SBVLIFA: Linguagens Formais e Autômatos

Aula 09: Propriedades das Linguagens Livres de Contexto

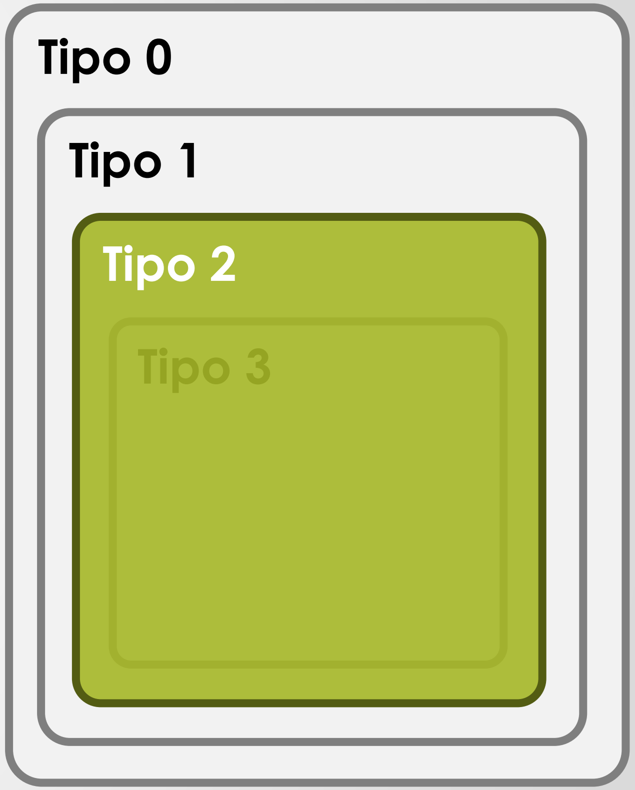
Bacharelado em Ciência da Computação
Prof. Dr. David Buzatto

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Linguagens Livres de Contexto

- Linguagens Livres de Contexto

Tipo	Classe de Linguagens	Modelo de Gramática	Modelo de Reconhecedor
0	Recursivamente enumeráveis	Irrestrita	Máquina de Turing
1	Sensíveis ao contexto	Sensível ao contexto	Máquina de Turing com fita limitada
2	**Livres de contexto**	**Livre de contexto**	**Autômato de pilha**
3	Regulares	Linear (direita ou esquerda)	Autômato finito



Propriedades das Linguagens Livres de Contexto

- Veremos nessa aula algumas propriedades importantes das linguagens livres de contexto, a saber:
  - **Formas Normais para CFGs:** simplificações;
  - **Lema do Bombeamento:** provar que certas linguagens não são livres de contexto;
  - **Propriedades de Fechamento;**
  - **Propriedades de Decisão.**

INSTITUTO FEDERAL DE
EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Propriedades das Linguagens Livres de Contexto

Formas Normais para CFGs

- Mostraremos que toda CFL, sem ε, é gerada por uma CFG na qual todas as produções tem a forma $A \rightarrow BC$ ou $A \rightarrow a$, onde A, B, e C são variáveis e a é um terminal, sendo que essa forma é chamada de *Forma Normal de Chomsky*;
- Para isso, precisaremos empregar diversas simplificações em uma CFG:
  - Eliminação de símbolos inúteis, ou seja, variáveis e terminais não alcançáveis a partir do símbolo de início;
  - Eliminação de ε-produções ($A \rightarrow \varepsilon$);
  - Eliminação de produções unitárias ($A \rightarrow B$).

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Um símbolo X é útil para uma gramática $G = (V, T, P, S)$ se existe alguma derivação na forma $S \overset{*}{\Rightarrow} \alpha X \beta \overset{*}{\Rightarrow} w$, onde w está em T^*;
  - X pode estar em V ou T;
  - A forma sentencial $\alpha X \beta$ pode ser a primeira ou a última derivação;
  - Se X não é útil, dizemos que ele é inútil;
- Todo símbolo gerador **e** alcançável é um símbolo útil:
  - X é gerador se $X \overset{*}{\Rightarrow} w$ para alguma string de terminais w. Todo terminal é gerador, pois w pode ser esse terminal derivado em zero etapas;
  - X é alcançável se existe uma derivação $S \overset{*}{\Rightarrow} \alpha X \beta$ para algum α e β;
- **Obtenção dos símbolos úteis:** eliminação dos símbolos que não são geradores e, posteriormente, dos símbolos que não são alcançáveis.

INSTITUTO FEDERAL DE
EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- **Exemplo:** $S \to AB \mid a$
  $A \to b$

- Todos os símbolos, com exceção de B, são geradores:
  - a e b geram a si mesmos (terminais);
  - S gera a;
  - A gera b.
- Eliminando B, temos: $S \to a$
  $A \to b$

- Apenas S e a são alcançáveis a partir de S, sendo assim, eliminamos A e b: $S \to a$
- Sendo assim, $L(G) = \{a\}$

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- **Exemplo:** $S \rightarrow AB \mid a$
  $A \rightarrow b$
- Todos os símbolos, com exceção de B, são geradores:
  - a e b geram a si mesmos (terminais);
  - S gera a;
  - A gera b.
- Eliminando B, temos: $S \rightarrow a$
  $A \rightarrow b$
- Apenas S e a são alcançáveis a partir de S, sendo assim, eliminamos A e b: $S \rightarrow a$
- Sendo assim, $L(G) = \{a\}$

Para pensar: se começássemos pela eliminação de símbolos não alcançáveis teríamos a mesma gramática resultante?

