Számítási modellek

1. előadás

Számítási modellek – a kezdetek

- Alex Thue 1910-es évek, sztringek algebrai, kombinatorikai tulajdonságai $\mapsto$ a sztringeket átíró *szabályrendszer*
- Alan Turing és kortársai 1930-as évek közepe: Turing gép, kiszámíthatóság; első számítási modellek (matematikai gépek) és kifejező erejük
- 40-es évek: véges automata formalizálása, 50-es évek vége: veremautomata formalizálása
- Noam Chomsky 1950-es évek a formális grammatika fogalma
- 1950-es évek: Backus-Naur Forma
- gépi fordítási projektek, matematikai nyelvészethez kapcsolódó munkák, 1950-es 60-as évek
- Rabin és Scott: formális eszközök kifejlesztése számítási modellek kifejezőerejének és korlátainak vizsgálatára
- Chomsky-féle nyelvhierarchia, 1956-64
- 60-as évek: időigény, tárigény, bonyolultsági osztályok, NP-teljesség
- új, nemkonvencionális modellek

Alapfogalmak, jelölések

ábécé, betű, szó, szó hossza

Ábécé: Egy véges, nemüres halmaz.
Az ábécé elemeit **betűk**nek nevezzük.

Egy V ábécé elemeiből képzett véges sorozatokat **V feletti szavaknak** vagy sztringeknek nevezzük. Egy $u = t_1 \cdots t_n$ szóban lévő betűk számát (n) a szó **hosszának** nevezzük. Jelölés: $|u| = n$.
A 0 hosszú sorozat jelölése ε, ezt **üres szónak** nevezzük $(|\varepsilon| = 0)$.

Példa:
Legyen $V = \{a, b\}$, ekkor a és b a V ábécé két betűje. $abba$ és $aabba$ egy-egy V feletti szó. $|aabba| = 5$. $aabba$ és $baaba$ különböző szavak, bár mindkettő 5 hosszú valamint 3 a-t és 2 b-t tartalmaz. A betűk sorrendje számít!

Alapfogalmak, jelölések

az összes szavak halmaza

V^* jelöli a V **ábécé feletti szavak halmazát**, beleértve az üres szót is.

$V^+ = V^* \setminus \{\varepsilon\}$ a V **ábécé feletti, nemüres szavak halmazát** jelöli.

Példa: $V = \{a, b\}$, ekkor
$V^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$.

V^* szavainak egy lehetséges felsorolása a **hosszlexikografikus (shortlex) rendezés** szerinti:

- Feltesszük, hogy az ábécé rendezett. (A fenti példában, mondjuk a előbb van, mint b.)
- A rövidebb szó mindig megelőzi a hosszabbat.
- Az azonos hosszúságú szavak az ábécé rendezettsége alapján meghatározott ábécésorrendben követik egymást.

Ez a rendezés egyértelműen meghatározza a szavak sorrendjét.

Műveletek szavakon

konkatenáció

Legyen V egy ábécé továbbá legyenek $u = s_1 \cdots s_n$ és $v = t_1 \cdots t_k$ V feletti szavak (azaz legyen $s_1, \ldots, s_n, t_1, \ldots, t_k \in V$). Ekkor az $uv := s_1 \cdots s_n t_1 \cdots t_k$ szót az u és v szavak **konkatenáltjának** nevezzük. (Az u szó betűi után írjuk a v szó betűit.)

Nyilván $|uv| = |u| + |v|$.

Példa:
Legyen $V = \{a, b\}$, valamint $u = abb$ és $v = aaaba$ egy-egy V feletti szó. Ekkor $uv = abbaaaba$ illetve $vu = aaabaabb$.

Azaz a konkatenáció **általában nem kommutatív**, de asszociatív (azaz $u(vw) = (uv)w$ teljesül minden $u, v, w \in V^*$-ra)

Műveletek szavakon

i-edik hatvány

Legyen i nemnegatív egész szám és legyen u egy V ábécé feletti szó ($u \in V^*$). Az u szó **i-edik hatványa** alatt az u szó i darab példányának konkatenáltját értjük és u^i -vel jelöljük.

Konvenció szerint $u^0 := \varepsilon$. Ekkor $u^{n+k} = u^n u^k$ teljesül $(k, n \in \mathbb{N})$

Példa:

Legyen $V = \{a, b\}$ és legyen $u = abb$. Ekkor $u^0 = \varepsilon$, $u^1 = abb$, $u^2 = abbabb$ és $u^3 = abbabbabb$.

$(ab)^3 \neq a^3 b^3$ **! ! ! !**

$(ab)^3 = ababab, \quad a^3 b^3 = aaabbb.$

Műveletek szavakon

tükörkép

Legyen u egy V ábécé feletti szó. Az u szó u^{-1}-gyel jelölt **tükörképén** vagy fordítottján azt a szót értjük, amelyet úgy kapunk, hogy u betűit fordított sorrendben írjuk le.
Azaz ha $u = a_1 \cdots a_n,\ a_i \in V,\ 1 \leqslant i \leqslant n$, akkor $u^{-1} = a_n \cdots a_1$.

