Számítási modellek

3: Alapok, Grammatikák, L-rendszerek

Alapok, terminológia

- **Ábécé**: szimbólumok/betűk egy véges, nemüres halmaza.
- **Szavak** vagy **sztringek** egy V ábécé felett: Egy V ábécé elemeiből képzett véges sorozatok.
- Egy $u = t_1 \ldots t_n$ **szó hossza**: u-ban lévő betűk száma n.
  Jelölés: $|u| = n$.
- Az **üres szó** ε: a 0 hosszú szó ($|\varepsilon| = 0$).
- V^* : szavak halmaza V felett, amely tartalmazza az üres szót (ε) is.
- $V^+ = V^* \setminus \{\varepsilon\}$: nemüres szavak halmaza V felett.

- Példa:
  Legyen $V = \{a,b\}$, ekkor *ab* és *baaabb* szavak V felett.

Alapok, terminológia

- Legyen V egy ábécé és legyenek u és v V feletti szavak. Ekkor az uv szót az u és v szavak **konkatenált**jának nevezzük. (Az u szó betűi után írjuk a v szó betűit.)
- $|uv| = |u| + |v|$.

- Példa:
  Legyen $V = \{a,b\}$, $u = abb$ és $v = aaab$.
  Ekkor $uv = abbaaab$, $vu = aaababb$.

Alapok, terminológia

Tulajdonságok:

- A konkatenáció asszociatív, de általában nem kommutatív.
    - ha $u, v \in V^*$, $u \neq v$, akkor uv különbözik vu szótól, kivéve ha V egyetlen betűt tartalmaz (nem kommutatív).
    - ha $u, v, w \in V^*$, akkor $u(vw) = (uv)w$ (asszociatív).
- V^* **zárt** a konkatenáció műveletre (azaz bármely $u, v \in V^*$ esetén teljesül $uv \in V^*$).
- A konkatenáció egységelemes művelet, az **egységelem** az ε (azaz minden $u \in V^*$ esetén teljesül $u = u\varepsilon = \varepsilon u$).

Alapok, terminológia

- Legyen *i* nemnegatív egész szám és legyen *u* egy *V* ábécé feletti szó ($u \in V^*$). Az *u* szó ***i*-edik hatványa** u^i az *u* szó *i* darab példányának konkatenáltja.
- Konvenció: $u^0 = \varepsilon$.
- Példa:
  Legyen $V = \{a,b\}$ és $u = abb$ egy szó *V* felett.
  Ekkor $u^0 = \varepsilon$, $u^1 = abb$, $u^2 = abbabb$, $u^3 = abbabbabb$, ...

Alapok, terminológia

- Legyenek u és v szavak V felett. A szavak u és v **azonosak**, ha mint szimbólumsorozatok elemről-elemre megegyeznek, azaz, $|u|=|v|$ ha minden $i = 1,\dots,|u|$, az u szó i-edik betűje megegyezik a v szó i-edik betűjével.
- Legyen V egy ábécé és legyenek u és v szavak V felett.
  Az u **részszava** v-nek,
  ha $v = xuy$ teljesül valamely $x,y \in V^*$ szavakra.
- Az u szó **valódi részszava** v-nek, ha x és y közül legalább az egyik nem üres, vagyis ha $xy \neq \varepsilon$.
- Ha $x = \varepsilon$, akkor az u szó a v szó **prefixe**.
- Ha $y = \varepsilon$, akkor az u szó a v szó **szuffixe**.

Alapok, terminológia

- Példa:
  Legyen $V = \{a,b\}$ és $u = abb$.
  - Az u részszavai: $\varepsilon, a, b, ab, bb, abb$.
  - Az u valódi részszavai: $\varepsilon, a, b, ab, bb$.
  - Az u prefixei: ε, a, ab, abb.
  - Az u szuffixei: ε, b, bb, abb.

Alapok, terminológia

- Legyen *u* egy szó *V* felett. Az *u* szó **tükörképe** vagy **fordítottja** az a u^{-1} szó, amelyet úgy kapunk, hogy *u* szimbólumait fordított sorrendben írjuk le.
- Legyen $u = a_1 \ldots a_n$, $a_i \in V$, $1 \leq i \leq n$. Ekkor $u^{-1} = a_n \ldots a_1$.
- $(u^{-1})^{-1} = u.$
- $(u^{-1})^i = (u^i)^{-1}$ teljesül, $i = 1, 2, \ldots$

- Példa:
  Legyen $V = \{a,b\}$, $u = abba$ és $v = aabbba$
  Ekkor $u^{-1} = abba$ (palindrom) és $v^{-1} = abbbaa$.

Alapok, terminológia

- Legyen V egy ábécé és L egy tetszőleges részhalmaza V^*-nek. Ekkor L-t egy V feletti nyelvnek nevezzük.
- Az **üres nyelv** (egy olyan nyelv, amely nem tartalmaz egyetlen szót sem) jelölése $\emptyset$.
- Egy V ábécé feletti L nyelv **véges nyelv**, ha véges számú szót tartalmaz, egyébként **végtelen**.

- Példa:
  Legyen $V = \{a,b\}$ egy ábécé.

  $L_1 = \{a, b, \varepsilon\}$.

  $L_2 = \{a^i b^i \mid i \geq 0\}$.

  $L_3 = \{uu^{-1} \mid u \in V^*\}$.

  $L_4 = \{(a^n)^2 \mid n \geq 1\}$.