Para os curiosos: **Teorema 7.2**, página 274 (HOPCROFT et al., 2003)

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos geradores de $G = (V, T, P, S)$:
  - **Base:** todo símbolo de T é sem dúvida gerador, pois ele gera a si mesmo;
  - **Indução:** suponha que exista uma produção $A \rightarrow \alpha$ e que todo símbolo de α já seja conhecido como gerador, então A é gerador. Essa regra inclui $\alpha = \varepsilon$.

$$S \rightarrow AB \mid a$$
$$A \rightarrow b$$

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos geradores de $G = (V, T, P, S)$:
  - **Base:** todo símbolo de T é sem dúvida gerador, pois ele gera a si mesmo;
  - **Indução:** suponha que exista uma produção $A \rightarrow \alpha$ e que todo símbolo de α já seja conhecido como gerador, então A é gerador. Essa regra inclui $\alpha = \varepsilon$.

$$S \rightarrow AB \mid a$$
$$A \rightarrow b$$

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos geradores de $G = (V, T, P, S)$:
  - **Base:** todo símbolo de T é sem dúvida gerador, pois ele gera a si mesmo;
  - **Indução:** suponha que exista uma produção $A \rightarrow \alpha$ e que todo símbolo de α já seja conhecido como gerador, então A é gerador. Essa regra inclui $\alpha = \varepsilon$.

$$S \rightarrow AB \mid a$$
$$A \rightarrow b$$

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos geradores de $G = (V, T, P, S)$:
  - **Base:** todo símbolo de T é sem dúvida gerador, pois ele gera a si mesmo;
  - **Indução:** suponha que exista uma produção $A \rightarrow \alpha$ e que todo símbolo de α já seja conhecido como gerador, então A é gerador. Essa regra inclui $\alpha = \varepsilon$.

$$S \rightarrow AB \mid a$$
$$A \rightarrow b$$

B não foi reconhecido como gerador, então $S \rightarrow AB$ não pode ser usada!

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos geradores de $G = (V, T, P, S)$:
  - **Base:** todo símbolo de T é sem dúvida gerador, pois ele gera a si mesmo;
  - **Indução:** suponha que exista uma produção $A \rightarrow \alpha$ e que todo símbolo de α já seja conhecido como gerador, então A é gerador. Essa regra inclui $\alpha = \varepsilon$.

$$S \rightarrow AB \mid a$$
$$A \rightarrow b$$

- Desse modo, o conjunto de símbolos geradores é $\{a, b, A, S\}$
- **Teorema 7.4**, página 275 (HOPCROFT et al., 2003)

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos alcançáveis de $G = (V, T, P, S)$:
  - **Base:** S é sem dúvida alcançável;
  - **Indução:** suponha que descobrimos que alguma variável A é alcançável. Então, para todas as produções com A na cabeça, os símbolos dos corpos dessas produções também são alcançáveis.

$$S \rightarrow AB \mid a$$
$$A \rightarrow b$$

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos alcançáveis de $G = (V, T, P, S)$:
  - **Base:** S é sem dúvida alcançável;
  - **Indução:** suponha que descobrimos que alguma variável A é alcançável. Então, para todas as produções com A na cabeça, os símbolos dos corpos dessas produções também são alcançáveis.

$$S \to AB \mid a$$
$$A \to b$$

Formas Normais para CFGs

Simplificações: remoção de símbolos inúteis

- Indução para calcular os símbolos alcançáveis de $G = (V, T, P, S)$:
  - **Base:** S é sem dúvida alcançável;
  - **Indução:** suponha que descobrimos que alguma variável A é alcançável. Então, para todas as produções com A na cabeça, os símbolos dos corpos dessas produções também são alcançáveis.

$$S \rightarrow AB \mid a$$
$$A \rightarrow b$$

- Desse modo, o conjunto de símbolos alcançáveis é $\{S, A, B, a, b\}$
- **Teorema 7.6**, página 276 (HOPCROFT et al., 2003)

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- Apesar de convenientes no projeto de gramáticas, as ε–produções não são essenciais;
- Note que, obviamente, caso uma gramática não tenha uma produção que tenha um corpo ε, é impossível gerar a string vazia como um elemento da linguagem;
- Sendo assim, se uma linguagem L tem uma CFG, então $L - \{\varepsilon\}$ tem uma CFG sem ε–produções.

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- Começaremos descobrindo quais variáveis são anuláveis. Uma variável A é anulável se $A \overset{*}{\Rightarrow} \varepsilon$. Se A é anulável, sempre que A aparecer em um corpo de produção, digamos $B \rightarrow CAD$, A poderá (ou não) derivar ε. Assim, criamos duas versões da produção, uma sem A no corpo ($B \rightarrow CD$), que corresponde ao caso em que A teria sido usada para derivar ε e a outra com A ainda presente ($B \rightarrow CAD$). Caso a versão com A no corpo seja usada, não poderemos permitir que A derive ε, mas como eliminaremos todas as produções com corpos ε, conseguiremos resolver essa questão.

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- Seja $G = (V, T, P, S)$ uma CFG, encontraremos os símbolos anuláveis usando o algoritmo abaixo:
  - **Base:** se $A \rightarrow \varepsilon$ é uma produção de G, então A é anulável;
  - **Indução:** se existe uma produção $B \rightarrow C_1 C_2 \ldots C_k$, onde cada C_i é anulável, então B é anulável. Cada C_i deve ser uma variável para ser anulável, sendo assim só teremos que considerar produções com corpos que contém apenas variáveis;
- **Teorema 7.7**, página 277 (HOPCROFT et al., 2003)

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:**
$$S \rightarrow AB$$
$$A \rightarrow aAA \mid \varepsilon$$
$$B \rightarrow bBB \mid \varepsilon$$

- **Primeiro passo:** quais símbolos são anuláveis?

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:** $S \rightarrow AB$
  $A \rightarrow aAA \mid \varepsilon$
  $B \rightarrow bBB \mid \varepsilon$

- **Primeiro passo:** quais símbolos são anuláveis?
  - A e B são diretamente anuláveis, pois têm produções com ε como corpo;

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:** $S \rightarrow AB$
  $A \rightarrow aAA \mid \varepsilon$
  $B \rightarrow bBB \mid \varepsilon$

- **Primeiro passo:** quais símbolos são anuláveis?
  - A e B são diretamente anuláveis, pois têm produções com ε como corpo;
  - S é anulável, porque a produção $S \rightarrow AB$ tem um corpo que consiste apenas em símbolos anuláveis;

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:** $S \rightarrow AB$

$$A \rightarrow aAA \mid \varepsilon$$
$$B \rightarrow bBB \mid \varepsilon$$

- **Primeiro passo:** quais símbolos são anuláveis?
  - A e B são diretamente anuláveis, pois têm produções com ε como corpo;
  - S é anulável, porque a produção $S \rightarrow AB$ tem um corpo que consiste apenas em símbolos anuláveis;
  - Desse modo, todas as três variáveis são anuláveis.

