3. Gramàtiques incontextuals

Teoria de la Computació

FIB

Antoni Lozano

Q1 2024–2025

Gramàtiques incontextuals

Esquema general

Classes de llenguatges ↑

- Tots
- Decidibles (TM)
- NP (TM)
- P (TM)
- Incontextuals (CFG, PDA)
- Regulars (FA, NFA, λ-NFA, expressions regulars)

Esquema general

Models de càlcul ↑

- TM: *Turing Machine*
- PDA: *Push-Down Automaton (autòmat amb pila)*
  - DPDA: *Deterministic PDA*
- CFG: *Context-Free Grammar (gramàtica incontextual)*
- FA: *Finite Automaton*
  - DFA: *Deterministic Finite Automaton*
  - NFA: *Nondeterministic Finite Automaton*
  - λ-NFA: *Lambda Nondeterministic FA*

Motivació

Sabem que:

Teorema

El llenguatge $\{0^n 1^n \mid n \geq 0\}$ no és regular.

Però sovint cal reconèixer aquest llenguatge, com ara en la parentització dels llenguatges de programació.

Ho farem introduint els llenguatges incontextuals.

Motivació

Sabem que:

Teorema

El llenguatge $\{0^n 1^n \mid n \geq 0\}$ no és regular.

Però sovint cal reconèixer aquest llenguatge, com ara en la parentització dels llenguatges de programació.

Ho farem introduint els llenguatges incontextuals.

Terminologia i notació

Exemple 1

Gramàtica incontextual G, amb $\Sigma = \{0, 1\}$:

$$S \to 0\ S\ \ 1$$
$$S \to \lambda$$

- Conté dues regles de substitució (produccions)
- Té una variable (S) i un terminal (0)
- Permet fer infinites derivacions de mots. Per exemple:

$$S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 0011.$$

- El llenguatge generat és $L(G) = \{0^n1^n \mid n \geq 0\}$.

Terminologia i notació

Exemple 2

Gramàtica incontextual G, amb $\Sigma = \{0, 1\}$:

$$\begin{aligned}
&S \rightarrow A \\
&S \rightarrow B \\
&A \rightarrow 0\ A\ 1 \\
&A \rightarrow \lambda \\
&B \rightarrow 1\ B\ 0 \\
&B \rightarrow \lambda
\end{aligned}$$

- La variable que genera el llenguatge se'n diu inicial
  i és la primera que escrivim: S
- Una derivació sempre comença per la variable inicial:
  $S \Rightarrow A \Rightarrow 0A1 \Rightarrow 0\lambda 1 = 01$
- El llenguatge generat és
  $L(G) = L_A \cup L_B = \{0^n 1^n \mid n \geq 0\} \cup \{1^n 0^n \mid n \geq 0\}$

Terminologia i notació

Exemple 2

Escriurem les gramàtiques de forma més compacta!

Convenció: Escriurem

$$A \to \alpha \mid \beta$$

en lloc de $A \to \alpha$ i $A \to \beta$.

La gramàtica anterior G l'escriurem com

$$\begin{aligned} &S \to A \mid B \\ &A \to 0\,A\,1 \mid \lambda \\ &B \to 1\,B\,0 \mid \lambda \end{aligned}$$

Gramàtiques incontextuals

Definició de gramàtica

Definició

Una gramàtica incontextual (CFG, de *context-free grammar*) és un quàdruple (V, Σ, P, S) on

- V és un conjunt finit d'elements anomenats variables (o no terminals)
- Σ és un alfabet, $\Sigma \cap V \neq \emptyset$, d'elements anomenats terminals
- $P \subseteq V \times (V \cup \Sigma)^*$ és un conjunt finit de produccions (o regles)
- $S \in V$ és la variable inicial

Notació

Donada una CFG (V, Σ, P, S), una producció $(X, \alpha) \in P$ l'escriurem sempre com $X \to \alpha$.

Definició de gramàtica

Definició

Una gramàtica incontextual (CFG, de *context-free grammar*) és un quàdruple (V, Σ, P, S) on

- V és un conjunt finit d'elements anomenats variables (o no terminals)
- Σ és un alfabet, $\Sigma \cap V \neq \emptyset$, d'elements anomenats terminals
- $P \subseteq V \times (V \cup \Sigma)^*$ és un conjunt finit de produccions (o regles)
- $S \in V$ és la variable inicial

Notació

Donada una CFG (V, Σ, P, S), una producció $(X, \alpha) \in P$ l'escriurem sempre com $X \to \alpha$.

