Arbori de derivare

O metodă grafică de descriere a unei derivări într-o gramatică independentă de context, cu scopul de a o vizualiza rapid, o reprezintă noțiunea de **arbore de derivare**. Această noțiune derivă din noțiunea de arborescență cu rădăcină din teoria grafurilor (Berge Claude, *Teoria grafurilor și aplicațiile ei*, Editura Tehnică, București, 1969) și este prin definiție un graf finit orientat, fără circuite, în care orice vârf diferit de rădăcină este extremitatea terminală a unui singur arc și rădăcina nu este extremitatea terminală a nici unui arc.

Definiția 1. Se numește **arbore de derivare** asociat unei derivări generate de o gramatică independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ o arborescență, având vârfurile așezate pe nivele și care are următoarele proprietăți:

1. rădăcina este etichetată cu simbolul ***S***;

2. vârfurile cu cel puțin un descendent se numesc **vârfuri interne** sau **vârfuri neterminale** și sunt etichetate cu simboluri neterminale;

3. vârfurile fără niciun descendent se numesc **vârfuri terminale** sau **frunze** și sunt etichetate cu simboluri terminale;

4. dacă în mulțimea producțiilor ***P*** există producția *A* → $A_1A_2 \ldots A_k$, atunci în arborele de derivare vârfurile etichetate cu simbolurile $A_1, A_2, \ldots, A_k$ sunt în ordine de la stânga la dreapta descendenți direcți ai vârfului etichetat cu simbolul *A*.

Definiția 2. Cuvântul format din etichetele vârfurilor terminale, în ordine de la stânga la dreapta, se numește **frontiera arborelui de derivare**.

Exemplul 1. Fie gramatică independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde $\boldsymbol{N} = \{S, A\}$, $\boldsymbol{T} = \{a, b\}$, iar ***P*** este alcătuită din următoarele producții:

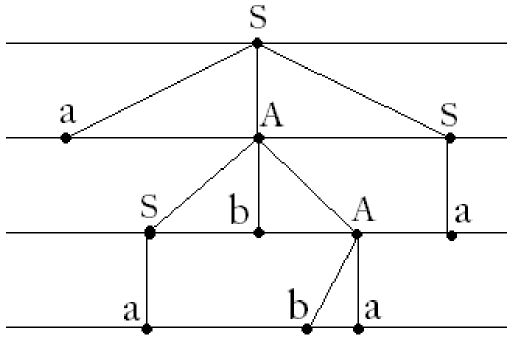
S → *aAS* (1)
S → *a* (2)
A → *SbA* (3)
A → *SS* (4)
A → *ba* (5)

Se consideră în gramatica ***G*** următoarea derivare:

$$S \underset{(1)}{\rightarrow} a\boldsymbol{A}S \underset{(3)}{\rightarrow} aSbA\boldsymbol{S} \underset{(2)}{\rightarrow} a\boldsymbol{S}bAa \underset{(2)}{\rightarrow} aab\boldsymbol{A}a \underset{(5)}{\rightarrow} aabbaa = a^2b^2a^2.$$

Acestei derivări i se asociază următorul arbore de derivare:

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu



Se observă că frontiera arborelui de derivare coincide cu $a^2b^2a^2$, adică cuvântul generat de derivare.

Între două vârfuri terminale dintr-un arbore de derivare, etichetate cu X_1 şi X_2 se pot considera drumurile care le leagă de rădăcina arborelui şi fie vârful etichetat cu *X* din care cele două drumuri se despart. Se spune că $\boldsymbol{X_1}$ **este la stânga lui** $\boldsymbol{X_2}$, dacă drumul de la *X* la X_1 se află la stânga celui de la *X* la X_2.

Legătura dintre limbajul generat de o gramatică independentă de context și arborii de derivare atașați derivărilor generate de ea este stabilită de următoarea teoremă, a cărei demonstrație utilizează următoarea notație:

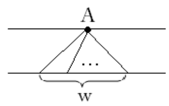
$\boldsymbol{G_A} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{A}, \boldsymbol{P})$ gramatica independentă de context obținută din gramatica $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ prin înlocuirea simbolului iniţial *S* cu simbolul *A.*

Teorema 1. Fie gramatică independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ și $w \in \boldsymbol{T}^*$. Atunci $w \in \boldsymbol{L}(\boldsymbol{G})$ dacă și numai dacă există un arbore de derivare cu frontiera *w*.