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:** $S \rightarrow AB$
  $A \rightarrow aAA \mid \varepsilon$
  $B \rightarrow bBB \mid \varepsilon$

- **Segundo passo:** construir a gramática G_1

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:** $S \rightarrow AB$
  $A \rightarrow aAA \mid \varepsilon$
  $B \rightarrow bBB \mid \varepsilon$

- **Segundo passo:** construir a gramática G_1
  - **Considerando $S \rightarrow AB$:** todos os símbolos do corpo são anuláveis e assim há quatro maneiras de escolher presente ou ausente para A e B (AB, A, B e ambos ausentes), mas não podemos optar em tornar todos os símbolos ausentes, sendo assim, teremos três produções:

$$G_1\colon\ S \rightarrow AB \mid A \mid B$$

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:**

$$S \rightarrow AB$$
$$A \rightarrow aAA \mid \varepsilon$$
$$B \rightarrow bBB \mid \varepsilon$$

- **Segundo passo:** construir a gramática G_1
  - **Considerando $A \rightarrow aAA$:** a segunda e a terceira posição contém símbolos anuláveis e assim há quatro maneiras de escolher presente ou ausente para A e A (aAA, aA, aA, a), sendo as quatro permitidas. Note que há duas opções que geram a mesma produção (aA), podendo então eliminar uma delas:

$$G_1\text{: } S \rightarrow AB \mid A \mid B$$
$$A \rightarrow aAA \mid aA \mid a$$

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Exemplo:** $S \to AB$
  $A \to aAA \mid \varepsilon$
  $B \to bBB \mid \varepsilon$

- **Segundo passo:** construir a gramática G_1
  - **Considerando $B \to bBB$:** a segunda e a terceira posição contém símbolos anuláveis e assim há quatro maneiras de escolher presente ou ausente para B e B (bBB, bB, bB, b), sendo as quatro permitidas. Note que há duas opções que geram a mesma produção (bB), podendo então eliminar uma delas:

$$
\begin{aligned}
G_1:\ S &\to AB \mid A \mid B \\
A &\to aAA \mid aA \mid a \\
B &\to bBB \mid bB \mid b
\end{aligned}
$$

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Formas Normais para CFGs

Simplificações: eliminação de ε-produções

- **Teorema 7.9:** se a gramática G_1 é construída a partir de G pela construção anterior para eliminar ε–produções, então:

$$L(G_1) = L(G) - \{\varepsilon\}$$

- **Prova do Teorema:** página 279 (HOPCROFT et al., 2003).

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Uma produção unitária é uma produção na forma $A \rightarrow B$, onde tanto A quanto B são variáveis;
- Lembre-se que esse tipo de produção pode ser útil, pois anteriormente vimos que o uso das produções unitárias $E \rightarrow T$ e $T \rightarrow F$ nos permitiu criar uma gramática não-ambígua para expressões aritméticas simples:

$$
\begin{array}{lcl}
E & \rightarrow & T \mid E + T \\
T & \rightarrow & F \mid T * F \\
F & \rightarrow & I \mid (E) \\
I & \rightarrow & a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{array}
$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

produção unitária

$$\begin{aligned} &\boxed{E \rightarrow T} \mid E + T \\ &T \rightarrow F \mid T * F \\ &F \rightarrow I \mid (E) \\ &I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \end{aligned}$$

$$E \rightarrow T$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

expansão de T

$$E \rightarrow F \mid T * F$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

produção unitária

$$\boxed{E \rightarrow F} \mid T * F$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

$$E \rightarrow T \mid E + T$$
$$T \rightarrow F \mid T * F$$
$$F \rightarrow I \mid (E)$$
$$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

expansão de F

$$E \rightarrow I \mid (E) \mid T * F$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

produção unitária

$$\boxed{E \rightarrow I} \mid (E) \mid T * F$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

expansão de I

$$E \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \mid (E) \mid T * F$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Apesar de úteis no contexto anterior, ou seja, na remoção de ambiguidade em uma CFG, as produções unitárias podem complicar certas provas, além de introduzirem etapas extras em derivações que não precisam estar lá;
- Como remover essas produções?

$$
\begin{array}{lcl}
E & \rightarrow & a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \mid (E) \mid T * F \mid E + T \\
T & \rightarrow & F \mid T * F \\
F & \rightarrow & I \mid (E) \\
I & \rightarrow & a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{array}
$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- Problema!
  - A técnica da expansão funciona com frequência, entretanto ela pode falhar em situações onde existem ciclos em produções unitárias, por exemplo;
  - $A \rightarrow B$
    $B \rightarrow C$
    $C \rightarrow A$
- O que fazer?
  - Primeiro, encontrar todos os pares A e B tais que $A \overset{*}{\Rightarrow} B$ que usam somente uma sequência de produções unitárias. Esses pares são chamados de pares unitários;
  - Pode ser que $A \overset{*}{\Rightarrow} B$ seja verdade, mas que não haja derivações com produções unitárias, por exemplo $A \rightarrow BC$ e $C \rightarrow \varepsilon$.

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \overset{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \overset{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Base: (E, E), (T, T), (F, F) e (I, I)

Indução:

1. (E, E) e a produção $E \rightarrow T$ nos dão o par unitário (E, T)
2. (E, T) e a produção $T \rightarrow F$ nos dão o par unitário (E, F)
3. (E, F) e a produção $F \rightarrow I$ nos dão o par unitário (E, I)
4. (T, T) e a produção $T \rightarrow F$ nos dão o par unitário (T, F)
5. (T, F) e a produção $F \rightarrow I$ nos dão o par unitário (T, I)
6. (F, F) e a produção $F \rightarrow I$ nos dão o par unitário (F, I)

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \overset{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Base: (E, E), (T, T), (F, F) e (I, I)

Indução:

1. (E, E) e a produção $E \rightarrow T$ nos dão o par unitário (E, T)
2. (E, T) e a produção $T \rightarrow F$ nos dão o par unitário (E, F)
3. (E, F) e a produção $F \rightarrow I$ nos dão o par unitário (E, I)
4. (T, T) e a produção $T \rightarrow F$ nos dão o par unitário (T, F)
5. (T, F) e a produção $F \rightarrow I$ nos dão o par unitário (T, I)
6. (F, F) e a produção $F \rightarrow I$ nos dão o par unitário (F, I)

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \overset{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Base: (E, E), (T, T), (F, F) e (I, I)

Indução:

