

Stefan Sokołowski

JĘZYKI FORMALNE I METODY KOMPILACJI

wykład 3

Inst. Informatyki Stosowanej, ANS
Elbląg, 2024/2025

Wykład 3, str. 1

Umowa notacyjna dla gramatyk bezkontekstowych

- ponieważ **nieterminale** oznaczają pojęcia — sugestywne nazwy
np. „wyrażenie”, „liczba”
- **nieterminale ujmujemy** w nawiasy kątowe: $\langle \dots \rangle$
np. $\langle \text{wyrażenie} \rangle$, $\langle \text{liczba} \rangle$
- wszystkie produkcje z pojedynczego nieterminalu **grupujemy**; np.:

piszemy	zamiast
$\langle \text{cyfra} \rangle ::= 0 \mid 1 \mid \dots \mid 9$	$\langle \text{cyfra} \rangle \rightarrow 0$ $\langle \text{cyfra} \rangle \rightarrow 1$ $\dots$ $\langle \text{cyfra} \rangle \rightarrow 9$
$\langle \text{liczba} \rangle ::= \langle \text{cyfra} \rangle \mid \langle \text{liczba} \rangle \langle \text{cyfra} \rangle$	$\langle \text{liczba} \rangle \rightarrow \langle \text{cyfra} \rangle$ $\langle \text{liczba} \rangle \rightarrow \langle \text{liczba} \rangle \langle \text{cyfra} \rangle$

Backus-Naur Form



John Backus
USA
1924–2007



Peter Naur
Dania
1928–2016

Backus-Naur Form

Sposób opisu składni języków programowania:

- **nieterminale** — sugestywne nazwy ujęte w nawiasy kątowne
- **produkcje** z jednego nieterminalu napisane razem

Prostacki przykład:

$\langle \text{program} \rangle$	$::=$	$\langle \text{deklaracje} \rangle \langle \text{komendy} \rangle$	
$\langle \text{deklaracje} \rangle$	$::=$	$\lambda \mid \langle \text{deklaracje} \rangle \langle \text{deklaracja} \rangle$	— ciąg $\langle \text{deklaracja} \rangle$
$\langle \text{deklaracja} \rangle$	$::=$	$\langle \text{typ} \rangle \langle \text{identyfikator} \rangle ;$	
$\langle \text{typ} \rangle$	$::=$	$\text{int} \mid \text{double} \mid \text{char}$	
$\langle \text{komendy} \rangle$	$::=$	$\lambda \mid \langle \text{komendy} \rangle \langle \text{komenda} \rangle$	— ciąg $\langle \text{komenda} \rangle$
$\langle \text{komenda} \rangle$	$::=$	$\langle \text{identyfikator} \rangle = \langle \text{wyrażenie} \rangle ;$	— przypisanie
		$\mid \{ \langle \text{program} \rangle \}$	— blok
$\langle \text{wyrażenie} \rangle$	$::=$	$\dots$	
$\langle \text{identyfikator} \rangle$	$::=$	$\dots$	

Taki opis nadaje się do automatycznego przetwarzania — do (prawie) automatycznej konstrukcji translatora.

Drzewa wyvodu

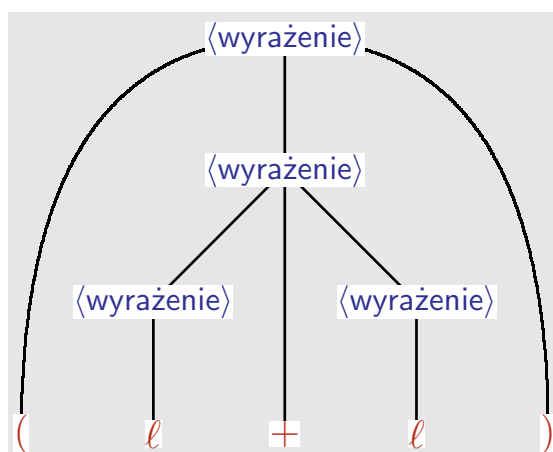
$\langle \text{wyrażenie} \rangle ::= \ell \mid (\langle \text{wyrażenie} \rangle) \mid \langle \text{wyrażenie} \rangle + \langle \text{wyrażenie} \rangle$

