

(Fraasistruktuuri)grammatika on nelik $G = (V, \Sigma, P, S)$, kus

- V on mingi tähestik, s.t. lihtsalt mingi lõplik hulk.
- $\Sigma \subseteq V$, kusjuures $\Sigma \neq V$ ja $\Sigma \neq \emptyset$. Hulka Σ kuuluvaid tähti nimetatakse terminalideks (terminaalsümboliteks, terminaalseteks sümboliteks). Hulka $V \setminus \Sigma$ kuuluvaid tähti nimetatakse mitteterminalideks.
- P on mingi lõplik sõnepaaride hulk, sõned on üle tähestiku V . Hulka P kuuluvaid paare nimetatakse *reegliteks*. Paari, mille vasakpoolne komponent $u \in V^*$ ja parempoolne komponent $v \in V^*$, kirjutatakse tavaliselt $u \rightarrow v$.
- $S \in V \setminus \Sigma$. Teda kutsutakse *algsümboliks*.

Reeglite hulk P defineerib seose $\Rightarrow$ kõigi sõnede hulgal V^* . Kui $x, y, u, v \in V^*$ ja $(u \rightarrow v) \in P$, siis loeme, et $xuy \Rightarrow xvy$. S.t. mingis sõnes võib alamsõne u asendada alamsõnega v . Seose $\Rightarrow$ refleksiivset transitiivset sulundit tähistame $\Rightarrow^*$.

Grammatika $G = (V, \Sigma, P, S)$ genereerib keele

$$L(G) = \{x \in \Sigma^* \mid S \Rightarrow^* x\}.$$

Näide: grammatika $(\{a, b, c, A, C, P, Q, S\}, \{a, b, c\}, \{S \rightarrow AP, S \rightarrow QC, A \rightarrow a, A \rightarrow Aa, P \rightarrow bc, P \rightarrow bPc, C \rightarrow c, C \rightarrow Cc, Q \rightarrow ab, Q \rightarrow aQb\}, S)$ genereerib keele $\{a^m b^n c^n \mid m, n \geq 1\} \cup \{a^m b^m c^n \mid m, n \geq 1\}$.

Grammatika on *kontekstivaba*, kui tema reeglite vasakud pooled koosnevad ühestainsast mitteterminalist. Eeltoodud näitegrammatika oli kontekstivaba.

Kontekstivaba grammatika on *Chomsky normaalkujul*, kui tema reeglite paremad pooled koosnevad kas täpselt kahest mitteterminalist või täpselt ühest terminalist. Eeltoodud grammatika ei olnud Chomsky normaalkujul. Grammatika, mis genereerib sama keele kui näitegrammatika ja on Chomsky normaalkujul, on näiteks $(\{a, b, c, T_a, T_b, T_c, A, C, P, Q, N_1, N_2, S\}, \{a, b, c\}, \{S \rightarrow AP, S \rightarrow QC, T_a \rightarrow a, T_b \rightarrow b, T_c \rightarrow c, A \rightarrow a, A \rightarrow AT_a, P \rightarrow T_b T_c, P \rightarrow T_b N_1, N_1 \rightarrow PT_c, C \rightarrow c, C \rightarrow CT_c, Q \rightarrow T_a T_b, Q \rightarrow T_a N_2, N_2 \rightarrow QT_b\}, S)$. Muuhulgas kehtib teoreem, et kui G on kontekstivaba grammatika ja $\varepsilon \notin L(G)$, siis leidub Chomsky normaalkujul olev kontekstivaba grammatika G' nii, et $L(G) = L(G')$.

Ülesanne. Kirjutada programm, mille sisenditeks on Chomsky normaalkujul kontekstivaba grammatika $G = (V, \Sigma, P, s)$ ning sõne $x \in \Sigma^*$. Programm peab kindlaks tegema, kas $x \in L(G)$.

Vihje. Vaata alamülesandeid kujul „kas $N \Rightarrow^* x[i \dots j]$?“ Siin N on mingi mitteterminal ja $1 \leq i \leq j \leq |x|$.