1. (E, E) e a produção $E \rightarrow T$ nos dão o par unitário (E, T)
2. (E, T) e a produção $T \rightarrow F$ nos dão o par unitário (E, F)
3. (E, F) e a produção $F \rightarrow I$ nos dão o par unitário (E, I)
4. (T, T) e a produção $T \rightarrow F$ nos dão o par unitário (T, F)
5. (T, F) e a produção $F \rightarrow I$ nos dão o par unitário (T, I)
6. (F, F) e a produção $F \rightarrow I$ nos dão o par unitário (F, I)

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \stackrel{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Base: (E, E), (T, T), (F, F) e (I, I)

Indução:

1. (E, E) e a produção $E \rightarrow T$ nos dão o par unitário (E, T)
2. (E, T) e a produção $T \rightarrow F$ nos dão o par unitário (E, F)
3. (E, F) e a produção $F \rightarrow I$ nos dão o par unitário (E, I)
4. (T, T) e a produção $T \rightarrow F$ nos dão o par unitário (T, F)
5. (T, F) e a produção $F \rightarrow I$ nos dão o par unitário (T, I)
6. (F, F) e a produção $F \rightarrow I$ nos dão o par unitário (F, I)

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \overset{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Base: (E, E), (T, T), (F, F) e (I, I)

Indução:

1. (E, E) e a produção $E \rightarrow T$ nos dão o par unitário (E, T)
2. (E, T) e a produção $T \rightarrow F$ nos dão o par unitário (E, F)
3. (E, F) e a produção $F \rightarrow I$ nos dão o par unitário (E, I)
4. (T, T) e a produção $T \rightarrow F$ nos dão o par unitário (T, F)
5. (T, F) e a produção $F \rightarrow I$ nos dão o par unitário (T, I)
6. (F, F) e a produção $F \rightarrow I$ nos dão o par unitário (F, I)

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \overset{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Base: (E, E), (T, T), (F, F) e (I, I)

Indução:

1. (E, E) e a produção $E \rightarrow T$ nos dão o par unitário (E, T)
2. (E, T) e a produção $T \rightarrow F$ nos dão o par unitário (E, F)
3. (E, F) e a produção $F \rightarrow I$ nos dão o par unitário (E, I)
4. (T, T) e a produção $T \rightarrow F$ nos dão o par unitário (T, F)
5. (T, F) e a produção $F \rightarrow I$ nos dão o par unitário (T, I)
6. (F, F) e a produção $F \rightarrow I$ nos dão o par unitário (F, I)

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Encontrando os pares unitários:**
  - **Base:** (A, A) é um par unitário para qualquer variável A, insto é, $A \overset{*}{\Rightarrow} A$ em zero etapas;
  - **Indução:** Suponha que descobrimos que (A, B) é um par unitário e $B \rightarrow C$ é uma produção onde C é uma variável. Sendo assim, (A, C) é um par unitário.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Base: (E, E), (T, T), (F, F) e (I, I)

Indução:

1. (E, E) e a produção $E \rightarrow T$ nos dão o par unitário (E, T)
2. (E, T) e a produção $T \rightarrow F$ nos dão o par unitário (E, F)
3. (E, F) e a produção $F \rightarrow I$ nos dão o par unitário (E, I)
4. (T, T) e a produção $T \rightarrow F$ nos dão o par unitário (T, F)
5. (T, F) e a produção $F \rightarrow I$ nos dão o par unitário (T, I)
6. (F, F) e a produção $F \rightarrow I$ nos dão o par unitário (F, I)

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Eliminação de produções unitárias:**
  - Dada uma CFG $G = (V, T, P, S)$, construa $G_1 = (V, T, P_1, S)$:
    1. Encontre todos os pares unitários de G; (slide anterior)
    2. Para cada par unitário (A, B), adicione a P_1 todas as produções $A \rightarrow \alpha$, onde $B \rightarrow \alpha$ é uma produção não-unitária em P. Observe que $A = B$ é possível; desse modo P_1 contém todas as produções não-unitárias em P.

Par	Produções
(E, E)	
(E, T)	
(E, F)	
(E, I)	
(T, T)	
(T, F)	
(T, I)	
(F, F)	
(F, I)	
(I, I)	

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Eliminação de produções unitárias:**
  - Dada uma CFG $G = (V, T, P, S)$, construa $G_1 = (V, T, P_1, S)$:
    1. Encontre todos os pares unitários de G;
    2. Para cada par unitário (A, B), adicione a P_1 todas as produções $A \rightarrow \alpha$, onde $B \rightarrow \alpha$ é uma produção não-unitária em P. Observe que $A = B$ é possível; desse modo P_1 contém todas as produções não-unitárias em P.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Par	Produções
(E, E)	$E \rightarrow E + T$
(E, T)	$E \rightarrow T * F$
(E, F)	$E \rightarrow (E)$
(E, I)	$E \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(T, T)	$T \rightarrow T * F$
(T, F)	$T \rightarrow (E)$
(T, I)	$T \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(F, F)	$F \rightarrow (E)$
(F, I)	$F \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(I, I)	$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Eliminação de produções unitárias:**
  - Dada uma CFG $G = (V, T, P, S)$, construa $G_1 = (V, T, P_1, S)$:
    1. Encontre todos os pares unitários de G;
    2. Para cada par unitário (A, B), adicione a P_1 todas as produções $A \rightarrow \alpha$, onde $B \rightarrow \alpha$ é uma produção não-unitária em P. Observe que $A = B$ é possível; desse modo P_1 contém todas as produções não-unitárias em P.

