A primitív és nemprimitív szavak nyelvei

doktori (PhD) értekezés tézisei

Horváth Géza

Debreceni Egyetem
Természettudományi Kar

Debrecen, 2002

Bevezetés

A disszertáció öt fejezetből áll.

Az első fejezetben [4], [5] és [21] alapján rövid áttekintést adunk a formális nyelvek elméletének azon fogalmairól, melyekre szükségünk van a további fejezetekben.

A második fejezet első részében ismertetünk néhány nyelvet, melyek esetében - többnyire a Bar-Hillel Lemma segítségével, egy esetben pedig az adott nyelv karakterizációjának segítségével - bizonyítható, hogy hol helyezkednek el a Chomsky-féle hierarchiában. A fejezet második részében a primitív szavak nyelvét vizsgáljuk, ismertetünk számos iterációs (pumpálós) lemmát, és Dömösi Pál, Horváth Sándor, Ito Masami, Kászonyi László, és Katsura Masashi [6] cikkének segítségével megmutatjuk, hogy egy adott ábécé feletti összes primitív szavak nyelve minden ismertetett iterációs lemma feltételét kielégíti, így ebben az esetben az iterációs lemmák nem alkalmasak annak a kérdésnek az eldöntésére, hogy a primitív szavak nyelve valódi környezetfüggő nyelv-e.

A harmadik fejezetben áttérünk a nemprimitív szavak nyelvére. Mivel a Bar-Hillel Lemmával igazolható, hogy ez a nyelv valódi környezetfüggő nyelv, ezért ebben az esetben a további vizsgálatok a nemprimitív szavakból álló környezetfüggetlen, lineáris és reguláris nyelvek karakterizációira koncentrálnak. Először néhány új fogalmat vezetünk be - sűrű nyelvek és diszjunktív nyelvek - Shyr [15] könyve és Shyr és Thierrin [23] cikke alapján, majd megadjuk a nemprimitív szavakból álló környezetfüggetlen nyelvek karakterizációját Ito és Katsura [18] cikkének segítségével.

A negyedik fejezet a [2] cikkben összefoglalt, Dömösi Pállal és Ito Masamival közös eredményeket tartalmaz. Ebben a fejezetben a harmadik fejezet folytatásaként megadjuk a nemprimitív szavakból álló lineáris és reguláris nyelvek karakterizációit. Végül ezen eredmények ismeretében megadjuk a Chomsky-féle hierarchia különböző szintjein elhelyezkedő, nemprimitív szavakból álló nyelvek hierarchiáját.

Az ötödik fejezetben a csak primitív szavakat generáló, kis környezetfüggetlen nyelvtanok vizsgálatára kerül sor. Ebben az esetben Chomsky-féle normál formát használunk környezetfüggetlen nyelvtanok helyett, a könnyebb kezelhetőség érdekében. Ezt megtehetjük, mivel minden λ-mentes környezetfüggetlen nyelvtanhoz létezik vele ekvivalens Chomsky-féle normál formájú nyelvtan, így amennyiben egy környezetfüggetlen nyelvtan generálja egy adott ábécé felett az összes primitív szót, akkor létezik Chomsky-féle normál formájú nyelvtan is, ami szintén generálja az összes primitív szót az adott ábécé felett. Továbbá elegendő csak azt igazolni egy-egy nyelvtanról, hogy az adott nemterminális ábécé felett nem generál nemprimitív szót, ezért csak a nyelvtanok vázát adjuk meg. 1 és 2 nemterminális esetén könnyű helyzetben vagyunk, a nyelvtanok karakterizációja egyszerű, viszont 3 nemterminális esetén a lehetséges vázak nagy száma miatt szükséges volt számítógép és megfelelő program használata, annak ellenőrzésére, hogy adott hosszúságig mely vázak generálnak csak primitív szót. Itt meg kell jegyeznünk, hogy a fejezet célja nem a fenti feladatot megoldó algoritmus kidolgozása és leprogramozása. Egy igen egyszerű programról van szó, melyet mindössze számítási eszközként használunk. A program által szolgáltatott eredményeket matematikai módszerekkel igazoltuk, ezek bizonyítják, hogy eredményeink helyesek. A program 3 nemterminális esetére 12 maximális vázat szolgáltatott, melyek 12 hosszúságig csak primitív szavakat generálnak, ezek mindegyikéről bizonyítottuk, hogy tetszőleges hosszúságig csak primitív szavakat generálnak, ezzel megadjuk a legfeljebb 3 nemterminális feletti összes, maximális, csak primitív szavakat generáló, Chomsky-féle normál formájú nyelvtan karakterizációját. Eredményeinket a Dömösi Pállal, Manfred Kudlekkel és Dirk Hauschildtel közös [1] dolgozatban foglaltuk össze. A nyelvtanok karakterizációjából kiderül, hogy minden - a fentebbi alaknak megfelelő - nyelvtan a primitív szavak végtelen, valódi részhalmazát generálja, valamint azt is igazolni tudjuk, hogy létezik olyan, csak primitív szavakból álló környezetfüggetlen nyelv, mely nem generálható reguláris nyelvtannal, így a csak primtív szavakból álló környezetfüggetlen nyelvek osztálya bővebb, mint a csak primitív szavakból álló reguláris nyelvek osztálya. 4 nemterminális esetén a maximális számítási kapacitást kihasználva 27 hosszúságig vizsgáltuk a lehetséges vázak által generált szavakat, de ennél a hossznál még mindig 413 maximális vázat kaptunk, melyek száma túl nagy ahhoz, hogy mindegyik esetről bizonyítsuk, hogy 27 hosszúság felett is csak primitív szavakat generálnak. Ugyanakkor megállapíthatjuk, hogy ha egy nyelvtan egy legalább kétbetűs ábécé felett generálja az összes primitív szót, és nem generál egyetlen nemprimtív szót sem, akkor a mondatszimbólum nem szerepelhet egyetlen, a nyelvtan által generált legalább 2 hosszúságú szóban sem. Ezek után azokra a nyelvtanokra korlátozzuk vizsgálatainkat, melyekben egyik szabály jobboldalán sem szerepel a mondatszimbólum, hisz csak ezek a nyelvtanok generálhatják az összes primitív szót 3 nemterminális betű felett. A program ebben az esetben - ismét 12 hosszúságig vizsgálva a fentebbi alaknak megfelelő lehetséges vázakat - 30 maximális vázat adott, melyek mindegyikéről bizonyítottuk, hogy 12 hosszúság felett is csak primitív szavakat generálnak, ezzel megadjuk a fentebbi alaknak megfelelő összes váz karakterizációját. Ezen eredményeket a [3] dolgozatban foglaljuk össze. A kapott vázak mindegyike ebben az esetben is végtelen, valódi részhalmazát generálja a 3 nemterminális feletti összes primitív szó által alkotott nyelvnek. Az 5 nemterminális esetének vizsgálatára jelenleg nem áll rendelkezésre a megfelelő számítási kapacitás, de várhatóan a program által szolgáltatott vázak száma tovább emelkedne a 4 nemterminális esetében - megszorítások nélül - kapott 413 vázhoz képest.

1. Alapfogalmak

Ebben a fejezetben áttekintést adunk az értekezésben felhasznált fogalmakról.

Ábécé alatt egy véges, nemüres halmazt értünk: $0 < |V| < \infty$. Az ábécé elemeit *betűknek* hívjuk. Az ábécét V-vel jelöljük. Betűk egy véges láncát *szónak* nevezzük. Egy P *szó hosszán* a szót alkotó betűk számát értjük ismétlődésekkel együtt. (Jele: $|P|$.) Az egyetlen betűt sem tartalmazó szót *üres szónak* nevezzük és λ-val jelöljük. V^* jelöli az V ábécé feletti összes szavak halmazát, és V^+ jelöli a V ábécé feletti összes, nemüres szavak halmazát. $V^+ = V^* \setminus \{\lambda\}$. A $p = x_1x_2\dots x_n$ és a $q = y_1y_2\dots y_m$ *szó egyenlő*, ha $n = m$ és $x_i = y_i$ $\forall i \in \{1\dots n\}$-re.

