

Gramatyki - przypomnienie

< T, N, S, P >

T - zbiór symboli terminalnych

N - zbiór symboli nieterminalnych

S - symbol startowy

P - lista produkcji

V - słownik

V = T plus N

V* - zbiór słów generowanych
ze słownika V

Gramatyka bezkontekstowa to taka,
której wszystkie produkcje są postaci:

$$A \rightarrow \gamma \quad \text{przy czym } A \in N, \gamma \in V^*$$

Notacja BNF: $A ::= aB \mid \{c\}D \mid [e]FG$

Notacja dla potrzeb wykładu: A - nieterminal, b - terminal, γ - słowo

Drzewo wyvodu

Definicja wyvodu

Drzewo wyvodu (parse tree) = graficzna reprezentacja wyvodu

Przykład:

Gramatyka postaci:

(1) $E ::= E + E$

(2) $E ::= E * E$

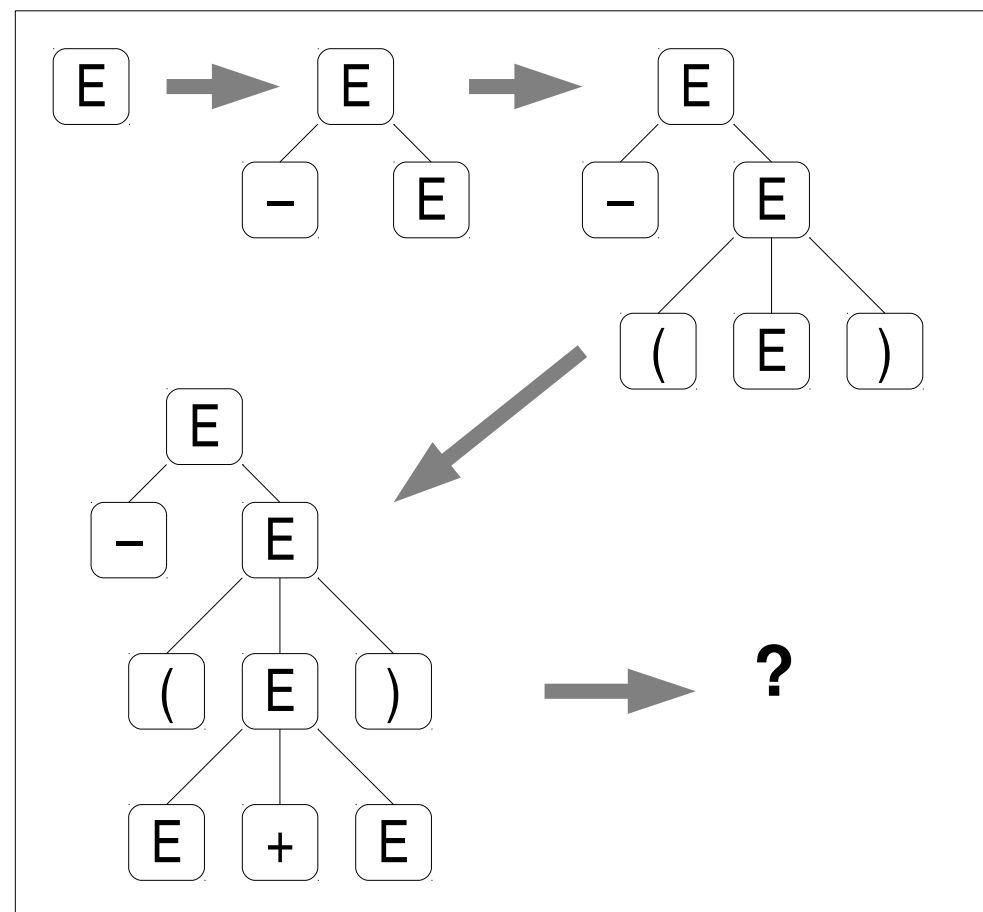
(3) $E ::= (E)$

(4) $E ::= - E$

(5) $E ::= id$

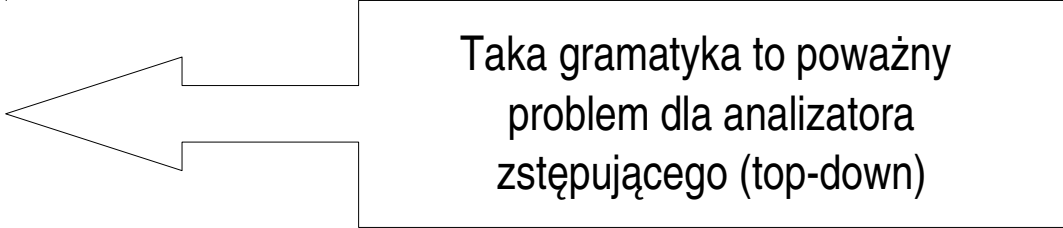
słowo: $- (id + id)$

Wywód: (4), (3), (1), (5), (5)



Lewostronna rekurencja (rekursja)

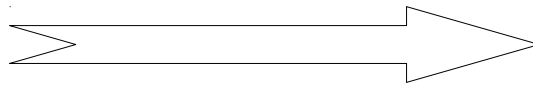
$E ::= E + T \mid T$
 $T ::= T * F \mid F$
 $F ::= \text{id}$



Taka gramatyka to poważny
problem dla analizatora
zstępującego (top-down)

Usuwanie rekurencji bezpośredniej:

$A \rightarrow A\alpha \mid \beta$



$A \rightarrow \beta A'$

$A' \rightarrow \alpha A' \mid \epsilon$

Co to jest cykl i epsilon-produkcja ?

Lewostronna rekurencja cd.

Usuwanie rekurencji pośredniej:

Założenia:

n - ilość symboli nieterminalnych

symbole nieterminalne są uporządkowane od A_1 do A_n