$E \rightarrow T \mid E + T$

$T \rightarrow F \mid T * F$

$F \rightarrow I \mid (E)$

$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

Par	Produções
(E, E)	$E \rightarrow E + T$
(E, T)	$E \rightarrow T * F$
(E, F)	$E \rightarrow (E)$
(E, I)	$E \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(T, T)	$T \rightarrow T * F$
(T, F)	$T \rightarrow (E)$
(T, I)	$T \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(F, F)	$F \rightarrow (E)$
(F, I)	$F \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(I, I)	$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

Formas Normais para CFGs

Simplificações: eliminação de produções unitárias

- **Eliminação de produções unitárias:**
  - Dada uma CFG $G = (V, T, P, S)$, construa $G_1 = (V, T, P_1, S)$:
    1. Encontre todos os pares unitários de G;
    2. Para cada par unitário (A, B), adicione a P_1 todas as produções $A \rightarrow \alpha$, onde $B \rightarrow \alpha$ é uma produção não-unitária em P. Observe que $A = B$ é possível; desse modo P_1 contém todas as produções não-unitárias em P.

$$
\begin{aligned}
E &\rightarrow T \mid E + T \\
T &\rightarrow F \mid T * F \\
F &\rightarrow I \mid (E) \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

eliminação de produções unitárias

$$
\begin{aligned}
E &\rightarrow E + T \mid T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \\
T &\rightarrow T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \\
F &\rightarrow (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Formas Normais para CFGs

Aplicação das Simplificações

- Para converter qualquer CFG G em uma CFG equivalente que não tenha símbolos inúteis, ε-produções e/ou produções unitárias, devemos aplicar as simplificações na seguinte ordem:
  1. Eliminar ε-produções;
  2. Eliminar produções unitárias;
  3. Eliminar símbolos inúteis.

Forma Normal de Chomsky

- Toda gramática G de uma CFL não-vazia sem ε está na Forma Normal de Chomsky (CNF[1]) se não possui símbolos inúteis e se todas as suas produções estão em uma dentre as duas formas simples:
  1. $A \rightarrow BC$, onde A, B e C são todas variáveis;
  2. $A \rightarrow a$, onde A é uma variável e a é um terminal.

Avram Noam Chomksy

- A conversão de uma gramática para a CNF consiste em primeiramente em simplificar a gramática eliminando ε-produções, produções unitárias e símbolos inúteis. Essa simplificação gerará uma gramática onde todas as produções terão a forma $A \rightarrow a$, permitida na CNF, ou terão um corpo de comprimento maior ou igual a 2. Após esse processo, nossa tarefa será:
  a) Organizar todos os corpos de comprimento maior ou igual a 2 que consistem apenas em variáveis;
  b) Desmembrar os corpos de comprimento 3 ou mais em uma cascata de produções, cada uma com um corpo consistindo em duas variáveis.

[1]*Chomsky Normal Form*

Forma Normal de Chomsky

a) Organizar todos os corpos de comprimento maior ou igual a 2 que consistem apenas em variáveis:

- Para todo terminal a que aparecer em um corpo de comprimento 2 ou mais, crie uma nova variável, digamos A. Essa variável terá apenas uma produção, $A \rightarrow a$. Agora usamos A em lugar de a em todo lugar em que a aparecer em um corpo de comprimento 2 ou mais. Nesse ponto, toda produção terá um corpo que será um único terminal ou pelo menos duas variáveis e nenhum terminal.

Forma Normal de Chomsky

b) Desmembrar os corpos de comprimento 3 ou mais em uma cascata de produções, cada uma com um corpo consistindo em duas variáveis:

- Desmembrar as produções $A \rightarrow B_1 B_2 \ldots B_k$, para $k \geq 3$, em um grupo de produções com duas variáveis em cada corpo. Com isso, introduzimos $k-2$ variáveis, $C_1, C_2, \ldots, C_{k-2}$. A produção original é substituída pelas $k-1$ produções:

$$
\begin{array}{lcl}
A & \rightarrow & B_1 C_1 \\
C_1 & \rightarrow & B_2 C_2 \\
 & \ldots & \\
C_{k-3} & \rightarrow & B_{k-2} C_{k-2} \\
C_{k-2} & \rightarrow & B_{k-1} B_k
\end{array}
$$

Forma Normal de Chomsky

- **Exemplo:**

$$
\begin{aligned}
E &\rightarrow E + T \mid T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \\
T &\rightarrow T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \\
F &\rightarrow (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \\
I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{aligned}
$$

Forma Normal de Chomsky

$$E \rightarrow E + T \mid T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$
$$T \rightarrow T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$
$$F \rightarrow (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$
$$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

- Parte a) Organizar todos os corpos de comprimento maior ou igual a 2 que consistem apenas em variáveis:
  - Oito terminais: a, b, 0, 1, $+$, $*$, (e), aparecendo em corpos que não tem um único terminal;
  - Criaremos oito novas variáveis que corresponderão à esses terminais e suas oito produções correspondentes ($A \rightarrow a$, $B \rightarrow b$, $Z \rightarrow 0$, $O \rightarrow 1$, $P \rightarrow +$, $M \rightarrow *$, $L \rightarrow ($, $R \rightarrow)$) e realizaremos as substituições necessárias:

$$E \rightarrow EPT \mid TMF \mid LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO$$
$$T \rightarrow TMF \mid LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO$$
$$F \rightarrow LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO$$
$$I \rightarrow a \mid b \mid IA \mid IB \mid IZ \mid IO$$
$$A \rightarrow a$$
$$B \rightarrow b$$
$$Z \rightarrow 0$$
$$O \rightarrow 1$$
$$P \rightarrow +$$
$$M \rightarrow *$$
$$L \rightarrow ($$
$$R \rightarrow)$$

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Forma Normal de Chomsky

$$E \rightarrow E + T \mid T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$
$$T \rightarrow T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$
$$F \rightarrow (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$
$$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

- Parte a) Organizar todos os corpos de comprimento maior ou igual a 2 que consistem apenas em variáveis:
  - Oito terminais: a, b, 0, 1, $+$, $*$, (e), aparecendo em corpos que não tem um único terminal;
  - Criaremos oito novas variáveis que corresponderão à esses terminais e suas oito produções correspondentes ($A \rightarrow a$, $B \rightarrow b$, $Z \rightarrow 0$, $O \rightarrow 1$, $P \rightarrow +$, $M \rightarrow *$, $L \rightarrow ($, $R \rightarrow)$) e realizaremos as substituições necessárias:

$$E \rightarrow \underline{EPT} \mid \underline{TMF} \mid \underline{LER} \mid a \mid b \mid \underline{IA} \mid \underline{IB} \mid \underline{IZ} \mid \underline{IO}$$
$$T \rightarrow \underline{TMF} \mid \underline{LER} \mid a \mid b \mid \underline{IA} \mid \underline{IB} \mid \underline{IZ} \mid \underline{IO}$$
$$F \rightarrow \underline{LER} \mid a \mid b \mid \underline{IA} \mid \underline{IB} \mid \underline{IZ} \mid \underline{IO}$$
$$I \rightarrow a \mid b \mid \underline{IA} \mid \underline{IB} \mid \underline{IZ} \mid \underline{IO}$$
$$A \rightarrow a$$
$$B \rightarrow b$$
$$Z \rightarrow 0$$
$$O \rightarrow 1$$
$$P \rightarrow +$$
$$M \rightarrow *$$
$$L \rightarrow ($$
$$R \rightarrow)$$

Agora, todas as produções estão na CNF, com exceção das que têm corpos com comprimento 3: EPT, TMF e LER

Forma Normal de Chomsky

$$
\begin{aligned}
E &\to EPT \mid TMF \mid LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
T &\to TMF \mid LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
F &\to LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
I &\to a \mid b \mid IA \mid IB \mid IZ \mid IO
\end{aligned}
$$

...