Demonstrație. Este suficient să se demonstreze că pentru orice simbol $A \in \boldsymbol{N}$ are loc $A \overset{*}{\Rightarrow} w$, dacă și numai dacă există un arbore de derivare cu frontiera $w \in \boldsymbol{T}^*$.

Implicația ($\Longrightarrow$), adică presupunem că pentru $A \in \boldsymbol{N}$ are loc derivarea $A \underset{k\ pași}{\overset{*}{\Longrightarrow}} w \in \boldsymbol{T}^*$, în *k* pași și se demonstrează prin inducție matematică după numărul de pași $k \in \mathbb{N}^*$ că există un arbore de derivare în gramatica $\boldsymbol{G_A}$ cu frontiera $w \in \boldsymbol{T}^*$.

Într-adevăr, pentru *k* = 1 se obține producția $A \to w \in \boldsymbol{P}$. În acest caz, arborele de derivare atașat producției în gramatica $\boldsymbol{G_A}$ are două nivele și este de forma:



Se observă că pe nivelul doi se află frontiera *w* și deci afirmația este adevărată pentru *k* = 1.

Ipoteza de inducție: presupunem că implicația este adevărată pentru orice simbol $A \in \boldsymbol{N}$ și orice derivare având *k* pași.

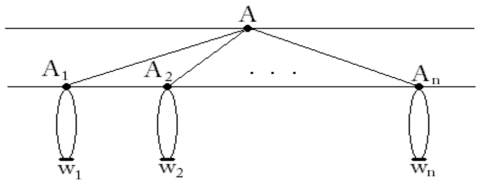
În baza ipotezei de inducție se demonstrează că implicația este adevărată pentru orice simbol $A \in \boldsymbol{N}$ și orice derivare având *k*+1 pași.

Fie derivarea în *k*+1 pași:

$$A \to A_1A_2 \cdots A_n \underset{k\ pași}{\overset{*}{\Longrightarrow}} w = w_1w_2 \cdots w_n, \text{ astfel încât } A_i \underset{\leq k\ pași}{\overset{*}{\Longrightarrow}} w_i, \forall i = \overline{1,n}.$$

Conform ipotezei de inducție, există în gramatica $\boldsymbol{G_{A_i}}$ arbore de derivare cu frontiera w_i, $\forall i = \overline{1,n}$. Atunci arborele de derivare asociat derivării în *k*+1 pași are forma:

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu

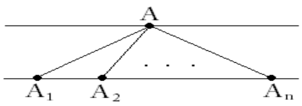


Rezultă că implicația ($\Longrightarrow$) este adevărată pentru orice $k \in \mathbb{N}^*$.

Implicația ($\Longleftarrow$), adică presupunem că există un arbore de derivare cu frontiera $w \in \boldsymbol{T}^*$, având k vârfuri neterminale.

Se demonstrează implicația prin inducție matematică relativă la numărul vârfurilor neterminale $k \in \mathbb{N}^*$.

Pentru k = 1 arborele de derivare are un singur vârf neterminal etichetat cu A de forma:

Atunci există în gramatica 

$\boldsymbol{G_A}$ producția $A \rightarrow A_1A_2 \cdots A_n \in \boldsymbol{P}$ și $w = A_1A_2 \cdots A_n \in \boldsymbol{T}^*$.

Ipoteza de inducție: presupunem implicația adevărată pentru orice $A \in \boldsymbol{N}$ și orice arbore de derivare cu rădăcina A care are cel mult k vârfuri neterminale cu etichete din $\boldsymbol{N}$ și frontiera $w \in \boldsymbol{T}^*$.

Fie un arbore de derivare cu rădăcina $A \in \boldsymbol{N}$ care are k+1 vârfuri neterminale cu etichete din $\boldsymbol{N}$ și fie $A_1, A_2, \cdots, A_n$ vârfurile descendente direct ale rădăcinii A. Atunci există în $\boldsymbol{P}$ producția $A \rightarrow A_1A_2 \cdots A_n$.