$\langle \text{wyrażenie} \rangle$

$\Rightarrow (\langle \text{wyrażenie} \rangle)$
 $\Rightarrow (\langle \text{wyrażenie} \rangle + \langle \text{wyrażenie} \rangle)$
 $\Rightarrow (\ell + \langle \text{wyrażenie} \rangle)$
 $\Rightarrow (\ell + \ell)$

$\langle \text{wyrażenie} \rangle$

$\Rightarrow (\langle \text{wyrażenie} \rangle)$
 $\Rightarrow (\langle \text{wyrażenie} \rangle + \langle \text{wyrażenie} \rangle)$
 $\Rightarrow (\langle \text{wyrażenie} \rangle + \ell)$
 $\Rightarrow (\ell + \ell)$



Drzewa wyvodu

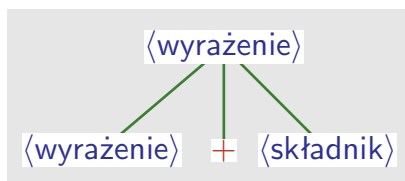
DEFINICJA:

Drzewo wyvodu słowa $w \in \Sigma^*$ w gramatyce bezkontekstowej $G = \langle \Sigma, N, P, \langle S \rangle \rangle$ to takie drzewo, które ma:

- w liściach symbole terminalne składające się na słowo w ,
- w pozostałych (wewnętrznych) węzłach nieterminale rozmieszczone wg takich zasad:
 - jeśli w jakimś węźle znajduje się nieterminał $\langle A \rangle$ a w jego dzieciach od lewej do prawej symbole $a_1, a_2, \dots, a_n \in \Sigma \cup N$, to $\langle A \rangle \rightarrow a_1 a_2 \dots a_n$ musi być produkcją z P ,
 - w korzeniu znajduje się nieterminał początkowy $\langle S \rangle$.

PRZYKŁAD:

Fragment drzewa



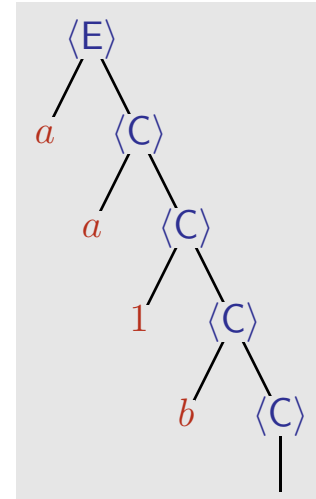
jest legalny jeśli

$\langle \text{wyrażenie} \rangle \rightarrow \langle \text{wyrażenie} \rangle + \langle \text{składnik} \rangle$ jest produkcją z gramatyki.

Drzewa wyvodu

Drzewa wyvodu słów w gramatyce prawoliniowej są „przeważone” w prawo tak, że stanowią prawie listy:

$$\begin{aligned}\langle E \rangle &::= a \langle C \rangle \mid b \langle C \rangle \\ \langle C \rangle &::= \lambda \mid a \langle C \rangle \mid b \langle C \rangle \mid 0 \langle C \rangle \mid 1 \langle C \rangle\end{aligned}$$



$$\langle E \rangle \Rightarrow a \langle C \rangle \Rightarrow aa \langle C \rangle \Rightarrow aa1 \langle C \rangle \Rightarrow aa1b \langle C \rangle \Rightarrow aa1b$$

Drzewa wyvodu

Drzewa wyvodu słów w gramatyce bezkontekstowej mogą być „rozczapierzone” we wszystkie strony.

$$\langle \text{zmienna} \rangle ::= a \mid b \mid \dots \mid z$$

$$\langle \text{cyfra} \rangle ::= 0 \mid 1 \mid \dots \mid 9$$

$$\langle \text{liczba} \rangle ::= \langle \text{liczba} \rangle \langle \text{cyfra} \rangle \mid \langle \text{cyfra} \rangle$$

$$\langle \text{atom} \rangle ::= \langle \text{zmienna} \rangle \mid \langle \text{liczba} \rangle \mid (\langle \text{wyrażenie} \rangle)$$

$$\langle \text{wyrażenie} \rangle ::= \langle \text{składnik} \rangle \mid \langle \text{wyrażenie} \rangle + \langle \text{składnik} \rangle$$

$$\langle \text{składnik} \rangle ::= \langle \text{atom} \rangle \mid \langle \text{składnik} \rangle * \langle \text{atom} \rangle$$

