Automaták és Formális Nyelvek – XII előadás

(2020 május 14)

(A) Rekurzív és rekurzívan felsorolható nyelvek zártsági tulajdonságai

1. Állítás (i) Ha $L \subseteq \Sigma_1^*$ és rekurzív nyelv, akkor $\overline{L}$ is az.
(ii) Ha $L_1, L_2 \subseteq \Sigma^*$ rekurzív nyelvek, akkor $L_1 \cap L_2$, $L_1 \cup L_2$ és $L_1 \cdot L_2$ is rekurzív nyelv.

Bizonyítás (i) Azt tanultuk hogy L pontosan akkor rekurzív nyelv, ha Σ_1^*-nak egy eldönthető részhalmaza, azaz ha van olyan algoritmus ami minden $w \in \Sigma_1^*$ szó esetén véges lépésben eldönti hogy w az L nyelvhez ~~ta~~tartozik-e vagy sem. De akkor az algoritmus ezzel azt is eldönti véges lépésben, hogy $w \in \overline{L}$ vagy $w \notin \overline{L}$. Tehát $\overline{L}$ is rekurzív nyelv.

(ii) Ha L_1 és L_2 rekurzív nyelvek, akkor a tanultak értelmében létezik egy olyan A_1 és egy olyan A_2 algoritmus ami minden $w \in \Sigma_1^*$-szóról véges lépésben eldönti hogy az L_1 nyelvhez, illetve az L_2 nyelvhez tartozik-e. Akkor a két algoritmust minden szóra egymás után alkalmazva, ugyancsak véges lépésben eldönthető ~~hogy~~ egy $w \in \Sigma_1^*$ szó az $L_1 \cap L_2$ nyelvhez, illetve egy $w_1 \cdot w_2$ szó az $L_1 \cdot L_2$ nyelvhez tartozik-e vagy sem. Tehát $L_1 \cap L_2$ és $L_1 \cdot L_2$ a Σ^* eldönthető részhalmazai, azaz rekurzív nyelvek. Mivel (i) szerint $\overline{L_1}$, $\overline{L_2}$ is rekurzív nyelvek, $L_1 \cup L_2 = \overline{\overline{L_1} \cap \overline{L_2}}$ ugyancsak rekurzív nyelv a fentiek és (i) szerint. Tehát (ii) minden állítása teljesül. □

Következmény 1 (i) Ha $L_1, L_2 \subseteq \Sigma^*$ rekurzív nyelvek, akkor $L_1 \setminus L_2$ is az.
(ii) Ha $L \subseteq \Sigma^*$ rekurzív nyelv, akkor minden $n \geq 1$-re L^n is az.

Bizonyítás. Az első állítás szerint $\overline{L_2}$ rekurzív. Mivel $L_1 \setminus L_2 = L_1 \cap \overline{L_2}$, ezért az Állítás 1(ii) szerint $L_1 \setminus L_2$ is rekurzív nyelv.
(ii) az Állítás 1(ii) felhasználásával n-szerinti teljes indukcióval bizonyítható: Ha ugyanis $L^n \subseteq \Sigma^*$ rekurzív nyelv, akkor az Állítás 1(ii) szerint $L^{n+1} = L^n \cdot L$ is rekurzív. □

A következő állítást legegyszerűbb módon ún. generatív nyelvtanok segítségével bizonyíthatjuk, amiket a köv. alfejezetben mutatunk be.

Állítás 2. Ha $L_1, L_2 \subseteq \Sigma^*$ rekurzívan felsorolható nyelvek, akkor $L_1 \cup L_2$ és $L_1 \cdot L_2$ is rekurzívan felsorolhatóak.

1. Tétel. Ha $L \subseteq \Sigma^*$ egy olyan rekurzívan felsorolható nyelv, aminek az $\overline{L} \subseteq \Sigma^*$ komplementuma is rekurzívan felsorolható, akkor L (és $\overline{L}$) egy rekurzív nyelv.