Pentru $i \in \{1,2,\dots,n\}$ există un arbore de derivare cu rădăcina $\boldsymbol{A_i}$ și cel mult k vârfuri neterminale, având frontiera w_i, unde $A_i = w_i$ dacă $A_i \in \boldsymbol{T}$.

În baza ipotezei de inducție, rezultă $A_i \overset{*}{\Rightarrow} w_i, \forall i = \overline{1,n}$, de unde se obţine:

$$A \overset{*}{\Rightarrow} A_1A_2 \cdots A_n \overset{*}{\Rightarrow} w_1w_2 \dots w_n = w \in \boldsymbol{L(G)},$$

care este derivarea căreia îi este asociat arborele de derivare cu rădăcina $A \in \boldsymbol{N}$ care are k+1 vârfuri neterminale cu etichete din $\boldsymbol{N}$, iar w este frontiera arborelui de derivare cu rădăcina A.

Rezultă implicația adevărată pentru orice $k \in \mathbb{N}^*$.

Definiția 3. Un arbore de derivare se numește **binar**, dacă fiecare vârf neterminal are cel mult doi descendenți direcți.

Consecința 1. Orice arbore de derivare asociat unei derivări generate de o gramatică independentă de context în forma normală Chomsky este binar.

Justificarea rezultă din forma particulară a producțiilor în forma normală Chomsky.

Exemplul 2. Fie gramatica independentă de context în forma normală Chomsky

$\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ unde $\boldsymbol{N} = \{S, A, B, C, D, E\}, \boldsymbol{T} = \{a, b\}$, iar mulțimea producțiilor $\boldsymbol{P}$ este:

$S \rightarrow BE$ (1)

$E \rightarrow AS$ (2)

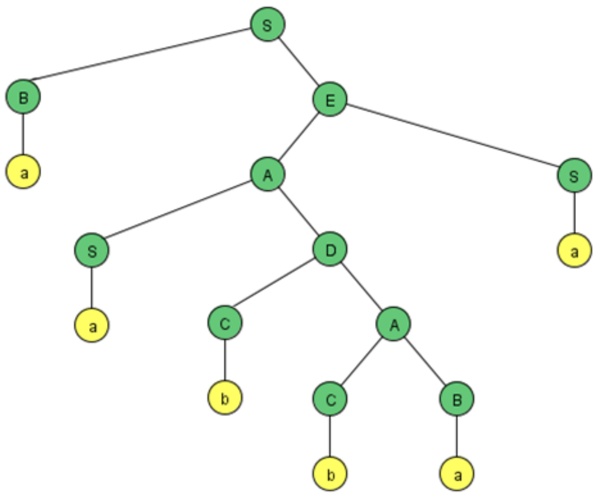
Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu

$S \rightarrow a$ (3)
$A \rightarrow SD$ (4)
$D \rightarrow CA$ (5)
$A \rightarrow SS$ (6)
$A \rightarrow CB$ (7)
$C \rightarrow b$ (8)
$B \rightarrow a$ (9)

Fie cuvântul *w* = *aabbaa,* a cărui derivare în gramatica ***G*** este:

$$S \underset{(1)}{\rightarrow} BE \overset{*}{\underset{(9,2)}{\Longrightarrow}} a\boldsymbol{AS} \overset{*}{\underset{(4,3)}{\Longrightarrow}} a\boldsymbol{SD}a \overset{*}{\underset{(3,5)}{\Longrightarrow}} aa\boldsymbol{CA}a \overset{*}{\underset{(8,7)}{\Longrightarrow}} aab\boldsymbol{CB}a \overset{*}{\underset{(8,9)}{\Longrightarrow}} aabbaa\,.$$

Arborele de derivare asociat derivării are următoarea formă binară:



Definiția 4. O gramatică independentă de context ***G*** se numește **ambiguă***,* dacă un cuvânt $w \in \boldsymbol{L}(\boldsymbol{G})$ se poate genera prin cel puţin două derivări stângi distincte.

Exemplul 3. Fie gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde $\boldsymbol{N} = \{S, A\}, \boldsymbol{T} = \{a, b\}$, iar mulțimea producțiilor ***P*** este:

$S \rightarrow ab$ (1)
$S \rightarrow aSb$ (2)
$S \rightarrow aA$ (3)
$A \rightarrow aSbb$ (4).