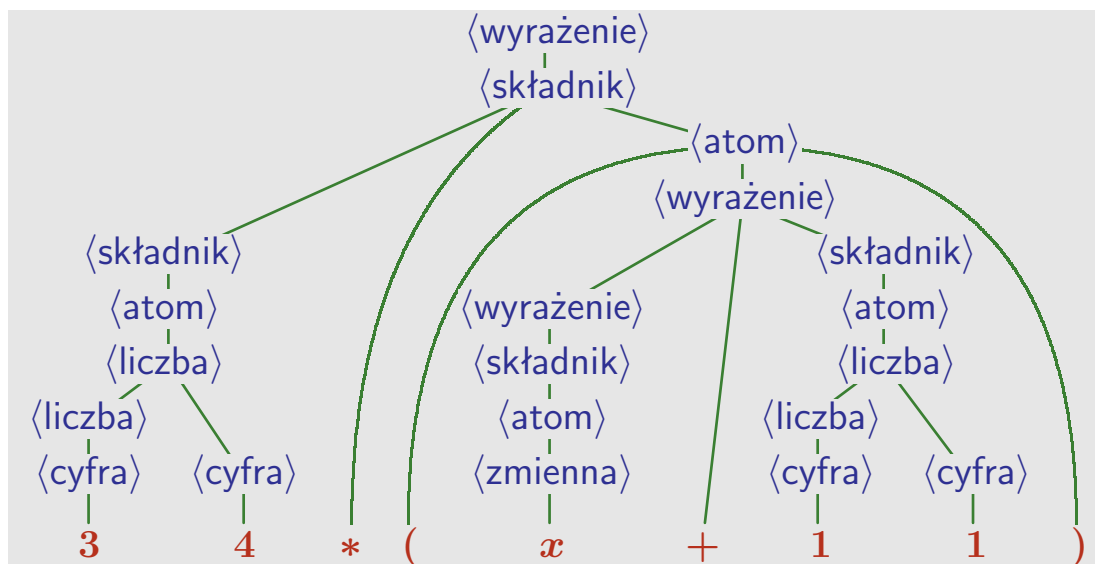
nieterminal
początkowy

Drzewa wyvodu

ciąg terminali:

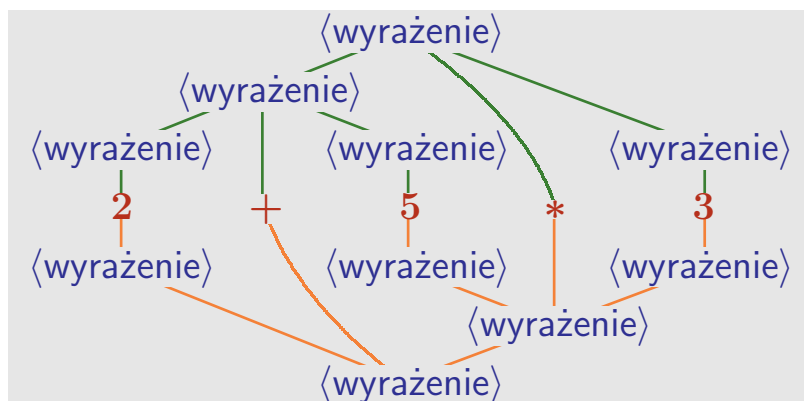
34 * (x + 11)

$\langle \text{zmienna} \rangle ::= a \mid b \mid \dots \mid z$
 $\langle \text{cyfra} \rangle ::= 0 \mid 1 \mid \dots \mid 9$
 $\langle \text{liczba} \rangle ::= \langle \text{liczba} \rangle \langle \text{cyfra} \rangle \mid \langle \text{cyfra} \rangle$
 $\langle \text{atom} \rangle ::= \langle \text{zmienna} \rangle \mid \langle \text{liczba} \rangle \mid (\langle \text{wyrażenie} \rangle)$
 $\langle \text{wyrażenie} \rangle ::= \langle \text{składnik} \rangle \mid \langle \text{wyrażenie} \rangle + \langle \text{składnik} \rangle$
 $\langle \text{składnik} \rangle ::= \langle \text{atom} \rangle \mid \langle \text{składnik} \rangle * \langle \text{atom} \rangle$



