Peter Štibraný, 2i4
Úloha 1, sada 2

Úloha: Dokážte ekvivalenciu Chomského definície a štandardnej definície kontextových gramatík. Pre kontextovú gramatiku $G = (N, T, P, \sigma)$ sú definície nasledovné:

— Chomského definícia je, že pravidlá gramatiky majú tvar $u\alpha v \to u\beta v$, kde $u, v \in (N \cup T)^*$, $\alpha \in N$ a $\beta \in (N \cup T)^+$.

— Štandardná definícia je, že pravidlá gramatiky majú tvar $u \to v$, pričom $u \in (N \cup T)^* N (N \cup T)^*$, $v \in (N \cup T)^+$ a zároveň $|u| \leq |v|$.

Riešenie: Dôkaz ekvivalencie dvoch definícií urobíme v dvoch krokoch. Najprv ukážeme, že pre každú Chomského kontextovú gramatiku existuje štandardná kontextová gramatika, pričom obe gramatiky generujú ten istý jazyk. V druhom kroku (tvrdenie 4 a pomocné tvrdenia 2, 3) dokážeme opačné tvrdenie, teda pre štandardnú kontextovú gramatiku nájdeme ekvivalentnú Chomského kontextovú gramatiku.

Pod *ekvivalenciou definícií* sa myslí ekvivalencia tried jazykov určených gramatikami podľa dvoch definícií zo zadania.

Tvrdenie 1: Nech $G = (N, T, P, \sigma)$ je Chomského kontextová gramatika. Potom existuje štandardná kontextová gramatika G' taká, že $L(G') = L(G)$.

Dôkaz: Dokonca platí $G' = G$. Nech $p = u\alpha v \to u\beta v$, $p \in P$ a tiež $p \in P'$. Musíme ukázať, že $|u\alpha v| \leq |u\beta v|$. Platí $|u\alpha v| = |u| + |\alpha| + |v|$ a tiež $|u\beta v| = |u| + |\beta| + |v|$. Aby platila požadovaná nerovnosť, stačí nám ukázať platnosť nerovnosti $|\alpha| \leq |\beta|$. Podľa definície $|\alpha| = 1$ a zároveň $1 \leq |\beta|$. Teda $|\alpha| \leq |\beta|$ platí. $\square$

Tvrdenie 2: Nech G je štandardná kontextová gramatika. Potom existuje štandardná kontextová gramatika G', pričom $P' \subseteq (N')^+ \times (N' \cup T')^+$, a platí $L(G') = L(G)$. G' je kontextová gramatika, ktorá má na ľavej strane iba neterminály.

Dôkaz: $G' = (N', T', P', \sigma')$, pričom $N' = N \cup \{\xi_t | t \in T'\}$, $T' = T$, $\sigma' = \sigma$. Na vytvorenie pravidiel P' si zaveďme homomorfizmus $h: (N \cup T)^* \to (N')^*$, pričom $h(\gamma) = \gamma$ pre $\forall \gamma \in N$ a $h(t) = \xi_t$ pre $\forall t \in T$. Pravidlá gramatiky G' budú $P' = \{h(u) \to h(v) | u \to v \in P\} \cup \{\xi_t \to t | t \in T'\}$. Vidíme, že ľavé strany pravidiel v P' tvoria iba neterminály z N'. Už stačí iba ukázať, že obe generujú rovnaký jazyk, teda $L(G') = L(G)$.

$L(G) \subseteq L(G')$: Nech $w \in L(G)$. Potom v gramatike G existuje odvodenie $\sigma \Rightarrow w_1 \Rightarrow \ldots \Rightarrow w_n \Rightarrow w$. Začiatok hľadaného odvodenia v gramatike G' bude $\sigma \Rightarrow h(w_1) \Rightarrow h(w_2) \Rightarrow \ldots \Rightarrow h(w_n) \Rightarrow h(w)$. V tomto momente bude slovo $h(w)$ pozostávať iba zo symbolov $\xi_t, t \in T'$ a už stačí iba použiť pravidlá $\xi_t \to t$ na odvodenie slova w zo slova $h(w)$.

$L(G') \subseteq L(G)$: Nech $w \in L(G')$ a teda existuje odvodenie slova w v gramatike G'. V tomto odvodení museli byť použité pravidlá typu $\xi_t \to t$, pretože inými pravidlami terminálne symboly v slove nevzniknú a slovo w musí byť zložené z terminálnych symbolov.