1.1. Definíció A $p = x_1\dots x_m$ és $q = y_1\dots y_n$ $(x_1,\dots,x_m,y_1,\dots,y_n \in V)$ szavak *szorzatán* a $pq = x_1\dots x_my_1\dots y_n$ szót értjük.

A szavak szorzása mint művelet általában nem kommutatív. Egy V^*-beli p szó és az üres szó $p\lambda$ szorzatán magát a p szót értjük. Ugyanígy, az üres szó és egy V^+-beli p szó λp szorzatán magát a p szót értjük. Továbbá $\lambda\lambda = \lambda$. Mivel $(pq)r = p(qr)$ minden $p, q, r \in V^+$ esetén teljesül, vagyis a szavak szorzása mint művelet a V^+ elemeire nézve asszociatív, V^+ a szavak szorzására nézve félcsoportot alkot, melyet V feletti szabad félcsoportnak is hívunk. Hasonlóan, tekintettel arra, hogy $(pq)r = p(qr)$ minden $p, q, r \in V^*$ esetén teljesül, azaz a szavak szorzása V^* elemeire nézve is asszociatív, továbbá arra, hogy $\lambda p = p\lambda = p$, azaz λ egységelemet alkot V^*-ban a szorzásra nézve, a V^* a szavak szorzására nézve egységelemes félcsoportot, más néven monoidot alkot, melyet V feletti egységelemes szabad félcsoportnak vagy V feletti szabad monoidnak is hívunk.

1.2. Definíció Legyen $q = p_1\dots p_n$ alakú, ahol $p_1,\dots,p_n$ egyenlő szavak. Ekkor a q szót a p szó n-edik *hatványának* nevezzük. (Jele: p^n.)

1.3. Definíció Egy szó *hatványszó*, ha bármely szónak legalább második hatványa.

1.4. Definíció Bármely szó 0. hatványa az *üres szó*.

1.5. Definíció A p szó *részszava* a q szónak, ha léteznek r, s szavak úgy, hogy $q = rps$.

1.6. Definíció A p szó *kezdőszelete* a q szónak, ha létezik olyan s szó, melyre $q = ps$.

1.7. Definíció A p szó *végződése* a q szónak, ha létezik r szó úgy, hogy $q = rp$.

1.8. Definíció Egy adott ábécé feletti összes szavak egy részhalmazát az adott ábécé feletti *formális nyelvnek*, vagy röviden csak *nyelvnek* nevezzük. Jele: L, $(L \subseteq V^*)$.

1.9. Definíció Azt a nyelvet, amelynek egyetlen szava sincs, *üres nyelvnek* nevezzük.

1.10. Definíció Egy nyelv *véges*, ha csak véges sok szót tartalmaz, egyébként *végtelen*.

Műveletek: A formális nyelvekre, mint szóhalmazokra közvetlenül értelmezhetők a halmazelméleti alapműveletek.

$$L_1 \cup L_2 = \{P \mid P \in L_1 \text{ vagy } P \in L_2\}$$
$$L_1 \cap L_2 = \{P \mid P \in L_1 \text{ és } P \in L_2\}$$
$$L_1 \setminus L_2 = \{P \mid P \in L_1 \text{ és } P \notin L_2\}$$
$$\overline{L} = V^* \setminus L$$

1.11. Definíció Két nyelv *konkatenációján* a következő nyelvet értjük:

$$L_1 \cdot L_2 = \{PQ \mid P \in L_1 \text{ és } Q \in L_2\}.$$

1.12. Definíció Legyen $i = 1, 2, \ldots$. Ekkor egy L nyelv i-edik hatványán a nyelv i-szer egymás utáni, önmagával való konkatenációját értjük. Jelölés: L^i. Megállapodás szerint $L^0 = \{\lambda\}$.

1.13. Definíció A *konkatenáció lezárását* az $L^* = \bigcup_{i=0}^{\infty} L^i$ összefüggéssel értelmezzük. Használatos még az $L^+ = L^* \setminus \{\lambda\}$ jelölés is.

1.14. Definíció A $G = (V_N, V_T, S, H)$ rendezett négyest generatív grammatikának vagy *generatív nyelvtannak* nevezzük, ha $V = V_N \cup V_T$, V_N és V_T diszjunkt véges ábécék, azaz $V_N \cap V_T = \emptyset$, $S \in V_N$, $H \subseteq (V_N \cup V_T)^* V_N (V_N \cup V_T)^* \times (V_N \cup V_T)^*$. A V_N elemeit *nemterminális jeleknek* vagy változóknak nevezzük, és általában nagybetűkkel jelöljük. A V_T elemeit *terminális jeleknek* vagy konstansoknak nevezzük, és általában kisbetűkkel jelöljük. Az S egy kitüntetett nemterminális jel, amely a G nyelvtanban a generálás kezdőeleme. S-et *mondatszimbólumnak* hívjuk. A H elemeit képező (P, Q) rendezett párokat *helyettesítési szabályoknak* nevezzük, és általában $P \rightarrow Q$ alakban írjuk. $0 < |H| < \infty$. A H elemei olyan helyettesítési szabályok, amelyek baloldala tartalmaz legalább egy nemterminálist.

1.15. Definíció A Q szó a P szóból *egy lépésben levezethető*, ha léteznek $A, B \in (V_N \cup V_T)^*$ szavak úgy, hogy $P = ARB$, $Q = ATB$ és létezik $R \rightarrow T \in H$ szabály. (Jele: $P \Rightarrow Q$.)

1.16. Definíció *Több lépésben levezethető* a Q szó a P szóból, ha léteznek $R_1, \ldots, R_n$ szavak úgy, hogy $R_1 = P$, $R_n = Q$, $R_i \Rightarrow R_{i+1}$, $i = 1, \ldots, n-1$. (Jele: $P \overset{+}{\Rightarrow} Q$.)

1.17. Definíció A Q szó a P szóból *levezethető*, ha $P \Rightarrow Q$ vagy $P \overset{+}{\Rightarrow} Q$. (Jele: $P \overset{*}{\Rightarrow} Q$.)

1.18. Definíció A $G = (V_N, V_T, S, H)$ generatív nyelvtan által *generált nyelven* az $L(G) = \{P \mid S \overset{*}{\Rightarrow} P,\ P \in {V_T}^*\}$ halmazt értjük.

1.19. Definíció Két generatív grammatika *ekvivalens*, ha az általuk generált nyelv megegyezik.

1.20. Definíció Chomsky-féle hierarchia

A $G = (V_N, V_T, S, H)$ *nyelvtan i típusú*, ha az alábbi megkötések közül az i. teljesül rá.

0. típusú: mondatszerkezetű nyelvtan

A generatív nyelvtan általános definícióján kívül nincs külön megkötés.

1. típusú: környezetfüggő nyelvtan
Minden H-beli szabály $PXQ \to PRQ$ alakú, ahol $P, Q \in (V_N \cup V_T)^*$, $X \in V_N$, $R \in (V_N \cup V_T)^+$, vagy pedig $S \to \lambda$ alakú, de ha $S \to \lambda \in H$, akkor S nem fordulhat elő egyetlen H-beli szabály jobboldalán sem.

2. típusú: környezetfüggetlen nyelvtan
Minden H-beli szabály $P \to Q$ alakú, ahol $P \in V_N$, $Q \in (V_N \cup V_T)^*$.

3. típusú: jobbról lineáris vagy reguláris nyelvtan
Minden H-beli szabály $P \to a$ vagy $P \to aQ$ alakú, ahol $P, Q \in V_N$, $a \in {V_T}^*$.

1.21. Definíció Az $i = 0, 1, 2, 3$ értékek esetén azt mondjuk, hogy egy L *nyelv* i *típusú*, ha L-hez van olyan i-típusú G nyelvtan, amely azt generálja: $L = L(G)$. Az i típusú nyelvek osztályát $\mathcal{L}_i$-vel jelöljük.

1.1. Tétel [4] *Minden 3. típusú grammatika egyben 2. típusú, és minden 1. típusú egyben 0. típusú is. Tehát* $\mathcal{L}_0 \supseteq \mathcal{L}_1$ *és* $\mathcal{L}_2 \supseteq \mathcal{L}_3$.