Példa:
Legyen $V = \{a, b\}$, valamint $u = abba$ és $v = baabba$ egy-egy V ábécé feletti szó. Ekkor $u^{-1} = abba$ és $v^{-1} = abbaab$.

Ha $u = u^{-1}$, akkor u-t **palindrom** tulajdonságúnak, vagy palindrómának nevezzük.

Tehát *abba* egy palindróma.

Alapfogalmak, jelölések

részszó, prefix, suffix

Legyen V egy ábécé és legyenek u és v szavak V felett. Az u szó a v szó **részszava**, ha $v = xuy$ teljesül, valamely $x, y \in V^*$ szavakra.

Ha $x = \varepsilon$, akkor u-t a v szó **prefixének**, ha pedig $y = \varepsilon$, akkor u-t a v szó **suffixének** nevezzük.

Példa:

Legyen $V = \{a, b\}$ és $u = abba$.

u részszavai: $\varepsilon, a, b, ab, bb, ba, abb, bba, abba$.

u prefixei: $\varepsilon, a, ab, abb, abba$.

u suffixei: $\varepsilon, a, ba, bba, abba$.

***abb* nem suffixe *u*-nak!!!** A suffix nem a tükörkép szó prefixe!!!

Alapfogalmak, jelölések

nyelv

Legyen V egy ábécé, V^* egy L részhalmazát V feletti **nyelvnek** nevezzük.
Az üres nyelv (nyelv, amely egyetlen szót sem tartalmaz) jelölése $\emptyset$.
Egy V ábécé feletti nyelv véges nyelv, ha véges számú szót tartalmaz, ellenkező esetben végtelen.

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

- $L_1 = \{a, ba, \varepsilon\}$.
- $L_2 = \{a^i b^i \mid i \geqslant 0\}$.
- $L_3 = \{uu^{-1} \mid u \in V^*\}$.
- $L_4 = \{u \mid u \in \{a, b\}^+, u\text{-ban ugyanannyi } a \text{ van, mint } b\}$.

L_1 véges nyelv, a többi végtelen.

Nyelvekre vonatkozó műveletek

halmazműveletek

Legyen V egy ábécé és legyenek L_1, L_2 nyelvek V felett (vagyis $L_1 \subseteq V^*$ és $L_2 \subseteq V^*$).

$L_1 \cup L_2 = \{u \mid u \in L_1 \text{ vagy } u \in L_2\}$ (az L_1 és az L_2 nyelv **uniója**)

$L_1 \cap L_2 = \{u \mid u \in L_1 \text{ és } u \in L_2\}$ (az L_1 és L_2 nyelv **metszete**)

$L_1 - L_2 = \{u \mid u \in L_1 \text{ és } u \notin L_2\}$ (az L_1 és az L_2 nyelv **különbsége**)

Az $L \subseteq V^*$ nyelv **komplementere** a V ábécére nézve az $\bar{L} = V^* - L$ nyelv.

Nyelvekre vonatkozó műveletek

konkatenáció

Legyen V egy ábécé és $L_1, L_2 \subseteq V^*$. Az L_1 és az L_2 nyelvek **konkatenációján** az $L_1L_2 = \{u_1u_2 \mid u_1 \in L_1, u_2 \in L_2\}$ nyelvet értjük. Asszociatív, de nyilván **nem kommutatív** művelet.

Példák:

$V = \{a, b\},\ L_1 = \{ab, bb\},\ L_2 = \{\varepsilon, a, bba\},$

$L_1L_2 = \{ab\varepsilon, bb\varepsilon, aba, bba, abbba, bbbba\}$
$= \{ab, bb, aba, bba, abbba, bbbba\}.$ (mindenkit mindenkivel!)

$L_3 = \{a^{2n}b^{2n} \mid n \in \mathbb{N}\}, L_4 = \{a^{3n}b^{3n} \mid n \in \mathbb{N}\}$

$L_3L_4 = \{a^{2n}b^{2n}a^{3k}b^{3k} \mid n, k \in \mathbb{N}\}.$

Nyelvekre vonatkozó műveletek

tükörkép nyelv, i-edik hatvány

Legyen V egy ábécé és legyen $L \subseteq V^*$. Ekkor $L^{-1} = \{u^{-1} \mid u \in L\}$ a **tükörképe** (megfordítása) az L nyelvnek.
Egy L nyelv ***i*-edik hatványát** $L^0 := \{\varepsilon\}$ és $L^i := LL^{i-1}$ $(i \geqslant 1)$ definiálják. Azaz L^i jelöli az L i-edik iterációját a konkatenáció műveletére nézve.

Példa: $L = \{a, bb\}$.

$L^0 = \{\varepsilon\}$,
$L^1 = \{a, bb\}$,
$L^2 = \{a\,a, a\,bb, bb\,a, bb\,bb\}$,
$L^3 = \{a\,a\,a, a\,a\,bb, a\,bb\,a, a\,bb\,bb, bb\,a\,a, bb\,a\,bb, bb\,bb\,a, bb\,bb\,bb\}$,

Nyelvekre vonatkozó műveletek

Kleene lezárt

Az L nyelv **iteratív lezártja** (vagy röviden lezártja vagy Kleene-lezártja) alatt az $L^* = \bigcup_{i \geqslant 0} L^i$ nyelvet értjük.

