

Stefan Sokołowski

JĘZYKI FORMALNE I METODY KOMPILACJI

wykład 3

Inst. Informatyki Stosowanej, ANS
Elbląg, 2025/2026

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

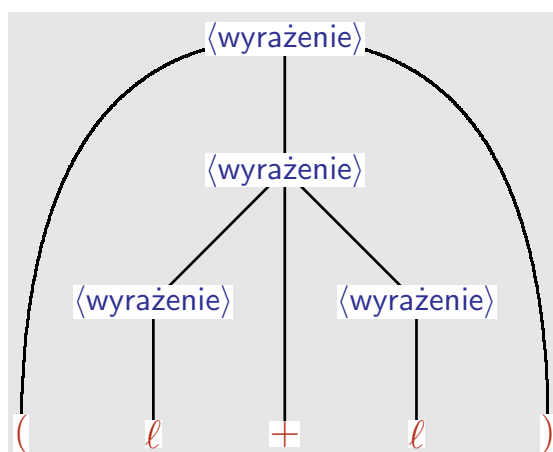
$\langle \text{wyrażenie} \rangle ::= \ell \mid (\langle \text{wyrażenie} \rangle) \mid \langle \text{wyrażenie} \rangle + \langle \text{wyrażenie} \rangle$

$\langle \text{wyrażenie} \rangle$

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

$\langle \text{wyrażenie} \rangle$

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



Drzewa wyvodu

DEFINICJA:

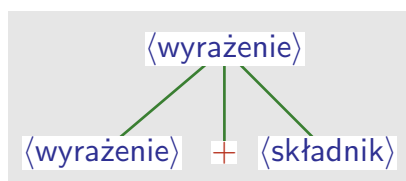
Drzewo wyvodu słowa $w \in \Sigma^*$ w gramatyce bezkontekstowej

$G = \langle \Sigma, N, P, \langle S \rangle \rangle$ 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 nieterminal $\langle A \rangle$ 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 nieterminal początkowy $\langle S \rangle$.

PRZYKŁAD:

Fragment drzewa



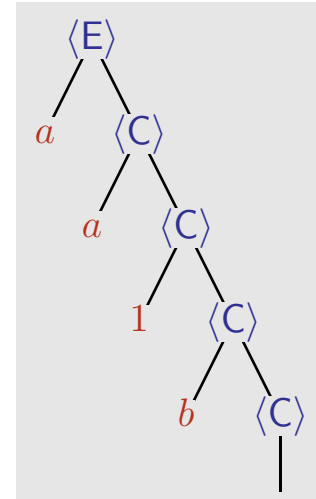
jest OK jeśli

$\langle \text{wyrażenie} \rangle \rightarrow \langle \text{wyrażenie} \rangle + \langle \text{składnik} \rangle$ 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.

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

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

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

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

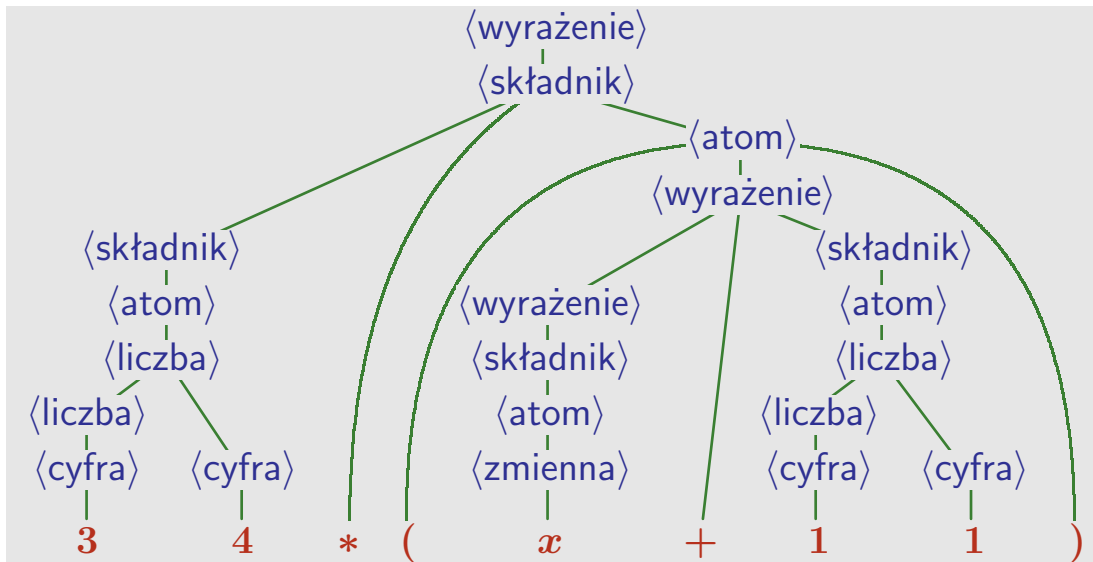
$$\langle \text{wyrażenie} \rangle ::= \langle \text{składnik} \rangle \mid \langle \text{wyrażenie} \rangle + \langle \text{składnik} \rangle$$

