A primitív és nemprimitív szavak nyelvei

doktori (PhD) értekezés

Horváth Géza

Debreceni Egyetem
Természettudományi Kar

Debrecen, 2002

Ezen értekezést a Debreceni Egyetem Matematika doktori program Informatika alprogramja keretében készítettem 1997–2002 között és ezúton benyújtom a Debreceni Egyetem doktori Ph.D. fokozatának elnyerése céljából.

Debrecen, 2002. június 10.

....
Horváth Géza
jelölt

Tanúsítom, hogy Horváth Géza doktorjelölt 1997–2002 között a fent megnevezett doktori alprogram keretében irányításommal végezte munkáját. Az értekezésben foglaltak a jelölt önálló munkáján alapulnak, az eredményekhez önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javaslom.

Debrecen, 2002. június 10.

....
Dr. Dömösi Pál
témavezető

Köszönetnyilvánítás

Ezúton is szeretnék köszönetet mondani mindazoknak, akik közvetlenül vagy közvetve hozzájárultak a disszertációm elkészítéséhez.

Témavezetőmnek, Dr. Dömösi Pálnak, aki megszerettette velem a formális nyelvek és automaták elméletét, aki elindított a tudományos pályán, és akitől a szakmán túl is számtalan dolgot tanultam,

Szüleimnek, akik felneveltek és mindig mindenben mellettem álltak,

Tanáraimnak, Dr. Pethő Attilának, aki bevezetett a kriptográfia tudományába, és Katsushi Inoue senseinek, aki megismertette velem a valószínűségi automaták elméletét,

Végezetül a Japán Oktatásügyi Minisztériumnak, hogy lehetővé tette, hogy másfél éven át Japánban folytathassam tanulmányaimat.

Tartalomjegyzék

Bevezetés

A formális nyelvek elméletének, mint önálló kutatási területnek a kialakulása a századfordulóra nyúlik vissza. Kezdetekben a nyelvészek próbáltak olyan rendszereket létrehozni, melyekkel megoldható az emberi nyelveknek a pontos, precíz leírása. A későbbiekben szándékuk volt egy olyan általános modell létrehozása, mely segítségével megoldható az élő nyelvek gépi egymásbafordítása. Ez az elképzelés a beszélt nyelvek összetettsége miatt túlzottnak bizonyult, viszont a számos kapott eredmény felhasználhatóvá vált a számítástechnikában alkalmazott magas szintű nyelvek gépi kódra történő fordításában. Innentől kezdve a formális nyelvek kutatása fokozatosan átkerült a matematikusok, informatikusok kezébe. A matematikai alapok lerakásában a legnagyobb jelentőssége Noam Chomskynak volt, a mondatszerkezetű nyelvtan modelljének elkészítésével (1959). Az azóta eltelt több mint négy évtizedben számos új eredmény látott napvilágot. A formális nyelvek elméletét igen sok területen alkalmazzák, a gyakorlatban és az elméletben egyaránt. Hatékony fordítóprogramok, interpreterek, szintaktikai elemzők készítése elképzelhetetlen a formális nyelvek magas szintű ismerete nélkül, de az algoritmusok bonyolultságának vizsgálatában és a nyelvészetben is jó szolgálatot tehet a megfelelő szintű jártasság a formális nyelvek elméletében.

A nyelvek analízisének egyik igen fontos kérdése, hogy az adott nyelv mely szinten szerepel a Chomsky-féle hierarchiában. Ennek a kérdésnek az eldöntéséhez számos ismert módszer áll rendelkezésre. A leggyakrabban használt eszközök a pumpálós lemmák, melyek olyan feltételeket támasztanak, melyek mindig teljesülnek az adott nyelvosztályba tartozó nyelvekre. Amennyiben egy nyelv nem elégíti ki az adott feltételt, abban az esetben a nyelv nem eleme a feltételhez tartozó nyelvosztálynak. Ugyanakkor meg kell jegyezni, hogy amennyiben egy nyelv kielégíti egy pumpálós lemma feltételeit, az még nem vonja maga után, hogy az adott nyelv a lemmához tartozó nyelvosztály eleme. Részletek megtalálhatóak többek között a [7],

[11] és [13] cikkekben. A nyelvek vizsgálatának egy másik fontos kérdése nyelvek illetve nyelvtanok karakterizációja. Ha meg tudjuk adni egy nyelv felépítését, illetve meg tudunk adni egy adott nyelvet generáló nyelvtant, akkor sokkal közelebb kerülünk a nyelv megismeréséhez, megkönnyítjük gyakorlati felhasználását, és - amennyiben nem volt ismert - meg tudjuk adni a nyelv Chomsky-féle hierarchiában elfoglalt helyét is.

A primitív szavak nyelvével kapcsolatos kutatások mintegy 15 évre nyúlnak vissza. Egy igen intenzíven vizsgált területről van szó, ezt mutatja, hogy számos dolgozat foglalkozik a primitív és a nemprimitív szavak által alkotott nyelvekkel, többek között: [6], [9], [18], [24], [25], [26] és [27]. Egy adott (legalább kétbetűs) ábécé feletti összes primitív szavak nyelve azok közé a nyelvek közé tartozik, melyekről a mai napig nem ismert, hogy a Chomsky-féle hierarchia mely szintjén helyezkednek el. Azaz nyitott kérdés, hogy az összes primitív szavak nyelve generálható-e környezetfüggetlen nyelvtannal, vagy pedig valódi környezetfüggő nyelv. Az összes nemprimitív szó által alkotott nyelvről a legismertebb pumpálós lemma, a Bar-Hillel Lemma segítségével sikerült igazolni, hogy valódi környezetfüggő nyelv. A továbbiakban érdekes kérdés, hogy a nemprimitív szavak mely részhalmaza generálható környezetfüggetlen, lineáris és reguláris nyelvtannal, valamint szintén érdekes kérdés a Chomsky-féle hierarchia különböző szintjein lévő, csak primitív szavakat generáló nyelvtanok karakterizációja. A disszertációban ezekre a kérdésekre keresünk választ.