Az L nyelv **pozitív lezártja** alatt az $L^+ = \bigcup_{i \geqslant 1} L^i$ nyelvet értjük.

Észrevétel:
Nyilvánvalóan, $L^+ = L^*$, ha $\varepsilon \in L$ és $L^+ = L^* - \{\varepsilon\}$ ha $\varepsilon \notin L$.

Példa: $L = \{a, bb\}$.

$L^0 = \{\varepsilon\}$,
$L^1 = \{a, bb\}$,
$L^2 = \{a\,a, a\,bb, bb\,a, bb\,bb\}$,
$L^3 = \{a\,a\,a, a\,a\,bb, a\,bb\,a, a\,bb\,bb, bb\,a\,a, bb\,a\,bb, bb\,bb\,a, bb\,bb\,bb\}$,
$L^4 = \{a\,a\,a\,a, a\,a\,a\,bb, a\,a\,bb\,a, \ldots\}, \ \ldots$

Ezen szavak uniója együttesen alkotja L^*-t. L^+ ettől csak annyiban tér el, hogy nincs benne az ε szó.

Nyelvcsaládok

Ha X egy halmaz, jelölje $\mathcal{P}(X)$ az X halmaz hatványhalmazát, azaz $\mathcal{P}(X) = \{A \,|\, A \subseteq X\}$.

Nyelvcsalád (vagy nyelvosztály) alatt nyelveknek egy halmazát értjük.

Tehát ha V egy ábécé:

- $a \in V$: betű
- $u \in V^*$: szó
- $L \subseteq V^*$ vagy $L \in \mathcal{P}(V^*)$: nyelv
- $\mathcal{L} \subseteq P(V^*)$ vagy $\mathcal{L} \in \mathcal{P}(\mathcal{P}(V^*))$: nyelvcsalád

Példa: $V = \{a, b\}$

- véges nyelvek nyelvcsaládja
- csak b betűvel kezdődő szavakat tartalmazó nyelvek nyelvcsaládja
- reguláris kifejezésekkel leírható nyelvek nyelvcsaládja

Grammatikák

Definíció

Egy $G = \langle N, T, P, S \rangle$ rendezett négyest **grammatikának** (generatív grammatikán vagy (generatív) nyelvtannak) nevezünk ha

- N és T diszjunkt véges ábécék (azaz $N \cap T = \emptyset$). N elemeit **nemterminális**, T elemeit pedig **terminális** szimbólumoknak nevezzük.
- $S \in N$ a grammatika **kezdőszimbóluma**.
- P rendezett (x, y) párok véges halmaza, ahol $x, y \in (N \cup T)^*$ és x legalább egy nemterminális szimbólumot tartalmaz.

A P halmaz elemeit **szabályoknak** (vagy átírási szabályoknak vagy produkcióknak) hívjuk. A gyakorlatban az (x, y) jelölés helyett szinte mindig az $x \to y$ jelölést használjuk amennyiben $\to$ nem eleme $N \cup T$-nek.

Grammatikák

Példák:

$G_1 = \langle \{S, A, B\}, \{a, b, c\}, \{S \to c, S \to AB, A \to aA, B \to \varepsilon,$
$abb \to aSb\}, S\rangle$ **nem** grammatika, mivel minden szabály baloldalának tartalmaznia kell legalább egy N-beli szimbólumot.

$G_2 = \langle \{S, A, B, C\}, \{a, b, c\}, \{S \to a, S \to AB, A \to Ab,$
$B \to \varepsilon, aCA \to aSc\}, S\rangle$ grammatika.

Generatív grammatikák

egylépéses levezetés

Definíció

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

Példa:

$$G_2 = \langle \{S, A, B, C\}, \{a, b, c\}, \{S \rightarrow a, S \rightarrow AB, A \rightarrow Ab, B \rightarrow \varepsilon, aCA \rightarrow aSc\}, S \rangle.$$

Ekkor $BB\mathbf{aCA}a \Rightarrow BB\mathbf{aSc}a$, $\mathbf{A}BB \Rightarrow \mathbf{Ab}BB$, $B\mathbf{B} \Rightarrow B$.

Reláció reflexív, tranzitív lezártja

- Legyenek $X_1, \dots X_n$ halmazok és $R \subseteq X_1 \times \dots \times X_n$, ekkor R-et **n-változós relációnak** nevezzük.
- $R \subseteq X \times Y$ és $S \subseteq Y \times Z$ bináris relációk **kompozíciója** $R \circ S = \{(x, z) \subseteq X \times Z \mid \exists y \in Y : (x, y) \in R \ \wedge \ (y, z) \in S\}$
- $R^1 := R$, $R^i := R \circ R^{i-1}$ $(i \geqslant 2)$ definiálja R **i-edik hatványát**
- Ha $R \subseteq X \times X$, akkor $S := R \cup \{(x, x) \mid x \in X\}$ az R reláció **reflexív lezártja**
- Ha $R \subseteq X \times X$, akkor $R^+ := \bigcup\limits_{i=1,2,\dots} R^i$ az R reláció **tranzitív lezártja**
- Ha $R \subseteq X \times X$, akkor $R^* := R^+ \cup \{(x, x) \mid x \in X\}$ az R reláció **reflexív, tranzitív lezártja**