- Parte b) Desmembrar os corpos de comprimento 3 ou mais em uma cascata de produções, cada uma com um corpo consistindo em duas variáveis:
  - Para cada um dos três corpos (EPT, TMF e LER), introduziremos uma variável extra;
    - Para EPT, introduzimos C_1 e substituímos $E \to EPT$ por $E \to EC_1$ e $C_1 \to PT$;
    - Para TMF, introduzimos C_2 e substituímos $E \to TMF$ e $T \to TMF$ por $E \to TC_2$, $T \to TC_2$ e $C_2 \to MF$;
    - Para LER, introduzimos C_3 e substituímos $E \to LER$, $T \to LER$ e $F \to LER$ por $E \to LC_3$, $T \to LC_3$, $F \to LC_3$ e $C_3 \to ER$

$$
\begin{aligned}
E &\to EC_1 \mid TC_2 \mid LC_3 \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
T &\to TC_2 \mid LC_3 \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
F &\to LC_3 \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
I &\to a \mid b \mid IA \mid IB \mid IZ \mid IO \\
A &\to a \\
B &\to b \\
Z &\to 0 \\
O &\to 1 \\
P &\to + \\
M &\to * \\
L &\to (\\
R &\to) \\
C_1 &\to PT \\
C_2 &\to MF \\
C_3 &\to ER
\end{aligned}
$$

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Forma Normal de Chomsky

- Parte b) Desmembrar os corpos de comprimento 3 ou mais em uma cascata de produções, cada uma com um corpo consistindo em duas variáveis:
  - Para cada um dos três corpos (EPT, TMF e LER), introduziremos uma variável extra;
    - Para EPT, introduzimos C_1 e substituímos $E \to EPT$ por $E \to EC_1$ e $C_1 \to PT$;
    - Para TMF, introduzimos C_2 e substituímos $E \to TMF$ e $T \to TMF$ por $E \to TC_2$, $T \to TC_2$ e $C_2 \to MF$;
    - Para LER, introduzimos C_3 e substituímos $E \to LER$, $T \to LER$ e $F \to LER$ por $E \to LC_3$, $T \to LC_3$, $F \to LC_3$ e $C_3 \to ER$

$$
\begin{aligned}
E &\to EPT \mid TMF \mid LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
T &\to TMF \mid LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
F &\to LER \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
I &\to a \mid b \mid IA \mid IB \mid IZ \mid IO \\
&\dots
\end{aligned}
$$

$$
\begin{aligned}
E &\to \underline{EC_1} \mid \underline{TC_2} \mid \underline{LC_3} \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
T &\to \underline{TC_2} \mid \underline{LC_3} \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
F &\to \underline{LC_3} \mid a \mid b \mid IA \mid IB \mid IZ \mid IO \\
I &\to a \mid b \mid IA \mid IB \mid IZ \mid IO \\
A &\to a \\
B &\to b \\
Z &\to 0 \\
O &\to 1 \\
P &\to + \\
M &\to * \\
L &\to (\\
R &\to) \\
C_1 &\to PT \\
C_2 &\to MF \\
C_3 &\to ER
\end{aligned}
$$

Com isso, obtemos a gramática final na CNF.

Forma Normal de Greibach

- Toda gramática G de uma CFL não-vazia sem ε está na Forma Normal de Greibach (GNF[2]) se cada uma das produções são da forma $A \rightarrow a\alpha$, onde a é um terminal e α é uma string de uma ou mais variáveis.
- O processo de conversão para a GNF não é simples;
- Pode-se iniciar realizando primeiro a conversão para a CNF;
- De modo geral, expande-se a primeira variável de cada produção até a obtenção de um terminal, entretanto como há a possibilidade de existirem ciclos onde nunca se alcançará um terminal, é necessário "provocar um curto-circuito no processo", criando uma produção que introduza um terminal como primeiro símbolo do corpo e que seja seguida por variáveis para gerar todas as sequências de variáveis que poderiam ter sido geradas no processo de geração desse terminal;
- Implica na derivação de w em exatamente n etapas, pois o uso de cada produção introduz exatamente um terminal em uma forma sentencial.

Sheila Greibach

[2]Greibach Normal Form

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Teorema: (*O lema do bombeamento para linguagens livres de contexto*) Seja L uma CFL. Então, existe uma constante n tal que, se z é qualquer string em L tal que $|z|$ é pelo menos n ($|z| \geq n$), podemos dividir z em cinco strings, $z = uvwxy$, tais que:

1. $|vwx| \leq n$. Ou seja, a porção intermediária não é muito longa;
2. $vx \neq \varepsilon$. Tendo em vista que v e x são os fragmentos a serem "bombeados", essa condição diz que pelo menos uma das strings que bombeamos não deve ser vazia;
3. Para todo $i \geq 0$, a string uv^iwx^iy está em L. Isto é, as duas strings v e x podem ser "bombeadas" qualquer número de vezes, incluindo 0, e a string resultante ainda será um elemento de L.

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Aplicação: Simular um jogo entre dois jogadores:

1. Escolhemos uma linguagem L que queremos mostrar que não é uma CFL;
2. O oponente escolhe n, que não conhecemos e, portanto, devemos fazer planos tendo em vista qualquer n possível;
3. Escolhemos z, podendo usar n como parâmetro ao fazê-lo;
4. O oponente desmembra z em $uvwxy$, sujeito apenas às restrições de que $|vwx| \geq n$ e $vx \neq \varepsilon$;
5. Ganhamos o jogo, se pudermos, escolhendo i e mostrando que uv^iwx^iy não está em L.