  $L_5 = \{u \mid u \in \{a, b\}^+ , N_a(u) = N_b(u)\}$, ahol $N_a(u)$ és $N_b(u)$ az a és a b szimbólumok u-beli előfordulásainak számát jelöli.

  L_1 egy véges nyelv, a többi végtelen.

Alapok, terminológia

- Egy **generatív grammatika** G egy (N, T, P, S) 4-es, ahol
  - N és T diszjunkt véges ábécék (azaz $N \cap T = \varnothing$).
  - N elemeit **nemterminális** szimbólumoknak nevezzük
  - T elemeit **terminális** szimbólumoknak nevezzük.
  - $S \in N$ a **kezdő szimbólum** (axióma).
  - P rendezett (x,y) párok véges halmaza, ahol $x,y \in (N \cup T)^*$ és x legalább egy nemterminális szimbólumot tartalmaz.
  - P elemeit **átírási szabályoknak** (**produkciós szabályoknak**) vagy röviden **szabályoknak** nevezik. $x \rightarrow y$ használható (x,y) helyett, ahol $\rightarrow \notin (N \cup T)$.

Alapok, terminológia

- Példa:
  - $G_1 = (\{S, A, B\}, \{a, b, c\}, \{S \to c, S \to AB, A \to aA, B \to \varepsilon, abb \to aSb\}, S)$ nem generatív grammatika.
  - $G_2 = (\{S, A, B, C\}, \{a, b, c\}, \{S \to a, S \to AB, A \to Ab, B \to \varepsilon, aCA \to aSc\}, S)$ generatív grammatika.

Alapok, terminológia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika és $u,v \in (N \cup T)^*$. A v szó **közvetlenül** vagy **egy lépésben levezethető** az u szóból G-ben, ha $u = u_1 x u_2$ és $v = u_1 y u_2$, ahol $u_1,u_2 \in (N \cup T)^*$ és $x \to y \in P$. Jelölés: $u \Rightarrow_G v$.

- Legyen $G = (N,T,P,S)$ egy generatív grammatika és $u\ v \in (N \cup T)^*$. A v szó **levezethető** az u szóból G-ben, jelölés $u \Rightarrow^*_G v$,
  - ha $u = v$, vagy
  - van olyan szó $z \in (N \cup T)^*$, amelyre $u \Rightarrow^*_G z$ és $z \Rightarrow_G v$.
  - $\Rightarrow^*$ a $\Rightarrow$ reflexív és tranzitív lezártja.
  - $\Rightarrow^+$ a $\Rightarrow$ tranzitív lezártja.

Alapok, terminológia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika és $u,v \in (N \cup T)^*$.
  A v szó ***k* lépésben levezethető** u szóból G-ben, $k \geq 1$, ha van olyan szavakból álló sorozat $u_1, \ldots, u_{k+1} \in (N \cup T)^*$, amelyre $u=u_1$, $v=u_{k+1}$, és $u_i \Rightarrow_G u_{i+1}$, $1 \leq i \leq k$.
- A v szó **levezethető** u szóból G-ben, ha
  - vagy $u = v$,
  - vagy van olyan $k \geq 1$ szám, amelyre v levezethető u-ból k lépésben.

Alapok, terminológia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika.
  A G által **generált nyelv** $L(G)$:
  $L(G) = \{w \mid S \Rightarrow^*_G w, \; w \in T^*\}$

- Azaz, $L(G)$ olyan szavakból áll, amelyek levezethetők S-ből és T^*-beliek.

Alapok, terminológia

- Példa:
  Legyen $G = (N,T,P,S)$ egy generatív grammatika, ahol
  $N = \{S,A,B\}$, $T = \{a,b\}$,
  $P = \{S \rightarrow aSb,\ S \rightarrow ab,\ S \rightarrow ba\}$.
  Ekkor $L(G) = \{a^n abb^n\ ,\ a^n bab^n \mid n \geq 0\}$.

- Példa:
  Legyen $G = (N,T,P,S)$ egy generatív grammatika, ahol
  $N = \{S,X,Y\}$, $T = \{a,b,c\}$,
  $P = \{S \rightarrow abc,\ S \rightarrow aXbc,\ Xb \rightarrow bX\ ,\ Xc \rightarrow Ybcc,\ bY \rightarrow Yb,$
  $aY \rightarrow aaX\ ,\ aY \rightarrow aa\}$.
  Ekkor $L(G) = \{a^n b^n c^n \mid n \geq 1\}$.

Alapok, terminológia

- Minden grammatika generál egy nyelvet, de ugyanaz a nyelv több különböző grammatika által is generálható.

- Két grammatika **ekvivalens**, ha ugyanazt a nyelvet generálják.
- Két nyelv **gyengén ekvivalens**, ha legfeljebb az üres szóban különböznek.

Chomsky hierarchia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika.
  A G grammatika i-típusú, i = 0, 1, 2, 3, ha P szabályhalmazra a következők teljesülnek:
  - i = 0: nincs korlátozás.
  - i = 1: P minden szabálya $u_1Au_2 \rightarrow u_1vu_2$ alakú, ahol $u_1, u_2, v \in (N \cup T)^*$, $A \in N$, és $v \neq \varepsilon$, kivéve a $S \rightarrow \varepsilon$ szabályt, ha létezik ilyen szabály P-ben.
    Ha P tartalmazza az $S \rightarrow \varepsilon$ szabályt, akkor S nem fordul elő egyetlen szabály jobb oldalán sem.
  - i = 2: P minden szabálya $A \rightarrow v$ alakú, ahol $A \in N$ és $v \in (N \cup T)^*$.
  - i = 3: P minden szabálya vagy $A \rightarrow uB$ vagy $A \rightarrow u$ alakú, ahol $A, B \in N$ és $u \in T^*$.