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 második részben 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. A bizonyításokban fontos szerepet játszanak a lineáris és reguláris nyelvekre vonatkozó iterációs lemmák, valamint Shyr és Thierrin illetve Lyndon és Schützenberger egy-egy tétele. 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 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 igazoljuk, ezek bizonyítják, hogy eredményeink helyesek. A program 3 nemterminális esetére 12 maximális vázat szolgáltatott, melyek 12 hosszúságig csak primitív szavakat generálnak, ezek mindegyikéről bizonyítjuk, 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 bizonyítjuk, 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. Fejezet

Alapfogalmak

Ebben a fejezetben áttekintést adunk az értekezésben felhasznált fogalmakról.

1.1. Az ábécé, a szavak és a nyelvek

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. $V^+ = V^* \setminus \{\lambda\}$.

1.1.1. Definíció A $p = x_1x_2\ldots x_n$ és a $q = y_1y_2\ldots y_m$ *szó egyenlő*, ha $n = m$ és $x_i = y_i$ $\forall i \in \{1\ldots n\}$-re.

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

1.1.3. 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.)

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

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

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

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

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

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

1.1.10. Definíció Azt a nyelvet, amelynek egyetlen szava sincs, *üres nyelvnek* nevezzük. Jelölés: $\emptyset$. Nem tévesztendő össze a $\{\lambda\}$ nyelvvel, amely egyedül az üres szót tartalmazza.

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

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

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

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

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

1.1.13. 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 $L^0 = \{\lambda\}$.

1.1.14. Definíció A *konkatenáció lezárását* az $L^* = \bigcup_{i=0}^{\infty} L^i$ összefüggéssel értelmezzük.

Használatos még az $L^+ = L^* \setminus \{\lambda\}$ jelölés is.

1.2. A generatív nyelvtan és a Chomsky-féle hierarchia

1.2.1. 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 $P \to Q$ alakban írjuk. $0 < |H| < \infty$. A H elemei olyan helyettesítési szabályok, amelyek baloldala tartalmaz legalább egy nemterminálist.

1.2.2. 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 $R \to T \in H$ szabály.
Jelölés: $P \Rightarrow Q$.

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

1.2.4. Definíció A Q szó a P szóból *levezethető*, ha $P \Rightarrow Q$ vagy $P \overset{+}{\Rightarrow} Q$. Jelölés: $P \overset{*}{\Rightarrow} Q$.

1.2.5. 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.

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

1.2.7. Definíció Chomsky-féle hierarchia

A $G = (V_N, V_T, S, H)$ *nyelvtan i tipusú*, 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 $P, Q \in V_N$, $a \in V_T{}^*$.

1.2.8. 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.

1.2.1. 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 $\mathcal{L}_2 \supseteq \mathcal{L}_3$.*

1.2.2. 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 $L(G) = L(G\,')$.*

1.2.1. 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 $\mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \mathcal{L}_3$.

1.3. Chomsky-féle normál alak környezetfüggetlen nyelvekre

1.3.1. 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.

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

1.3.1. Következmény *Minden nyelvtanhoz megadható vele ekvivalens olyan nyelvtan, melynek szabályai baloldalán terminális betű nem fordul elő.*

1.3.2. 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.

1.3.3. 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 $a \in V_T$.

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

2. Fejezet

Szavak és nyelvek kombinatorikus tulajdonságai

2.1. Bevezetés

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.

2.2. Palindromikus- és repetitív nyelvek

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.

2.2.1. 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.

2.2.2. 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.

2.2.3. 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.

2.2.4. 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:

2.2.5. 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:

2.2.6. 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.

2.3. A primitív szavakból álló nyelvek

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.

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

A determinisztikus környezetfüggetlen nyelvek valódi részhalmazát alkotják a környezetfüggetlen nyelveknek.

2.3.1. 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:

2.3.2. 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:

2.3.3. 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. Először néhány lemmát adunk meg, melyek fontos szerepet játszottak a további tételek bizonyításában. Jelölje az $lnko(a_1, ..., a_n)$ az $a_1, ..., a_n$ pozitív egészek legnagyobb közös osztóját.