1.2. Tétel (Üres szó lemma) [4] *Minden* $G = (V_N, V_T, S, H)$ *2. típusú nyelvtanhoz létezik olyan 1. típusú* $G' = ({V_N}', V_T, S', H')$ *nyelvtan, amelyre* $L(G) = L(G\,')$.

1.1. Következmény *A Chomsky féle nyelvosztályok hierarchiát alkotnak, azaz* $\mathcal{L}_0 \supseteq \mathcal{L}_1 \supseteq \mathcal{L}_2 \supseteq \mathcal{L}_3$.

Megmutatható, hogy a fenti hierarchia valódi, azaz $\mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \mathcal{L}_3$.

1.22. Definíció Azt mondjuk, hogy egy $G = (V_N, V_T, S, H)$ nyelvtan *normális alakban* van, ha a helyettesítési szabályokban terminális jelek csak $X \to a$ ($X \in V_N$, $a \in V_T$) alakú szabályokban fordulnak elő.

1.3. Tétel [4] *Minden nyelvtanhoz megadható vele ekvivalens normális alakú nyelvtan.*

1.23. Definíció Egy nyelvtant λ - *mentesnek* nevezünk, ha a szabályok jobboldalán egyáltalán nem fordul elő a λ, ezért a nyelvtan által generált nyelv nem tartalmazza az üres szót.

1.24. Definíció Egy $G = (V_N, V_T, S, H)$ környezetfüggetlen nyelvtanról azt mondjuk, hogy *Chomsky-féle normál alakú*, ha minden szabálya $X \to a$ vagy $X \to YZ$ alakú, ahol $X, Y, Z \in V_N$ és $a \in V_T$.

1.4. Tétel [4] *Minden* λ*-mentes környezetfüggetlen* G *grammatikához megadható vele ekvivalens Chomsky-féle normál alakú* G' *grammatika.*

2. Szavak és nyelvek kombinatorikus tulajdonságai

A nyelvek kombinatorikus tulajdonságai vizsgálatának középpontjában áll az adott nyelvek Chomsky-féle hierarchiában elfoglalt helye. Vannak olyan nyelvek, melyekről már bizonyított, hogy pontosan melyik szinten helyezkednek el a hierarchiában, ilyen például egy adott ábécé feletti összes palindrómák- vagy reptitív szavak nyelve. Megoldatlan probléma, hogy az összes primitív szavak nyelve egy adott legalább kétbetűs ábécé felett környezetfüggő-e, - mint azt feltételezzük, - vagy pedig környezetfüggetlen nyelvtannal is generálható. Ismertetni fogjuk a palindromikus- és repetitív nyelvek pontos helyét a Chomsky-hierarchiában, valamint a primitív szavak vizsgálatánál használt számos eljárást, melyek sok esetben segítenek annak a kerdésnek az eldöntésében, hogy egy adott nyelv generálható-e környezetfüggetlen nyelvtannal.

Legyen X egy ábécé. A $p = x_1...x_n$ $(x_1, ..., x_n \in X)$ *szó inverze* a $p^R = x_n...x_1$ szó. A p szó *palindróma*, ha $p = p^R$. (A λ üres szó triviálisan palindróma.) Ezek után az $L \subseteq X^*$ nyelv *palindromikus*, ha minden szava palindróma.

2.1. Tétel [12] *Az $L \subseteq X^*$ reguláris nyelv palindromikus akkor és csak akkor, ha véges uniója az alábbi formájú nyelveknek:*

$$L_p = \{p\},\ L_{q,r,s} = qr(sr)^*q^R \quad (p, q, r, s \in X^*)$$

ahol p, r és s palindróma.

2.2. Tétel [12] *Az $L \subseteq X^*$ környezetfüggetlen nyelv palindromikus akkor és csak akkor, ha az alábbi formájú:*

$$L = \bigcup_{a \in X \cup \{\lambda\}} \{pap^R \mid p \in L(a)\}$$

ahol az $L(a)$ $(a \in X \cup \{\lambda\})$ nyelvek reguláris nyelvek.

2.3. Tétel [6] *Ha $|X| > 1$ akkor az X feletti összes palindrómák halmaza környezetfüggetlen.*

Az $|X| = 1$ esetben az X^* nyelv minden szava palindróma és X^* egy igen egyszerű reguláris nyelv.

A p nemüres szó *négyzetszó*, ha létezik x szó, hogy $p = x^2$. A p szó *repetitív*, ha tartalmaz négyzetszót. A p szó *nemrepetitív* (vagy *négyzetmentes*), ha nem tartalmaz négyzetszót. Látható, hogy a négyzetmentes szavak egy kétbetűs ábécé felett maximum három hosszúságúak lehetnek. Thue a [17] cikkében bizonyította, hogy egy legalább hárombetűs ábécéből végtelen sok nemrepetitív szó képezhető.

Az alábbi - Bar-Hillel Lemma néven ismertté vált - tétel klasszikus példája az iterációs lemmáknak. Egy olyan - legtöbb esetben könnyen ellenőrizhető - megkötést tesz a nyelv bizonyos szavainak alakjára, mely teljesül minden környezetfüggetlen nyelvre, ezért számos esetben jól alkalmazható egy adott nyelv nem környezetfüggetlen voltának igazolására.

2.4. Tétel (Bar-Hillel Lemma) [8] *Minden L környezetfüggetlen nyelvhez léteznek n, m természetes számok úgy, hogy $\forall\ p \in L$, $|p| > n$ szóra $p = uvwxy$ alakban írható, ahol $|vwx| \leq m$, $|vx| > 0$ és $uv^iwx^iy \in L$ minden $i \geq 0$ egész számra.*

A Bar-Hillel Lemma segítségével könnyen belátható, hogy a nemrepetitív szavak egy tetszőleges halmaza egy adott ábécé felett környezetfüggetlen akkor és csak akkor, ha véges. Figyelembe véve, hogy egy legalább hárombetűs ábécé felett végtelen sok nemrepetitív szó található, következik az alábbi tétel:

2.5. Tétel [6] *Az összes nemrepetitív szavak halmaza egy legalább hárombetűs ábécé felett nem környezetfüggetlen.*

Mivel az $X^* \setminus L$ nyelv reguláris, ha $L \subseteq X^*$ véges, ezért látható, hogy az összes repetitív szavak halmaza egy egybetűs vagy kétbetűs ábécé felett reguláris nyelvet alkot. A több betűs esetre az alábbi tétel vonatkozik:

2.6. Tétel [14] *Egy adott, legalább hárombetűs ábécé feletti összes repetitív szavak halmaza nem környezetfüggetlen.*

Végezetül meg kell jegyeznünk, könnyen bizonyítható, hogy környezetfüggő nyelvtannal generálható egy adott ábécé feletti összes repetitív és nemrepetitív szavak nyelve is.

A p szó *primitív* szó, ha nem áll elő egy másik szó hatványaként. Pontosabban megfogalmazva bármely w szóra és $i \geq 2$ egészre $p \neq w^i$. A p szó *nemprimitív* szó (vagy *hatványszó*,) ha létezik w szó és $i \geq 2$ egész, hogy $p = w^i$. Így az üres szó $\lambda\lambda = \lambda$ miatt nemprimitív. Jelölje Q az összes primitív szavak halmazát az X ábécé felett.

2.1. Definíció Egy L nyelv determinisztikus környezetfüggetlen, ha felismerhető determinisztikus veremautomatával.

A determinisztikus környezetfüggetlen nyelvek valódi részhalmazát alkotják a környezetfüggetlen nyelveknek.

2.7. Tétel [6] *Az L nyelv determinisztikus környezetfüggetlen akkor és csak akkor, ha $X^* \setminus L$ determinisztikus környezetfüggetlen nyelv.*

A fenti tétel értelmében annak bizonyításához, hogy Q nem eleme a determinisztikus környezetfüggetlen nyelveknek, elegendő bebizonyítani, hogy az $X^* \setminus Q$ nem elégíti ki a Bar-Hillel feltételt. Mivel ez igaz $|X| > 1$ esetén, ezért igaz az alábbi tétel:

2.8. Tétel [6] *Q nem eleme a determinisztikus környezetfüggetlen nyelvek halmazának, ha $|X| > 1$.*

Ezek után felmerül a kérdés, hogy $|X| > 1$ esetén Q valódi környezetfüggő nyelv-e, vagy pedig (nem determinisztikus) környezetfüggetlen nyelv?