Chomsky hierarchia

- Egy L **nyelv *i*-típusú**, $i = 0, 1, 2, 3$, ha i-típusú grammatikával generálható.
- $\mathcal{L}_i$, $i = 0, 1, 2, 3$, jelöli az i-típusú nyelvek osztályát (családját).

Chomsky hierarchia

- 0-típusú grammatikákat **mondatszerkezetű** (**phrase-structured**) grammatikáknak is nevezzük.
- 1-típusú grammatikák a **környezetfüggő** (**context-sensitive**) grammatikák. Egy *A* nemterminális valamely előfordulása *v* szóval csak u_1 és u_2 kontextus jelenlétében helyettesíthető.
- 2-típusú grammatikákat **környezetfüggetlen** (**context-free**) grammatikáknak nevezzük. Egy *A* nemterminális *v*-vel való helyettesítése bármely kontextusban megengedett.
- 3-típusú grammatikákat **reguláris** (**regular**) vagy **véges állapotú** (**finite state**) grammatikáknak hívjuk, a véges állapotú automatákkal való kapcsolatuk miatt.

- A 0,1,2,3-típusú nyelvek osztályait rendre **rekurzívan felsorolható**, **környezetfüggő**, **környezetfüggetlen**, valamint **reguláris nyelvosztály**nak nevezzük.

Chomsky hierarchia

Nyelvészeti háttér

”The cunning fox hastily ate the leaping frog.”

- $S \rightarrow A + B$ (S: sentence, A: noun phrase, B: verb phrase)
- $A \rightarrow C + D + E$ (C : article, D: adjective, E : noun)
- $B \rightarrow G + B$ (G : adverb)
- $B \rightarrow F + A$ (F : verb)
- $C \rightarrow$ the
- $D \rightarrow$ cunning
- $E \rightarrow$ fox
- $G \rightarrow$ hastily
- $F \rightarrow$ ate
- $D \rightarrow$ leaping
- $E \rightarrow$ frog

Chomsky hierarchia

Nyelvészeti háttér

- \+ (space)
- cunning – leaping , fox – frog (felcserélhető, de más lesz a jelentés)
- szintaktikailag helyes mondat

Chomsky hierarchia

- Nyilvánvaló, hogy $\mathcal{L}_3 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_0$ and $\mathcal{L}_1 \subseteq \mathcal{L}_0$.
- Megmutatható, hogy a következő is fennáll (Chomsky hierarchia): $\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0$.

- Az $\mathcal{L}_2$ és $\mathcal{L}_1$ nyelvosztály közötti tartalmazási reláció nem látható azonnal a grammatikák definíciójából.
- Szintén $\mathcal{L}_1$-et generálják az u.n. hossznemcsökkentő grammatikák. Ezek $p \to q$ szabályaira $|p| \leq |q|$ teljesül, kivéve az $S \to \varepsilon$ alakú szabályt, feltéve, hogy P-ben létezik ilyen szabály.
  Ha $S \to \varepsilon \in P$, akkor S nem fordul elő P egyetlen szabályának jobb oldalán sem.

Nyelvekre vonatkozó műveletek

- Legyen V egy ábécé és legyenek L_1, L_2 nyelvek V felett (azaz, $L_1 \subseteq V^*$, $L_2 \subseteq V^*$)
  - **unió:** $L_1 \cup L_2 = \{u \mid u \in L_1 \text{ vagy } u \in L_2\}$.
  - **metszet:** $L_1 \cap L_2 = \{u \mid u \in L_1 \text{ és } u \in L_2\}$.
  - **különbség:** $L_1 - L_2 = \{u \mid u \in L_1 \text{ és } u \notin L_2\}$.

- Példa:

  Legyen $V = \{a,b\}$ egy ábécé és $L_1 = \{a,b\}$ és $L_2 = \{\varepsilon,a,bbb\}$ nyelvek V felett. Ekkor

  $L_1 \cup L_2 = \{\varepsilon, a, b, bbb\}$

  $L_1 \cap L_2 = \{a\}$

  $L_1 - L_2 = \{b\}$

Nyelvekre vonatkozó műveletek

- Az $L \subseteq V^*$ nyelv **komplementere** a V-re az $\overline{L} = V^* - L$ nyelv.
- Példa:
  Legyen $V = \{a\}$ egy ábécé és legyen $L = \{a^{4n} \mid n \geq 0\}$.
  Ekkor $\overline{L} = V^* - \{a^{4n} \mid n \geq 0\}$.

Nyelvekre vonatkozó műveletek

- Legyen V egy ábécé és L_1, L_2 nyelvek V felett (azaz, $L_1 \subseteq V^*$, $L_2 \subseteq V^*$). L_1 és L_2 **konkatenációja** $L_1L_2 = \{u_1u_2 \mid u_1 \in L_1 , u_2 \in L_2 \}$.

- Megjegyzés:
  A következő egyenlőségek minden L nyelvre érvényesek:
  - $\emptyset L = L\emptyset = \emptyset$ és
  - $\{\varepsilon\}L = L\{\varepsilon\} = L$.