Biz. A rekurzívan felsorolható nyelvek definíciója szerint létezik egy olyan M_1 és egy olyan M_2 Turing gép, amelyekre $L(M_1) = L$ és $L(M_2) = \overline{L}$. Tekintsük most azt az A algoritmust ami azt jelenti hogy egy $w \in \Sigma^*$ szót egyszerre vizsgálunk meg mind az M_1 mind az M_2 géppel, úgy hogy váltva lépünk mindkettő szalagjára. Mivel minden egyes szó vagy L-hez, vagy $\overline{L}$-hez tartozik, ezért az A algoritmus minden $w \in \Sigma^*$ szó esetén véges lépésben megáll, méghozzá úgy, hogy vagy a $w \in L$ vagy a $w \in \overline{L}$ eredménnyel végződik. Valóban, vagy $w \in L$ és akkor M_1 véges lépés után elfogadó állapotban megáll és A algoritmus véget ér, vagy $w \in \overline{L}$ és akkor M_2 áll meg véges lépés után elfogadó állapotban és A ugyancsak véget ér. Így mind L mind $\overline{L}$ eldönthető részhalmazai Σ^*-nak, tehát rekurzív nyelvek. □

Az alábbi következmény nyilvánvaló:

<u>Következmény 2.</u> Ha $L \subseteq \Sigma^*$ egy olyan rekurzívan felsorolható nyelv, ami nem rekurzív, akkor $\bar{L}$ nem rekurzívan felsorolható.

<u>Következmény 3.</u> Van olyan $L \subseteq \Sigma^*$ nyelv ami még csak nem is rekurzívan felsorolható.

<u>Bizonyítás.</u> Az előző előadáson, Turing eredményének a bizonyításakor láttuk, hogy az az $L \subseteq \Sigma^* = (0+1)^*$ nyelv, ami az összes standard Turing gép bináris kódjait tartalmazza egy rekurzívan felsorolható de nem rekurzív nyelv. De akkor a <u>Következmény 2</u>-nek megfelelően, $\bar{L}$ még csak rekurzívan felsorolható nyelv sem lehet. □

(B) <u>Generatív nyelvtanok</u>

Olyan $G=(V,T,P,S)$ alakban felírt („nyelvtanok") amelynek produkciós szabályai $\gamma_1 \to \gamma_2$ alakúak, ahol $\gamma_1, \gamma_2 \in (T \cup V)^*$.

Példák: 1)
$$S \to 0A1, \quad 0A \to 00A1, \quad A \to \varepsilon$$

2)
$$S \to aCaB, \quad Ca \to aaC, \quad CB \to DB, \quad CD \to E,$$
$$aD \to Da, \quad AD \to AC, \quad AE \to Ea, \quad AE \to \varepsilon$$

A generatív nyelvtan fogalma a környezetfüggetlen nyelvtanénál sokkal általánosabb hiszen egy környezetfüggetlen nyelvtan produkciós szabályai csak $A \to \alpha$ alakúak lehetnek, ahol $A \in V$ és $\alpha \in (T \cup V)^*$. Például az 1) nyelvtanból a $0A \to 00A1$ szabály, vagy a 2) nyelvtanból a $Ca \to aaC$ szabályok nem szerepelhetnek egy k. független nyelvtanban!

Azt mondjuk hogy G segítségével a $\gamma \in (V \cup T)^*$ szavakból (közvetlenül) deriválható a $\delta \in (V \cup T)^*$ szavak, ha van olyan <u>$\alpha \to \beta \in P$</u> produkciós szabály és olyan $\sigma, \nu \in (T \cup V)^*$ szavak, hogy $\gamma = \sigma\alpha\nu$ és $\delta = \sigma\beta\nu$ alakba írható. Ezt a fajta derivációt a $\gamma \Rightarrow \delta$ írásmóddal jelöljük, akárcsak a k. független nyelvtanoknál. Pl. az 1) nyelvtan egy közvetlen derivációja

$1\underline{0A}A0 \Rightarrow 1\underline{00A1}A0$ (itt $\sigma=1, \gamma=A0$), vagy a 2)-es nyelvtan esetén az $A\underline{Ca}B \Rightarrow A\underline{aaC}B$ deriváció.