Első lehetőség természetesen a Bar-Hillel Lemma alkalmazása a Q nyelvre, ám ez nem vezetett eredményre, mivel ismert az alábbi tétel:

2.9. Tétel [6] *Q eleget tesz a Bar-Hillel Lemma feltételeinek.*

Mivel a Bar-Hillel Lemma nem vezetett eredményre, más, a környezetfüggetlen nyelvekre teljesülő, erősebb, vagy legalábbis nem gyengébb iterációs és egyéb feltételeket próbáltak alkalmazni a Q nyelvre. A jelenleg rendelkezésünkre álló, minden környezetfüggetlen nyelvre teljesülő, a Bar-Hillel Lemmától erősebb feltételek közül először az Ogden feltételt ismertetjük.

2.2. Definíció (Ogden feltétel) [13] Legyen $L \subseteq X^*$ egy adott nyelv. Tegyük fel, hogy létezik csak az L-től függő $n \geq 2$ egész úgy, hogy ha $z \in L$ és megjelölünk több mint n "kitüntetett" pozíciót z-ben, akkor z felírható $z = uvwxy$ alakban úgy, hogy teljesülnek az alábbi feltételek:

1. vagy u, v, w vagy pedig w, x, y mindegyike tartalmaz "kitüntetett" pozíciót,
2. vwx legfeljebb n "kitüntetett" pozíciót tartalmaz,
3. $uv^m wx^m y \in L$ minden $m \geq 0$ egészre.

Ekkor azt mondjuk, hogy L kielégíti az Ogden feltételt.

2.10. Tétel [13] *Minden környezetfüggetlen nyelv kielégíti az Ogden feltételt.*

2.11. Tétel [6] *Q kielégíti az Ogden feltételt.*

Mivel bizonyításra került, hogy az Ogden feltételt is kielégíti a Q nyelv, ezért a környezetfüggetlen nyelvekre ismert pumpálós lemmák közül egy még erősebb, az erős Bader-Moura feltétel került alkalmazásra.

2.3. Definíció (Erős Bader-Moura feltétel) [11] Legyen $L \subseteq X^*$ egy adott nyelv. Tegyük fel, hogy létezik csak az L-től függő $n \geq 2$ egész úgy, hogy ha $z \in L$ és megjelölünk d "kitüntetett" pozíciót, valamint e "kizárt" pozíciót z-ben, (egy pozíció lehet egyszerre "kitüntetett" és "kizárt" is,) ahol $d > n^{e+1}$, akkor z felírható $z = uvwxy$ alakban úgy, hogy teljesülnek az alábbi feltételek:

1. vagy u, v, w vagy pedig w, x, y mindegyike tartalmaz "kitüntetett" pozíciót, valamint vx nem tartalmaz kizárt pozíciót,
2. Ha d' és e' a "kitüntetett" és "kizárt" pozíciók számát jelöli vwx-ben, akkor $d' \leq n^{e'+1}$,

3. $uv^m wx^m y \in L$ minden $m \geq 0$ egészre.

Ekkor azt mondjuk, hogy L kielégíti az erős Bader-Moura feltételt.

2.12. Tétel [11] *Minden környezetfüggetlen nyelv eleget tesz az erős Bader-Moura feltételnek.*

2.13. Tétel [6] *Q eleget tesz az erős Bader-Moura feltételnek.*

Vannak az ismertetett lemmáktól erősebb pumpálós lemmák is, de Q azon lemmák feltételeinek is eleget tesz, ezért ezzel a módszerrel nem lehet bizonyítani Q környezetfüggőségét.

2.4. Definíció (Sokolowski feltétel) [16] Az L nyelv kielégíti a Sokolowski feltételt, ha minden $Y \subseteq X$, $|Y| \geq 2$ ábécére és minden $u_1, u_2, u_3 \in X^*$ szóra, ha $\{u_1xu_2xu_3 \mid x \in Y^+\} \subseteq L$, akkor léteznek $x', x'' \in Y^+$, $x' \neq x''$ szavak úgy, hogy $u_1x'u_2x''u_3 \in L$.

2.14. Tétel [16] *Minden környezetfüggetlen nyelv kielégíti a Sokolowski feltételt.*

A Sokolowski feltételről bebizonyosodott, hogy az erős Bader-Moura feltétel teljesülése maga után vonja a Sokolowski feltétel teljesülését is, ezért igaz az alábbi tétel:

2.15. Tétel [6] *Q kielégíti a Sokolowski feltételt.*

Klasszikus lehetőség, hogy keresünk egy R reguláris nyelvet, melyre $Q \cap R$ nem környezetfüggetlen, ezzel igazoljuk, hogy Q sem az. Mindezidáig egyetlen ilyen reguláris nyelvet sem sikerült találni. A vizsgálatok során alkalmazták az $R_n = (a^*b)^n$, $a, b \in X$, $a \neq b$, $n \geq 1$ nyelvet, de bizonyításra került, hogy ha n legfeljebb 4 különböző prim hatvány szorzataként, vagy négynél több olyan $p_1, ..., p_n$ prim hatvány szorzataként előáll, hogy $1/p_1 + ... + 1/p_n \leq 4/5$, akkor $Q \cap R$ ebben az esetben környezetfüggetlen. Nyitott kérdés, hogy ez minden n-re fennáll-e.

3. Nemprimitív szavakból álló környezetfüggetlen nyelvek

Ebben a fejezetben megadjuk a nemprimitív szavakból álló környezetfüggetlen nyelvek felépítését.

Először néhány új fogalommal ismerkedünk meg. Legyen A egy adott nyelv. Jelölje $A^{(i)}$ a következő nyelvet: $A^{(i)} = \{u^i \mid u \in A\}$, $i \geq 1$ egész. Azt mondjuk, hogy az u és v szavak *konjugáltak*, ha $u = xy$ és $v = yx$ valamely $x, y \in X^*$ szavakra. Egy $A \subseteq X^*$ nyelvet *sűrű nyelvnek* hívunk, ha $X^*uX^* \cap A \neq \emptyset$ igaz minden $u \in X^*$ szóra. A definícióból következik, hogy a sűrű nyelvek végtelenek. Megjegyezzük még, hogy bármely $u \in X^+$ szó egyértelműen felírható $u = p^n$, $p \in Q$, $n \geq 1$ alakban, ezért $X^+ = \bigcup_{n\geq 1} Q^{(n)}$. Könnyen látható, hogy Q és $X^+ \setminus Q$ is eleme a sűrű nyelvek osztályának.

Legyen $A \subseteq X^*$. Az X^* feletti P_A ekvivalencia relációt az A által adott *fő kongruenciának* hívjuk, és a következőképp definiáljuk: $u \equiv v$ (P_A) akkor és csak akkor, ha $xuy \in A \Leftrightarrow xvy \in A$ teljesül az összes $x, y \in X^*$ párra. Egy $A \subseteq X^*$ nyelv *reguláris*, ha a P_A relációhoz tartozó ekvivalencia osztályok száma véges. A *diszjunktív nyelvek* osztályát a következőképpen definiáljuk: az $A \subseteq X^*$ nyelv diszjunktív, ha teljesül minden $u, v \in X^*$ szóra, hogy ha $u \equiv v$ (P_A), akkor $u = v$. Egy tipikus példa a diszjunktív nyelvekre a Q.

A továbbiakban ismertetjük azon környezetfüggetlen nyelvek felépítését, melyek $X^+ \setminus Q$ valamint a $Q^{(2)}$ nyelv egy részhalmazát tartalmazzák.