Drzewo wyvodu a struktura słowa

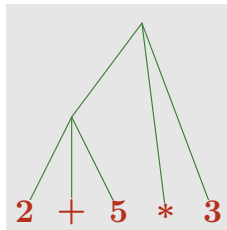
$\langle \text{wyrażenie} \rangle ::= 0 \mid 1 \mid \dots \mid 9 \mid$
 $\langle \text{wyrażenie} \rangle + \langle \text{wyrażenie} \rangle \mid$
 $\langle \text{wyrażenie} \rangle * \langle \text{wyrażenie} \rangle$



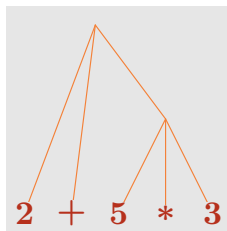
Niejednoznaczność wyvodu: istnieją dwa różne drzewa wyvodu danego słowa. W tym przypadku sugerują one kolejność działań:

- **drzewo górne** odpowiada wykonaniu dodawania przed mnożeniem
- **drzewo dolne** odpowiada wykonaniu mnożenia przed dodawaniem

Drzewo wyvodu a struktura słowa



← dla takiego rozbioru
wartość wyrażenia wynosi **21**



← dla takiego rozbioru
wartość wyrażenia wynosi **17**

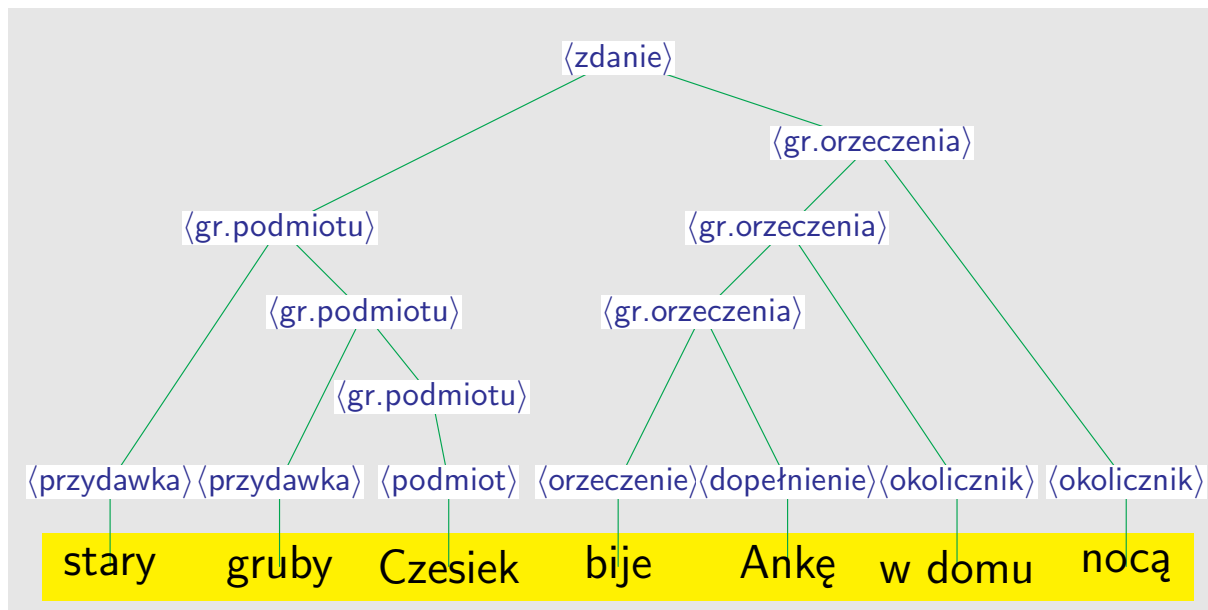
Znaczenie wyrażenia zależy od { przeprowadzenia jego rozbioru
zbudowania jego drzewa rozbioru