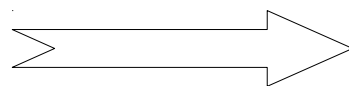
```
for ( i = 1; i <= n ; i++)
```

```
  for ( j =1; j <= i -1; j++)
```

a) Wszystkie produkcje postaci: $A_i \rightarrow A_j \gamma$

zamień na: $A_i \rightarrow \delta_1 \gamma \mid \dots \mid \delta_k \gamma$ gdzie $A_j \rightarrow \delta_1 \mid \dots \mid \delta_k$
są wszystkimi **aktualnymi** produkcjami dla A_j

b) Usuń rekurencję bezpośrednią z produkcji dla A_i

$$S ::= Aa \mid b$$
$$A ::= Ac \mid Sd \mid \epsilon$$

$$S ::= Aa \mid b$$
$$A ::= bdA' \mid A' \quad A' ::= cA' \mid adA' \mid \epsilon$$

Lewostronna faktoryzacja

Dla każdego nieterminala A i pewnej ilości produkcji n , znajdź najdłuższy prefix α taki, że produkcje A możemy zapisać w postaci:

$$A \rightarrow \alpha \beta_1 \mid \dots \mid \alpha \beta_n \mid \gamma$$

γ oznacza wszystkie produkcje, które nie zaczynają się od α

Produkcje takie zamień na:

$$A \rightarrow \alpha A' \mid \gamma \quad A' \rightarrow \beta_1 \mid \dots \mid \beta_n$$

Przykład: **$A ::= Bc \mid Bd \mid Be \mid f$**

Gramatyki niejednoznaczne

$A ::= a B b A \mid a B b A c A \mid d$

$B ::= e$

Analizator leksykalny

Rozpoznawanie ciągów znaków

Koncepcja atomu leksykalnego

Deterministyczny automat skończony

Wyrażenia regularne

Automat niedeterministyczny

Diagramy przejść

Przykłady:

liczba

napis

wyrażenie regularne

Implementacja: organizacja bufora, preprocesor, informacje o leksemach