3.1. Tétel [18] *Legyen $L \subseteq X^*$ egy olyan környezetfüggetlen nyelv, melyre $L \subseteq X^+ \setminus Q$. Ekkor $L_1 = L \cap Q^{(2)}$ környezetfüggetlen nyelv és $L_2 = L \cap (\bigcup_{i\geq 3} Q^{(i)})$ reguláris nyelv. Pontosan megadva, L_2 felírható az alábbi formában:*

$$L_2 = (\bigcup_{1\leq i\leq r} f_i^{m_i}(f_i^{k_i})^*) \cup F$$

ahol $f_i \in Q$, $m_i \geq 3$, $k_i \geq 1$ $(1 \leq i \leq r)$ és $F \subseteq \bigcup_{i\geq 3} Q^{(i)}$ véges halmaz.

3.1. Következmény [18] *Ha $i \geq 3$, akkor nem létezik $L \subseteq X^*$ végtelen környezetfüggetlen nyelv úgy, hogy $L \subseteq Q^{(i)}$.*

3.2. Tétel [18] *Legyen $L \subseteq X^*$ a $Q^{(2)}$ elemeiből álló környezetfüggetlen nyelv. Ekkor*

$$L = F \cup (\bigcup_{1\leq i\leq r} \{(a_i^n b_i a_i^m)^2 \mid n, m \geq 1\}) \cup (\bigcup_{1\leq j\leq s} \{(g_j h_j^n f_j)^2 \mid n \geq 1\})$$

ahol $b_i a_i^2 \in Q$ $(1 \leq i \leq r)$, $f_j g_j h_j \in Q$ $(1 \leq j \leq s)$ és $F \subseteq Q^{(2)}$ véges halmaz.

Ezek után belátható, hogy az $\{(a_i^n b_i a_i^m)^2 \mid n, m \geq 1\}$ $(1 \leq i \leq r)$ nyelv, és a $\{(g_j h_j^n f_j)^2 \mid n \geq 1\}$ $(1 \leq j \leq s)$ nyelv is környezetfüggetlen.

3.2. Következmény [18] *Nem létezik olyan $L \subseteq X^*$ sűrű környezetfüggetlen nyelv, melyre $L \subseteq Q^{(2)}$.*

Tudjuk, hogy az $L \subseteq X^*$ sűrű reguláris nyelvre $L \cap Q$ diszjunktív nyelv lesz. Ugyanakkor a sűrű környezetfüggetlen nyelvekre csak az alábbi megállapítást tehetjük:

3.3. Tétel [18] *Legyen $L \subseteq X^*$ sűrű környezetfüggetlen nyelv. Ekkor $L \cap Q$ sűrű nyelv.*

Továbbra is megoldatlan probléma, hogy az $L \cap Q$ diszjunktív nyelv-e, ha $L \subseteq X^*$ diszjunktív környezetfüggetlen nyelv.

4. Nemprimitív szavakból álló lineáris és reguláris nyelvek

Könnyen belátható, hogy megfelelő lineárisan korlátolt (determinisztikus) automatával eldönthető egy tetszőleges szóról, hogy primitív szó-e. Szintén egyszerűen bizonyítható, hogy minden $i \geq 2$ esetén lineárisan korlátolt (determinisztikus) automatával eldönthető, hogy egy adott szó eleme-e a $Q^{(i)}$ nyelvnek. Ugyancsak lineárisan korlátolt (determinisztikus) automata használatával eldönthető minden szóról, hogy eleme-e az $\bigcup_{i\geq 2} Q^{(i)}$ nyelvnek. Mindezekből következik, hogy ezek a nyelvek környezetfüggőek. Továbbá ismert, hogy a környezetfüggő nyelvek metszete is környezetfüggő, ezért minden L környezetfüggő nyelvre az $L \cap Q$, $L \cap Q^{(i)}$, $i \geq 2$ és $L \cap (\bigcup_{i\geq 2} Q^{(i)})$ nyelvek mindegyike környezetfüggő lesz. Ezekből az egyszerű tényekből következik, hogy a nemprimitív szavakból álló környezetfüggő nyelvek mindegyike felírható az alábbi formában: $L = L' \cap (\bigcup_{i\geq 2} Q^{(i)})$, ahol L' környezetfüggő nyelv. Másképp fogalmazva, minden L' környezetfüggő nyelvre az $L = L' \cap (\bigcup_{i\geq 2} Q^{(i)})$ nyelv is környezetfüggő lesz. Mindezeket figyelembe véve látható, hogy a nemprimitív szavakból álló környezetfüggő (vagy bővebb) nyelvek felépítésének további vizsgálatára nincs szükség.

Ito és Katsura megadta a nemprimitív szavakból álló környezetfüggetlen nyelvek felépítését, melyet összefoglaltunk a 3. fejezetben. Ebben a fejezetben megadjuk a nemprimitív szavakból álló lineáris és reguláris nyelvek karakterizációját, ezzel teljessé téve a Chomsky hierarchia különböző szintjein elhelyezkedő, nemprimitív szavakból álló nyelvek felépítéseinek meghatározását. A fejezetben kidolgozott eredmények a szerző és a [2] dolgozatban közreműködő társszerzők közös munkájának eredménye.

Elsőként meg kell határoznunk, mely nyelveket tekintünk lineáris nyelveknek. Lineáris nyelvek azok a nyelvek, melyek generálhatóak olyan nyelvtannal, melyben minden H-beli szabály $P \to p$ vagy $P \to aRb$ alakú, ahol $P, R \in V_N$, $p, a, b \in {V_T}^*$. Például egy legalább kétbetűs ábécé feletti összes palindrómák nyelve lineáris nyelv.

Elsőként megadjuk a nemprimitív szavakból álló lineáris nyelvek karakterizációját:

4.1. Tétel *Legyen L olyan lineáris nyelv, melyre $L \subseteq X^+ \setminus Q$. Ekkor L felírható $L = L_1 \cup L_2$ alakban, ahol $L_1 = L \cap Q^{(2)}$ lineáris nyelv, és $L_2 = L \cap (\bigcup_{i\geq 3} Q^{(i)})$ reguláris nyelv. Pontosan megadva:*

$$L_1 = F_1 \cup (\bigcup_{1\leq j\leq s} \{(f_j g_j^n h_j)^2 \mid n \geq 1\})$$

ahol $F_1 \subseteq Q^{(2)}$ véges halmaz, valamint $f_j g_j h_j \in Q$, $1 \leq j \leq s$.
L_2 felépítése megegyezik a környezetfüggetlen esettel, azaz

$$L_2 = F_2 \cup (\bigcup_{1\le i\le r} f_i^{m_i}(f_i^{k_i})^*)$$

ahol $F_2 \subseteq \bigcup_{i\ge 3} Q^{(i)}$ *véges halmaz és* $f_i \in Q,\ m_i \ge 3,\ k_i \ge 1,\ 1 \le i \le r$.

A következőkben a reguláris esetre fogalmazunk meg egy tételt.

4.2. Tétel *Legyen* L *reguláris nyelv úgy, hogy* $L \subseteq X^+ \setminus Q$. *Ekkor* $L_1 = L \cap Q^{(2)}$ *véges nyelv, és* $L_2 = L \cap (\bigcup_{i\ge 3} Q^{(i)})$ *reguláris nyelv, melynek felépítése megegyezik a környezetfüggetlen és a lineáris esettel.*

A továbbiakban összefoglaljuk a Chomsky hierarchia különböző szintjein elhelyezkedő, nemprimitív szavakból álló nyelvek felépítéseit:

4.1. Következmény *Legyenek* $\mathcal{L}_0,\ \mathcal{L}_1,\ \mathcal{L}_2,$ *és* $\mathcal{L}_3$ *az alábbi nyelvosztályok:*

$\mathcal{L}_0$ *az* $X^+ \setminus Q$ *elemeiből álló véges nyelvek osztálya,*

$\mathcal{L}_1$ *azon nyelvek osztálya, melyek az alábbi alakúak:*

$$\bigcup_{1\le i\le r} f_i^{m_i}(f_i^{k_i})^*,\ f_i \in Q,\ m_i \ge 3,\ k_i \ge 1,\ 1 \le i \le r,$$

$\mathcal{L}_2$ *az alábbi formájú nyelvek osztálya:*

$$\bigcup_{1\le j\le s} \{(f_j g_j^n h_j)^2 \mid n \ge 1\},\ f_j g_j h_j \in Q,\ 1 \le j \le s,$$