Reláció reflexív, tranzitív lezártja

Példa:

$X = \{1, 2, 3, 4, 5\}$, $R = \{(1, 2), (2, 3), (3, 4), (2, 5)\}$ ekkor

$S = \{(1, 2), (2, 3), (3, 4), (2, 5), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}$ az R reflexív lezártja

$R^1 = R = \{(1, 2), (2, 3), (3, 4), (2, 5)\}$

$R^2 = \{(1, 3), (2, 4), (1, 5)\}$

$R^3 = \{(1, 4)\}$

$R^4 = R^5 = \ldots = \emptyset$

$R^+ = \{(1, 2), (2, 3), (3, 4), (2, 5), (1, 3), (2, 4), (1, 5), (1, 4)\}$ az R tranzitív lezártja

$R^* = \{(1, 2), (2, 3), (3, 4), (2, 5), (1, 3), (2, 4), (1, 5), (1, 4), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}$ az R reflexív, tranzitív lezártja

Reláció reflexív, tranzitív lezártja

Észrevételek:

- A bináris relációk éppen az irányított gráfok (a reláció elemei megfeleltethetők az irányított éleknek).
- A bináris relációk reflexív, tranzitív lezártja egy olyan gráfnak felel meg, melyben u-ból v-be pontosan akkor van él, ha az eredeti gráfban van (véges) irányított út.
- $\Rightarrow_G$ egy bináris reláció a végtelen $X = (N \cup T)^*$ halmaz felett.
- A $\Rightarrow_G$ egylépéses levezetés reláció tehát természetes módon megad egy végtelen irányított gráfot az $(N \cup T)^*$ csúcshalmazon.
- A $\Rightarrow_G^*$ reflexív tranzitív lezárt gráfban u-ból v-be pontosan akkor van irányított él, ha véges sok egylépéses levezetésen át el lehet jutni u-ból v-be

Grammatikák

többlépéses levezetés

Definíció

Legyen $G = \langle N, T, P, S \rangle$ egy grammatika és legyen $u, v \in (N \cup T)^*$. u-ból (több lépésben)**levezethető** v, ha $(u, v) \in \Rightarrow_G^*$. ($\Rightarrow_G^*$ a $\Rightarrow_G$ reflexív tranzitív lezártja)

u-ból **k lépésben levezethető** v, ha $(u, v) \in \Rightarrow_G^k$. ($\Rightarrow_G^k$ a $\Rightarrow_G$ k-adik hatványa).

A kezdőszimbólumból levezethető szavakat **mondatformának** nevezzük.
Megadunk egy alternatív, közvetlen definíciót is:

Definíció

Legyen $G = \langle N, T, P, S \rangle$ egy grammatika és legyen $u, v \in (N \cup T)^*$. u-ból (több lépésben) **levezethető** v, ha $u = v$ vagy van olyan $n \geqslant 1$ és $w_0, \dots w_n \in (N \cup T)^*$, hogy $w_{i-1} \Rightarrow_G w_i$ $(1 \leqslant i \leqslant n)$ és $w_0 = u$ és $w_n = v$. Jelölés: $u \Rightarrow_G^* v$.

Grammatikák

többlépéses levezetés

Ha egyértelmű melyik nyelvtanról van szó, akkor $\Rightarrow_G$ helyett röviden $\Rightarrow$-t írhatunk.

Példa:

$$G_2 = \langle \{S, A, B, C\}, \{a, b, c\}, \{S \rightarrow a, S \rightarrow AB, A \rightarrow Ab, B \rightarrow \varepsilon, aCA \rightarrow aSc\}, S\rangle.$$

Ekkor $BB\mathbf{aCA}a \Rightarrow BB\mathbf{aSc}a$ és $BBa\mathbf{S}ca \Rightarrow BBa\mathbf{AB}ca$, tehát $BBaCAa \Rightarrow^* BBaABca$.

$S \Rightarrow AB \Rightarrow AbB \Rightarrow AbbB$, tehát $AbbB$ egy mondatforma.

Generált nyelv

Definíció

Legyen $G = \langle N, T, P, S \rangle$ egy tetszőleges generatív grammatika. A G grammatika által **generált** $L(G)$ **nyelv** alatt az $L(G) = \{w \mid S \Rightarrow_G^* w, w \in T^*\}$ szavakból álló halmazt értjük.

Példák:

- $G = \langle N, T, P, S \rangle$, ahol $N = \{S, A, B\}$, $T = \{a, b\}$ és $P = \{S \to aSb, S \to ab, S \to ba\}$.

  Ekkor $L(G) = \{a^n abb^n, a^n bab^n \mid n \geqslant 0\}$.
- $G_1 = \langle N, T, P, S \rangle$, ahol $N = \{S, A, B\}$, $T = \{a, b\}$ és $P = \{S \to ASB, S \to \varepsilon, AB \to BA, BA \to AB, A \to a, B \to b\}$.

  Ekkor $L(G_1) = \{u \in \{a, b\}^* \mid u\text{-ban ugyanannyi } a \text{ van, mint } b\}$.