Nyelvekre vonatkozó műveletek

- L^i jelöli az L nyelv ***i*-edik iterációját** (a konkatenáció műveletre nézve), ahol $i \geq 1$. Konvenció: $L^0 = \{\varepsilon\}$.
- Az L nyelv **iteratív lezártja** (vagy **Kleene lezártja**): $L^* = \bigcup_{i \geq 0} L^i$.
- L **pozitív lezártja**: $L^+ = \bigcup_{i \geq 1} L^i$.

- Megjegyzés:
  Ha $\varepsilon \in L$, akkor $L^+ = L^*$. Máskülönben, $L^+ = L^* - \{\varepsilon\}$.

Nyelvekre vonatkozó műveletek

- Példa (konkatenáció):
  Legyen $V = \{a, b\}$ és $L_1 = \{a, b\}$, $L_2 = \{\varepsilon, a, bbb\}$,
  $L_3 = \{a^{4n}b^{4n} \mid n \geq 0\}$ és $L_4 = \{a^{7n}b^{7n} \mid n \geq 0\}$. Ekkor
  - $L_1L_2 = \{a, b, aa, ba, abbb, bbbb\}$,
  - $L_3L_4 = \{a^{4n}b^{4n}a^{7m}b^{7m} \mid n \geq 0, m \geq 0\}$.

Nyelvekre vonatkozó műveletek

- Legyen V egy ábécé és $L \subseteq V^*$. Ekkor az L nyelv **tükörképe** (vagy **fordítottja**) $L^{-1} = \{u^{-1} \mid u \in L\}$.

- Megjegyzések:
  - $(L^{-1})^{-1} = L$,
  - $(L_1 L_2 \ldots L_n)^{-1} = L_n^{-1} \ldots L_2^{-1} L_1^{-1}$,
  - $(L^i)^{-1} = (L^{-1})^i$, ahol $i \geq 0$,
  - $(L^*)^{-1} = (L^{-1})^*$.

Nyelvekre vonatkozó műveletek

- Példa (tükör, fordított):
  Legyen $V = \{a,b\}$ és $L = \{\varepsilon, a, abb\}$ egy nyelv V felett.
  Ekkor $L^{-1} = \{\varepsilon, a, bba\}$.

Nyelvekre vonatkozó műveletek

- Egy $L \subseteq V^*$ nyelv **prefix nyelve**
  $\mathrm{PRE}(L) = \{ u \mid u \in V^*, uv \in L$ valamely $v \in V^*$-re $\}$.

- Megjegyzés:
  Definíció szerint, $L \subseteq \mathrm{PRE}(L)$ minden $L \in V^*$ nyelvre.

- Egy $L \subseteq V^*$ nyelv **szuffix nyelve**
  $\mathrm{SUF}(L) = \{ u \mid u \in V^*, vu \in L$ valamely $v \in V^*$-re $\}$.

Nyelvekre vonatkozó műveletek

- Legyen V_1 és V_2 két ábécé. A $h : V_1^* \rightarrow V_2^*$ leképezést **homomorfizmusnak** nevezzük, ha a következő feltételek teljesülnek:
  - minden $u \in V_1^*$ szóra pontosan egy $v \in V_2^*$ szó létezik, amelyre $h(u) = v$.
  - $h(uv) = h(u)h(v)$, minden $u, v \in V_1^*$-ra.

- Megjegyzések:
  - A fenti feltételekből következik, hogy $h(\varepsilon) = \varepsilon$.
    Valamint, minden $u \in V_1^*$-ra $h(u) = h(\varepsilon u) = h(u\varepsilon)$.
  - Minden $u = a_1a_2 \ldots a_n$, $a_i \in V_1$, $1 \leq i \leq n$, szóra teljesül, hogy $h(u) = h(a_1)h(a_2) \ldots h(a_n)$.
    Ez azt jelenti, hogy elegendő a h leképezést V_1 elemeire megadni, ez automatikusan kiterjesztődik V_1^*-ra.

Nyelvekre vonatkozó műveletek

- Egy homomorfizmus $h : V_1^* \rightarrow V_2^*$ **ε-mentes** ha minden $u \in V_1^+$-ra $h(u) \neq \varepsilon$.

- Legyen $h : V_1^* \rightarrow V_2^*$ egy homomorfizmus.
Egy $L \subseteq V_1^*$ nyelv ***h*-homomorf képe** a következő nyelv:
$h(L) = \{w \in V_2^* \mid w = h(u), u \in L\}$

- Példa (homomorfizmus):
Legyen $V_1 = V_2 = \{a,b\}$ két ábécé. Legyen $h : V_1^* \rightarrow V_2^*$ egy homomorfizmus, ahol $h(a) = bbb$, $h(b) = ab$ és $L = \{a, abba\}$.
Ekkor $h(L) = \{bbb, bbbababbbb\}$.

Nyelvekre vonatkozó műveletek

- Egy h homomorfizmus egy **izomorfizmus**, ha teljesül:
  $\forall\, u,v \in V_1^*$: ha $h(u) = h(v)$, akkor $u = v$.