Definujme homomorfizmus $h': (N' \cup T')^* \to (N \cup T)^*$ takto: $h'(\xi_t) = t$ pre všetky $t \in T'$, ďalej $h'(x) = x$ pre $x \in N \cup T$ (pripomímam, že $N \cup T \subseteq N' \cup T'$).

Nech $w_{i-1} \Rightarrow w_i \Rightarrow w_{i+1} \Rightarrow w_{i+2}$ je postupnosť krokov odvodenia, pričom v kroku $w_i \Rightarrow w_{i+1}$ bolo použité pravidlo typu $\xi_t \to t$, tj. ak $w_i = u\xi_t v$, potom $w_{i+1} = utv$. Potom $h'(w_i) = h'(w_{i+1})$. Platí $h'(w_i) = h'(u).h'(\xi_t).h'(v)$, čo je rovné $h'(u).t.h'(v)$, čo je to isté ako $h'(u).h'(t).h'(v) = h'(w_{i+1})$.

Vezmime odvodenie slova w v gramatike G' a vynechajme z neho tie kroky odvodenia, kde sa použili pravidlá typu $\xi_t \to t$. Takto dostaneme postupnosť krokov odvodenia, ktorá ale už nie je odvodením v gramatike G'. Keď však v týchto krokoch odvodenia zameníme slová ich homomorfnými obrazmi (pomocou h'), dostaneme korektné odvodenie slova w v gramatike G.

Keď pôvodné odvodenie bolo $\sigma \Rightarrow w_1 \Rightarrow \ldots \Rightarrow w_n \Rightarrow w$, tak vynechaním uvedených krokov odvodenia a použitím homomorfizmu dostaneme $h'(sigma) \Rightarrow h'(w_{i_1}) \Rightarrow \ldots \Rightarrow h'(w_{i_m}) \Rightarrow w$, pričom $i_1 < i_2 < \ldots < i_m$ a $m < n$.

Už sme ukázali, že vynechaním uvedených krokov odvodenia si nové odvodenie nepokazíme. Ukážeme ešte, že toto nové odvodenie používa iba pravidlá z gramatiky G. Predpokladajme že $w_{i_k} \Rightarrow w_{i_{k+1}}$ bol krok odvodenia v pôvodnom odvodení a že tu nebolo použité pravidlo typu $\xi_t \to t$. Potom to ale muselo byť pravidlo typu $h(u) \to h(v)$, pričom $u \to v \in P$. Vezmime si $h'(h(u)) \to h'(h(v))$. Čo dostaneme? $h'(h(t)) = h'(\xi_t) = t$ pre $t \in T$. Rovnako $h'(h(n)) = h'(n) = n$ pre $n \in N$. Vidíme že $h'(h(u)) \to h'(h(v)) = u \to v \in P$ a teda že aj v novom odvodení sa používajú iba pravidlá z P.

Našli sme odvodenie slova $w \in L(G')$ v gramatike G, preto $L(G') \subseteq L(G)$. $\square$

Tvrdenie 3: Nech G je štandardná kontextová gramatika, pričom $P \subseteq N^+ \times (N \cup T)^+$. Keďže počet pravidiel v gramatike G je konečný, možno každému pravidlu jednoznačne priradiť číslo 1 až n, kde n je počet pravidiel. Pravidlo s číslom l budem označovať p_l a ďalej budeme predpokladať, že $p_l = \alpha_1 \dots \alpha_k \to \beta_1 \dots \beta_k$, pričom k je dĺžka ľavej strany pravidla p_l, $\alpha_i \in N, i \in \{1, \dots, k\}$ a $\beta_i \in N \cup T$ pre $i \in \{1, \dots, k-1\}$, $\beta_k \in (N \cup T)^+$.