$\alpha \overset{*}{\Rightarrow} \beta$-val jelöljük, ha az α szóból a β szót megkapható 0 vagy több (de véges számú) derivációs lépéssel.
A G nyelvtan által előállított $L(G)$ nyelv alatt mindazon $w \in T^*$ szavak összességét értjük amelyek megkaphatók az S „startszimbólumból" véges lépésben, azaz

$$L(G) = \{w \in T^* \mid S \overset{*}{\Rightarrow} w\}$$

A generatív nyelvtanokat <u>0-típusú</u> nyelvtanoknak is nevezzük, az általuk előállított nyelveket pedig <u>0-típusú nyelveknek.</u>

Bizonyítás nélkül ismertetjük az alábbi eredményt:

<u>Tétel 2</u> Az $L \subseteq \Sigma^*$ nyelv akkor és csak akkor állítható ~~elő~~ elő az 0-típusú nyelvtannal ha rekurzíva felsorolható.

A tétel értelmében a 0-típusú nyelvek nem mások mint a rekurzíva felsorolható nyelvek - azaz a Turing gépek által előállítható nyelvek.

<u>Definíció 1</u> A 0-típusú nyelvek azon sajátos osztályát amelyek megadhatók olyan G generatív nyelvtannal amelyeknek minden produkciós szabálya: $\gamma_1 \to \gamma_2$ olyan hogy γ_2 legalább olyan hosszú mint γ_1 (azaz $|\gamma_1| \le |\gamma_2|$) <u>1-típusú nyelveknek</u>, vagy környezetérzékeny nyelveknek nevezzük. A megfelelő nyelvtanok pedig az 1-típusú nyelvtanok.

<u>Az elnevezés eredete</u>: Ezeknél a nyelvtanoknál a szabályai mind $\alpha_1 A \alpha_2 \to \alpha_1 B \alpha_2$ alakba írhatók (ahol $B \neq \varepsilon$). Egy ilyen szabályt úgy is értelmezhetjük hogy az $A \to B$ behelyettesítés csak az α_1, α_2 „környezetben" végezhető el.

<u>Állítás 3</u> Minden környezetérzékeny nyelv rekurzív (de nem minden rekurzív nyelv környezetérzékeny!)

Példák környezetfüggő nyelvekre: a beszélt nyelvek, magyar, angol, stb., szaknyelvek: pl. a mechanikai szaknyelv.

Lineárisan korlátos Turing gép (TK)

Ez egy olyan $TK=(Q, \Sigma, \Gamma, \delta, q_0, ¢, \$, B, F)$ Turing gép, ahol Σ tartalmaz két speciális szimbólumot ¢-t és \$-t, amik a bal és jobb széljelző szerepét töltik be az input szalagon. Sem ¢-t, sem \$ nem helyettesíthető egy elfogadott szó részeként – ezek a jelek nem felülírhatók és nem törölhetők.
A TK akkor fogad el egy $w \in \Sigma^*$ szót ha a $q_0 ¢ w \$$ ~~helyzetből~~ helyzetből véges lépés után a $¢ \alpha q \$$ helyzetbe kerül és ott megáll, ahol q végállapot. Azaz

$$q_0 ¢ w \$ \vdash^* ¢ \alpha q \$.$$

Tétel 3 $L \subseteq \Sigma^*$ akkor és csak akkor egy környezetérzékeny nyelv ha valamilyen korlátos TK gép elfogadja.