- Példa (izomorfizmus, decimális számok bináris reprezentációja):
  $V_1 = \{0, 1, 2, \ldots, 9\}$, $V_2 = \{0, 1\}$,
  $h(0) = 0000$, $h(1) = 0001$, $\ldots$, $h(9) = 1001$

Kontrollált környezetfüggetlen grammatikák

- Kérdés: Lehet-e nem környezetfüggetlen nyelveket generálni környezetfüggetlen grammatikákkal azáltal, hogy a produkciós szabályok alkalmazhatóságára feltételeket szabunk.
- Válasz: igen, pl.
  - programozott grammatikák
  - mátrixgrammatikák
  - szövegfeltételekkel adott grammatikák

Programozott Grammatika

- Egy $G = (N,T,P,S)$ **környezetfüggetlen programozott grammatika** egy 4-es, ahol
  - N és T diszjunkt véges ábécék,
  - $S \in N$ a kezdőszimbólum (axióma),
  - P rendezett 3-asok véges halmaza, amelyek $r = (p, \sigma, \varphi)$ alakúak, ahol p egy környezetfüggetlen szabály, $\sigma, \varphi \subseteq P$.
  - σ az r szabály **siker-halmaz**a, φ az r szabály **kudarc-halmaz**a.
  - Ha $r = (p, \sigma, \emptyset)$ minden $r \in P$ szabályra fennáll, akkor a G grammatika **előfordulás-ellenőrzés nélküli**, egyébként **előfordulás-ellenőrzéses**.

Programozott Grammatika

- $G = (N,T,P,S)$ környezetfüggetlen programozott grammatika
- Ha $u,v \in (N \cup T)^*$ egy levezetésben két egymást követő mondatforma (az $i-1$. és i. ahol $i \geq 0$) és az i.-ként alkalmazott szabály $r_i = (A \rightarrow w, \sigma, \varphi)$, akkor a következők egyike teljesül
  - ha $u = xAy$ valamely $x,y \in (N \cup T)^*$-ra, akkor $v = xwy$ és a levezetés i+1.-ként alkalmazott r_{i+1} szabályára (ha van) $r_{i+1} \in \sigma$ teljesül.
    (Azaz a következő következő szabály a siker-halmaz beli.)
  - ha a u nem tartalmazza A-t, akkor $v = u$ és a levezetés i+1.-ként alkalmazott r_{i+1} szabályára (ha van) $r_{i+1} \in \varphi$ teljesül.
    (Azaz a következő következő szabály a kudarc-halmaz beli.)
- Jelölés: $u \Rightarrow v$

Programozott Grammatika

- Legyen $G = (N,T,P,S)$ egy programozott grammatika.
  G által **generált nyelv** $L(G)$:

  $L(G) := \{ w \in T^* \mid S \Rightarrow^* w \}$,
  ahol $\Rightarrow^*$ a $\Rightarrow$ reláció reflexív, tranzitív lezártja

Programozott Grammatika

- Példa:
  Legyen $G = (N,T,P,S)$ egy programozott grammatika, ahol
  $N = \{S, A\}$, $T = \{a\}$, $P = \{r_1, r_2, r_3\}$, ahol
  - $r_1 = (S \to AA, \{r_1\}, \{r_2\})$,
  - $r_2 = (A \to S, \{r_2\}, \{r_1, r_3\})$,
  - $r_3 = (S \to a, \{r_3\}, \emptyset)$.

  Ekkor $L(G) = \{a^{2^n} \mid n \geq 0\}$.

Programozott Grammatika

Let $G = (\{S, A\}, \{a\}, S, \{r_1, r_2, r_3\})$, where

$$r_1 = (S \rightarrow AA, \{r_1\}, \{r_2\})$$
$$r_2 = (A \rightarrow S, \{r_2\}, \{r_1, r_3\})$$
$$r_3 = (S \rightarrow a, \{r_3\}, \emptyset)$$

The derivation for the string *aaaa* is as follows:

$$\begin{aligned} S &\Rightarrow_{r_1} AA \Rightarrow_{r_1} AA \Rightarrow_{r_2} SA \Rightarrow_{r_2} SS \Rightarrow_{r_2} SS \\ &\Rightarrow_{r_1} AAS \Rightarrow_{r_1} AAAA \Rightarrow_{r_1} AAAA \\ &\Rightarrow_{r_2} SAAA \Rightarrow_{r_2} SSAA \Rightarrow_{r_2} SSSA \Rightarrow_{r_2} SSSS \Rightarrow_{r_2} SSSS \\ &\Rightarrow_{r_3} aSSS \Rightarrow_{r_3} aaSS \Rightarrow_{r_3} aaaS \Rightarrow_{r_3} aaaa \Rightarrow_{r_3} aaaa \end{aligned}$$

As can be seen from the derivation and the rules, each time r_1 and r_2 succeed, they feed back to themselves, which forces each rule to continue to rewrite the string over and over until it can do so no more. Upon failing, the derivation can switch to a different rule. In the case of r_1, that means rewriting all *S*s as *AA*s, then switching to r_2. In the case of r_2, it means rewriting all *A*s as *S*s, then switching either to r_1, which will lead to doubling the number of *S*s produced, or to r_3 which converts the *S*s to *a*s then halts the derivation. Each cycle through r_1 then r_2 therefore either doubles the initial number of *S*s, or converts the *S*s to *a*s. The trivial case of generating *a*, in case it is difficult to see, simply involves vacuously applying r_1, thus jumping straight to r_2 which also vacuously applies, then jumping to r_3 which produces *a*.