2.3.1. Lemma [10] *Legyenek $x, y \in X^+$ szavak. Ha az x^i és az y^j szavaknak legalább $|x| + |y| - lnko(|x|, |y|)$ hosszú azonos kezdőszelete van, akkor x és y ugyanazon szónak a hatványa.*

2.3.2. Lemma [15] *Legyen $i \geq 1$ egész szám és $uv \in \{p^i \mid p \in Q\}$. Ekkor $vu \in \{p^i \mid p \in Q\}$. Másképp fogalmazva, a $\{p^i \mid p \in Q\}$ $(i \geq 1)$ halmaz zárt a ciklikus permutációra nézve.*

2.3.3. Lemma [15] *Ha $a \in X$, $pq \in X^+ \setminus a^+$ és $paq \in X^+ \setminus Q$, akkor $pq \in Q$.*

2.3.4. Lemma [6] *Ha $a, b \in X$, $a \neq b$, $pq \in X^+$ és $paq \in X^+ \setminus Q$, akkor $pbq \in Q$.*

A jelenleg rendelkezésünkre álló, minden környezetfüggetlen nyelvre teljesülő, a Bar-Hillel Lemmától erősebb feltételek közül először az Ogden feltételt ismertetjük.

2.3.2. 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. $uv^mwx^my \in L$ minden $m \geq 0$ egészre.

Ekkor azt mondjuk, hogy L kielégíti az Ogden feltételt.

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

2.3.5. 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.

2.3.3. 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. $uv^mwx^my \in L$ minden $m \geq 0$ egészre.

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

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

2.3.7. 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.

2.3.4. 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 $u_1x'u_2x''u_3 \in L$.

2.3.8. 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:

2.3.9. 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. Fejezet

Nemprimitív szavakból álló környezetfüggetlen nyelvek

3.1. Bevezetés

Ebben a fejezetben megismerünk számos problémát, melyek felmerülnek a környezetfüggetlen, csak nemprimitív szavakat tartalmazó nyelvek kapcsán, és ezekre a problémákra keresünk választ.

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.

3.2. Megoldott problémák és nyitott kérdések a környezetfüggetlen nyelvek témaköréből

Az alábbiakban állítások és hozzájuk kapcsolódó kérdések együtt szerepelnek. A problémák megoldását a fejezet hátralévő részében ismertetjük.

(C.1) Létezik $L \subseteq X^*$ végtelen környezetfüggetlen nyelv úgy, hogy $L \subseteq Q$.

(C.2) Létezik $L \subseteq X^*$ végtelen környezetfüggetlen nyelv úgy, hogy $L \subseteq Q^{(2)}$.

(C.3) Legyen $i \geq 3$. Létezik $L \subseteq X^*$ végtelen környezetfüggetlen nyelv úgy, hogy $L \subseteq Q^{(i)}$?

(C.4) Minden $n \geq 1$ számra létezik $L \subseteq X^*$ végtelen környezetfüggetlen nyelv úgy, hogy $L \subseteq \bigcup\limits_{1 \leq i \leq n} Q^{(i)}$ és $|L \cap Q^{(n)}| = \infty$.

A (C.1) állításra jó példa az $L = \{ab^+\}$ nyelv, ahol $X = \{a, b, ...\}$ és $a \neq b$.

A (C.2) állítás igaz az $L = \{ab^i ab^i \mid i \geq 1\}$ nyelvre, ahol $X = \{a, b, ...\}$ és $a \neq b$.

A (C.4) állításra példa az $L = \{(ab^+)^n\}$ nyelv, ahol $n \geq 1$ egész, valamint $X = \{a, b, ...\}$ és $a \neq b$.

A továbbiakban sűrű (vagy diszjunktív) nyelvekre teszünk megállapításokat. Megjegyezzük, hogy ezen állítások egy része nem igaz reguláris nyelvek esetén, például nem létezik $L \subseteq X^*$ sűrű reguláris nyelv, melyre teljesül, hogy $L \subseteq Q$.

(DC.1) Létezik $L \subseteq X^*$ diszjunktív környezetfüggetlen nyelv úgy, hogy $L \subseteq Q$.

(DC.2) Létezik $L \subseteq X^*$ sűrű (vagy diszjunktív) környezetfüggetlen nyelv úgy, hogy $L \subseteq Q^{(2)}$?

(DC.3) Legyen $i \geq 3$. Létezik $L \subseteq X^*$ sűrű (vagy diszjunktív) környezetfüggetlen nyelv úgy, hogy $L \subseteq Q^{(i)}$?

(DC.4) Létezik $L \subseteq X^*$ diszjunktív környezetfüggetlen nyelv úgy, hogy $L \subseteq Q \cup Q^{(2)}$, $|L \cap Q| = \infty$ és $|L \cap Q^{(2)}| = \infty$.

(DC.5) Létezik $L \subseteq X^*$ diszjunktív környezetfüggetlen nyelv úgy, hogy $|L \cap Q^{(i)}| = \infty$ minden $i \geq 1$ egész számra.

A (C.3), (DC.2) valamint a (DC.3) problémák megoldásához meg kellett határozni az $X^+ \setminus Q$ valamint a $Q^{(2)}$ nyelvek struktúráját, melyet külön részekben ismertetünk.

3.3. Az $X^+ \setminus Q$ elemeiből álló környezetfüggetlen nyelvek felépítése

Ebben a részben ismertetjük azon környezetfüggetlen nyelvek felépítését, melyek $X^+ \setminus Q$ egy részhalmazát tartalmazzák. Elsőként egy új lemmát adunk meg.

3.3.1. Lemma [20] *Legyenek $u, v \in X^*$ konjugáltak, és legyen $u \in Q^{(i)}$ valamely $i \geq 1$ egészre. Ekkor $v \in Q^{(i)}$.*

Ezután következhet a tétel, mely megadja az $X^+ \setminus Q$ struktúráját.

3.3.1. 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.

