguagem $\{0^n 1^{2n} \mid n \geq 0\} \cup \{0\}^\star$ não é regular mas é independente de contexto.

5. Considere a linguagem $L = \{\mathtt{a}^n \mathtt{b}^m \mid m > 3n \geq 0\}$ de alfabeto $\Sigma = \{\mathtt{a}, \mathtt{b}\}$.

a) Prove que não existe uma expressão regular que descreva L.

b) Prove que L é uma linguagem independente de contexto não ambígua. Para isso, determine uma GIC G que gere L e seja não ambígua (e demonstre que G gera L e é não ambígua).

6. Considere a gramática independente de contexto $G = (\{A\}, \{\mathtt{a}, \mathtt{b}\}, P, A)$ com P dado por

$$A \ \rightarrow \ \mathtt{a}A\mathtt{b} \mid A\mathtt{b} \mid \mathtt{b}$$

a) Prove que G é ambígua.

b) Determine a forma geral das palavras de $\mathcal{L}(G)$. Explique usando $\Rightarrow^n$, com $n > 0$.

c) Indique uma gramática não ambígua equivalente a G. Justifique sucintamente.

d) Indique uma gramática não ambígua que gere o fecho de Kleene de $\mathcal{L}(G)$. Justifique sucintamente como garante a não ambiguidade.

7. Considere a gramática independente de contexto $G = (\{A, B, C\}, \{\mathtt{a}, \mathtt{b}\}, P, C)$ com P dado por

$$\begin{aligned} C &\rightarrow B\mathtt{b}A \mid A \\ B &\rightarrow \mathtt{b}B \mid \mathtt{aaa}B \mid \varepsilon \\ A &\rightarrow \mathtt{a}A \mid \varepsilon \end{aligned}$$

a) Indique uma derivação pela esquerda, uma derivação pela direita e uma derivação que não seja nem pela esquerda nem pela direita para a palavra $\mathtt{aaabbaaabaa}$ de $\mathcal{L}(G)$. Determine as árvores de derivação correspondentes. Desta análise, pode concluir que G é ambígua ou que G não é ambígua?

b) Prove que a palavra $\mathtt{aaabaab}$ não pertence a $\mathcal{L}(G)$.

c) Por indução sobre o número de passos da derivação, prove que qualquer que seja $n \geq 1$ e qualquer que seja $x \in \{A, B, C, \mathtt{a}, \mathtt{b}\}^\star$ se tem $B \Rightarrow^n x$ se e só se $x \in \{\mathtt{aaa}, \mathtt{b}\}^{n-1}$ ou $x \in \{\mathtt{aaa}, \mathtt{b}\}^n\{B\}$.

Sugestão: *Por definição, $L^{k+1} = L^k L = LL^k$, qualquer que seja a linguagem L, para $k \in \mathbb{N}$. Por definição de derivação em $k+1$ passos, tem-se $\alpha \Rightarrow^{k+1} \gamma$ se e só se existe β tal que $\alpha \Rightarrow^k \beta$ e $\beta \Rightarrow \gamma$, com $k \in \mathbb{N}$.*

d) Prove por indução que todo $x \in \{\mathtt{aaa}, \mathtt{b}\}^n$ admite uma única derivação a partir de B (e, portanto, uma única derivação pela esquerda). Conclua que a gramática formada pelas regras para B não é ambígua.

e) Indique a linguagem gerada a partir de cada variável da gramática. Conclua que $\mathcal{L}(G)$ é regular.

f) Na continuação das alíneas anteriores, averigue se a gramática G é ambígua. Justifique.

8. Considere a linguagem $L = \{\mathtt{a}^{2k} w \mathtt{a}^{3k} z \mid zw \in \mathcal{L}(\mathtt{bb}^\star),\ k \in \mathbb{N}\}$ de alfabeto $\Sigma = \{\mathtt{a}, \mathtt{b}\}$.

a) Prove que L não satisfaz a condição do lema da repetição para linguagens regulares, para nenhum $n > 0$.

b) Justifique que $L = \mathcal{L}(G)$ para $G = (\{S, A, T, B\}, \{\mathtt{a}, \mathtt{b}\}, P, S)$ com P dado por

$$\begin{aligned} S &\rightarrow \mathtt{aa}T\mathtt{aaa}B \mid A\mathtt{b}B & \qquad T &\rightarrow \mathtt{aa}T\mathtt{aaa} \mid \mathtt{b}B \\ A &\rightarrow \mathtt{aaaaa}A \mid \varepsilon & \qquad B &\rightarrow \mathtt{b}B \mid \varepsilon \end{aligned}$$

c) Averigue se a gramática G definida na alínea anterior é ambígua.