Nech $G^i = (N^i, T^i, P^i, \sigma)$ pre $i \in \{0, \dots, n\}$ je gramatika, ktorú dostaneme z G nasledovným spôsobom: $T^i = T$, $N^i = N^{i-1} \cup \{\gamma_1^i, \dots, \gamma_m^i\}$, kde m je dĺžka ľavej strany pravidla p_i. Nová množina pravidiel gramatiky G^i bude

$$
\begin{aligned}
P^i = (P^{i-1} \setminus \{p_i\}) \cup \{ & \\
\alpha_1 \dots \alpha_k \to \gamma_1^i \alpha_2 \dots \alpha_k, & \qquad (*_1) \\
\gamma_1^i \alpha_2 \dots \alpha_k \to \gamma_1^i \gamma_2^i \alpha_3 \dots \alpha_k, & \qquad (*_2) \\
\dots & \\
\gamma_1^i \dots \gamma_{k-2}^i \alpha_{k-1} \alpha_k \to \gamma_1^i \dots \gamma_{k-2}^i \gamma_{k-1}^i \alpha_k, & \qquad (*_{k-1}) \\
\gamma_1^i \dots \gamma_{k-1}^i \alpha_k \to \gamma_1^i \dots \gamma_{k-1}^i \gamma_k^i, & \qquad (*_k) \\
\gamma_1^i \dots \gamma_k^i \to \beta_1 \gamma_2^i \dots \gamma_k^i, & \qquad (*_{k+1}) \\
\beta_1 \gamma_2^i \dots \gamma_k^i \to \beta_1 \beta_2 \gamma_3^i \dots \gamma_k^i, & \qquad (*_{k+2}) \\
\dots & \\
\beta_1 \dots \beta_{k-1} \gamma_k^i \to \beta_1 \dots \beta_{k-1} \beta_k \} & \qquad (*_{2k})
\end{aligned}
$$

G^0 definujme ako G. Potom $L(G^i) = L(G)$. *Poznámka:* pod γ_k^i sa nemyslí i–násobné použitie γ_k, je to iba označenie symbolu, použité preto, aby sa nové neterminály v jednotlivých gramatikách G^i navzájom odlíšili.

Dôkaz: Dôkaz indukciou vzhľadom na i. Pre $i = 0$ máme $G^i = G^0 = G$ a $L(G^0) = L(G)$ platí triviálne.

Nech platí $L(G^j) = L(G)$. Ukážeme, že potom platí aj $L(G^{j+1}) = L(G)$.

Najprv ukážeme $L(G) \subseteq L(G^{j+1})$. Vezmime si slovo w z jazyka $L(G)$. Podľa indukčného predpokladu potom $w \in L(G^j)$, a teda existuje v G^j odvodenie tohto slova. Ak v tomto odvodení nebolo použité pravidlo p_j, potom je toto odvodenie zároveň aj odvodením v gramatike G^{j+1} a teda $w \in L(G^{j+1})$. Ak v odvodení slova w v jazyku G^j bolo použité pravidlo p_j, potom tie kroky odvodenia, kde bolo toto pravidlo použité, možno nahradiť takou postupnosťou krokov odvodenia, pri ktorej sa použijú pravidlá $*_1$ až $*_{2k}$ (k je dĺžka ľavej strany pravidla p_{j+1}). Ľahko nahliadneme, že takto odvodíme rovnaké slovo ako pri odvodení v G. Nové pravidlá sme totiž tvorili takým spôsobom, aby boli funkčne rovnaké ako pravidlo, ktoré nahrádzajú. Ak je v pôvodnom odvodení viac krokov odvodenia s použitím pravidla p_{j+1}, tak sa nahradia všetky.

Chceme ešte ukázať $L(G^{j+1}) \subseteq L(G)$. Nech $w \in L(G^{j+1})$. Opäť máme dve možnosti: ak sa v odvodení slova w v gramatike G^{j+1} nepoužilo ani jedno z nových pravidiel $*_1$ až $*_{2k}$, potom je toto odvodenie zároveň aj odvodením v G^j a teda podľa indukčného predpokladu $w \in L(G)$.

Ostáva nám ukázať existenciu odvodenia slova w v gramatike $L(G)$, keď vieme, že existuje odvodenie w v $L(G^{j+1})$, pričom v tomto odvodení bolo použité jedno z pravidiel $*_1$ až $*_{2k}$. Nech bolo použité pravidlo $*_l$, pričom $1 < l \leq 2k$. To znamená, že existuje krok odvodenia $xuy \Rightarrow xvy$, pričom $u \to v$ je pravidlo l. Avšak podslovo u mohlo vzniknúť jedine použitím pravidla $*_{l-1}$, pretože v pravidlách P^j sa s neterminálmi γ_i^{j+1} nepracuje a pomocou nových pravidiel sa dá u odvodiť jedine pomocou pravidla $*_{l-1}$. Rovnakou argumentáciou vieme ukázať, že museli byť použité všetky pravidlá $*_1$ až $*_l$! Ak $l < 2k$, potom nato aby sme sa zbavili neterminálov γ_i^{j+1} sa musia použiť (postupne) aj pravidlá $*_{l+1}$ až $*_{2k}$.

