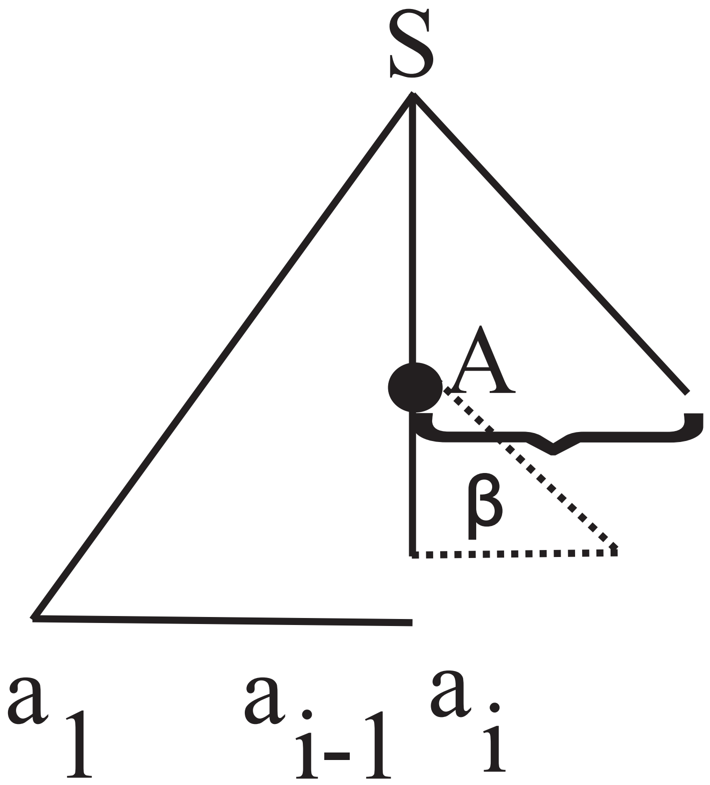
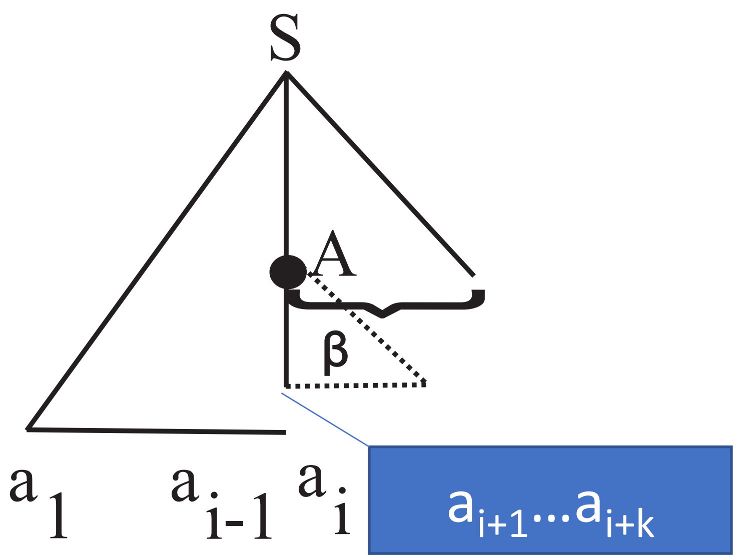
LL(1) Parser



Linear algorithm

LL(k)

- L = left (sequence is read from left to right)
- L = left (use leftmost derivation)
- Prediction of length **k**



LL(k) Principle

- In any moment of parsing, <u>acțion is uniquely determinde </u>by:
- Closed part ($a_1...a_i$)
- Current symbol A
- Prediction $a_{i+1}...a_{i+k}$ (length k)

FIRST_k

- ≈ first k terminal symbols that can be generated from α
- **Definition**:

$$FIRST_k : (N \cup \Sigma)^* \rightarrow \mathcal{P}(\Sigma^k)$$
$$FIRST_k(\alpha) = \{u | u \in \Sigma^k, \alpha \overset{*}{\Rightarrow} ux, |u| = k \text{ sau } \alpha \overset{*}{\Rightarrow} u, |u| \leq k\}$$