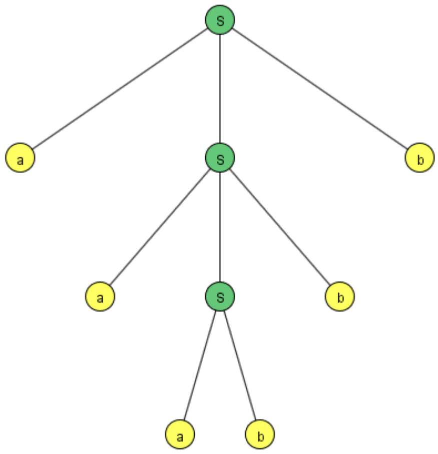
Această gramatică este ambiguă. Într-adevăr, fie cuvântul *w* = *aaabbb* $\in \boldsymbol{L}(\boldsymbol{G})$ care se poate genera prin două prin următoarele două derivări stângi distincte:

$$(i)\ S \underset{(2)}{\rightarrow} a\boldsymbol{S}b \underset{(2)}{\rightarrow} aa\boldsymbol{S}bb \underset{(1)}{\rightarrow} aaabbb;$$

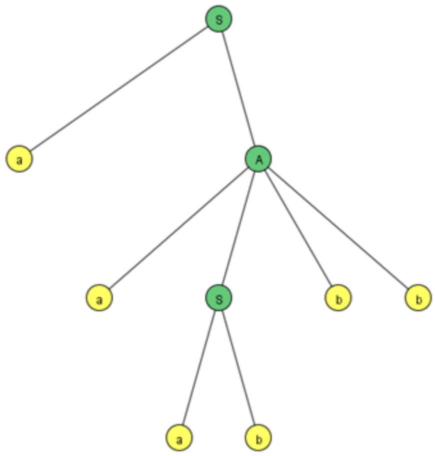
$$(ii)\ S \underset{(3)}{\longrightarrow} a\boldsymbol{A} \underset{(4)}{\longrightarrow} aa\boldsymbol{S}bb \underset{(1)}{\longrightarrow} aaabbb.$$

Arborii de derivare asociați celor două derivări ajută la vizualizarea rapidă a ambiguității și sunt prezentați mai jos.

Arborele de derivare asociat derivării (*i*):



Arborele de derivare asociat derivării (*ii*):



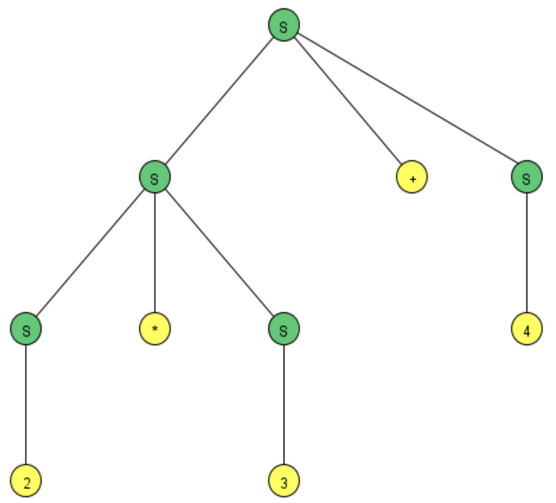
Exemplu: Considerăm următoarea gramatică independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde $\boldsymbol{N} = \{S\}$, $\boldsymbol{T} = \{+, *, 0,1,2,3,4,5,6,7,8,9\}$, iar mulțimea producțiilor $\boldsymbol{P}$ este următoarea:

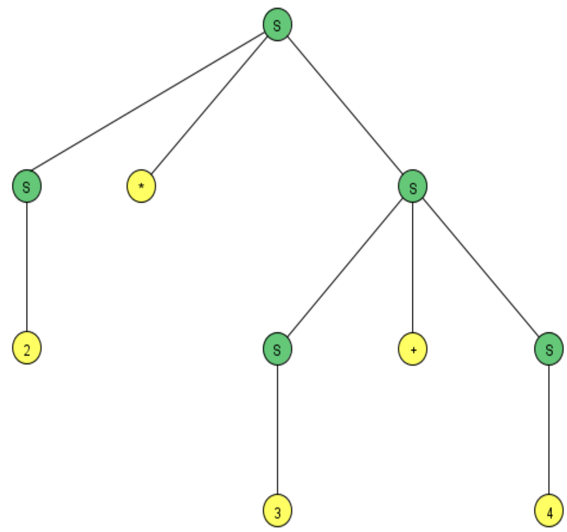
$$S \to S + S \qquad (1)$$
$$S \to S * S \qquad (2)$$
$$S \to 0|\ 1|\ 2|\ 3|\ 4|\ 5|\ 6|\ 7|\ 8\ |9 \qquad (3)$$