Source: https://en.wikipedia.org/wiki/Controlled_grammar

Mátrix Grammatika

- Egy **előfordulás-ellenőrzéses környezetfüggetlen mátrix grammatika** egy 5-ös $G = (N, T, M, S, \mathcal{F})$, ahol
  - N és T diszjunkt ábécék,
  - $S \in N$ a kezdőszimbólum (axióma),
  - $M = \{m_1, m_2, \ldots, m_n\}$, $n \geq 1$, sorozatok véges halmaza $m_i = (p_{i1}, \ldots, p_{ik(i)})$, $k(i) \geq 1$, $1 \leq i \leq n$, ahol minden p_{ij}, $1 \leq i \leq n$, $1 \leq j \leq k(i)$, egy környezetfüggetlen szabály,
  - $\mathcal{F} \subseteq \{p_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq k(i)\}$ (M-beli sorozatok szabályainak egy részhalmaza).

- M elemeit **mátrix**oknak hívjuk.

Mátrix Grammatika

- A $G = (N, T, M, S, \mathcal{F})$ **mátrix grammatika előfordulás-ellenőrzés nélküli**, akkor és csak akkor, ha $\mathcal{F} = \emptyset$.

Mátrix Grammatika

- Legyen $G = (N,T,M,S,\mathcal{F})$ egy mátrix grammatika és $w,w' \in (N \cup T)^*$. Ekkor w' levezethető w-ből valamely m_i mátrix szerint,
  $m_i : (A_{i1} \rightarrow v_{i1}, \ldots, A_{ik(i)} \rightarrow v_{ik(i)}) \in M$, $1 \le i \le n$, $k(i) \ge 1$,
  (jelölés: $w \Rightarrow_{mi} w'$),
  akkor és csak akkor, ha léteznek olyan
  $w_{i1}, \ldots, w_{ik(i)+1} \in (N \cup T)^*$ szavak, amelyekre
  $w = w_{i1}$, $w' = w_{ik(i)+1}$ és
  minden i-re és j-re, $1 \le i \le n$, $1 \le j \le k(i)$,
  - vagy $w_{ij} = w'_{ij}A_{ij}w''_{ij}$ és $w_{ij+1} = w'_{ij}v_{ij}w''_{ij}$,
  - vagy A_{ij} nem fordul elő w_{ij}-ben és $w_{ij} = w_{ij+1}$, valamint $A_{ij} \rightarrow v_{ij} \in \mathcal{F}$.

Mátrix Grammatika

- Legyen $G = (N,T,M,S,\mathcal{F})$ egy mátrix grammatika.
  A G által generált nyelv $L(G)$:
  $L(G) = \{w \in T^* \mid S \Rightarrow_{m_{j1}} y_1 \Rightarrow_{m_{j2}} y_2 \Rightarrow_{m_{j3}} \ldots \Rightarrow_{m_{js}} w,\ 1 \le j_i \le r,\ 1 \le i \le s\}$.

- Példa:
  Legyen $G = (N,T,M,S,\varnothing)$ egy előfordulás ellenőrzés nélküli mátrix grammatika, ahol $N = \{S, A, B\}$, $T = \{a, b\}$, és
  $M = \{m_1, m_2, m_3, m_4, m_5\}$, ahol
  $m_1 = (S \to AB)$,
  $m_2 = (A \to bA, B \to bB)$,
  $m_3 = (A \to b, B \to b)$,
  $m_4 = (A \to aA, B \to aB)$, and
  $m_5 = (A \to a, B \to a)$.
  Ekkor $L(G) = \{ww \mid w \in \{a, b\}^+\}$.

Véletlen szövegfeltételekkel adott grammatika

- Egy **véletlen szövegfeltételekkel adott grammatika** (random context grammar) $G = (N,T,P,S)$ egy 4-es, ahol
  - N és T diszjunkt véges ábécék,
  - $S \in N$ a kezdő szimbólum (axióma),
  - P rendezett hármasok véges halmaza, amelyek (p, Q, R), alakúak, ahol p egy környezetfüggetlen szabály, $Q, R \subseteq N$.

Véletlen szövegfeltételekkel adott grammatika

- Legyen $G = (N, T, P, S)$ egy véletlen szövegfeltételekkel adott grammatika. Ekkor az y szó levezethető x-ből, $x,y \in (N \cup T)^*$, (jelölés: $x \Rightarrow y$), ha
  - $x = x'Ax''$, $y = x'wx''$, valamely $x',x'' \in (N \cup T)^*$ szavakra és
  - minden $(A \rightarrow w, Q, R) \in P$ 3-asra, ha Q minden szimbóluma előfordul $x'x''$-ben, és R egyetlen szimbóluma sem szerepel $x'x''$-ben.

- Megjegyzés:

  Q a **megengedő kontextusa** (permitting context) $(A \rightarrow w, Q, R)$-nak,
  R a **tiltó kontextusa** (forbidding context) $(A \rightarrow w, Q, R)$-nak.

Véletlen szövegfeltételekkel adott grammatika

- Legyen $G = (N,T,P,S)$ egy véletlen szövegfeltételekkel adott grammatika.
  A G által **generált nyelv** $L(G)$:
  $L(G) = \{w \in T^* \mid S \Rightarrow^* w\}$.

- Példa:
  Legyen $G = (N,T,PS)$ egy véletlen szövegfeltételekkel adott grammatika, ahol
  $N = \{S, X, Y, A\}$, $T = \{a\}$, and $P = \{r_1,r_2, r_3, r_4, r_5\}$, ahol
  $r_1 = (S \rightarrow XX, \varnothing, \{Y, A\})$,
  $r_2 = (X \rightarrow Y, \varnothing, \{S\})$,
  $r_3 = (Y \rightarrow S, \varnothing, \{X\})$,
  $r_4 = (S \rightarrow A, \varnothing, \{X, Y\})$,
  $r_5 = (A \rightarrow a, \varnothing, \{S\})$.
  Ekkor $L(G) = \{a^{2^n} | n \geq 0\}$.