Vieme už, že v odvodení slova w v gramatike G^{j+1} boli použité všetky pravidlá $*_1$ až $*_{2k}$. Čo potrebujeme, je odvodenie v gramatike G. Ak však odstránime kroky odvodenia, kde boli použité pravidlá $*_1$ až $*_{2k}$, a na miesto kde bol predtým krok odvodenia s použitím pravidla $*_{k+1}$ dáme krok odvodenia s použitím pravidla p_{j+1}, dostaneme odvodenie slova w v gramatike G^j. Podľa indukčného predpokladu ($L(G^j) = L(G)$) potom už môžeme tvrdiť, že $w \in L(G)$ a teda $L(G^{j+1}) \subseteq L(G)$.

Ešte však musíme ukázať, že po hore uvedenej náhrade dostaneme skutočne odvodenie slova w v G^j. Čo nám môže toto odvodenie pokaziť sú „paralelné“ odvodenia v podslovách slova w, ktorých kroky odvodenia sú medzi krokmi odvodenia s pravidlami $*_1$ až $*_{2k}$. Ak však v kroku odvodenia $uw_mv \Rightarrow uw_{m+1}v$ je použité pravidlo $*_l$, $l < 2k$, a v kroku $u'w_{m'}v' \Rightarrow u'w_{m'+1}v'$ je použité pravidlo $*_{l+1}$, tak potom musí platiť $w_{m+1} = w_{m'}$, pretože ináč by sa pravidlo $*_{l+1}$ nedalo použiť. Teda ľubovoľné paralelné odvodenia ktoré medzitým prebehli mohli zmeniť iba podslová u a v.

Prečo sme pravidlo p_{j+1} použili pri kroku odvodenia, ktorým sme nahradili pôvodný krok odvodenia s použitím $*_{k+1}$? Nech $uw_mv \Rightarrow uw_{m+1}v$ je krok odvodenia s použitím pravidla $*_{k+1}$. Potom pred týmto krokom nemohli prebiehať žiadne paralelné odvodenia, kde by sa vyskytovala pravá časť pravidla p_{j+1}, a obdobne za týmto pravidlom nemohli prebiehať odvodenia na podslove, ktoré je rovné ľavej časti pravidla p_{j+1}. Keď miesto kroku odvodenia s pravidlom $*_{k+1}$ bude krok odvodenia s pravidlom p_{j+1}, potom paralelné odvodenia to neovplyvní.

Ak sa v odvodení slova w v gramatike G^{j+1} vyskytne postupnosť krokov odvodenia s novými pravidlami viackrát, tak treba zameniť každú takúto postupnosť za pravidlo p_{j+1}, aby sme získali odvodenie v gramatike G^j. □

Tvrdenie 4: Nech G je štandardná kontextová gramatika. Potom existuje Chomského kontextová gramatika G' a platí $L(G') = L(G)$.

Dôkaz: Ku gramatike G vieme podľa tvrdenia 2 nájsť gramatiku G'' ktorá má na ľavej strane pravidiel iba neterminály a $L(G'') = L(G)$. Nech $n = |P''|$, tj. n je počet pravidiel gramatiky G''. Nech $G' = (G'')^n$, pričom uvažujeme označenie z tvrdenia 3. Z tvrdenia 3 potom dostávame platnosť $L(G') = L(G'')$ a keďže $L(G'') = L(G)$ tak platí aj $L(G') = L(G)$. Potrebujeme ešte ukázať, že pravidlá v gramatike G' spĺňajú podmienku z Chomského definície. Do pravidiel gramatiky G' sme však pridávali iba také pravidlá, ktoré spĺňajú Chomského podmienku. Navyše v $G' = (G'')^n$ už neostali žiadne pôvodné pravidlá gramatiky G''. Preto G' je Chomského kontextová gramatika a $L(G') = L(G)$. □

Tvrdenie 5: Trieda jazykov generovaná Chomského kontextovými gramatikami a trieda jazykov generovaná štandardnými kontextovými gramatikami sú ekvivalentné.

Dôkaz: Tvrdenie je dôsledkom tvrdení 1 a 4. □