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

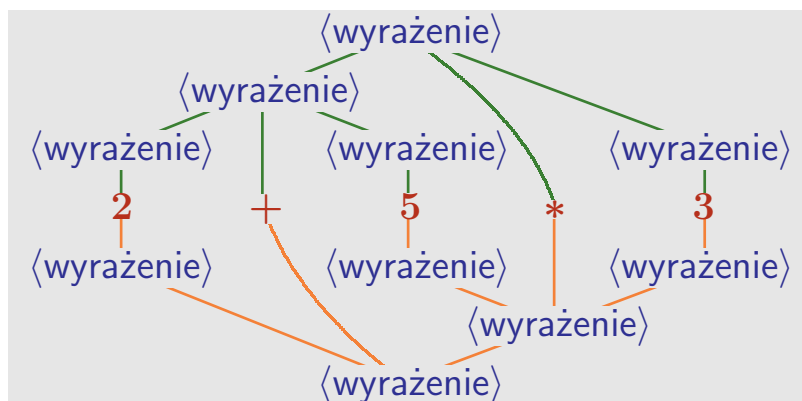
nieterminal
początkowy

Drzewa wywodu

ciąg terminali:

$$34 * (x + 11)$$
$$\begin{aligned} \langle \text{zmienna} \rangle &::= a \mid b \mid \dots \mid z \\ \langle \text{cyfra} \rangle &::= 0 \mid 1 \mid \dots \mid 9 \\ \langle \text{liczba} \rangle &::= \langle \text{liczba} \rangle \langle \text{cyfra} \rangle \mid \langle \text{cyfra} \rangle \\ \langle \text{atom} \rangle &::= \langle \text{zmienna} \rangle \mid \langle \text{liczba} \rangle \mid (\langle \text{wyrażenie} \rangle) \\ \langle \text{wyrażenie} \rangle &::= \langle \text{składnik} \rangle \mid \langle \text{wyrażenie} \rangle + \langle \text{składnik} \rangle \\ \langle \text{składnik} \rangle &::= \langle \text{atom} \rangle \mid \langle \text{składnik} \rangle * \langle \text{atom} \rangle \end{aligned}$$


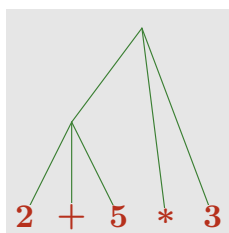
Drzewo wyvodu a struktura słowa

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


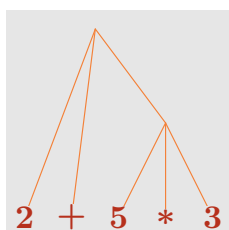
Niejednoznaczność wyvodu: istnieją dwa różne drzewa wyvodu danego słowa. W tym przypadku sugerują one kolejność działań:

- **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**



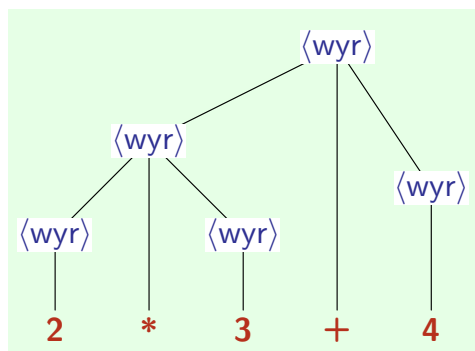
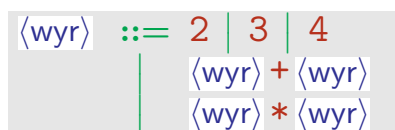
← dla takiego rozbioru
wartość wyrażenia wynosi **17**