A bizonyításban a 3.3.1. lemmát, az előző fejezet 2.3.1. lemmáját és a Bar-Hillel Lemmát használták.

Ezek után már meg tudjuk adni a választ a (C.3) és a (DC.3) kérdésekre.

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

3.4. A $Q^{(2)}$ elemeiből álló környezetfüggetlen nyelvek meghatározása

A $Q^{(2)}$ részeként előálló környezetfüggetlen nyelvek meghatározásához szükséges volt az előző fejezetben tárgyalt Ogden Lemma, valamint az alábbi lemmák:

3.4.1. Lemma [20] *Legyen $x, y \in X^+$. Ha $xy = yx$, akkor létezik $p \in Q$ szó, valamint $i, j \geq 1$ számok úgy, hogy $x = p^i$ és $y = p^j$.*

3.4.2. Lemma [18] *Legyenek $u, v \in Q$ konjugált szavak, és legyen $u = xy$, $v = yx$, $x, y \in X^+$. Ekkor (x, y) egyértelműen meghatározott.*

3.4.3. Lemma [18] *Legyen $z = uvwxy \in X^+$ úgy, hogy $vx \neq \lambda$. Ha $uv^n wx^n y \in Q^{(2)}$ minden $n \geq 1$-re, akkor v és x konjugáltak.*

3.4.4. Lemma [18] *Legyen $\alpha \in Q$ és $\beta \notin \alpha X^* \cup X^* \alpha$, $\beta \neq \lambda$. Ekkor ha $\alpha^i \beta \alpha^j \beta \alpha^k \in Q^{(2)}$ valamely $i, j, k \geq 1$ egészekre, akkor $j = i + k$.*

3.4.5. Lemma [18] *Legyen $L \subseteq X^*$ a $Q^{(2)}$ elemeiből álló környezetfüggetlen nyelv, és legyen $\alpha \in Q$, $\beta \notin \alpha X^* \cup X^* \alpha$, $\beta \neq \lambda$. Továbbá legyen $L_{\alpha,\beta} = L \cap \alpha^+ \beta \alpha^+ \beta \alpha^+$. Ekkor léteznek T és M pozitív egészek, melyekre teljesülnek a következők:*

1. *$(\alpha^{p+nT} \beta \alpha^q)^2 \in L_{\alpha,\beta}$ minden $n \geq 0$ egészre, ha $(\alpha^p \beta \alpha^q)^2 \in L_{\alpha,\beta}$ és $p \geq M$,*

2. *$(\alpha^p \beta \alpha^{q+nT})^2 \in L_{\alpha,\beta}$ minden $n \geq 0$ egészre, ha $(\alpha^p \beta \alpha^q)^2 \in L_{\alpha,\beta}$ és $q \geq M$,*

3. *$(\alpha^{p+nT} \beta \alpha^{q+mT})^2 \in L_{\alpha,\beta}$ minden $n, m \geq 0$ egészekre, ha $(\alpha^p \beta \alpha^q)^2 \in L_{\alpha,\beta}$ és $p, q \geq M$.*

Legyenek ezek után $u, v \in X^*$ és T, p, q pozitív egészek. Jelölje $L_l[u, v; T, p, q]$ az $\{(u^{p+nT} v u^q)^2 \mid n \geq 0\}$ nyelvet, $L_r[u, v; T, p, q]$ az $\{(u^p v u^{q+nT})^2 \mid n \geq 0\}$ nyelvet, valamint $L[u, v; T, p, q]$ az $\{(u^{p+nT} v u^{q+mT})^2 \mid n, m \geq 0\}$ nyelvet.

3.4.1. Következmény [18] *Legyen* $L \subseteq X^*$ *a* $Q^{(2)}$ *elemeiből álló környezetfüggetlen nyelv, és legyen* $\alpha \in Q$, $\beta \notin \alpha X^* \cup X^*\alpha$, $\beta \neq \lambda$. *Ekkor*

$$L_{\alpha,\beta} = F \cup (\bigcup_{1\leq i\leq d} L_l[\alpha,\beta;T_i,p_i,q_i]) \cup (\bigcup_{1\leq j\leq e} L_r[\alpha,\beta;T'_j,p'_j,q'_j]) \cup (\bigcup_{1\leq k\leq f} L[\alpha,\beta;T''_k,p''_k,q''_k])$$

ahol F *véges halmaz, és* T_i,p_i,q_i, $1 \leq i \leq d$, T'_j,p'_j,q'_j, $1 \leq j \leq e$, T''_k,p''_k,q''_k, $1 \leq k \leq f$ *pozitív egészek.*

3.4.6. Lemma [18] *Legyen* $L \subseteq X^*$ *a* $Q^{(2)}$ *elemeiből álló környezetfüggetlen nyelv, és legyen* $\alpha \in Q$, $\beta \notin \alpha X^* \cup X^*\alpha$, $\beta \neq \lambda$, $\beta = \beta_1\beta_2$. *Továbbá legyen* $L_{\alpha,\beta_1,\beta_2} = L \cap \beta_2\alpha^+\beta\alpha^+\beta_1$. *Ekkor léteznek* T *és* M *pozitív egészek úgy, hogy* $\beta_2\alpha^{p+nT}\beta\alpha^{p+nT}\beta_1 \in L_{\alpha,\beta_1,\beta_2}$ *minden* $n \geq 0$ *egészre, ha* $\beta_2\alpha^p\beta\alpha^p\beta_1 \in L_{\alpha,\beta_1,\beta_2}$ *és* $p \geq M$.