Exemple

Una gramàtica que, a la pràctica, descrivim com

$$\begin{aligned} S &\to a\ S \mid a\ T \\ T &\to a\ T\ b \mid \lambda \end{aligned}$$

és formalment un quàdruple $(\{S, T\}, \{a, b\}, P, S)$, on

$$P = \{(S, aS), (S, aT), (T, aTb), (T, \lambda)\}.$$

La primera notació és completa:

- Les variables hi apareixen en majúscula
- Els terminals hi apareixen en minúscula
- Les produccions hi són totes
- La variable de la part esquerra de la primera producció és la inicial

Exemple

Una gramàtica que, a la pràctica, descrivim com

$$S \to a\,S \mid a\,T$$
$$T \to a\,T\,b \mid \lambda$$

és formalment un quàdruple $(\{S, T\}, \{a, b\}, P, S)$, on

$$P = \{(S, aS), (S, aT), (T, aTb), (T, \lambda)\}.$$

La primera notació és completa:

- Les variables hi apareixen en majúscula
- Els terminals hi apareixen en minúscula
- Les produccions hi són totes
- La variable de la part esquerra de la primera producció és la inicial

Derivació

Definició

Donada una gramàtica $G = (V, \Sigma, P, S)$ i dues cadenes $\alpha, \beta \in (V \cup \Sigma)^*$, escrivim $\alpha \Rightarrow_G \beta$ i diem que hi ha una derivació directa de α a β, si existeix una producció $A \to \gamma$ a P tal que

$$\alpha = \alpha_1 A \alpha_2 \ \text{ i } \ \beta = \alpha_1 \gamma \alpha_2$$

per a $\alpha_1, \alpha_2, \gamma \in (V \cup \Sigma)^*$.

Exemple

En la CFG G vista abans

$$S \to a\,S \mid a\,T$$
$$T \to a\,T\,b \mid \lambda$$

tenim les derivacions directes per a tot $k \geq 0$

$$a^k S \Rightarrow_G a^{k+1} S, \quad a^k S \Rightarrow_G a^{k+1} T, \quad T \Rightarrow_G \lambda.$$

Derivació

Definició

Donada una gramàtica $G = (V, \Sigma, P, S)$ i dues cadenes $\alpha, \beta \in (V \cup \Sigma)^*$, escrivim $\alpha \Rightarrow_G \beta$ i diem que hi ha una derivació directa de α a β, si existeix una producció $A \to \gamma$ a P tal que

$$\alpha = \alpha_1 A \alpha_2 \text{ i } \beta = \alpha_1 \gamma \alpha_2$$

per a $\alpha_1, \alpha_2, \gamma \in (V \cup \Sigma)^*$.

Exemple

En la CFG G vista abans

$$\begin{aligned} S &\to a\,S \mid a\,T \\ T &\to a\,T\,b \mid \lambda \end{aligned}$$

tenim les derivacions directes per a tot $k \geq 0$

$$a^k S \Rightarrow_G a^{k+1} S, \quad a^k S \Rightarrow_G a^{k+1} T, \quad T \Rightarrow_G \lambda.$$

Definició

Donada una gramàtica $G = (V, \Sigma, \delta, S)$ i dues cadenes $\alpha, \beta \in (V \cup \Sigma)^*$, escrivim

$$\alpha \Rightarrow_G^n \beta$$

i diem que hi ha una derivació de n passos de α a β, si

- $n = 0$ i $\alpha = \beta$ o bé
- $n > 0$ i existeix γ tal que $\alpha \Rightarrow_G \gamma \Rightarrow_G^{n-1} \beta$.

Escrivim $\alpha \Rightarrow_G^* \beta$ si $\alpha \Rightarrow_G^n \beta$ per a algun $n \geq 0$.

(S'elideix la G quan és clara pel context)

Exemple

En la gramàtica anterior,

$S \rightarrow a\,S \mid a\,T$
$T \rightarrow a\,T\,b \mid \lambda$

tenim la derivació $S \Rightarrow aT \Rightarrow aaTb \Rightarrow aab$, que abreugem $S \Rightarrow^3 aab$.