$\mathcal{L}_3$ *azon nyelvek osztálya, melyek struktúrája a következő:*

$$\bigcup_{1\le i\le r} \{(a_i^n b_i a_i^m)^2 \mid n, m \ge 1\},\ \text{ahol } a_i^2 b_i \in Q,\ 1 \le i \le r.$$

Ekkor a következőket mondhatjuk:
(a) L *nemprimitív szavakból álló környezetfüggetlen nyelv akkor és csak akkor, ha* $L = L_0 \cup L_1 \cup L_2 \cup L_3,\ L_i \in \mathcal{L}_i,\ i = 0, 1, 2, 3.$
(b) L *nemprimitív szavakból álló lineáris nyelv akkor és csak akkor, ha* $L = L_0 \cup L_1 \cup L_2,\ L_i \in \mathcal{L}_i,\ i = 0, 1, 2.$
(c) L *nemprimitív szavakból álló reguláris nyelv akkor és csak akkor, ha* $L = L_0 \cup L_1,\ L_i \in \mathcal{L}_i,\ i = 0, 1.$
(d) L *nemprimitív szavakból álló véges nyelv akkor és csak akkor, ha* $L = L_0,\ L_0 \in \mathcal{L}_0.$

Az (a) állítás Ito és Katsura eredménye, melyet az előző fejezetben ismertettünk, a (d) állítás pedig triviális. Megemlítjük még, hogy néhány nevezetes, speciális alakú, nemprimitív szavakból álló környezetfüggő nyelv karakterizációja még megoldatlan probléma.

5. Primitív szavakat generáló, Chomsky-féle normál formájú kis nyelvtanok

A második fejezetben felmerült a kérdés, hogy egy adott ábécé feletti összes primitív szavak nyelve valódi környezetfüggő nyelv-e, vagy pedig generálható környezetfüggetlen nyelvtannal is. Ez a kérdés továbbra is nyitott. A kutatások segítése érdekében, valamint mivel önmagában is érdekes a kérdés, ebben a fejezetben megadjuk az összes, legfeljebb három nemterminálisból álló, Chomsky-féle normál formájú, csak primitív szavakat generáló nyelvtant. Ezek a nyelvtanok kicsik és maximálisak. Kicsik abban az értelemben, hogy legfeljebb három nemterminálist tartalmaznak, és maximálisak, mivel bármely újabb szabályt hozzájuk véve már generálnak nemprimitív szót is. Chomsky-féle normál formát használunk környezetfüggetlen nyelvtanok helyett, a könnyebb kezelhetőség érdekében. Mivel minden Chomsky-féle normál formájú nyelvtanhoz létezik vele ekvivalens Chomsky-féle normál formájú nyelvtan, melyben minden nemterminális betűből legalább egy, csak terminálisokból álló szó levezethető, ezért elegendő ezekkel a nyelvtanokkal foglalkoznunk. Ebben az esetben ahhoz, hogy egy nyelvtan csak primitív szavakat generáljon a terminálisok ábécéje felett, elengedhetetlenül szükséges, hogy a nemterminálisok ábécéje felett is csak primitív szavakat generáljon, ezért elegendő megadnunk a nyelvtanok vázát. A kapott nyelvtanok vizsgálatából kiderül, hogy létezik olyan Chomsky-féle normál formájú, három nemterminálisból álló nyelvtan, mely által generált nyelv nem reguláris, ezért a környezetfüggetlen, csak primitív szavakból álló nyelvek osztálya bővebb, mint a reguláris, csak primitív szavakból álló nyelvek osztálya. Az is látható, hogy minden, a fentebbi alaknak megfelelő nyelvtan a primitív szavak végtelen halmazát generálja.

A fejezet hátralévő részében *nyelvtan* alatt $G = (N, \Sigma, S, P\}$ alakban adott, λ-mentes, Chomsky-féle normál formájú nyelvtant értünk. A G nyelvtan *mondatformáinak halmaza* az $S(G) = \{W \mid W \in (N \cup \Sigma)^*,\ S \stackrel{*}{\Rightarrow} W\}$ halmaz.

A $G_1 = (N_1, \Sigma_1, S_1, P_1)$ nyelvtan *betű-izomorf* a $G_2 = (N_2, \Sigma_2, S_2, P_2)$ nyelvtannal, ha létezik $\varphi : N_1 \cup \Sigma_1 \to N_2 \cup \Sigma_2$ bijektív leképezés úgy, hogy $\varphi(S_1) = S_2$, $\{\varphi(A) \mid A \in N_1\} = N_2$, $\{\varphi(a) \mid a \in \Sigma_1\} = \Sigma_2$, valamint $\{\varphi(x_1)...\varphi(x_s) \to \varphi(y_1)...\varphi(y_t) \mid x_1...x_s \to y_1...y_t \in P_1\} = P_2$. Ebben a fejezetben nem teszünk különbséget a betű-izomorf nyelvtanok között.

Minden x terminális betűre legyen $N(x) = \{X \in N \mid X \to x \in P\}$. A G nyelvtan *redukált*, ha kielégíti az alábbi feltételeket:

(I.) Bármely x, y terminális szimbólumok esetén ha $N(x) = N(y)$, akkor $x = y$.

(II.) Minden $x \in N \cup \Sigma$ esetén létezik $W_1, W_2 \in (N \cup \Sigma)^*$ pár úgy, hogy $W_1 x W_2 \in S(G)$.

A továbbiakban redukált nyelvtanokat használunk.

Minden $X \in N$ nemterminálisra legyen $\Sigma(X) = \{x \in \Sigma \mid X \to x \in P\}$, ahol $\Sigma(X) = \emptyset$ is lehetséges.

Ezek után legyen a $G = (N, \Sigma, S, P)$ nyelvtan *váza* a $G_0 = (N, S, P_0)$, ahol $P_0 = \{A \to BC \in P \mid A, B, C \in N\}$. Az $S(G_0) = \{W \in N^+ \mid S \overset{*}{\Rightarrow} W\}$ halmazt a G_0 *váz által generált (mondatformájú) nyelvnek* nevezzük. A G_0 váz *maximális (a primitív szavakra nézve,)* ha $S(G_0)$ csak primitív szavakat tartalmaz, valamint minden $X, Y, Z \in N$, $X \to YZ \notin P_0$ esetén a $G_0' = (N, S, P_0')$, $P_0' = P_0 \cup \{X \to YZ\}$ vázhoz tartozó $S(G_0')$ nyelv tartalmaz nemprimitív szót is.

Jelölje Q_Σ a Σ feletti primitív szavak halmazát, és Q_N az N feletti primitív szavak halmazát. Ekkor $L(G) \subseteq Q_\Sigma \Rightarrow S(G_0) \subseteq Q_N$.

Az $S(G_0) \subseteq Q_N \Rightarrow L(G) \subseteq Q_\Sigma$ állítás igaz, amennyiben $\Sigma(X) \cap \Sigma(Y) = \emptyset$ teljesül minden $X, Y \in N$, $X \neq Y$ esetén.

Ezen állítás alapján rögzített számú nemterminális esetén a maximális vázak karakterizációinak felhasználásával meg tudjuk adni a redukált nyelvtanok karakterizációit. ($|N| > 2$ esetén figyelni kell a terminálisok felcserélhetőségére is.)

Maximális váz 1 nemterminálissal

Ha $|N| = 1$, akkor az egyetlen maximális váz a $G_0 = (N, S, \emptyset)$, és az egyetlen redukált nyelvtan a $G = (\{S\}, \{s\}, S, \{S \to s\})$.

Maximális váz 2 nemterminálissal

Ha $|N| = 2$, akkor az egyetlen maximális váz G_0, ahol

$P_0 = \{S \to SX,\ S \to XS,\ X \to XX\}$

$S(G_0) = \{X\}^* \cdot \{S\} \cdot \{X\}^*$ (csak az $\{S \to SX,\ S \to XS\}$ szabályok szükségesek,) és $S(G_0) \subset Q_N$ nyilvánvaló.