Legyen $u,v,w \in X^*$, és legyenek T,p pozitív egészek. Ezek után $L[u,v,w;T,p]$ jelölje a $\{(wu^{p+nT}v)^2 \mid n \geq 0\}$ nyelvet.

3.4.2. Következmény [18] *Legyen* $L \subseteq X^*$ *a* $Q^{(2)}$ *elemeiből álló környezetfüggetlen nyelv, és legyen* $\alpha \in Q$, $\beta \notin \alpha X^* \cup X^*\alpha$, $\beta \neq \lambda$, $\beta = \beta_1\beta_2$. *Ekkor*

$$L_{\alpha,\beta_1,\beta_2} = F \cup (\bigcup_{1\leq i\leq d} L[\alpha,\beta_1,\beta_2;T_i,p_i])$$

ahol F *véges halmaz és* T_i,p_i *pozitív egészek.*

3.4.1. 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.

Az $\{(a_i^n b_i a_i^m)^2 \mid n,m \geq 1\}$ $(1 \leq i \leq r)$ nyelvre tekintsük az alábbi környezetfüggetlen nyelvtant:

$$G_i = (V, X, S, P_i),\ 1 \leq i \leq r,\ V = \{S, T, U\},$$
$$P_i = \{S \to TT,\ T \to a_i T a_i,\ T \to U,\ U \to b_i\}.$$

A $\{(g_j h_j^n f_j)^2 \mid n \geq 1\}$ $(1 \leq j \leq s)$ nyelvet pedig az alábbi környezetfüggetlen nyelvtan generálja:

$$G_j = (V, X, S, P_j),\ 1 \leq j \leq s,\ V = \{S, T, U\},$$
$$P_j = \{S \to g_j T f_j,\ T \to h_j T h_j,\ T \to U,\ U \to f_j g_j\}.$$

Mindezek felhasználásával meg tudjuk válaszolni a (DC.2) kérdést:

3.4.3. 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:

3.4.2. 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. Fejezet

Nemprimitív szavakból álló lineáris és reguláris nyelvek

4.1. Bevezetés

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 $\bigcup\limits_{i\geq 2} Q^{(i)}$ 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 $L \cap (\bigcup\limits_{i\geq 2} Q^{(i)})$ nyelvek mindegyike környezetfüggő lesz. Ezekből az egyszerű tényekből következik, hogy a nemprimitív szavakból álló környezetfüggő nyelvek mindegyike felírható az alábbi formában: $L = L' \cap (\bigcup\limits_{i\geq 2} Q^{(i)})$, ahol L' környezetfüggő nyelv. Másképp fogalmazva, minden L' környezetfüggő nyelvre az $L = L' \cap (\bigcup\limits_{i\geq 2} Q^{(i)})$ 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.

4.2. Iterációs tételek lineáris és reguláris nyelvekre

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.

A továbbiakban három segédtételt ismertetünk, melyek fontos szerepet játszanak bizonyításainkban. Az alábbiakon túl használni fogjuk még a 2.3.2. lemmát is.

4.2.1. Tétel [22] *Legyen $f, g \in Q$, $f \neq g$. Ekkor $f^m g^n \in Q$ minden $m \geq 2$, $n \geq 2$ egészekre.*

4.2.2. Tétel (Iterációs Lemma reguláris nyelvekre) [19] *Legyen L reguláris nyelv. Ekkor létezik n konstans, hogy ha z tetszőleges szava az L nyelvnek és $|z| \geq n$, akkor z felírható $z = uvw$ alakban úgy, hogy $|uv| \leq n$, $|v| \geq 1$ és minden $i \geq 0$ esetén uv^iw is eleme az L nyelvnek. Továbbá n nem nagyobb, mint az L nyelvet felismerő minimális véges automata állapotainak a száma.*

4.2.3. Tétel (Iterációs Lemma lineáris nyelvekre) [19] *Minden L lineáris nyelvhez létezik n konstans, hogy az L nyelv bármely legalább n hosszúságú z szava felírható $z = uvwxy$ alakban úgy, hogy $|uvxy| \leq n$, $|vx| \geq 1$ és minden $i \geq 0$ esetén uv^iwx^iy is eleme az L nyelvnek.*

Megjegyezzük még, hogy bármely $u \in X^+$ szó egyértelműen megadható $u = p^n$, $p \in Q$, $n \geq 1$ alakban, így X^* előáll a $Q^{(i)}$, $i \geq 0$ halmazok uniójaként: $X^* = \bigcup_{i \geq 0} Q^{(i)}$.

4.3. Az $X^+ \setminus Q$ elemeiből álló lineáris és reguláris nyelvek felépítése

Elsőként megadjuk a nemprimitív szavakból álló lineáris nyelvek karakterizációját:

4.3.1. 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

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

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

Bizonyítás: Közismert tény, hogy minden L lineáris nyelvre $L \cap R$ is lineáris nyelv, ha R reguláris nyelv. Továbbá a 3.3.1. tétel értelmében $L_2 = L \cap (\bigcup_{i \geq 3} Q^{(i)})$ reguláris nyelv minden L környezetfüggetlen nyelvre. Mivel a lineáris nyelvek egyben környezetfüggetlen nyelvek, ezért L_2 reguláris lesz akkor is, ha L lineáris. Ezután, mivel $L_1 = L \setminus L_2$, ahol L lineáris és L_2 reguláris, ezért L_1 lineáris nyelv. Amit még bizonyítanunk kell, hogy $L_1 = F_1 \cup (\bigcup_{1 \leq j \leq s} \{(f_j g_j^n h_j)^2 \mid n \geq 1\})$ ahol $F_1 \subseteq Q^{(2)}$ véges halmaz, és $f_j g_j h_j \in Q,\ 1 \leq j \leq s$.