Definició

Donada una gramàtica $G = (V, \Sigma, \delta, S)$ i dues cadenes $\alpha, \beta \in (V \cup \Sigma)^*$, escrivim

$$\alpha \Rightarrow_G^n \beta$$

i diem que hi ha una derivació de n passos de α a β, si

- $n = 0$ i $\alpha = \beta$ o bé
- $n > 0$ i existeix γ tal que $\alpha \Rightarrow_G \gamma \Rightarrow_G^{n-1} \beta$.

Escrivim $\alpha \Rightarrow_G^* \beta$ si $\alpha \Rightarrow_G^n \beta$ per a algun $n \geq 0$.
(S'elideix la G quan és clara pel context)

Exemple

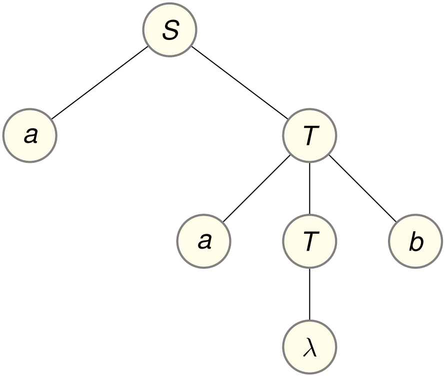
En la gramàtica anterior,

$$S \to a\,S \mid a\,T$$
$$T \to a\,T\,b \mid \lambda$$

tenim la derivació $S \Rightarrow aT \Rightarrow aaTb \Rightarrow aab$, que abreugem $S \Rightarrow^3 aab$.

A tota derivació li correspon un arbre de derivació.

La derivació $S \Rightarrow aT \Rightarrow aaTb \Rightarrow aab$ es representa amb l'arbre



Les fulles de l'arbre llegides d'esquerra a dreta formen el mot generat.

Llenguatge generat

Llenguatge d'una gramàtica

El llenguatge generat per una CFG $G = (V, \Sigma, P, S)$ és

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow_G^* w\}.$$

Exemple

Considerem la gramàtica G

$$S \to a\,S \mid a\,T$$
$$T \to a\,T\,b \mid \lambda$$

Una derivació des de S sempre és de la forma

$$S \Rightarrow^i a^i S \Rightarrow^1 a^{i+1} T \Rightarrow^j a^{i+1} a^j T b^j \Rightarrow^1 a^{i+1} a^j \lambda a^j = a^{i+j+1} b^j$$

amb $i, j \geq 0$. Per tant, tenim $L(G) = \{a^m b^n \mid m > n\}$.

Llenguatge generat

Llenguatge d'una gramàtica

El llenguatge generat per una CFG $G = (V, \Sigma, P, S)$ és

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^*_G w\}.$$

Exemple

Considerem la gramàtica G

$$\begin{aligned} S &\to a\,S \mid a\,T \\ T &\to a\,T\,b \mid \lambda \end{aligned}$$

Una derivació des de S sempre és de la forma

$$S \Rightarrow^i a^i S \Rightarrow^1 a^{i+1} T \Rightarrow^j a^{i+1} a^j T b^j \Rightarrow^1 a^{i+1} a^j \lambda a^j = a^{i+j+1} b^j$$

amb $i, j \geq 0$. Per tant, tenim $L(G) = \{a^m b^n \mid m > n\}$.

Llenguatge generat

Exemple: CFG per a $L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| \neq |y|\} \subseteq \{0,1,\#\}^*$

$$
\begin{aligned}
G: \quad & S \to B\,E \\
& S \to D\,B \\
& E \to B\,E \mid A \\
& D \to D\,B \mid A \\
& A \to B\,A\,B \mid \# \\
& B \to 0 \mid 1
\end{aligned}
$$

Llenguatge generat

Exemple: CFG per a $L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| \neq |y|\} \subseteq \{0,1,\#\}^*$

$$
\begin{aligned}
G: \quad & S \rightarrow B\ E \\
& S \rightarrow D\ B \\
& E \rightarrow B\ E \mid A \\
& D \rightarrow D\ B \mid A \\
& A \rightarrow B\ A\ B \mid \# \\
& B \rightarrow 0 \mid 1
\end{aligned}
$$