Ekvivalens grammatikák

Minden grammatika generál egy nyelvet, de ugyanazat a nyelvet több különböző grammatika is generálhatja.

Két **grammatika ekvivalens**, ha ugyanazt a nyelvet generálják.

Példa:
$G_2 = \langle N, T, P, S\rangle$, ahol $N = \{S, A, B\}$, $T = \{a, b\}$ és
$P = \{S \to SS, S \to aSb, S \to bSa, S \to \varepsilon\}$.

Belátható, hogy
$L(G_2) = \{u \in \{a, b\}^* \mid u\text{-ban ugyanannyi } a \text{ van, mint } b\}$. Így ez a grammatika ekvivalens az előzővel.

Grammatikák Chomsky féle osztályozása

Legyen $G = \langle N, T, P, S \rangle$ egy generatív grammatika. A G generatív grammatika i-típusú ($i = 0, 1, 2, 3$), ha P szabályhalmazára teljesülnek a következők:

- $i = 0$: nincs korlátozás,
- $i = 1$: P minden szabálya $u_1 A u_2 \rightarrow u_1 v u_2$ alakú, ahol $u_1, u_2, v \in (N \cup T)^*, A \in N$, és $v \neq \varepsilon$, kivéve az $S \rightarrow \varepsilon$ alakú szabályt, feltéve, hogy P-ben ilyen szabály létezik. Ha P tartalmazza az $S \rightarrow \varepsilon$ szabályt, akkor S nem fordul elő P egyetlen szabályának 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 $u \in T^*$.

Grammatikaosztályba sorolás

Példa: Legyen $G = \langle N, T, P, S \rangle$, ahol $N = \{S, A, B\}$, $T = \{a, b\}$ és $P = \{S \rightarrow ASB, S \rightarrow \varepsilon, AB \rightarrow BA, BA \rightarrow AB, A \rightarrow a, B \rightarrow b\}$. Szabályról szabályra nézzük meg, hogy az adott szabály melyik típusba engedi sorolni a grammatikát.

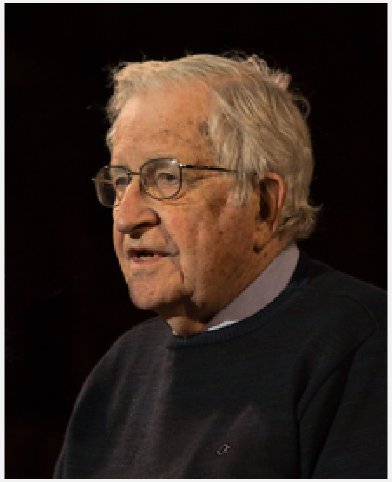
$S \rightarrow ASB$	0, 1, 2
$S \rightarrow \varepsilon$	0, 1, 2, 3
$AB \rightarrow BA$	0
$BA \rightarrow AB$	0
$A \rightarrow a$	0, 1, 2, 3
$B \rightarrow b$	0, 1, 2, 3

Az első 2 szabály nem lehet együtt 1-típusú grammatikában.

A 3. és 4. szabály a 0-son kívül mindent kizár, így a grammatika csak a 0-típusba skatulyázható.

Nyelvek Chomsky féle osztályozása

Noam Chomsky (1928 –) amerikai nyelvész, filozófus, politikai aktivista. A generatív nyelvtan elméletének megalkotója, kidolgozója a róla elnevezett Chomsky-hierarchiának. (Wikipédia)



Egy L nyelvet ***i*-típusúnak** mondunk, ahol $i = 0, 1, 2, 3$, ha i-típusú grammatikával generálható. $\mathcal{L}_i$ $(i = 0, 1, 2, 3)$ jelöli az i-típusú nyelvek nyelvosztályát.

Példa: A korábbi 0-típusú G_1 és 2-típusú G_2 grammatikák egyaránt $L = \{u \in \{a, b\}^* \mid u\text{-ban ugyanannyi } a \text{ van, mint } b\}$-t generálják, így $L \in \mathcal{L}_2$.

A Chomsky féle grammatikaosztályok elnevezése

- A 0-típusú grammatikákat **mondatszerkezetű** (phrase-structure) grammatikáknak is nevezzük, ami nyelvészeti eredetükre utal.
- A 1-típusú grammatikák a **környezetfüggő** (context-sensitive) grammatikák, mert szabályai olyanok, hogy, egy A nemterminális valamely előfordulása egy v szóval csak u_1 és u_2 kontextus jelenlétében helyettesíthető.
- 2-típusú grammatikákat **környezetfüggetlen** (context-free) grammatikáknak mondjuk, mert szabályai olyanok, hogy az A nemterminális v-vel való helyettesítése bármely kontextusban megengedett.
- A 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 **rekurzíven felsorolható**, **környezetfüggő**, **környezetfüggetlen**, valamint **reguláris** nyelvosztálynak is mondjuk.

A Chomsky-féle hierarchia

A grammatikák alakjából nyilvánvaló, hogy $\mathcal{L}_3 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_0$ és $\mathcal{L}_1 \subseteq \mathcal{L}_0$.