A redukált nyelvtanok az alábbi alakúak:
$G = (\{S, X\}, \{s, x\}, P_1 \cup P_2 \cup \{S \to s,\ X \to x\}, S)$, ahol
$P_1 \subseteq \{S \to SX,\ S \to XS\}$, $P_2 \subseteq \{X \to XX\}$, $P_1 \neq \emptyset$.

Maximális vázak 3 nemterminálissal

Három nemterminális esetén a lehetséges vázak nagy száma miatt szükség volt számítógép, és a lehetséges maximális vázak meghatározására képes program használatára. A szükséges programot a szerző készítette Pascal nyelven, majd - mivel nem állt rendelkezésre megfelelő gépkapacitás, - ezt a programot Dirk Hauschildt írta át C nyelvre, és a végső verziót a Hamburgi Egyetemen rendelkezésre álló gépeken futtatta. A program előállította az összes maximális vázat, melyek által generált összes, legfeljebb 12 hosszú szó primitív szó. A kapott vázakon túl - a mondatszimbólumtól különböző két nemterminális felcserélésével vagy pedig az összes $X \to YZ$ alakú szabály $X \to ZY$ alakú szabályra cserélésével kapható szimmetrikus vázaktól eltekintve - nincs több, csak primitív szavakat generáló maximális váz. A kapott 12 váz egyike nem redukált, és a maradék 11 váz mindegyikéről sikerült bebizonyítani, hogy 12 hosszúság felett is csak primitív szavakat generálnak. A bizonyítások alapötlete az első 10 esetben a szerzőtől származik, a 11. eset bizonyítása a szerző és az [1] dolgozatban közreműködő társszerzők közös munkájának eredménye.

A program feladata alapvetően a következő volt:
Előállította az összes lehetséges vázat, majd ellenőrizte a vázakról, hogy 12

hosszúságig csak primitív szavakat generálnak-e, vagy pedig generálnak nemprimitív szót is. A kapott, csak primitív szavakat generáló vázak közül kiválasztotta a maximálisakat.

Az alábbiakban megadjuk a 11 redukált vázat. A szimmetrikus esetektől eltekintve a felsoroltakon túl nincs a fenti feltételeknek megfelelő váz.

A továbbiakban legyen $N = \{S, X, Y\}$ és S a mondatszimbólum.

1. eset
$P_0 = \{S \to XY,\ S \to SX,\ S \to XS,\ S \to YX,\ X \to XX,$
$Y \to SX,\ Y \to XS,\ Y \to XY,\ Y \to YX\}.$

2. eset
$P_0 = \{S \to XY,\ S \to SX,\ S \to XS,\ S \to YX,\ X \to XX,\ Y \to YY\}.$

3. eset
$P_0 = \{S \to XY,\ S \to SY,\ S \to XS,\ X \to XX,\ Y \to YY\}.$

4. eset
$P_0 = \{S \to XS,\ S \to YS,\ S \to SX,\ S \to SY,\ X \to XX,\ X \to XY,$
$X \to YX,\ X \to YY,\ Y \to XX,\ Y \to XY,\ Y \to YX,\ Y \to YY\}.$

5. eset
$P_0 = \{S \to XS,\ X \to SY,\ X \to XX,\ X \to YY\}.$

6. eset
$P_0 = \{S \to XS,\ X \to SY,\ X \to XX,\ X \to XY,\ Y \to YX,\ Y \to YY\}.$

7. eset
$P_0 = \{S \to XY,\ S \to SY,\ Y \to XS,\ Y \to YY\}.$

8. eset
$P_0 = \{S \to XY,\ S \to YX,\ X \to SS\}.$

9. eset
$P_0 = \{S \to XY,\ X \to XS,\ Y \to SY\}.$
Ebben az esetben $S(G_0)$ nem reguláris.

10. eset
$P_0 = \{S \to XS,\ S \to SX,\ X \to YS,\ X \to SY,\ X \to XX,$
$Y \to XY,\ Y \to YX\}.$

11. eset
$P_0 = \{S \to XY,\ X \to SY,\ Y \to XS\}.$

A fentebbi 11 redukált maximális vázhoz hozzáadhatjuk az egyetlen nem redukált maximális vázat, mely annyiban tér el a 2 nemterminális $\{S, X\}$ esetétől, hogy hozzávesszük mind a 9 lehetséges szabályt, melyek bal oldalán Y szerepel.

Maximális vázak 4 nemterminálissal

A programot futtattuk 4 nemterminális esetére is. A program először 6 hosszúságig vizsgálta a vázak által generált szavakat, majd növeltük a hosszat 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, és végül 27 hosszúságig. Ennél a hossznál még mindig 413 maximális vázat kaptunk, melyek száma túl nagy ahhoz, hogy mindegyik esetről igazoljuk, hogy csak primitív szavakat generálnak.

Ugyanakkor megállapíthatjuk, hogy ha egy nyelvtan egy legalább kétbetűs ábécé felett generálja az összes primitív szót, és nem generál egyetlen nemprimtív szót sem, akkor a mondatszimbólum nem szerepelhet egyetlen, a nyelvtan által generált legalább 2 hosszúságú szóban sem.

Ezek után azokra a nyelvtanokra korlátoztuk vizsgálatainkat, melyekben egyik szabály jobboldalán sem szerepel a mondatszimbólum, hisz csak ezek a nyelvtanok generálhatják az összes primitív szót 3 nemterminális betű felett. A programot lefuttattuk a fentebbi alaknak megfelelő vázakra 12 hosszúságig. 28 redukált maximális vázat kaptunk, valamint 2 nem redukált maximális vázat, melyek mindegyikéről sikerült bizonyítani, hogy 12 hosszúság felett is csak primitív szavakat generálnak. A bizonyítások mindegyike a szerző munkája, melyet a [3] dolgozatban foglalunk össze.

A továbbiakban legyen $N = \{S, X, Y, Z\}$, S a mondatszimbólum, és a G_0 váz által generált $S(G_0)$ nyelven az egyszerűség kedvéért a G_0 váz által generált legalább 2 hosszú szavak halmazát értjük. Jelölje Q az $\{X, Y, Z\}$ ábécé feletti összes, legalább 2 hosszú primitív szavak halmazát.

1. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\ S \to ZY,\ X \to XX,\ Y \to YY,\ Z \to ZZ\}.$$

2. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\ S \to ZY,\ X \to XX,\ Y \to YY,\ Z \to XY,\ Z \to YX\}.$$

3. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\ S \to ZY,\ X \to XX,\ Y \to YY,\ Z \to XY,\ Z \to XZ,\ Z \to ZY\}.$$

4. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\ S \to ZY,\ X \to XX,\ Y \to YY,\ Z \to XZ,\ Z \to ZX,\ Z \to YZ,\ Z \to ZY\}.$$

5. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\ S \to ZY,\ X \to XX,\ X \to YZ,\ X \to ZY\}.$$

6. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\ S \to ZY,\ X \to XX,\ X \to YZ,\ Y \to YX,\ Z \to XZ\}.$$

7. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\\ S \to ZY,\ X \to XX,\ Y \to XY,\ Y \to YX,\ Y \to YZ,\\ Y \to ZY,\ Z \to XZ,\ Z \to ZX\}.$$

8. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ S \to YZ,\\ S \to ZY,\ X \to YZ,\ X \to ZY,\ Y \to XZ,\ Y \to ZX\}.$$

9. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\\ Y \to YY,\ Y \to ZZ,\ Z \to YY,\ Z \to ZZ,\ Y \to YZ,\\ Y \to ZY,\ Z \to YZ,\ Z \to ZY\}.$$

10. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\ Y \to YY,\\ Z \to YY,\ Z \to XY,\ Z \to YX,\ Z \to XZ,\ Z \to ZX\}.$$

11. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\\ Y \to YY,\ Z \to YY,\ Z \to XY,\ Z \to XZ,\ Z \to ZY\}.$$

12. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\\ Y \to YY,\ Y \to YZ,\ Z \to XZ\}.$$

13. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\ X \to YZ,\\ X \to ZY,\ Y \to XY,\ Y \to YX,\ Z \to XZ,\ Z \to ZX\}.$$

14. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\\ X \to YZ,\ Y \to XY,\ Y \to XZ\}.$$

15. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\\ Y \to XY,\ Y \to YX,\ Y \to XZ,\ Y \to ZX,\ Z \to XY,\\ Z \to YX,\ Z \to XZ,\ Z \to ZX\}.$$

16. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ Y \to YY,\ Y \to ZZ,\\ Z \to YY,\ Z \to ZZ,\ X \to XY,\ X \to YX,\ X \to XZ,\\ X \to ZX,\ Y \to YZ,\ Y \to ZY,\ Z \to YZ,\ Z \to ZY\}.$$

17. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ Y \to YY,\\ Y \to ZZ,\ Y \to XZ\}.$$

18. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ Y \to YY,\ Z \to ZZ,\ Y \to XZ,\ Y \to YZ\}.$$

19. eset

$$P_0 = \{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ Y \to ZZ,\ Z \to XY\}.$$

20. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ X \to ZZ,\ X \to YZ,\ Y \to XY\}.$$

21. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ Y \to YY,\ Z \to ZZ,\ X \to XZ,\ X \to ZX,\ Y \to YZ,\ Y \to ZY\}.$$

22. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ Y \to YY,\ Z \to ZZ,\ X \to XZ,\ Y \to ZY,\ Z \to ZX,\ Z \to YZ\}.$$

23. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ Z \to XX,\ Z \to YY,\ Z \to ZZ,\ X \to XZ,\ X \to ZX,\ Z \to XZ,\ Z \to ZX\}.$$

24. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ Z \to YY,\ Z \to ZZ,\ X \to XZ,\ X \to ZX,\ Y \to YZ,\ Y \to ZY\}.$$

25. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ Z \to ZZ,\ X \to XZ,\ X \to YZ,\ Y \to XY,\ Z \to ZX\}.$$

26. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to ZZ,\ Z \to XY,\ Z \to YX\}.$$

27. eset

$$P_0 = \{S \to XY,\ S \to YX,\ Z \to XX,\ Z \to YY,\ Z \to ZZ,\ X \to XZ,\ X \to ZX,\ Y \to YZ,\ Y \to ZY\}.$$

28. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to YZ,\ Y \to ZX,\ Z \to YX\}.$$

Végezetül tekintsük a nem redukált eseteket:

29. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ Y \to YY,\ Z \to XX,\ Z \to YY,\ Z \to ZZ,\ Z \to XY,\ Z \to YX,\ Z \to XZ,\ Z \to ZX,\ Z \to YZ,\ Z \to ZY\}.$$

30. eset

$$P_0 = \{S \to XY,\ S \to YX,\ X \to XX,\ Z \to XX,\ Z \to YY,\ Z \to ZZ,\ Y \to XY,\ Y \to YX,\ Z \to XY,\ Z \to YX,\ Z \to XZ,\ Z \to ZX,\ Z \to YZ,\ Z \to ZY\}.$$

Hivatkozások

[1] Dömösi, P., Horváth, G., Kudlek, M., Hauschildt, D., Some Results on Small Context-free Grammars Generating Primitive Words, *Technical Report FBI-HH-B-187/96*, University of Hamburg, (1996) ; *Publ. Math. Debrecen* **54**, (1999), 667-686.

[2] Dömösi, P., Horváth, G., Ito, M., A Small Hierarchy of Languages Consisting of Non-Primitive Words, *Preprints No. 282, Technical Reports No. 2002/6*, Debreceni Egyetem, (2002) ; *Publ. Math. Debrecen*, beadva.

[3] Horváth, G., Small Grammars and Primitive Words, *Workshop on Algebraic Systems and Conventional and Unconventional Computation Theory*, Kyoto Sangyo University, Kyoto, Japán, Sept., (2002) elfogadva.

[4] Dömösi, P., Formális Nyelvek és Automaták, *Egyetemi jegyzet*, Kossuth Lajos Tudományegyetem, Debrecen, (1995-1996).

[5] Révész, Gy., Bevezetés a formális nyelvek elméletébe I-II., *Tankönyvkiadó*, Budapest, (1989).

[6] Dömösi, P., Horváth, S., Ito, M., Kászonyi, L., Katsura, M., Some combinatorial properties of words, and the Chomsky hierarchy, *Proc. 2nd Int. Coll. Words, Languages and Combinatorics*, Kyoto, Japan, Aug., (1992), 25-28., ed.: M. Ito and H. Jürgensen ; *World Scientific Publishers*, Singapore, (1994), 105-123.

[7] Bader, Ch., Moura, A., A generalizatoin of Ogden's lemma, *Journal of the ACM*, **29**, (1982), 404-407.

[8] Bar-Hillel, Y., Perles, M., Shamir, E., On formal properties of simple phrase structure grammars, *Z. Phonetik. Sprachwiss. Komm.*, **14**, (1961), 143-172.

[9] Dömösi, P., Horváth, S., Ito, M., On the connection between formal languages and primitive words, *Proc. First Conference on Scientific Communications,* Univ. of Oradea, (1991). ; *Publ. Math. Debrecen* **42**, (1993), 315-321.

[10] Fine, N. J., Wilf, H. S., Uniqueness theorems for periodic functions, *Proc. Am. Math. Soc.*, **16**, (1965), 109-114.

[11] Horváth, S., A comparsion of iteration conditions of formal languages, *Proc. Colloq. Algebra, Combinatorics and Logic in Comp. Sci.*, Győr, Hungary, (1983) ; *Coll. Math. Soc. J. Bolyai*, **42**, *J. Bolyai Math. Soc.*, Budapest and North-Holland, Holland, Amsterdam, (1986), 453-463.

[12] Horváth, S., Karhumäki, J., Kleijn, J., Results concerning palindromicity, *J. Inf. Process. Cybern. EIK*, **23**, (1987), 441-451.

[13] Ogden, W., A helpful result for poving inherent ambiguity, *Math. Syst. Theory*, **2**, (1968), 191-194.

[14] Ross, R., Winklamann, K., Repetitive strings are not context-free, *RAIRO Informatique théorique*, **16**, (1982), 192-199.

[15] Shyr, H. J., Free Monoids and Languages, 3rd ed., *Lecture Notes*, Inst. of Applied Math., National Chung-Hsing Univ., Taichung, Taiwan, (2001).

[16] Sokolowski, S., A method for proving programming languages non context-free, *Inf. Proc. Lett.* **7**, (1978), 151-153.

[17] Thue, A., Über unendliche Zeichenreihen, *Norske Videnskabers Selskabs Skrifter Mat.-Nat. Kl.*, Kristiania, **7** (1906), 1-11.

[18] Ito, M., Katsura, M., Context-free languages consisting of non-primitive words, *Semigroup Forum* **37** (1988), 45-52.

[19] Hopcroft, J. E., Ullman, J. D., Introduction to Automata Theory, languages, and Computation, *Addison-Wesley, Reading, Mass.,* (1979).

[20] Lothaire, M., Combinatorics on Words, *Addison-Wesley, Reading, Mass.,* (1983).

[21] Salomaa, A., Jewels of Formal Language Theory, *Pitman*, London, (1981).

[22] Lyndon, R. C., Schützenberger, M. P., The equation $a^M = b^N c^P$ in a free group, *Michigan Math. J.* **9** (1962), 289-298.

[23] Shyr, H. J., Thierrin, G., Disjunctive languages and codes, *FCT'77, LNCS* **56**, *Springer-Verlag* (1977), 171–176.

[24] Dömösi, P., Horváth, S., Ito, M., Kászonyi, L., Katsura, M., Formal languages consisting of primitive words, *Proc. Conf. FCT'93, ed.: Z. Ésik, Springer LNCS 710*, (1993), 194-203.

[25] Horváth, S., Strong interchangeability, nonlinearity and related properties of primitive words, *manuscript,* Budapest, Aug., (1994) ; *Report FBI-HH 183/96*, FB Informatik, Universität Hamburg, (1996).

[26] Ito, M., Katsura, M., Shyr, H. J., Yu, S. S., Automata accepting primitive words, *Semigroup Forum*, **37** (1988), 45-52.

[27] Petersen, H., The ambiguity of primitive words, *Proc. STACS'94, Springer LNCS 775*, (1994), 679-690.