$B \Rightarrow^* 0, B \Rightarrow^* 1$

Llenguatge generat

Exemple: CFG per a $L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| \neq |y|\} \subseteq \{0,1,\#\}^*$

$$
\begin{aligned}
G: \quad & S \rightarrow B\ E \\
& S \rightarrow D\ B \\
& E \rightarrow B\ E \mid A \\
& D \rightarrow D\ B \mid A \\
& A \rightarrow B\ A\ B \mid \# \\
& B \rightarrow 0 \mid 1
\end{aligned}
$$

$$A \Rightarrow^* x\#y,\ |x| = |y|$$

$$B \Rightarrow^* 0, B \Rightarrow^* 1$$

Llenguatge generat

Exemple: CFG per a $L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge\ |x| \neq |y|\} \subseteq \{0,1,\#\}^*$

$$
\begin{array}{lll}
G: & S \to B\ E & \\
& S \to D\ B & \\
& E \to B\ E \mid A & \\
& D \to D\ B \mid A & D \Rightarrow^* x\#y,\ |x| \leq |y| \\
& A \to B\ A\ B \mid \# & A \Rightarrow^* x\#y,\ |x| = |y| \\
& B \to 0 \mid 1 & B \Rightarrow^* 0, B \Rightarrow^* 1
\end{array}
$$

Llenguatge generat

Exemple: CFG per a $L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge\ |x| \neq |y|\} \subseteq \{0,1,\#\}^*$

$$
\begin{array}{lll}
G: & S \rightarrow B\ E & \\
 & S \rightarrow D\ B & \\
 & E \rightarrow B\ E \mid A & E \Rightarrow^* x\#y,\ |x| \geq |y| \\
 & D \rightarrow D\ B \mid A & D \Rightarrow^* x\#y,\ |x| \leq |y| \\
 & A \rightarrow B\ A\ B \mid \# & A \Rightarrow^* x\#y,\ |x| = |y| \\
 & B \rightarrow 0 \mid 1 & B \Rightarrow^* 0,\ B \Rightarrow^* 1
\end{array}
$$

Llenguatge generat

Exemple: CFG per a $L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge\ |x| \neq |y|\} \subseteq \{0,1,\#\}^*$

$$
\begin{array}{lll}
G: & S \to B\ E & \\
 & S \to D\ B & DB \Rightarrow^* x\#y,\ |x| < |y| \\
 & E \to B\ E \mid A & E \Rightarrow^* x\#y,\ |x| \geq |y| \\
 & D \to D\ B \mid A & D \Rightarrow^* x\#y,\ |x| \leq |y| \\
 & A \to B\ A\ B \mid \# & A \Rightarrow^* x\#y,\ |x| = |y| \\
 & B \to 0 \mid 1 & B \Rightarrow^* 0,\ B \Rightarrow^* 1
\end{array}
$$

Llenguatge generat

Exemple: CFG per a $L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| \neq |y|\} \subseteq \{0,1,\#\}^*$

$$
\begin{array}{lll}
G: & S \to B\ E & BE \Rightarrow^* x\#y,\ |x| > |y| \\
 & S \to D\ B & DB \Rightarrow^* x\#y,\ |x| < |y| \\
 & E \to B\ E \mid A & E \Rightarrow^* x\#y,\ |x| \geq |y| \\
 & D \to D\ B \mid A & D \Rightarrow^* x\#y,\ |x| \leq |y| \\
 & A \to B\ A\ B \mid \# & A \Rightarrow^* x\#y,\ |x| = |y| \\
 & B \to 0 \mid 1 & B \Rightarrow^* 0, B \Rightarrow^* 1
\end{array}
$$

Classes de llenguatges

Llenguatges incontextuals

Un llenguatge se'n diu incontextual (*context-free language*) si existeix una CFG que el genera.

És a dir, L és incontextual si existeix una CFG tal que $L = L(G)$.

Classe dels incontextuals

Es defineix CFL com la classe dels llenguatges incontextuals.

Regulars i incontextuals

REG $\subseteq$ CFL

De forma equivalent: tot regular és incontextual.

La construcció és la següent. Suposem que L és reconegut per un DFA

$$A = (\{q_0, \ldots, q_{n-1}\}, \Sigma, \delta, q_0, F).$$

Ara definim una CFG $G = (\{R_0, \ldots, R_{n-1}\}, \Sigma, P, R_0)$ tal que

$$P = \{R_i \to aR_j \mid \delta(q_i, a) = q_j\} \cup \{R_i \to \lambda \mid q_i \in F\}.$$

Aleshores, L(G) = L(A) = L. Per tant, L és incontextual.

Exercici

Demostreu per inducció que la construcció anterior és correcta (és a dir, que $L(G) = L(A)$).

Regulars i incontextuals

REG $\subseteq$ CFL

De forma equivalent: tot regular és incontextual.