Gramatica G este ambiguă, deoarece cuvântul $w = 2 * 3 + 4\ \in \boldsymbol{T}^*$ poate fi obținut prin două derivări stângi distincte din simbolul inițial S, după cum se poate cu uşurinţă observa din figurile de mai jos:

$$S \underset{(1)}{\rightarrow} \boldsymbol{S}+S \underset{(2)}{\rightarrow} \boldsymbol{S}*S+S \underset{(3)}{\rightarrow} 2*\boldsymbol{S}+S \underset{(3)}{\rightarrow} \underset{(3)}{\rightarrow} 2*3+\boldsymbol{S} \underset{(3)}{\rightarrow} 2*3+4$$

$$S \underset{(2)}{\rightarrow} \boldsymbol{S}*S \underset{(3)}{\rightarrow} 2*\boldsymbol{S} \underset{(1)}{\rightarrow} 2*\boldsymbol{S}+S \underset{(3)}{\rightarrow} \underset{(3)}{\rightarrow} 2*3+\boldsymbol{S} \underset{(3)}{\rightarrow} 2*3+4$$





Eliminarea acestei ambiguități din procesul de evaluare a unei expresii aritmetice se poate realiza fie prin utilizarea parantezelor, fie prin asocierea unor priorități operatorilor aritmetici.

Temă.

1. Fie gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde $\boldsymbol{N} = \{S, X, Y\}$, $\boldsymbol{T} = \{a, b\}$ și mulțimea producțiilor $\boldsymbol{P}$ este următoarea:

$$\begin{array}{rl} S \rightarrow X & (1) \\ X \rightarrow XY \mid aYX \mid a \mid \lambda & (2) \\ Y \rightarrow X \mid XYb \mid b & (3) \end{array}$$

a) Desenați un arbore de derivare în $\boldsymbol{G}$ pentru un cuvânt de lungime cel puțin 5, indicând și cuvântul corespunzător arborelui respectiv.
b) Simplificați gramatica $\boldsymbol{G}$.
c) Aduceți gramatica $\boldsymbol{G}$ la forma normală Chomsky.

2. Fie gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde $\boldsymbol{N} = \{S, X, Y\}$, $\boldsymbol{T} = \{a, b\}$ și mulțimea producțiilor $\boldsymbol{P}$ este următoarea:

$$\begin{array}{rl} S \rightarrow X \mid Y \mid XY & (1) \\ X \rightarrow aXYX \mid bXY \mid Y \mid a & (2) \\ Y \rightarrow XYXY \mid aXYb \mid b \mid \lambda & (3) \end{array}$$

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu

a) Desenați un arbore de derivare în $\boldsymbol{G}$ pentru un cuvânt de lungime cel puțin 7, indicând și cuvântul corespunzător arborelui respectiv.

b) Simplificați gramatica $\boldsymbol{G}$.

c) Aduceți gramatica $\boldsymbol{G}$ la forma normală Chomsky.

3. Fie gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde $\boldsymbol{N} = \{S, A, B, C\}$, $\boldsymbol{T} = \{a, b, c\}$ și mulțimea producțiilor $\boldsymbol{P}$ este următoarea:

$$S \to A \mid B \mid aA \mid bB \qquad (1)$$
$$A \to aAa \mid bAb \mid B \mid \lambda \qquad (2)$$
$$B \to bBb \mid aBa \mid A \mid \lambda \qquad (3)$$
$$C \to cCc \mid aABb \mid \lambda \qquad (4)$$

a) Desenați un arbore de derivare în $\boldsymbol{G}$ pentru un cuvânt de lungime cel puțin 7, indicând și cuvântul corespunzător arborelui respectiv.

b) Simplificați gramatica $\boldsymbol{G}$.

c) Aduceți gramatica $\boldsymbol{G}$ la forma normală Chomsky.