Tétel (Chomsky-féle nyelvhierarchia)

$$\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0.$$

Itt az $\mathcal{L}_2$ és az $\mathcal{L}_1$ nyelvosztályok közötti tartalmazási reláció nem látható azonnal a megfelelő grammatikák definíciójából.

Szintén $\mathcal{L}_1$-et tudják generálni a **hossznemcsökkentő** grammatikák. Ezek $p \to q$ szabályaira $|p| \leqslant |q|$ teljesül kivéve az $S \to \varepsilon$ alakú szabályt, feltéve, hogy P-ben ilyen szabály létezik. Ha P tartalmazza az $S \to \varepsilon$ szabályt, akkor S nem fordul elő P egyetlen szabályának jobb oldalán sem.

Grammatikák a gyakorlatban

BNF (Backus-Naur Form) **John Backus, Peter Naur ALGOL 60**

A BNF egy széles körben használt metanyelv melynek segítségével szabályok alkothatók meg (például egy programozási nyelv szintaktikai szabályai).Építőkövei:

- $\langle$név$\rangle$, **fogalmak** (vagy más néven **nemterminálisok**)
- $::=$, a szabályok bal- és jobboldalának elválasztászára
- a sztringek alkotóelemei (**terminálisok**)

Egy szabály bal- és jobboldalból áll, köztük $::=$, baloldalon pontosan 1 fogalom, jobboldalon terminálisok és fogalmak véges sorozata állhat.

A BNF megfelel a környezetfüggetlen grammatikának. Példa:

$\langle$azonosító$\rangle$::= $\langle$betű$\rangle \mid \langle$betű$\rangle\langle$azvég$\rangle$
$\langle$azvég$\rangle$::= $\langle$betű$\rangle \mid \langle$számjegy$\rangle \mid \langle$betű$\rangle\langle$azvég$\rangle \mid \langle$számjegy$\rangle\langle$azvég$\rangle$
$\langle$betű$\rangle$::=$a \mid \cdots \mid z$
$\langle$számjegy$\rangle$::=$0 \mid \cdots \mid 9$

Kontrollált környezetfüggetlen grammatikák

BNF-fel nem írható le teljeskörűen a programozási nyelvek szintaxisa (lásd pl. deklaráció fogalma, az $\{uu \mid u \in V^*\}$ nyelv nem környezetfüggetlen).

Módosítható-e úgy a környezetfüggetlen grammatikák modellje, hogy a természetes, egyszerű szabályalakot megtartsuk, de a modell kifejező ereje nőjön?

Tudnánk-e nem környezetfüggetlen nyelveket generálni környezetfüggetlen grammatikákkal azáltal, hogy például a produkciós szabályok alkalmazhatóságára feltételeket szabunk.

A válasz igen, nézzünk erre néhány példát!

3 példát nézünk, a **programozott grammatikák** és a **mátrixgrammatikák** esetén az alkalmazott szabályok sorrendiségére vonatkozik a kontroll, míg a **véletlen szövegfeltételekkel adott grammatikák** esetében a szabályok alkalmazhatósága az aktuális mondatforma tulajdonságaitól függ.

Progamozott grammatikák

Környezetfüggetlen programozott grammatikán egy $G = \langle N, T, P, S \rangle$ négyest értünk, ahol

- N és T diszjunkt véges ábécék,
- $S \in N$ a kezdőszimbólum
- P rendezett hármasok véges halmaza, melyek $r = (p, \sigma, \phi)$ alakúak, ahol p egy környezetfüggetlen szabály, $\sigma, \phi \subseteq P$.

σ-t az r szabály **siker-halmazának**, ϕ a szabály **kudarc-halmazának** nevezzük.

Programozott grammatikák

Jelölje Q a **megengedett** szabályok halmazát. Kezdetben $Q := P$.

Egy $u \in (N \cup T)^*$ mondatformából **egy lépésben levezethető** a $v \in (N \cup T)^*$ mondatforma, ha van olyan $r = (A \to w, \sigma, \phi) \in Q$, melyre a következők egyike teljesül

- $u = xAy$ és $v = xwy$ valamely $x, y \in (N \cup T)^*$-ra, ekkor a szabályalkalmazás **sikeres** és $Q := \sigma$,
- u nem tartalmazza A-t és $v = u$, ekkor a szabályalkalmazás **sikertelen** és $Q := \phi$.

Tehát adott mindig megengedett szabályok egy $Q \subseteq P$ részhalmaza. Választunk egy $r \in Q$ szabályt. Amennyiben sikerül a kiválasztott szabályt alkalmazni, akkor az adott szabály siker-halmazából, ha nem járunk sikerrel, akkor pedig a kudarc-halmazából kell a következő szabályt választani.

Az így kapott $\Rightarrow$ reláció reflexív, tranzitív lezártjaként definiáljuk $\Rightarrow^*$-t. $L(G) := \{w \in T^* \mid S \Rightarrow^* w\}$.

Programozott grammatikák

Ha $r = (p, \sigma, \emptyset)$ minden $r \in P$ szabályra fennáll, akkor a programozott grammatika **előfordulásellenőrzés nélküli**, egyébként **előfordulásellenőrzéses**.