A tételből a 3. Állítás a következik. Ez korlátos TK gép szalagja ugyanis minden csak véges konfigurációs ciklust tartalmazhat. Így véges számú lépés után megadható hogy a Turing gép végtelen ismétlődő ciklusba került-e vagy sem. Ezért az w véges lépésben eldönthető hogy milyen szavakat fogad el – vagyis a nyelv rekurzív.

A környezetfüggetlen nyelvtanokat 2-es típusú nyelvtanoknak is nevezik. Mivel a környezetfüggetlen nyelvtanok tartalmazhatnak $A \to \varepsilon$ típusú szabályt, ahol $|\varepsilon|=0$, $|A|=1$, nem minden 2-es típusú nyelvtan 1-es típusú is és nem minden k. független nyelv környezetérzékeny is. Mivel minden olyan L k. független ~~nyelvtan~~ nyelv ami nem tartalmaz ε-t előállítható egy olyan k. f. nyelvtannal amiben nincs $A \to \varepsilon$ alakú szabály, igaz az alábbi:

Állítás 4. Ha $L \subseteq \Sigma^*$ környezetfüggetlen nyelv nem tartalmaz ε-t akkor L környezetérzékeny nyelv is.

<u>Definíció 2</u> A G k. független nyelvtant <u>jobb</u> (<u>bal</u>) <u>lineárisnak</u> nevezzük, ha minden produkciós szabálya

$$\begin{matrix} A \to wB \\ A \to w \end{matrix} \quad \left(\text{illetve} \quad \begin{matrix} A \to Bw \\ A \to w \end{matrix}\right) \text{ alakú.}$$

Közös néven a bal és jobblineáris nyelvtanokat <u>reguláris</u> nyelvtanoknak nevezzük.

Például az $\begin{matrix} S \to 0A \\ A \to 10A \mid \varepsilon \end{matrix}$ nyelvtan jobbreguláris és a $0(10)^*$ nyelvet állítja elő.

<u>Tétel 3</u>. Az L nyelv reguláris $\Longleftrightarrow$ Ha van olyan $G=(V,T,P,S)$ reguláris nyelvtan ami L-t előállítja (azaz $L(G)=L$).

(c) Hierarchia tétel (Chomsky)

A bevezetett 0-típusú, 1-típusú, 2 típusú és 3-típusú nyelvek osztályait Chomsky-féle nyelvosztályoknak is nevezzük és a továbbiakban őket rendre az $\mathcal{L}_0$, $\mathcal{L}_1$, $\mathcal{L}_2$ és $\mathcal{L}_3$ szimbólumokkal jelöljük. Az előbbi definíciók és összefüggések értelmében:

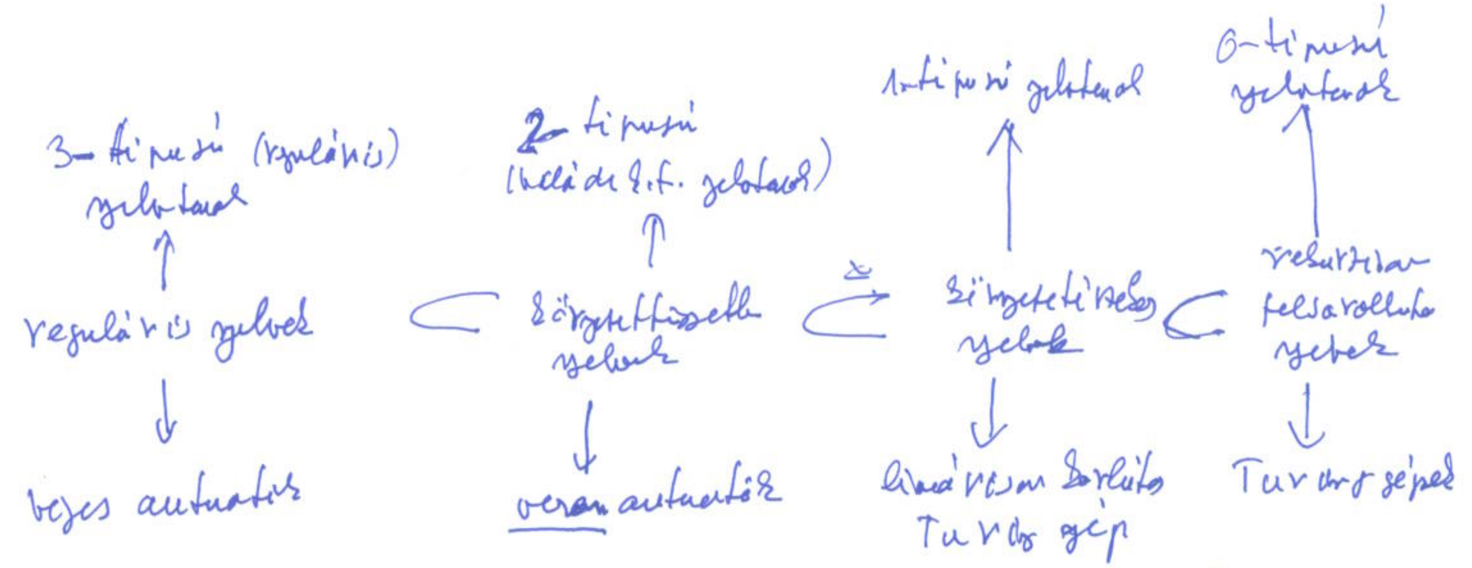
$$\mathcal{L}_3 \subseteq \mathcal{L}_2 \overset{*}{\subseteq} \mathcal{L}_1 \subseteq \mathcal{L}_0,$$

ahol a * jel arra utal hogy $\mathcal{L}_1$ csak az olyan $L \in \mathcal{L}_2$ k.f. nyelveket tartalmazza amelyeknek ε nem eleme.
Valójában, ebben a sorozatban mindenütt valódi tartalmazások is írhatunk:

$$\mathcal{L}_3 \subset \mathcal{L}_2 \overset{*}{\subset} \mathcal{L}_1 \subset \mathcal{L}_0$$

Valóban, láttuk hogy van olyan k.f nyelv $\mathcal{L}_2$-ben ami nem reguláris, így nem tartozik $\mathcal{L}_3$-hoz. Az $\mathcal{L}_2 \overset{*}{\subset} \mathcal{L}_1$ tartalmazásról már beszéltünk és azt is megmutattuk hogy van olyan $\mathcal{L}_0$-beli - tehát rekurzívan felsorolható nyelv ami <u>nem</u> rekurzív, tehát nem lehet $\mathcal{L}_1$-ben hiszen $\mathcal{L}_1$-ben csak rekurzív nyelvek vannak.

A nyelvosztályoknak ezt a növekvő sorozatát nevezzük Chomsky-féle hierarchiának. Minden nyelvosztályt egyértelműen meghatároz egy nyelvtani típus és minden nyelvosztálynak egyértelműen megfeleltethetők az automatatípusok. Ezt ábrázoltuk az alábbi diagramon:



Ez a hierarchia tovább is ~~finomítható~~ finomítható, hiszen tudjuk hogy L_2-nek, a k. független nyelvek osztályának valódi részhalmazát képezik azok a nyelvek amelyek determinisztikus veremautomatákkal felismerhetők, ennek pedig valódi részhalmaza a reguláris nyelvek osztálya. Hasonlóan, az is megmutatható hogy az 1-típusú nyelvek a rekurzív nyelveknek egy valódi részhalmazát képezik. Ezt ábrázoltuk az alábbi rajzon.

* Például a determinisztikus veremautomatákkal felismerhető nyelvek osztályába tartoznak az ún. LR(k) nyelvek is – ide sorolhatók be a különböző programozási nyelvek is.