Definition

- *A cfg is LL(k) if for any 2 leftmost derivation we have:*

$$1.\ S \overset{*}{\Rightarrow}_{st} wA\alpha \Rightarrow_{st} w\beta\alpha \overset{*}{\Rightarrow}_{st} wx;$$

$$2.\ S \overset{*}{\Rightarrow}_{st} wA\alpha \Rightarrow_{st} w\gamma\alpha \overset{*}{\Rightarrow}_{st} wy;$$

such that $FIRST_k(x) = FIRST_k(y)$ *then* $\beta = \gamma.$

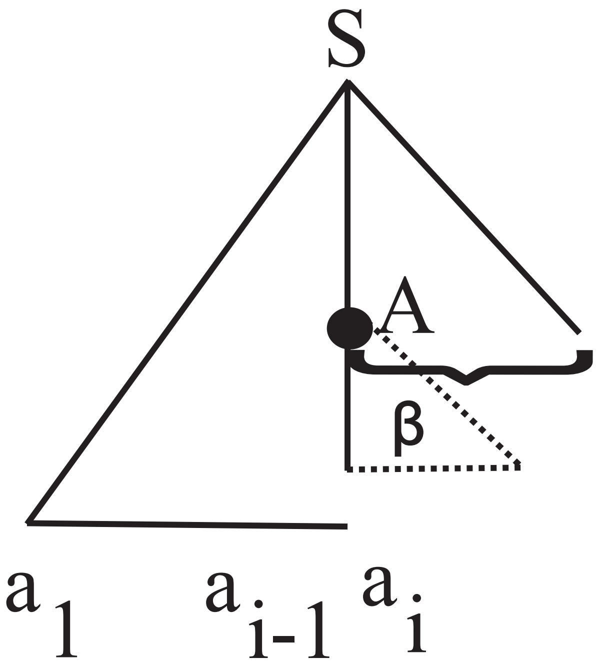
Theorem

The necessary and sufficient condition for a grammar to be LL (k) is that for any pair of distinct productions of a nonterminal ($A \rightarrow \beta$, $A \rightarrow \gamma, \beta \neq \gamma$) the condition holds:

$$FIRST_k(\beta\alpha) \cap FIRST_k(\gamma\alpha) = \Phi, \forall\alpha \quad \text{such that} \quad S \overset{*}{=>} uA\alpha$$

FOLLOW

$A \to \boldsymbol{\varepsilon}$



- $FOLLOW_k(A) \approx$ next k symbols generated after/ following A A

$$FOLLOW : (N \cup \Sigma)^* \to \mathcal{P}(\Sigma)$$
$$FOLLOW(\beta) = \{w \in \Sigma | S \overset{*}{\Rightarrow} \alpha\beta\gamma, w \in FIRST(\gamma)\}$$

LL(1) Parser

- Prediction of length 1

- Steps:
  1) construct FIRST, FOLLOW
  2) Construct LL(1) parse table
  3) Analyse sequence based on moves between configurations

Executed 1 time

Theorem: A grammar is LL(1) if and only if for any nonterminal A with productions $A \rightarrow \alpha_1 | \alpha_2 | \ldots | \alpha_n$, $FIRST(\alpha_i) \cap FIRST(\alpha_j) = \emptyset$ and if $\alpha_i \Rightarrow \varepsilon$, $FIRST(\alpha_i) \cap FOLLOW(A) = \emptyset$, $\forall i,j = 1,n, i \neq j$

Construct FIRST

Concatenation of length 1

- $FIRST_1$ denoted FIRST
- Remarks:
  - If L_1, L_2 are 2 languages over alphabet Σ, then : $L_1 \oplus L_2 = \{w | x \in L_1, y \in L_2, xy = w, |w| \leq 1 \text{ sau } xy = wz, |w| = 1\}$ and
  - $FIRST(\alpha\beta) = FIRST(\alpha) \oplus FIRST(\beta)$