Drzewo wyvodu a struktura słowa

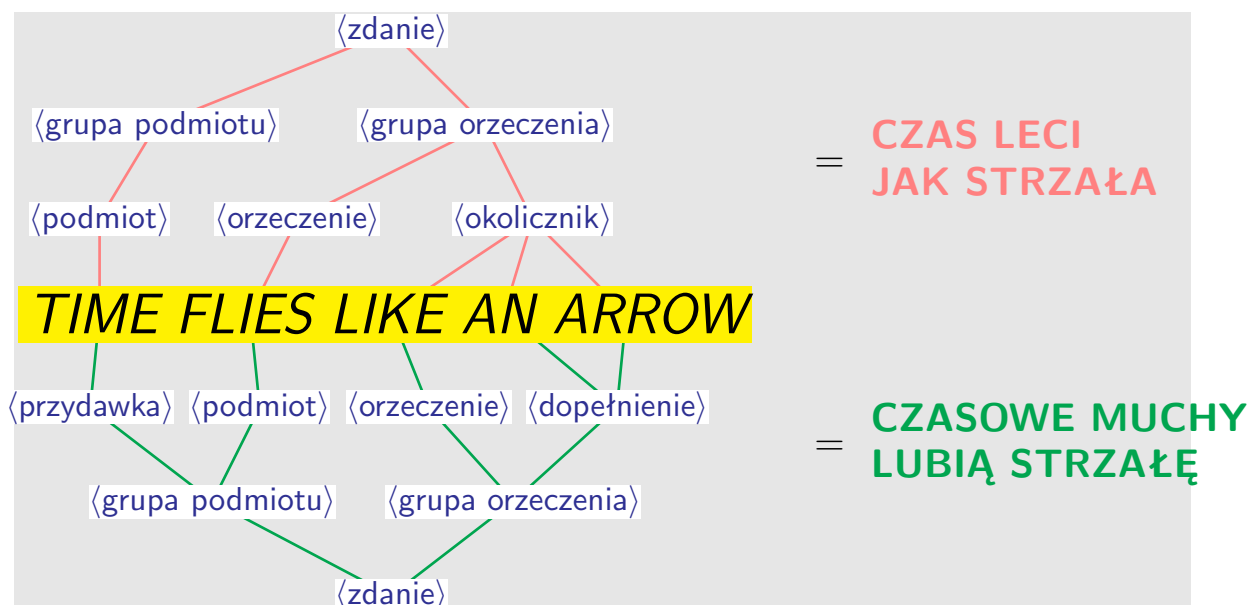
Gramatyka fragmentu języka polskiego:

⟨zdanie⟩	::=	⟨gr.podmiotu⟩	⟨gr.orzeczenia⟩
⟨gr.podmiotu⟩	::=	⟨podmiot⟩	
		⟨przydawka⟩	⟨gr.podmiotu⟩
⟨gr.orzeczenia⟩	::=	⟨orzeczenie⟩	
		⟨orzeczenie⟩	⟨dopełnienie⟩
		⟨gr.orzeczenia⟩	⟨okolicznik⟩
⟨podmiot⟩	::=	Anka	Basia Czesiek
⟨przydawka⟩	::=	gruby	stary
⟨orzeczenie⟩	::=	bije	kocha
⟨dopełnienie⟩	::=	Ankę	Basię Cześka
⟨okolicznik⟩	::=	w domu	nocą

Drzewo wyvodu a struktura słowa



Drzewo wyvodu a struktura słowa



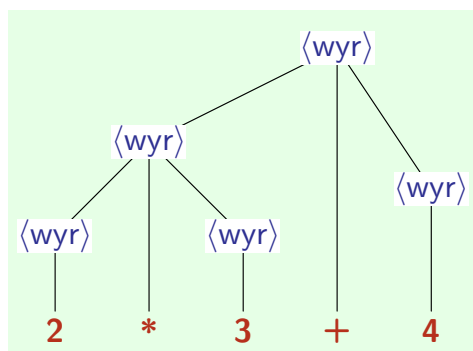
Znaczenie zależy od drzewa rozbioru również w językach naturalnych.

Drzewo wyvodu a struktura słowa

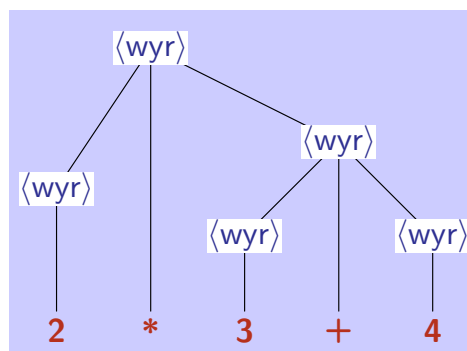
Z punktu widzenia **należenia słowa do języka** postać drzewa wyvodu nie ma znaczenia.