Először megmutatjuk, hogy az $\{(fg^n h)^2 \mid n \geq 1\}$ nyelv lineáris nyelv. Tekintsük az alábbi lineáris nyelvtant, mely pontosan az $\{(fg^n h)^2 \mid n \geq 1\}$ nyelvet generálja: $S \to fTh,\ T \to gTg,\ T \to hf$. Ebből látszik, hogy az L_1 nyelv generálható lineáris nyelvtannal.

A 3.4.1. tételben láthattuk, hogy környezetfüggetlen esetben a $Q^{(2)}$ elemeiből álló nyelv felépítése a következő: $F \cup (\bigcup_{1 \leq i \leq r} \{(a_i^n b_i a_i^m)^2 \mid n, m \geq 1\}) \cup (\bigcup_{1 \leq j \leq s} \{(g_j h_j^n f_j)^2 \mid n \geq 1\})$, 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. A továbbiakban még be kell látnunk, hogy az $L_{kf} = \bigcup_{1 \leq i \leq r} \{(a_i^n b_i a_i^m)^2 \mid n, m \geq 1\}$ nyelv felírható három nyelv uniójaként, melyek közül az első kettő $\bigcup_{1 \leq j \leq s} \{(g_j h_j^n f_j)^2 \mid n \geq 1\}$ alakú, a harmadik rész pedig nem generálható lineáris nyelvtannal.

Az L_{kf} nyelv felírható $\{a^j b a^{j+k} b a^k \mid j, k \geq 1\}$ alakú nyelvek véges uniójaként. Minden $a, b \in X^+$ esetén az $L_A = \{a^j b a^{j+k} b a^k \mid j < n,\ k \geq 1\}$ nyelv és az $L_B = \{a^j b a^{j+k} b a^k \mid j \geq 1,\ k < n\}$ nyelv is véges uniója az $\{(f g^n h)^2 \mid n \geq 1\}$ alakú nyelveknek, amennyiben L_A esetén $f = a^j b$, $g = a$, $h = \lambda$ és L_B esetén $f = \lambda$, $g = a$, $h = ba^k$. $L_{kf} \setminus \{L_A \cup L_B\} = \{a^j b a^{j+k} b a^k \mid j, k \geq n\}$. A továbbiakban megmutatjuk, hogy minden $L \subseteq Q^{(2)}$ lineáris nyelv esetén létezik $n \geq 1$ egész, melyre L nem tartalmaz elemet az $\{a^j b a^{j+k} b a^k \mid j, k \geq n\}$ nyelvből.

Indirekt feltételként legyen $L \subseteq Q^{(2)}$ olyan lineáris nyelv, melyre bármely n esetén $L \cap \{a^j b a^{j+k} b a^k \mid j, k \geq n\} \neq \emptyset$. Az n legyen a 4.2.3. tételnek megfelelő. Válasszunk egy $z = a^s b a^{s+t} b a^t$ szót az L nyelvből úgy, hogy $s, t \geq n$. Ekkor a 4.2.3. tétel értelmében z felírható $z = uvwxy$, $|uvxy| \leq n$, $|vx| \geq 1$ alakban, úgy, hogy minden $i \geq 0$ egészre $uv^i w x^i y \in L$. Itt meg kell jegyeznünk, hogy a lineáris nyelvek zártak a homomorfizmusra és az inverz homomorfizmusra nézve, ezért feltehetjük, hogy $u, v, x, y \in a^*$. Például tekintsük az alábbi homomorfizmust: $\psi : \{c, d\} \to X^*$, $\psi(c) = a$, $\psi(d) = b$ és az $u'v'w'x'y'$ a $z' = c^s d c^{s+t} d c^t$ szó felbontása a 4.2.3. tételnek megfelelően, $|u'v'x'y'| \leq n'$, $|v'x'| \geq 1$, $u'v'^i w' x'^i y' \in \{c^j d c^{j+k} d c^k \mid j, k \geq 1\}$, $i \geq 0$, ahol n' egy megfelelő pozitív egész. Természetesen ekkor $\psi(u') = u$, $\psi(v') = v$, $\psi(w') = w$, $\psi(x') = x$, $\psi(y') = y$, ahol $u, v, x, y \in a^*$.

Ugyanakkor $z = a^s b a^{s+t} b a^t \in Q^{(2)}$, amiből következik, hogy $a^s b a^t \in Q$. Mivel a 2.3.2. lemma alapján $Q^{(n)}$, $n \geq 1$ zárt a ciklikus permutációra nézve, ezért $b a^{s+t} \in Q$. Legyen $f = b a^{s+t}$, és $g = h$ úgy, hogy $h \in Q$, $h^k = a$. Ekkor a 4.2.1. tétel értelmében $f^2 g^m \in Q$ minden $m \geq 2$ egészre. Ismét alkalmazva a 2.3.2. lemmát azt kapjuk, hogy $uv^i w x^i y \in Q$, $i \geq 2$, ami ellentmond azon feltevésünknek, hogy $uv^i w x^i y \in Q^{(2)}$, $i \geq 0$.