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

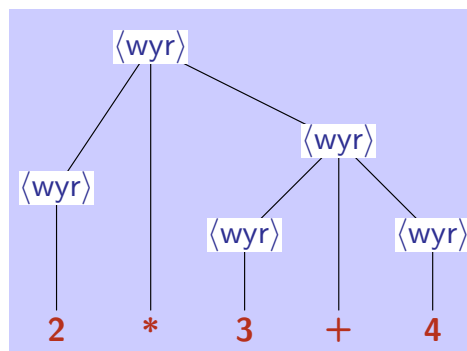
Drzewo wyvodu a struktura słowa

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

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



$$2 * 3 + 4 = 10$$



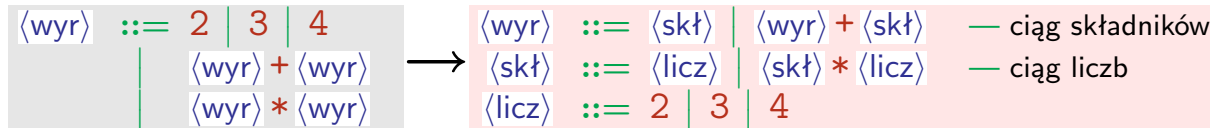
$$2 * 3 + 4 = 14$$

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

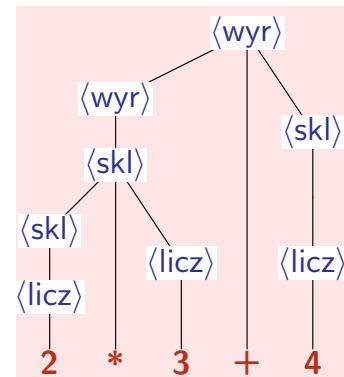
Drzewo wyvodu a struktura słowa

Potrzebujemy gramatyki

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



- w drzewie wygenerowanym przez czerwoną gramatykę pod nieterminaliem $\langle \text{skł} \rangle$ 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 $+$ wiązałby silniej niż $*$



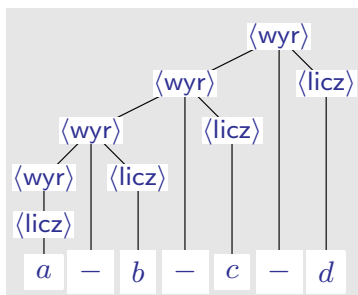
Działanie w gramatyce — im **dalej** od nieterminalu początkowego, tym **głębiej** występuje w drzewie, więc **silniej łączy**.

Drzewo wyvodu a struktura słowa

Jak wyrażać działania łączące w lewo, a jak w prawo?

ODEJMOWANIE

$$\underline{a - b} - c - d$$

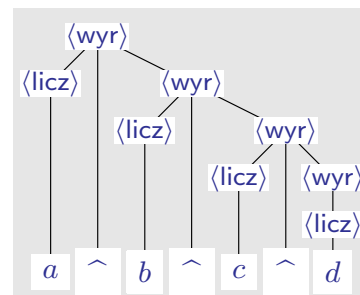


$$\langle \text{wyr} \rangle ::= \langle \text{licz} \rangle \mid \langle \text{wyr} \rangle - \langle \text{licz} \rangle$$

działanie łączy w **lewo** — rekursja **lewostronna**

POTĘGOWANIE

$$a \wedge \underline{b \wedge c} \wedge d$$



$$\langle \text{wyr} \rangle ::= \langle \text{licz} \rangle \mid \langle \text{licz} \rangle \wedge \langle \text{wyr} \rangle$$

działanie łączy w **prawo** — rekursja **prawostronna**

Drzewo wyvodu a struktura słowa

Powtarzanie fragmentu wyvodu — w gramatyce rekursja.

Rekursja lewostronna czy prawostronna ?

— odpowiedź na ogół

- nie ma wpływu na sam generowany język, ale
- ma wpływ na kształt drzew wyvodu; więc zależy od niej **znaczenie** słów.

Wg przyjętych umów większość działań arytmetycznych łączy do lewej, więc należy je opisywać rekursją lewostronną.