Véletlen szövegfeltételekkel adott grammatika

- A random context grammar generating the language $\{a^{2^n} | n \geq 0\}$.

Let $G = (\{S, X, Y, A\}, \{a\}, S, \{r_1, r_2, r_3, r_4, r_5\})$, where

$$r_1 = (S \to XX, \emptyset, \{Y, A\})$$
$$r_2 = (X \to Y, \emptyset, \{S\})$$
$$r_3 = (Y \to S, \emptyset, \{X\})$$
$$r_4 = (S \to A, \emptyset, \{X, Y\})$$
$$r_5 = (A \to a, \emptyset, \{S\})$$

Consider now the production for *aaaa*:

$$\begin{aligned} S &\Rightarrow_{r_1} XX \Rightarrow_{r_2} YX \Rightarrow_{r_2} YY \Rightarrow_{r_3} SY \Rightarrow_{r_3} SS \\ &\Rightarrow_{r_1} XXS \Rightarrow_{r_1} XXXX \Rightarrow_{r_2} YXXX \Rightarrow_{r_2} YYXX \Rightarrow_{r_2} YYYX \Rightarrow_{r_2} YYYY \\ &\Rightarrow_{r_3} SYYY \Rightarrow_{r_3} SSYY \Rightarrow_{r_3} SSSY \Rightarrow_{r_3} SSSS \\ &\Rightarrow_{r_4} ASSS \Rightarrow_{r_4} AASS \Rightarrow_{r_4} AAAS \Rightarrow_{r_4} AAAA \\ &\Rightarrow_{r_5} aAAA \Rightarrow_{r_5} aaAA \Rightarrow_{r_5} aaaA \Rightarrow_{r_5} aaaa \end{aligned}$$

Source: https://en.wikipedia.org/wiki/Controlled_grammar

Nyelvek családjai

- $\mathcal{L}(PR_{ac})$ jelöli az **előfordulás ellenőrzéses, ε-mentes környezetfüggetlen** szabályokból álló **programozott grammatikák** által generált nyelvek osztályát.
- $\mathcal{L}(PR^{\varepsilon}_{ac})$ jelöli a tetszőleges **előfordulás ellenőrzéses, környezetfüggetlen** szabályokból álló **programozott grammatikák** által generált nyelvek osztályát.
- Ha a grammatika **előfordulás ellenőrzés nélküli**, az ac indexet elmarad.
- $\mathcal{L}(MAT_{ac})$, $\mathcal{L}(MAT^{\varepsilon}_{ac})$ a **mátrix grammatikál osztálya**, ε-mentes vagy nem ε-mentes szabályokkal előfordulás ellenőrzéssel, vagy anélkül.
- $\mathcal{L}(RC_{ac})$, $\mathcal{L}(RC^{\varepsilon}_{ac})$ a **véletlen szövegfeltételekkel adott grammatikák osztálya**, ε-mentes vagy nem ε-mentes szabályokkal előfordulás ellenőrzéssel, vagy anélkül.

- **Tétel1** [Dassow, Paun, 2012]:
  A következő relációk állnak fenn:
  $\mathcal{L}_2 \subset \mathcal{L}(PR_{ac}) = \mathcal{L}(MAT_{ac}) = \mathcal{L}(RC_{ac}) \subset \mathcal{L}_1$ és
  $\mathcal{L}(PR^{\varepsilon}_{ac}) = \mathcal{L}(MAT^{\varepsilon}_{ac}) = \mathcal{L}(RC^{\varepsilon}_{ac}) = \mathcal{L}_0$.

Fejlődő rendszerek, L-rendszerek

- Egy **0L-rendszer** (**interakció nélküli Lindenmayer-rendszer, L-rendszer**) egy $G = (V, P, w)$ hármas, ahol
  - V egy véges ábécé,
  - P környezetfüggetlen V feletti átírási szabályok véges halmaza,
  - $w \in V^+$ az axióma.
  - Minden $a \in V$-re létezik $a \rightarrow x \in P$ szabály (P-t teljesnek mondjuk).

- Megjegyzés: Minden $a \in V$ szimbólumra, amely nem szerepel a bal oldalon egy P-beli produkciós szabályban, az $a \rightarrow a$ szabályt feltételezzük; az ilyen szimbólumot konstansnak, vagy terminál szimbólumnak nevezzük.

0L-átírás

- Legyen $z_1, z_2 \in V^*$ két szó, z_1 **átírható** z_2-re G-nek megfelelően (jelölés: $z_1 \Rightarrow z_2$), ha $z_1 = a_1 a_2 \ldots a_r$, $z_2 = x_1 x_2 \ldots x_r$, valamely $a_i \to x_i \in P$-re, $1 \leq i \leq r$.

- Ahány szabály csak lehetséges, szimultán kerül alkalmazásra. Ez különböztet meg egy 0L-rendszert egy klasszikus formális grammatika által generált nyelvtől.

0L-rendszer, generált nyelv

- Legyen $G = (V, P, w)$ egy 0L-rendszer. A G által **generált nyelv** $L(G)$: $L(G) = \{z \in V^* \mid w \Rightarrow^* z\}$, ahol $\Rightarrow^*$ a $\Rightarrow$ reflexív tranzitív lezártja.

