R O Z D Z I A Ł   V

Gramatyki struktur frazowych

Przypomnijmy, że gramatyka generatywna $G = \langle V_N, V_T, S, F \rangle$ jest typu 0 (lub równoważnie struktur frazowych), gdy wszystkie jej reguły produkcji mają postać $P \rightarrow Q$, gdzie $P \in (V_N \cup V_T)^* \, V_N \, (V_N \cup V_T)^*$, a $Q \in (V_N \cup V_T)^*$.

1. Postać normalna

Twierdzenie 5.1.

Dla każdej gramatyki G typu 0 istnieje równoważna jej gramatyka G' typu 0 o regułach produkcji następującej postaci:

1. $S \rightarrow \lambda$,	5. $AB \rightarrow AC$,
2. $A \rightarrow a$,	6. $AC \rightarrow BC$,
3. $A \rightarrow B$,	7. $AB \rightarrow B$,
4. $A \rightarrow BC$,	

gdzie $A, B, C \in V_N$, $a \in V_T$, zaś S jest symbolem początkowym.

Zauważmy, że w postaci normalnej z powyższego twierdzenia:

1) reguły 2 - 4 są identyczne z regułami 1 - 3 postaci normalnej Kurody,
2) reguły 5 i 6 mają kształt reguł ($*$) z ostatniego wniosku z poprzedniego rozdziału,
3) λ występuje jedynie w 1 regule,
4) jeżeli jest to gramatyka λ - wolna, to wszystkie reguły niemonotoniczne są w niej jedynie kształtu 7,
5) ze złożenia reguł 4 i 7 otrzymujemy regułę 3, a ze złożenia reguł 7 i 4 otrzymujemy reguły równoważne regułom 5 i 6. Z powodzeniem moglibyśmy więc usunąć z powyższego twierdzenia reguły 3, 5 i 6, pozostawiając w nim jedynie reguły kształtu 1, 2, 4 i 7 (oczywiście w konkretnej gramatyce, krotność reguł tego typu zapewne by wówczas wzrosła).

Dowód.

Niech G będzie gramatyką typu 0. Zatem wszystkie jej reguły produkcji mają postać: $P \rightarrow Q$, gdzie $P \in (V_N \cup V_T)^* \, V_N \, (V_N \cup V_T)^*$, a $Q \in (V_N \cup V_T)^*$.

Gdy $Q = \lambda$, to:

a) gdy $P = S$, to jest to reguła kształtu 1, więc ją zostawiamy;

b) gdy $P \neq S$, to każdą z takich reguł zastępujemy zbiorem reguł

$$\bigcup_{X \in V_N \cup V_T} \{PX \to X,\ XP \to X\}.$$

Gdy $\lambda \in L(G)$, a w G nie ma reguły produkcji $S \to \lambda$, to ją dodajemy (jest to bowiem jedyny sposób wyprowadzenia λ w gramatyce z twierdzenia).
W ten sposób z λ już się uporaliśmy.

Z I twierdzenia o postaci normalnej każdą gramatykę G typu 0 możemy zamienić na równoważną jej gramatykę G' o postaci normalnej, w której regułach produkcji F' symbole terminalne mogą wystąpić jedynie w regułach produkcji postaci $A \to a$, gdzie A jest nieterminałem, zaś a terminałem. Są to więc reguły kształtu 2. dowodzonego twierdzenia.

Poza tym w F' mogą wystąpić jeszcze jedynie reguły produkcji kształtu $P \to Q$, gdzie P i Q są niepustymi ciągami nieterminałów.

1) Gdy $|P| \le |Q|$, to postępując identycznie jak w dowodzie twierdzenia 4.3 o postaci normalnej Kurody - otrzymujemy reguły produkcji o kształcie zgodnym z nią (tj. zgodnie z pierwszą uwagą po twierdzeniu 5.1 są to reguły żądanego kształtu z dowodzonego twierdzenia).

2) Gdy $|P| > |Q|$, to będzie to reguła kształtu $X_1\, X_2 \ldots X_m \to Y_1 \ldots Y_n$, gdzie $m > n \ge 1$. Każdą z nich możemy zastąpić następującym zbiorem reguł:

$$\begin{cases} X_{m-1} \quad X_m \to Z_m \quad U_m \\ Z_m \quad U_m \to U_m \end{cases}$$ (złożenie reguł kształtu 5 i 6) (reguła kształtu 7)

$$\begin{cases} X_{m-2}\ U_m \to Z_{m-1}\ U_{m-1} \\ Z_{m-1}\quad U_{m-1} \to U_{m-1} \end{cases}$$ (złożenie reguł kształtu 5 i 6) (reguła kształtu 7)

$$\begin{cases} X_{m-3}\ U_{m-1} \to Z_{n+1}\ U_{n+1} \\ Z_{m-2}\quad U_{m-2} \to U_{m-2} \end{cases}$$ (złożenie reguł kształtu 5 i 6) (reguła kształtu 7)

$$\vdots$$

$$\begin{cases} X_n \quad U_{n+2} \to Z_{n+1}\ U_{n+1} \\ Z_{n+1} \quad U_{n+1} \to U_n\ Y_n \end{cases}$$ (złożenie reguł kształtu 5 i 6) (złożenie reguł kształtu 5 i 6)