Z punktu widzenia **przekazywanego sensu** postać drzewa wyvodu jest bardzo ważna.

$\langle \text{wyr} \rangle$	$::=$	2	3	4
		$\langle \text{wyr} \rangle + \langle \text{wyr} \rangle$		
		$\langle \text{wyr} \rangle * \langle \text{wyr} \rangle$		



$$2 * 3 + 4 = 10$$



$$2 * 3 + 4 = 14$$

Jak wyrazić w gramatyce, że „mnożenie ma pierwszeństwo przed dodawaniem” (czyli wykluczyć niebieskie drzewo)?

Drzewo wyvodu a struktura słowa

Potrzebujemy gramatyki

- **równoważnej** — definiującej ten sam język, oraz
- **jednoznacznej** — żeby każde słowo z języka miało tylko jedno drzewo wyvodu

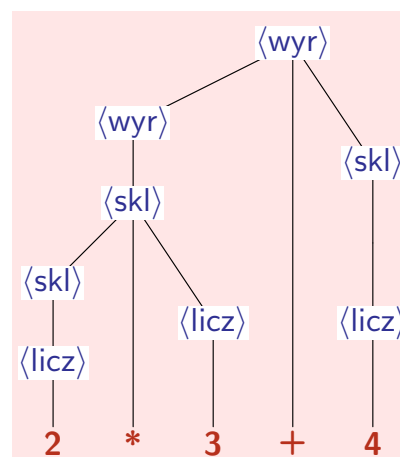
$\langle \text{wyr} \rangle$	$::=$	2	3	4
		$\langle \text{wyr} \rangle + \langle \text{wyr} \rangle$		
		$\langle \text{wyr} \rangle * \langle \text{wyr} \rangle$		

→

$\langle \text{wyr} \rangle$	$::=$	$\langle \text{skł} \rangle$		$\langle \text{wyr} \rangle + \langle \text{skł} \rangle$	
$\langle \text{skł} \rangle$	$::=$	$\langle \text{licz} \rangle$		$\langle \text{skł} \rangle * \langle \text{licz} \rangle$	
$\langle \text{licz} \rangle$	$::=$	2	3	4	

— ciąg składników
— ciąg liczb

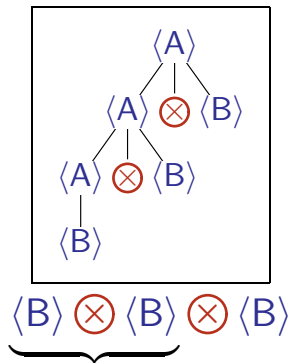
- w drzewie wygenerowanym przez czerwoną gramatykę pod nieterminal $\langle \text{skł} \rangle$ nie może już pojawić się nieterminal $\langle \text{wyr} \rangle$
- dlatego pod terminalem $*$ nie może pojawić się terminal $+$
- wobec tego terminale $*$ zawsze są głębiej w drzewie niż terminale $+$
- czyli czerwona gramatyka nie generuje drzew, w których $+$ wiązałyby silniej niż $*$



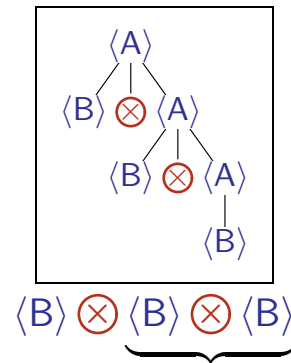
Drzewo wyvodu a struktura słowa

Jak wyrażać w gramatyce działania z priorytetami?

$\langle A \rangle ::= \langle B \rangle \mid \langle A \rangle \otimes \langle B \rangle$ lub $\langle A \rangle ::= \langle B \rangle \mid \langle B \rangle \otimes \langle A \rangle$



„rekursja” **lewostronna** —
działanie łączy do **lewej**



„rekursja” **prawostronna** —
działanie łączy do **prawej**

Działanie w gramatyce — im **dalej** od nieterminala początkowego, tym **głębiej** występuje w drzewie, więc **silniej łączy**.