La construcció és la següent. Suposem que L és reconegut per un DFA

$$A = (\{q_0, \ldots, q_{n-1}\}, \Sigma, \delta, q_0, F).$$

Ara definim una CFG $G = (\{R_0, \ldots, R_{n-1}\}, \Sigma, P, R_0)$ tal que

$$P = \{R_i \to aR_j \mid \delta(q_i, a) = q_j\} \cup \{R_i \to \lambda \mid q_i \in F\}.$$

Aleshores, L(G) = L(A) = L. Per tant, L és incontextual.

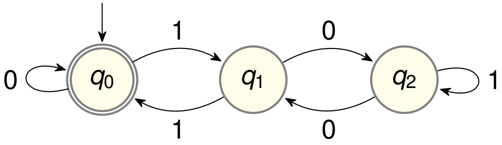
Exercici

Demostreu per inducció que la construcció anterior és correcta
(és a dir, que $L(G) = L(A)$).

Regulars i incontextuals

Exemple: múltiples de 3 en binari

Havíem vist que el llenguatge dels mots sobre $\{0, 1\}$ que interpretats com a nombres binaris són múltiples de 3 és regular:



La conversió de DFA a CFG vista abans produeix la gramàtica:

$$
\begin{aligned}
R_0 &\to 0\ R_0 \mid 1\ R_1 \mid \lambda \\
R_1 &\to 0\ R_2 \mid 1\ R_0 \\
R_2 &\to 0\ R_1 \mid 1\ R_2
\end{aligned}
$$

Gramàtiques incontextuals

Reunió

CFL és tancada per reunió

Suposem que $G_1 = (V_1, \Sigma, P_1, S_1)$ i $G_2 = (V_2, \Sigma, P_2, S_2)$ són dues CFGs tals que $L(G_1) = L_1$ i $L(G_2) = L_2$. Definim la CFG

$$G = (\{S\} \cup V_1 \cup V_2, \Sigma, \{S \to S_1 \mid S_2\} \cup P_1 \cup P_2, S)$$

on S és un símbol nou ($S \notin V_1 \cup V_2$).

Qualsevol derivació d'un mot en G és de la forma

$$S \Rightarrow S_1 \Rightarrow^* w \quad \text{o} \quad S \Rightarrow S_2 \Rightarrow^* w.$$

En el primer cas, $w \in L(G_1)$ i en el segon $w \in L(G_2)$. A l'inrevés, tot mot de $L(G_1) \cup L(G_2)$ té una derivació en G. Per tant,

$$L(G) = L(G_1) \cup L(G_2) = L_1 \cup L_2$$

i el llenguatge $L_1 \cup L_2$ és incontextual.

Reunió

Exemple: reunió

Abans hem vist que

$$L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| \neq |y|\}$$

es pot generar amb una gramàtica que combina dues CFGs que generen

$$L_1 = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| < |y|\} \text{ i}$$
$$L_2 = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| > |y|\}.$$

Exercici

Completeu l'exemple donant les CFGs per a L, L_1 i L_2.

Reunió

Exemple: reunió

Abans hem vist que

$$L = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| \neq |y|\}$$

es pot generar amb una gramàtica que combina dues CFGs que generen

$$L_1 = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| < |y|\} \text{ i}$$
$$L_2 = \{x\#y \mid x, y \in \{0,1\}^* \ \wedge \ |x| > |y|\}.$$

Exercici

Completeu l'exemple donant les CFGs per a L, L_1 i L_2.

Concatenació

Suposem que $G_1 = (V_1, \Sigma, P_1, S_1)$ i $G_2 = (V_2, \Sigma, P_2, S_2)$ són dues CFGs tals que $L(G_1) = L_1$ i $L(G_2) = L_2$.

CFL és tancada per concatenació

La concatenació $L_1 L_2$ és generada per la CFG

$$(\{S\} \cup V_1 \cup V_2, \Sigma, \{S \to S_1 S_2\} \cup P_1 \cup P_2, S)$$

on S és un símbol nou ($S \notin V_1 \cup V_2$).

Exercici

Demostreu que $\{a^i b^{i+j} c^j \mid i, j \geq 0\}$ és incontextual sabent que $\{a^n b^n \mid n \geq 0\}$ i $\{b^n c^n \mid n \geq 0\}$ ho són.

Estrella de Kleene

Suposem que $G = (V, \Sigma, P, S)$ és una CFG tal que $L(G) = L$.