Drzewo wyvodu a struktura słowa

Oznaczenie w gramatyce dla powtarzania z łączeniem lewostronnym:

$\langle S \rangle ::= \{\text{ciąg-symb}\}^*$ oznacza $\langle S \rangle ::= \lambda \mid \langle S \rangle \{\text{ciąg-symb}\}$

czyli: nieterminal $\langle S \rangle$ generuje „ciąg-symb” powtórzony wielokrotnie (ew. 0-krotnie)

Przykład:

$\langle \text{liczba_bin} \rangle ::= \{0 \mid 1\}^*$

czyli: nieterminal $\langle \text{liczba_bin} \rangle$ generuje: 0 lub 1 powtórzone wielokrotnie
z **łączeniem lewostronnym**

Przykład:

$\langle \text{suma} \rangle ::= \langle \text{składnik} \rangle \{+ \langle \text{składnik} \rangle\}^*$

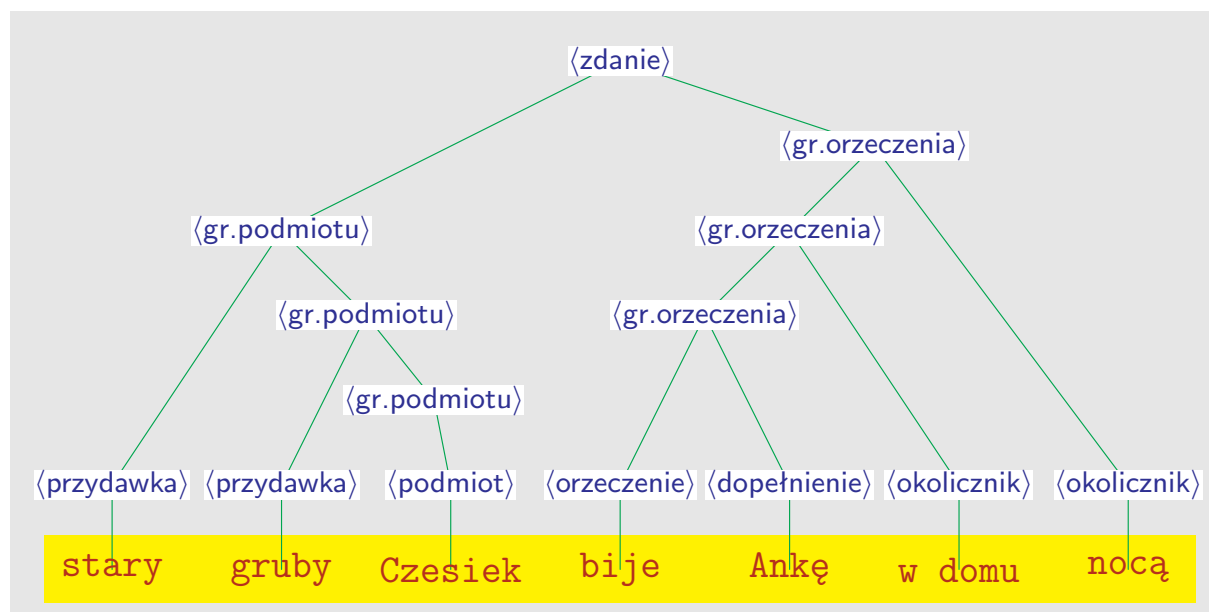
czyli: nieterminal $\langle \text{suma} \rangle$ generuje: $\langle \text{składnik} \rangle$, a po nim kilka razy „+ $\langle \text{składnik} \rangle$ ”
z **łączeniem lewostronnym**

Drzewo wyvodu a struktura słowa

Gramatyka fragmentu języka polskiego:

$\langle \text{zdanie} \rangle$	$::=$	$\langle \text{gr. podmiotu} \rangle$	$\langle \text{gr. orzeczenia} \rangle$
$\langle \text{gr. podmiotu} \rangle$	$::=$	$\langle \text{podmiot} \rangle$	
		$\langle \text{przydawka} \rangle$	$\langle \text{gr. podmiotu} \rangle$
$\langle \text{gr. orzeczenia} \rangle$	$::=$	$\langle \text{orzeczenie} \rangle$	
		$\langle \text{orzeczenie} \rangle$	$\langle \text{dopełnienie} \rangle$
		$\langle \text{gr. orzeczenia} \rangle$	$\langle \text{okolicznik} \rangle$
$\langle \text{podmiot} \rangle$	$::=$	Anka	Basia Czesiek
$\langle \text{przydawka} \rangle$	$::=$	gruby	stary
$\langle \text{orzeczenie} \rangle$	$::=$	bije	kocha
$\langle \text{dopełnienie} \rangle$	$::=$	Ankę	Basię Cześka
$\langle \text{okolicznik} \rangle$	$::=$	w domu	nocą

Drzewo wyvodu a struktura słowa



Drzewo wyvodu a struktura słowa



Znaczenie zależy od drzewa rozbioru również w językach naturalnych.