(ich stosowanie pozwoli nam zamienić wyrażenie $X_1 \ldots X_{n-1} X_n \ldots X_{m-1} X_m$ na wyrażenie $X_1 \ldots X_{n-1} U_n Y_n$) oraz:

b) $X_{n-1}U_n \to U_{n-1}Y_{n-1}$ (złożenie reguł kształtu 5 i 6)

$$\vdots$$

$X_1U_2 \to U_1Y_1$ (złożenie reguł kształtu 5 i 6)

(co przekształci wyrażenie $X_1 \ldots X_{n-1} U_n Y_n$ w wyrażenie $U_1 Y_1 \ldots Y_n$), aż w końcu stosując regułę

c) $U_1Y_1 \to Y_1$ (reguła kształtu 7)

- otrzymujemy wymagane wyrażenie $Y_1 \ldots Y_n$.

Na koniec musimy jeszcze stosowane tu reguły, będące złożeniem reguł kształtu 5 i 6, rozłożyć właśnie na reguły kształtu 5 i 6. I tak, każdą regułę postaci $AB \rightarrow CD$ zastępujemy parą reguł $AB \rightarrow AC$ (reguła kształtu 5) i $AC \rightarrow DC$ (reguła kształtu 6). □

Zauważmy, że powyższy dowód nie przedstawia sposobu konstrukcji gramatyki w postaci normalnej z powyższego twierdzenia. Nie jesteśmy bowiem w stanie efektywnie określić dla dowolnej gramatyki G, czy $\lambda \in L(G)$, czy też $\lambda \notin L(G)$. Wiemy jedynie, że zawsze ALBO $\lambda \in L(G)$ ALBO $\lambda \notin L(G)$.

2. Drzewa derywacji

Niech $G = \langle V_N, V_T, S, F \rangle$. Poszczególne jej reguły produkcji oznaczamy odpowiednio przez $f_1, f_2, \ldots, f_n$.

Przykład 5.1.

Niech $G = \langle V_N, V_T, S, F \rangle$, gdzie:

$V_N = \{S, A, B\}$, $V_T = \{a, b\}$,

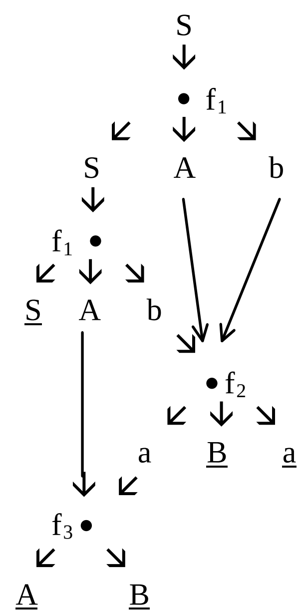
a $F = \{S \rightarrow SAb, bAb \rightarrow aBa, Aa \rightarrow AB, BB \rightarrow aBb\}$.

Mamy więc reguły:

$f_1 : S \rightarrow SAb$, $f_3 : Aa \rightarrow AB$,

$f_2 : bAb \rightarrow aBa$, $f_4 : BB \rightarrow aBb$.

Derywację $S \xrightarrow{1} SAb \xrightarrow{1} SAbAb \xrightarrow{2} SAaBa \xrightarrow{3} SABBa$ (w której nad strzałkami zaznaczono dodatkowo numery stosowanych reguł produkcji f_i) można przedstawić w postaci drzewa z rys 5.1.



Rys. 5.1.

Podkreśleniami dodatkowo zaznaczono tu litery końcowego słowa „SABBa”. □

Spostrzeżenia

1) Końcowe słowo tworzą czytane po kolei od lewej wszystkie „wolno zwisające” literki (są one nazywane też „l i ś ć m i” tego drzewa).
2) Zarówno każde drzewo opisuje pewną derywację w gramatyce typu 0 (niekoniecznie w sposób jednoznaczny), jak i na odwrót - każda derywacja da się opisać za pomocą pewnego drzewa derywacji.
3) Strzałki w drzewach derywacji nigdy nie będą się krzyżowały (bo w poszczególnych słowach danej derywacji zmieniamy na nowy jedynie ciąg <u>sąsiadujących</u> symboli alfabetu $V_N \cup V_T$). □

Zadanie 5.1: Podaj drzewo derywacji słowa "sasza" w gramatyce $G = \langle V_N, V_T, S, F \rangle$, w której $V_N = \{S, A, F, K, L, O, R, T, Z\}$, $V_T = \{s, a, z\}$, $F = \{S \rightarrow FAL$, $L \rightarrow ZA$, $L \rightarrow \lambda$, $F \rightarrow a$, $aA \rightarrow za$, $aZ \rightarrow asK$, $FA \rightarrow sK$, $K \rightarrow a$, $F \rightarrow OT$, $OT \rightarrow s$, $KA \rightarrow Ra$, $R \rightarrow z\}$, a S jest symbolem początkowym. Jakie jeszcze słowa można wygenerować tą gramatyką? Narysuj drzewa ich derywacji. □