$\square$

A következőkben a reguláris esetre fogalmazunk meg egy tételt.

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

Bizonyítás: A 4.3.1. tétel értelmében elegendő bizonyítani, hogy minden $a, b, c \in X^+$ esetén az $L \subseteq Q^{(2)}$ reguláris nyelv csak véges számú elemet tartalmaz az $\{(ab^m c)^2 \mid m \geq 1\}$ nyelvből.

Ismét indirekten bizonyítunk, az indirekt feltétel a következő: létezik $L \subseteq Q^{(2)}$ reguláris nyelv, melyre minden n pozitív egész esetén $L \cap \{(ab^m c)^2 \mid m \geq n\} \neq \emptyset$. Ekkor a 4.2.2. tétel értelmében létezik n pozitív egész úgy, hogy $m > n$ és $z = ab^m cab^m c$ felírható $z = uvw$ alakban úgy, hogy $|uv| \leq n$, $|v| \geq 1$ és $uv^i w \in L$, $i \geq 0$. A reguláris nyelvek szintén zártak a homomorfizmusra és az inverz homomorfizmusra, ezért az előző tétel bizonyításában használt megfontolások alapján most is feltehetjük, hogy

1. $v = a$, vagy
2. $v \in b^+$, vagy
3. $v \in ab^+$.

Tekintsük először a $v = a$ esetet. Ekkor a 4.2.2. tétel értelmében $a^i b^m cab^m c \in L$ minden $i \geq 0$ egészre. Használjuk a 4.2.1. tételt úgy, hogy legyen $f = ab^m c$ és $g = d$, $d \in Q$, $d^j = a$. Ekkor azt kapjuk, hogy $ab^m cab^m ca^i \in Q$, $i \geq 2$. Innen a 2.3.2. lemma értelmében $a^i b^m cab^m c \in Q$, $i \geq 3$, ami ellentmond annak, hogy $a^i b^m cab^m c \in Q^{(2)}$, $i \geq 0$.

Tekintsük most a $v = b^+$ esetet. Ekkor a 4.2.2. tétel értelmében létezik k pozitív egész, hogy $ab^{m+i*k} cab^m c \in L$ minden $i \geq 0$ egészre. Használjuk a 4.2.1. tételt úgy, hogy legyen $f = cab^m$ és $g = d$, $d \in Q$, $d^j = b$. Ekkor azt kapjuk, hogy $cab^m cab^{m+i} \in Q$, $i \geq 2$. Innen a 2.3.2. lemma értelmében $ab^{m+i} cab^m c \in Q$, $i \geq 2$, ami ellentmond annak, hogy $ab^{m+i*k} cab^m c \in Q^{(2)}$, $i \geq 0$.

Tekintsük végül a $v = ab^+$ esetet. Ekkor a 4.2.2. tétel értelmében létezik k pozitív egész, melyre $(ab^k)^i b^{m-k} cab^m c \in L$ minden $i \geq 0$ egészre. Használjuk a 4.2.1. tételt úgy, hogy legyen $f = b^{m-k} cab^k$ és $g = d$, $d \in Q$, $d^j = ab^k$. Ekkor azt kapjuk, hogy $b^{m-k} cab^m cab^k (ab^k)^i \in Q$, $i \geq 2$. Innen a 2.3.2. lemma értelmében $(ab^k)^i b^{m-k} cab^m c \in Q, i \geq 3$, ami ellentmond annak, hogy $(ab^k)^i b^{m-k} cab^m c \in Q^{(2)}, i \geq 0$.

□

4.4. Az $X^+ \setminus Q$ elemeiből álló nyelvek hierarchiája

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

4.4.1. 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:*
$\bigcup\limits_{1\le i\le r} f_i^{m_i}(f_i^{k_i})^*,\ f_i \in Q,\ m_i \ge 3,\ k_i \ge 1,\ 1 \le i \le r,$

$\mathcal{L}_2$ *az alábbi formájú nyelvek osztálya:*
$\bigcup\limits_{1\le j\le s} \{(f_j g_j^n h_j)^2 \mid n \ge 1\},\ f_j g_j h_j \in Q,\ 1 \le j \le s,$

$\mathcal{L}_3$ *azon nyelvek osztálya, melyek struktúrája a következő:*
$\bigcup\limits_{1\le i\le r} \{(a_i^n b_i a_i^m)^2 \mid n, m \ge 1\},$ *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. Fejezet

Primitív szavakat generáló, Chomsky-féle normál formájú kis nyelvtanok

5.1. Bevezetés

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, és az általuk generált nyelvek karakterizációját. 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, valamint a vázakhoz tartozó mondatformák halmazát. 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.

5.2. Betű-izomorf nyelvtanok, redukált nyelvtanok, vázak

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 $S(G) = \{W \mid W \in (N \cup \Sigma)^*,\ S \overset{*}{\Rightarrow} W\}$ 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 bizonyításához tekintsük egy tetszőleges $w \in L(G)$ szó bináris levezetési fáját. Vágjuk le az összes $x \in \Sigma$ levelet, melyhez $X \to x$ alakú szabályok segítségével jutottunk. Ekkor a $W \in N^+$ szó bináris levezetési fáját kapjuk, ahol $W \stackrel{*}{\Rightarrow} w$. A $\Sigma(X) \cap \Sigma(Y) = \emptyset$ feltétel következtében minden $x \in \Sigma$ terminálishoz egyértelműen létezik $X \in N$ nemterminális úgy, hogy $x \in \Sigma(X)$. Definiáljuk a betűt betűbe képező $c : \Sigma \to N$ homomorfizmust a következőképp: $c(x) = X$, ha $x \in \Sigma(X)$. Amennyiben $w = u^k$, $k > 1$, akkor $W = c(w) = (c(u))^k$, ami ellentmondás.