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Exemplo 1 (CFLs podem fazer a correspondência entre a igualdade ou desigualdade de dois grupos de símbolos, mas não de três grupos): Seja L a linguagem $\{0^n1^n2^n \mid n \geq 1\}$. Suponha que L seja livre de contexto. Então existe um inteiro n dado pelo lema do bombeamento (não é o n da definição da linguagem). Vamos escolher $z = 0^n1^n2^n$.

Suponha que o oponente desmembrou z como $z = uvwxy$, onde $|vwx| \leq n$ e v e x não são ambos ε. Sabemos assim que vwx não pode envolver ao mesmo tempo 0's e 2's, pois o último 0 e o primeiro 2 estão separados por $n + 1$ posições.

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Exemplo 1 (continuação): Provaremos que L contém alguma string conhecida por não estar em L, contradizendo assim a hipótese que L é uma CFL. Os casos são:

1. vwx não tem nenhum 2. Então, vx consiste em apenas 0's e 1's e tem pelo menos um desses símbolos. Portanto uwy, que teria que estar em L pelo lema do bombeamento, tem n 2's, mas tem menos de n 0's ou menos de n 1's, ou ambos. Assim, ela não pertence a L e concluímos que L não é uma CFL nesse caso.
2. vwx não tem nenhum 0. De modo semelhante, uwy tem n 0's, mas tem um número menor de 1's ou 2's. Portanto, ele não está em L.

Qualquer que seja o caso, concluímos que L tem uma string que sabemos que não está em L. Essa contradição nos permite concluir que nossa hipótese estava errada; L não é uma CFL. $\square$

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Exemplo 2 (CFLs não podem comparar dois pares com números iguais de símbolos quando esses pares se intercalam): Seja L a linguagem $\{0^i 1^j 2^i 3^j \mid i \geq 1 \text{ e } j \geq 1\}$. Se L é livre de contexto, seja n a constante para L e escolhemos $z = 0^n 1^n 2^n 3^n$. Podemos escrever $z = uvwxy$ sujeita às restrições usuais $|vwx| \leq n$ e $vx \neq \varepsilon$. Então vwx está contido na substring de um símbolo ou inclui dois símbolos adjacentes.

Se vwx consiste em apenas um símbolo, então uwy tem n de três diferentes símbolos e menos de n do quarto símbolo. Desse modo, ele não pode estar em L. Se vwx inclui dois símbolos, digamos 1's e 2's, então uwy está omitindo alguns 1's ou alguns 2's, ou ambos. Supondo que ela esteja omitindo 1's, como existem n 3's, essa string não pode estar em L. De modo semelhante, se ele estiver omitindo 2's, como ele tem n 0's, uwy não pode estar em L. Contestamos assim a hipótese de que L é uma CFL e concluímos que ela não é. $\square$

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Exemplo 3 (CFLs não podem comparar duas strings de comprimento arbitrário, se essas strings forem escolhidas a partir de um alfabeto com mais de um símbolo – implica na impossibilidade de aplicação de CFGs na análise semântica de compiladores): Seja $L = \{ww \mid w \text{ está em } \{0,1\}^*\}$, isto é, L consiste em strings repetitivos como ε, 0101, 00100010 ou 110110. Se L é livre de contexto, seja n a constante para o seu lema do bombeamento. Considere $z = 0^n1^n0^n1^n$. Essa string é 0^n1^n repetido e assim z está em L.

Podemos escrever $z = uvwxy$ tal que $|vwx| \leq n$ e $vx \neq \varepsilon$. Mostraremos que uwy não está em L e, portanto, mostraremos por contradição que L não deve ser uma CFL.

Primeiro, observe que, como $|vwx| \leq n$, $|uwy| \geq 3n$. Desse modo, se uwy é alguma string repetitiva, digamos tt, então t tem comprimento pelo menos igual a $3n/2$. Existem vários casos a considerar, dependendo de onde vwx está no interior de z:

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Exemplo 3 (continuação):

1. Suponha que vwx esteja dentro dos n primeiros 0's. Em particular, considere que vx consiste em k 0's, onde $k > 0$. Então, uwy começa com $0^{n-k}1^n$. Como $|uwy| = 4n - k$, sabendo que, se $uwy = tt$, então $|t| = 2n - k/2$. Desse modo, t não termina até depois do primeiro bloco de 1's; isto é, t termina em 0's. Porém, uwy termina em 1, e assim ele não pode ser igual a tt.
2. Suponha que vwx inclua o primeiro bloco de 0's e o primeiro bloco de 1's. Pode ser que vx consista apenas em 0's, se $x = \varepsilon$. Então, o argumento de que uwy não é da forma tt é o mesmo do caso (1). Se vx tem pelo menos um 1, então notamos que t, que tem comprimento pelo menos $3n/2$, deve terminar em 1^n, porque uwy termina em 1^n. Porém, não existe nenhum bloco de n 1's exceto o bloco final, e então t não pode se repetir em uwy.
3. Se vwx está contido no primeiro bloco 1's, então o argumento de que uwy não está em L é semelhante à segunda parte do caso (2).

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Propriedades das Linguagens Livres de Contexto

O Lema do Bombeamento para Linguagens Livres de Contexto

Exemplo 3 (continuação):

4. Suponha que vwx inclua o primeiro bloco de 1's e o segundo bloco de 0's. Se vx realmente não tem nenhum 0, então o argumento é o mesmo que teríamos se vwx estivesse contido no primeiro bloco de 1's. Se vx tem pelo menos um 0, então uwy começa com um bloco de n 0's, e o mesmo ocorreria com t se $uwy = tt$. No entanto, não existe nenhum outro bloco de n 0's em uwy para a segunda cópia de t. Concluímos que também nesse caso uwy não está em L.
5. Nos outros casos, em que vwx está na segunda metade de z, o argumento é simétrico ao dos casos em que vwx está contido na primeira metade de z.