    $FIRST(X_1 \ldots X_n) = FIRST(X_1) \oplus \ldots \oplus FIRST(X_n)$

Algoritmul 3.3 FIRST

INPUT: G
OUTPUT: $FIRST(X), \forall X \in N \cup \Sigma$
for $\forall a \in \Sigma$ **do**
  $F_i(a) = \{a\}, \forall i \geq 0$
end for
i := 0;
$F_0(A) = \{x | x \in \Sigma, A \to x\alpha \text{ sau } A \to x \in P\}$; {*inițializare*}
repeat
  i := i+1;
  for $\forall X \in N$ **do**
    if F_{i-1} au fost calculate $\forall X \in N \cup \Sigma$ **then**
      {*dacă* $\exists Y_j, F_{i-1}(Y_j) = \emptyset$ *atunci nu se poate aplica*}
      $F_i(A) = F_{i-1}(A) \cup$
      $\{x | A \to Y_1 \ldots Y_n \in P, x \in F_{i-1}(Y_1) \oplus \ldots \oplus F_{i-1}(Y_n)\}$
    end if
  end for
until $F_{i-1}(A) = F_i(A)$
$FIRST(X) := F_i(X), \forall X \in N \cup \Sigma$

Algoritmul 3.4 FOLLOW

INPUT: G, FIRST(X), $\forall X \in N \cup \Sigma$
OUTPUT: FOLLOW(X), $\forall X \in N \cup \Sigma$
$F(X) = \emptyset, \forall X \in N - \{S\}$; {*inițializare*}
$F(S) = \{\epsilon\}$; {*corespunzător simbolului \$ folosit în analiză*}
repeat
  for $B \in N$ **do**
    for $A \to \alpha B y \in P$ **do**
      if $\epsilon \in FIRST(y)$ **then**
        $F'(B) = F(B) \cup F(A)$;
      else
        $F'(B) = F(B) \cup FIRST(y)$
      end if
    end for
  end for
until $F'(X) = F(X), \forall X \in N$
$FOLLOW(X) = F(X), \forall X \in N$.

Correction:
In order for the algorithm to function correctly, consider:
For ∀a ∈ FIRST(y) do

Step 2: Construct LL(1) parse table

- Possible action depend on:
  - Current symbol $\in \mathbf{N} \cup \mathbf{\Sigma}$
  - Possible prediction $\in \mathbf{\Sigma}$
- Add a special character "\$" ($\notin \mathbf{N} \cup \mathbf{\Sigma}$) – marking for "empty stack"

= > table:

- One line for each symbol $\in \mathbf{N} \cup \mathbf{\Sigma} \cup \{\$\}$
- One column for each symbol $\in \mathbf{\Sigma} \cup \{\$\}$

Rules LL(1) table

1. $M(A, a) = (\alpha, i), \forall a \in FIRST(\alpha), a \neq \epsilon, A \rightarrow \alpha$ production in P with number i

   $M(A, b) = (\alpha, i),$ if $\epsilon \in FIRST(\alpha), \forall b \in FOLLOW(A), A \rightarrow \alpha$ production in P with number i

2. $M(a, a) = pop, \forall a \in \Sigma;$

3. $M(\$, \$) = acc;$

4. M(x,a)=err (error) otherwise

Remark

A grammar is LL(1) if the LL(1) parse table does NOT contain conflicts – there exists at most one value in each cell of the table M(A,a)

Step 3: Definire configurations and moves

- INPUT:
  - Language grammar $G = (N, \boldsymbol{\Sigma}, P,S)$
  - LL(1) parse table
  - Sequence to be parsed $w = a_1 \ldots a_n$
- OUTPUT:

  If ($w \in L(G)$) *then* **string of productions**

  else **error** & **location of error**

LL(1) configurations

$$(\alpha, \beta, \pi)$$

where:

- α = input stack
- β = working stack
- π = output (result)

Initial configuration:
$(w\$, S\$, \varepsilon)$

Final configuration:
$(\$, \$, \pi)$

Moves

1. Push – put in stack

$$(ux, A\alpha\$, \pi) \vdash (ux, \beta\alpha\$, \pi i), \quad \text{if} \quad M(A,u) = (\beta, i);$$

(pop A and push symbols of β)

2. Pop – take off from stack (from both stacks)

$$(ux, a\alpha\$, \pi) \vdash (x, \alpha\$, \pi), \quad \text{if} \quad \mathrm{M(a,u)=pop}$$

3. Accept

$$(\$, \$, \pi) \vdash acc$$

4. Error - otherwise

Algorithm LL(1) parse

- here