Példa: Legyen $G = \langle \{S, A\}, \{a\}, S, \{r_1, r_2, r_3\} \rangle$, ahol
$r_1 = (S \to AA, \{r_1\}, \{r_2\})$
$r_2 = (A \to S, \{r_2\}, \{r_1, r_3\})$
$r_3 = (S \to a, \{r_3\}, \emptyset)$

$S \overset{r_1}{\Rightarrow} AA$ (siker:$\{r_1\}$) $\overset{r_1}{\Rightarrow} AA$ (kudarc:$\{r_2\}$) $\overset{r_2}{\Rightarrow} SA$ (siker:$\{r_2\}$)
$\overset{r_2}{\Rightarrow} SS$ (siker:$\{r_2\}$) $\overset{r_2}{\Rightarrow} SS$ (kudarc:$\{r_1, r_3\}$) $\overset{r_3}{\Rightarrow} aS$ (siker:$\{r_3\}$)
$\overset{r_3}{\Rightarrow} aa$.

Az SS-en kudarccal végződött $A \to S$ átírás után választhattuk volna az r_1 szabályt is, ez újra megduplázta volna az S-ek számát.

G egy előfordulásellenőrzéses programozott grammatika és $L(G) = \{a^{2^n} \mid n \geqslant 0\}$.

Mátrixgrammatikák

Előfordulásellenőrzéses környezetfüggetlen mátrixgrammatikán egy $G = \langle N, T, M, S, F\rangle$ ötöst értünk, ahol

- N és T diszjunkt véges ábécék,
- $S \in N$ a kezdőszimbólum,
- $M = \{m_1, m_2, \dots, m_n\}$, $n \geqslant 1$, sorozatok véges halmaza
- $m_i = (p_{i_1}, \dots p_{i_{k(i)}}), k(i) \geqslant 1, 1 \leqslant i \leqslant n$, ahol minden $p_{i_j}, 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant k(i)$, egy környezetfüggetlen szabály, és
- F M-beli sorozatok szabályainak egy részhalmaza, azaz $F \subseteq \{p_{i_j} \mid 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant k(i)\}$.

M elemeit **mátrixoknak** nevezzük.

Ha $F = \emptyset$, akkor a mátrixgrammatikát **előfordulásellenőrzés nélkülinek** nevezzük.

Mátrixgrammatikák

Legyen $G = \langle N, T, M, S, F \rangle$ egy tetszőleges mátrixgrammatika. Ekkor $u \in (N \cup T)^*$-ból **egy lépésben levezethető** a $v \in (N \cup T)^*$ szó az $m_i = (A_{i_1} \to v_{i_1}, \ldots A_{i_{k(i)}} \to v_{i_{k(i)}}) \in M$ mátrix szerint, ha léteznek $w_1, \ldots, w_{k(i)+1} \in (N \cup T)^*$ szavak a következő feltételekkel:

- $w_1 = u$ és $w_{k(i)+1} = v$
- ha $w_j = x_j A_{i_j} y_j$, valamely $x_j, y_j \in (N \cup T)^*$-ra, akkor $w_{j+1} = x_j v_{i_j} y_j$ $(1 \leqslant j \leqslant k(i))$
- ha A_{i_j} nem fordul elő w_j-ben akkor $w_{j+1} = w_j$ és $A_{i_j} \to v_{i_j} \in F$ $(1 \leqslant j \leqslant k(i))$

Jelölés: $\Rightarrow_{m_i}$ vagy röviden $\Rightarrow$,

$\Rightarrow^*$ (többlépéses levezetés) és $L(G)$ (generált nyelv) definíciója a szokásos.

Mátrixgrammatikák

1. Példa: $u = XYAX$, $v = XYCX$.

- F bármi, $m = (A \to B, B \to C)$, ekkor $u \Rightarrow v$,
- $F = \emptyset$, $m = (A \to B, E \to F, B \to C)$, ekkor $u \not\Rightarrow v$,
- $F \supseteq \{E \to F\}$, $m = (A \to B, E \to F, B \to C)$, ekkor $u \Rightarrow v$.

2. Példa:

Legyen $G = \langle N, T, M, S, \emptyset \rangle$ egy mátrixgrammatika előfordulásellenőrzés nélkül, ahol $N = \{S, A, B\}$, $T = \{a, b\}$, és $M = \{m_1, m_2, m_3, m_4, m_5\}$, ahol

$$m_1 = (S \to AB),$$
$$m_2 = (A \to bA, B \to bB), \qquad m_4 = (A \to aA, B \to aB),$$
$$m_3 = (A \to b, B \to b), \qquad m_5 = (A \to a, B \to a).$$

$$S \overset{m_1}{\Longrightarrow} AB \overset{m_4}{\Longrightarrow} aAaB \overset{m_4}{\Longrightarrow} aaAaaB \overset{m_2}{\Longrightarrow} aabAaabB \overset{m_4}{\Longrightarrow} aabaAaabaB \overset{m_3}{\Longrightarrow} aababaabab.$$

Könnyen látható, hogy $L(G) = \{ww \mid w \in \{a, b\}^+\}$.