Portanto, em nenhum caso uwy está em L, e concluímos que L não é livre de contexto. $\square$

INSTITUTO FEDERAL DE
EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Propriedades das Linguagens Livres de Contexto

Propriedades de Fechamento das Linguagens Livres de Contexto

- Substituições (generalização do homomorfismo):
  - Substituição de cada símbolo de uma linguagem por uma linguagem inteira;
  - **Teorema 7.23:** Se L é uma linguagem livre de contexto sobre o alfabeto Σ e s é uma substituição em Σ tal que $s(a)$ é uma CFL para cada a em Σ, então $s(L)$ é uma CFL.
- Fechamento sob operações:
  - União;
  - Concatenação;
  - Fechamento $(*)$ e Fechamento Positivo $(+)$;
  - Homomorfismo e Homomorfismo Inverso;
  - Reversão;
- CFLs não são fechadas sob intersecção ou diferença, enquanto as linguagens regulares o são. Entretanto, essas duas operações aplicadas à uma CFL e uma linguagem regular é sempre uma CFL.

Propriedades das Linguagens Livres de Contexto

Propriedades de Decisão das Linguagens Livres de Contexto

- Conversão entre representações:
  - CFG → PDA;
  - PDA que aceita pelo estado final → PDA que aceita por pilha vazia;
  - PDA que por pilha vazia → PDA que aceita pelo estado final;
- Teste de caráter vazio de CFLs;
  - Dada uma gramática G, decidir se o símbolo de início S é gerador, ou seja, se S deriva pelo menos uma string. L é vazia se e somente se S não é gerador;
- Teste de pertinência em CFLs;
  - Obter uma CNF para L, o que implica em derivações em uma árvore binária. Se w tiver comprimento n, haverá exatamente $2n - 1$ nós rotulados por variáveis. Outra forma mais eficiente é a aplicação do algoritmo CYK baseado em programação dinâmica, apresentado nas páginas 320 a 324 (HOPCROFT et al., 2003).

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Propriedades das Linguagens Livres de Contexto

Estudo dos Problemas Indecidíveis

- Nas próximas aulas estudaremos a teoria que nos permitirá provar formalmente que existem problemas que não podermos resolver por qualquer algoritmo que possa funcionar em um computador!
- No momento podemos apenas enumerar algumas questões indecidíveis mais significativas relacionadas às CFLs e CFGs;
- Não existem algoritmos para:
  - Determinar se uma CFG é ambígua;
    - A gramática é ambígua?
  - Determinar se uma CFL L é inerentemente ambígua;
    - Existe uma gramática G' equivalente a G que não é ambígua?
  - Verificar se a intersecção de duas CFLs é vazia;
  - Verificar se duas CFLs são iguais;
  - Verificar se uma CFL é igual a Σ^*, onde Σ é o alfabeto da linguagem.

pergunta sobre a gramática

perguntas sobre as linguagens,
pressupondo a existência
de um PDA ou de uma CFG

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
SÃO PAULO
Campus São João da Boa Vista

Propriedades das Linguagens Livres de Contexto

Exercícios Escritos

Exercício e9.1: Encontre uma gramática equivalente a gramática abaixo que não tenha símbolos inúteis.

$$
\begin{array}{lcl}
S & \rightarrow & AB \mid CA \\
A & \rightarrow & a \\
B & \rightarrow & BC \mid AB \\
C & \rightarrow & aB \mid b
\end{array}
$$

Propriedades das Linguagens Livres de Contexto

Exercícios Escritos

Exercício e9.2: Usando a gramática abaixo:

$$
\begin{aligned}
S &\rightarrow ASB \mid \varepsilon \\
A &\rightarrow aAS \mid a \\
B &\rightarrow SbS \mid A \mid bb
\end{aligned}
$$

a) Elimine as ε-produções.
b) Elimine quaisquer produções unitárias na gramática resultante de a).
c) Elimine quaisquer símbolos inúteis na gramática resultante de b).
d) Coloque a gramática resultante de c) na Forma Normal de Chomsky.

Propriedades das Linguagens Livres de Contexto

Exercícios Escritos

Exercício e9.3: Usando a gramática abaixo:

$$
\begin{aligned}
S &\rightarrow 0A0 \mid 1B1 \mid BB \\
A &\rightarrow C \\
B &\rightarrow S \mid A \\
C &\rightarrow S \mid \varepsilon
\end{aligned}
$$

a) Elimine as ε-produções.

b) Elimine quaisquer produções unitárias na gramática resultante de a).

c) Elimine quaisquer símbolos inúteis na gramática resultante de b).

d) Coloque a gramática resultante de c) na Forma Normal de Chomsky.

Propriedades das Linguagens Livres de Contexto

Exercícios Escritos

Exercício e9.4: Usando a gramática abaixo:

$$
\begin{aligned}
S &\rightarrow AAA \mid B \\
A &\rightarrow aA \mid B \\
B &\rightarrow \varepsilon
\end{aligned}
$$

a) Elimine as ε-produções.

b) Elimine quaisquer produções unitárias na gramática resultante de a).

c) Elimine quaisquer símbolos inúteis na gramática resultante de b).

d) Coloque a gramática resultante de c) na Forma Normal de Chomsky.

Propriedades das Linguagens Livres de Contexto

Exercícios Escritos

Exercício e9.5: Usando a gramática abaixo:

$$
\begin{aligned}
S &\rightarrow aAa \mid bBb \mid \varepsilon \\
A &\rightarrow C \mid a \\
B &\rightarrow C \mid b \\
C &\rightarrow CDE \mid \varepsilon \\
D &\rightarrow A \mid B \mid ab
\end{aligned}
$$

a) Elimine as ε-produções.

b) Elimine quaisquer produções unitárias na gramática resultante de a).

c) Elimine quaisquer símbolos inúteis na gramática resultante de b).

d) Coloque a gramática resultante de c) na Forma Normal de Chomsky.

Propriedades das Linguagens Livres de Contexto

Exercícios Escritos

Exercício e9.6: Projete uma gramática na CNF para o conjunto de strings de parênteses balanceados. **Dica:** projete uma gramática sem necessariamente estar na CNF e depois realize a normalização.

Bibliografia

HOPCROFT, J. E.; ULLMAN, J. D.; MOTWANI, R. **Introdução à Teoria de Autômatos, Linguagens e Computação**. 2. ed. Rio de Janeiro: Elsevier, 2002. 560 p.

RAMOS, M. V. M.; JOSÉ NETO, J.; VEGA, I. S. **Linguagens Formais: Teoria, Modelagem e Implementação**. Porto Alegre: Bookman, 2009. 656 p.

SIPSER, M. **Introdução à Teoria da Computação**. 2. ed. São Paulo: Cengage Learning, 2017. 459 p.