CFL és tancada per l'estrella de Kleene

L'estrella de Kleene de L, L^*, és generada per la CFG

$$(\{S\} \cup V_1, \Sigma, \{S \to SS_1 | \lambda\} \cup P_1, S)$$

on S és un símbol nou ($S \notin V_1$).

Exercici

Argumenteu que les construccions anteriors justifiquen el tancament de CFL per les propietats respectives.

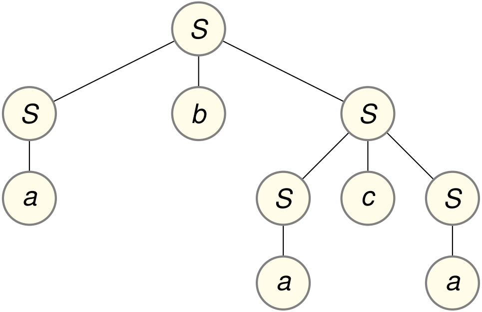
Gramàtiques incontextuals

Exemple

Sigui $G = (\{S\}, \{a, b, c\}, P, S)$ on P conté les produccions

$$S \to S\ b\ S \mid S\ c\ S \mid a$$

Llavors, el mot *abaca* té l'arbre de derivació



Una de les derivacions corresponents a l'arbre de dalt seria

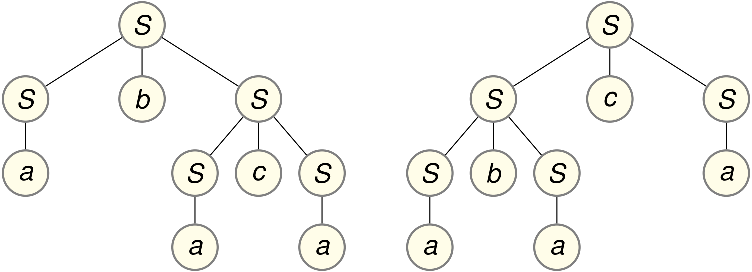
$$S \Rightarrow SbS \Rightarrow SbScS \Rightarrow abScS \Rightarrow abSca \Rightarrow abaca.$$

Ambigüitat

Una gramàtica incontextual és ambigua si existeix un mot $x \in L(G)$ per al qual es poden trobar dos arbres de derivació. En cas contrari, es diu que G és inambigua.

Exemple d'ambigüitat 1

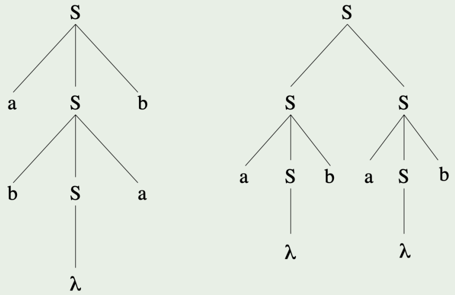
En la gramàtica de l'exemple anterior, *abaca* té dos arbres de derivació:



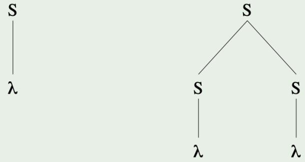
Per tant, la gramàtica és ambigua.

Exemple d'ambigüitat 2

La gramàtica $S \rightarrow \mathtt{a}S\mathtt{b} \mid \mathtt{b}S\mathtt{a} \mid SS \mid \lambda$ és ambigua perquè té un mot amb més d'un arbre de derivació. Per exemple, per al mot $\mathtt{abab}$:



De fet, també es poden trobar dos arbres de derivació per al mot λ:



El fet que una gramàtica sigui ambigua pot ser degut a la forma en què s'ha definit o bé pot degut a una. propietat del llenguatge que genera.

Definició

Un llenguatge incontextual és inherentment ambigu si tota gramàtica que el genera és ambigua.

Exemples

1. El llenguatge $\{x \in \{a, b\}^* \mid |x|_a = |x|_b\}$ no és inherentment ambigu perquè té una gramàtica inambigua que el genera:

$$
\begin{aligned}
S &\to a\,B\,S \mid b\,A\,S \mid \lambda \\
A &\to b\,A\,A \mid a \\
B &\to a\,B\,B \mid b
\end{aligned}
$$

2. El llenguatge $\{a^i b^j c^k \mid i = j \ \vee j = k\}$ només és generat per gramàtiques ambigües i, per tant, és inherentment ambigu.