Véletlen szövegfeltételekkel adott grammatika

Környezetfüggetlen véletlen szövegfeltételekkel adott grammatikán (random context grammar) egy $G = \langle N, T, P, S \rangle$ négyest értünk, ahol

- N és T diszjunkt véges ábécék,
- $S \in N$ a kezdőszimbólum (axióma),
- P rendezett hármasok véges halmaza, amelyek (p, Q, R), alakúak, ahol p egy környezetfüggetlen szabály, $Q, R \subseteq N$.

elnevezés: Q-t az (p, Q, R) hármas **megengedő** (permitting), R-et pedig **tiltó** (forbidding) kontextusának nevezzük.

Véletlen szövegfeltételekkel adott grammatika

Egy környezetfüggetlen véletlen szövegfeltételekkel adott grammatika **előfordulásellenőrzés nélküli**, ha minden $(p, Q, R) \in P$ esetén $R = \emptyset$. **Előfordulásellenőrzéses**: amennyiben $R \neq \emptyset$ is megengedett.

Más elnevezés az előfordulásellenőrzés nélküli esetre: környezetfüggetlen megengedő véletlen szövegfeltételekkel adott grammatika (permitting random context grammar, pRC grammar).

Hasonlóan definálható a környezetfüggetlen tiltó véletlen szövegfeltételekkel adott grammatika (forbidding random context grammar, fRC grammar).

Véletlen szövegfeltételekkel adott grammatika

Legyen $G = \langle N, T, P, S\rangle$ egy tetszőleges véletlen szövegfeltételekkel adott grammatika. Ekkor az v szó **egy lépésben levezethető** u-ból ($u, v \in (N \cup T)^*$), ha létezik $(A \to w, Q, R) \in P$ $x, y \in (N \cup T)^*$, hogy

- $u = xAy$ és $v = xwy$
- minden Q-beli nemterminális előfordul xy-ban
- semelyik R-beli nemterminális sem fordul elő xy-ban

Jelölés: $\Rightarrow$. $\Rightarrow^*$ (többlépéses levezetés) és $L(G)$ (generált nyelv) definíciója a szokásos.

Példa:

- A $BABaC$ mondatformára az $(A \to XY, \{C\}, \{A\})$ szabály alkalmazható, mivel C van, A viszont nincs a $BBaC$ környezetben. $BABaC \Rightarrow BXYBaC$.
- A $BABaC$ mondatformára az $(A \to XY, \{A, B\}, \{X\})$ szabály nem alkalmazható, mert nincs az átírandón felül további A a szóban, azaz a $BBaC$ környezetben.

Véletlen szövegfeltételekkel adott grammatika

Példa: Legyen $G = \langle N, T, P, S \rangle$ egy véletlen szövegfeltételekkel adott grammatika, ahol $N = \{S, X, Y, A\}$, $T = \{a\}$, és $P = \{r_1, r_2, r_3, r_4, r_5\}$, ahol

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

$S \Rightarrow XX \Rightarrow XY \Rightarrow YY \Rightarrow YS \Rightarrow SS \Rightarrow SXX \Rightarrow XXXX \Rightarrow XYXX \Rightarrow XYYX \Rightarrow XYYY \Rightarrow YYYY \Rightarrow YYSY \Rightarrow SYSY \Rightarrow SSSY \Rightarrow SSSS \Rightarrow SSAS \Rightarrow SSAA \Rightarrow ASAA \Rightarrow AAAA \Rightarrow AAAa \Rightarrow aAAa \Rightarrow aaAa \Rightarrow aaaa.$

Végig kényszerpályán a levezetés, egyedül csupa S-nél van opció, ez mindig 2-hatványnál fordul elő.

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

Nyelvcsaládok, kifejező erő

$\mathcal{L}(\mathrm{PR}_{\mathrm{ac}})$, illetve $\mathcal{L}(\mathrm{PR}_{\mathrm{ac}}^{\varepsilon})$ jelöli rendre az előfordulásellenőrzéses ε-mentes környezetfüggetlen szabályokból, valamint a tetszőleges környezetfüggetlen szabályokból álló programozott grammatikák által generált nyelvek osztályát.

Ha a grammatika előfordulásellenőrzés nélküli, akkor az ac index hiányzik.

Hasonlóképpen: $\mathcal{L}(\mathrm{MAT}_{\mathrm{ac}})$, $\mathcal{L}(\mathrm{MAT}_{\mathrm{ac}}^{\varepsilon})$, $\mathcal{L}(\mathrm{RC}_{\mathrm{ac}})$, illetve $\mathcal{L}(\mathrm{RC}_{\mathrm{ac}}^{\varepsilon})$.

Tétel

Fennállnak a következők

$$\mathcal{L}_2 \subset \mathcal{L}(\mathrm{PR}_{\mathrm{ac}}) = \mathcal{L}(\mathrm{MAT}_{\mathrm{ac}}) = \mathcal{L}(\mathrm{RC}_{\mathrm{ac}}) \subset \mathcal{L}_1$$

és

$$\mathcal{L}(\mathrm{PR}_{\mathrm{ac}}^{\varepsilon}) = \mathcal{L}(\mathrm{MAT}_{\mathrm{ac}}^{\varepsilon}) = \mathcal{L}(\mathrm{RC}_{\mathrm{ac}}^{\varepsilon}) = \mathcal{L}_0.$$

Nyelvcsaládok, kifejező erő