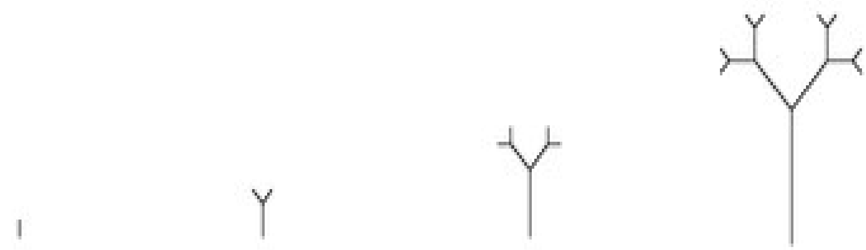
0L-rendszer, generált nyelv

- Példa:
  Legyen $G = (V, P, w)$ egy 0L-rendszer, ahol
  $V = \{a\}$,
  $P = \{a \rightarrow a^2\}$,
  $w = a^3$.
  Ekkor $L(G) = \{a^{3 \cdot 2^n} | i \geq 0\}$

0L-rendszer, generált nyelv

- Példa (fraktál, bináris fa): Legyen $G = (V, P, w)$ egy 0L-rendszer, ahol
  $V = \{0, 1, [,]\}$,
  $P = \{1 \rightarrow 11, 0 \rightarrow 1[0]0\}$,
  $w = 0$.

  Ez a következő sorozatot állítja elő:
  $w = w_0 = 0$
  $w_1 = 1[0]0$
  $w_2 = 11[1[0]0]1[0]0$
  $w_3 = 1111[11[1[0]0]1[0]0]11[1[0]0]1[0]0$
  …
  Ez a sztring képként rajzolható a szimbólum
  az alábbiak szerinti értelmezésével:
  0: húzzunk egy szakaszt,
  amely levéllel végződik
  1: húzzunk egy szakaszt
  [: “push position and angle”, forduljunk balra 45 fokot
  ]: “pop position and angle”, forduljunk jobbra 45 fokot



0L-rendszerek által generált nyelvek családja

- $\mathcal{L}$(0L) jelöli a 0L-rendszerek által generált nyelvek családját.

E0L-rendszer

- Egy **E0L-rendszer (Extended 0L-system**) egy $G = (V,T,P,w)$ 4-es, ahol $G = (V, P, w)$ egy 0L-rendszer és T terminális szimbólumok ábécéje.
- **Átírás** $\Rightarrow_G$ (röviden $\Rightarrow$) és $\Rightarrow^*$ hasonlóan definiált, mint 0L-rendszereknél.
- A G által **generált nyelv** $L(G)$:
  $L(G) = \{z \in T^* \mid w \Rightarrow^* z\}$.
- $\mathcal{L}$(E0L) jelöli az E0L-rendszerek által generált nyelvek családját.

E0L-rendszer, generált nyelv

- Példa:
  Legyen $G = (V,T,P,w)$ egy E0L-rendszer, ahol
  $V = \{a, b\}$,
  $T = \{b\}$,
  $P = \{a \rightarrow b, a \rightarrow bb, b \rightarrow b\}$,
  $w = a$.
  Ekkor $L(G) = \{b, bb\}$.

D0L-rendszer

- Egy **D0L-rendszer** (**deterministic 0L-system**) egy 0L-rendszer, ha minden $a \in V$ esetén pontosan egy szabály $a \to x$, $x \in V^*$ létezik.

- Ha az axiómát véges nyelvvel helyettesítjük, akkor egy 0L-rendszert (D0L-rendszert) kapunk véges számú axiómával, más néven egy **F0L-rendszert** (**FD0L-rendszert**).

- Megjegyzés: Mivel egy $G = (V, P, w)$ D0L-rendszer P szabályhalmaza egy $h : V \to V^*$ homomorfizmust definiál, ezért gyakran használják a $G = (V, h, w)$ jelölést is.

T0L-rendszerek

- Egy **T0L-rendszer** $G = (V, P_1, \ldots, P_n, w)$ egy $(n+2)$-es, $n \geq 1$, ahol minden $G_i = (V, P_i, w)$, $1 \leq i \leq n$, egy 0L-rendszer.

- A G által **generált nyelv** $L(G)$:
$L(G) = \{z \in V^* \mid w \Rightarrow_{G_{i1}} w_1 \Rightarrow_{G_{i2}} \ldots \Rightarrow_{G_{im}} w_m = z,$
$1 \leq i_j \leq n, 1 \leq j \leq m\}$.

- $\mathcal{L}$(T0L) jelöli a T0L-rendszerek által generált nyelvek családját

T0L-rendszerek

- Példa: Legyen $G = (V, P_1, P_2, w)$ egy T0L-rendszer, ahol
  $V = \{a\}$,
  $P_1 = \{a \to a^2\}$, $P_2 = \{a \to a^3\}$, and
  $w = a$.
- Ekkor $L(G) = \{a^i \mid i = 2^m 3^n, m,n \geq 0\}$.

Irodalom

- Handbook of Formal Languages, G. Rozenberg, A. Salomaa, (eds.), Springer–Verlag, Berlin–Heidelberg, 1997.
- Gy. E. Révész, Introduction to Formal Languages, Dover Publications, Inc., New York, 2012.
- G. Rozenberg, A. Salomaa, The mathematical theory of L systems, Vol. 90., Academic Press, 1980.
- J. Dassow, Gh. Paun. Regulated rewriting in formal language theory, Springer Publishing Company, Inc., 2012.