Az előző 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.)

5.3. Maximális vázak 1 és 2 nemterminálissal

Maximális váz 1 nemterminálissal

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

$\square$

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
$P_1 \subseteq \{S \to SX,\ S \to XS\}$, $P_2 \subseteq \{X \to XX\}$, $P_1 \neq \emptyset$.

$\square$

5.4. 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.

A futási sebesség növelése érdekében az alábbi módosítások történtek:
A program több lépésben futott. Első lépésben 6 hosszúságig vizsgálta a vázak által generált szavakat, és létrehozta az $\mathcal{N}$ halmazt, melybe azok a minimális vázak kerültek, melyek nemprimitív szót is generálnak, valamint a $\mathcal{P}$ halmazt, melybe a maximális, 6 hosszúságig csak primitív szavakat generáló vázak kerültek. Ezek után $\mathcal{P}$ elemeit megvizsgálva sikerült bizonyítani néhány vázról, hogy nem generálnak nemprimitív szót. Ezeket megtartottuk a $\mathcal{P}$ halmazban, a többi vázat elhagytuk a $\mathcal{P}$ halmazból. A követhező futtatás már 8 hosszúságig vizsgálta a lehetséges vázak által generált szavakat, de ekkor már csak azokat a vázakat kellett vizsgálni, melyek nem tartalmazzák $\mathcal{N}$ elemeinek egyikét sem, és a $\mathcal{P}$ halmazban található vázak egyikének sem részei. Az újabb eredmények ismeretében módosítottuk a $\mathcal{P}$ halmazt, majd ezt az eljárást ismételtük 9, 10 és 12 hosszúságig, míg végül eljutottunk a 12 maximális vázhoz, melyek közül egy nem redukált.

Az alábbiakban bizonyítjuk mind a 11 redukált vázról, hogy csak primitív szavakat generálnak. A szimmetrikus esetektől eltekintve az alábbiakon túl nincs a fenti feltételeknek megfelelő váz.

A továbbiakban legyen $N = \{S, X, Y\}$, S a mondatszimbólum, és jelölje $Q = Q_N$ az N feletti összes primitív szavak halmazát.

1. eset

$$\begin{aligned} P_0 = \{&S \to XY,\ S \to SX,\ S \to XS,\ S \to YX,\ X \to XX, \\ &Y \to SX,\ Y \to XS,\ Y \to XY,\ Y \to YX\}. \end{aligned}$$

$$S(G_0) = (X^* \cdot \{S, Y\} \cdot X^*) \setminus \{Y\} \subset Q.$$

Erre a vázra tekintsük az alábbi bizonyítást:

Legyen $L = X^*(\{S\} \cup \{Y\})X^* \setminus \{Y\}$. Megmutatjuk, hogy $S(G_0) = L$. Először teljes indukciót alkalmazunk. $S \in L$, és bármely P_0-beli szabályt alkalmazva egy tetszőleges $W \in L$ szóra, $W' \in L$ szót kapunk. Ebből következik, hogy $S(G_0) \subseteq L$.

Másrészt megmutatjuk, hogy minden $W \in L$ szó levezethető az S mondatszimbólumból. $S, XY, YX \in L$ triviális. $X^mSX^n \in L$ levezethető az $S \overset{m}{\Rightarrow} X^mS \overset{n}{\Rightarrow} X^mSX^n$ levezetéssel, ahol az $\{S \to XS,\ S \to SX\}$ szabályokat használtuk, és az $X^mYX^n \in L$ $(m > 0\ )$ alakú szavak levezethetőek az $S \overset{m-1}{\Rightarrow} X^{m-1}S \overset{n}{\Rightarrow} X^{m-1}SX^n \overset{1}{\Rightarrow} X^mYX^n$ levezetéssel, az $\{S \to XS,\ S \to SX,\ S \to XY\}$ szabályok alkalmazásával. Végezetül az YX^m alakú szavakhoz használjuk az $S \overset{m-1}{\Rightarrow} SX^{m-1} \overset{1}{\Rightarrow} YX^m$ levezetést, és az $\{S \to SX,\ S \to YX\}$ szabályokat. Mindezekből következik, hogy $L \subseteq S(G_0)$.

Megjegyezzük még hogy csak az $\{S \to XS,\ S \to SX,\ S \to XY,\ S \to YX\}$ szabályok szükségesek $S(G_0)$ generálásához.

$S(G_0) \subseteq Q$ triviális, hiszen minden $W \in S(G_0)$-beli szó pontosan egy S vagy Y betűt tartalmaz.

$SY \notin S(G_0)$ következtében $S(G_0) \subset Q$.

□

2. eset

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

$$S(G_0) = (X^* \cdot (\{S\} \cup Y^+) \cdot X^*) \setminus Y^+ \subset Q.$$

A bizonyítás hasonló az 1. eset bizonyításához. $S(G_0) \subseteq L$ igazolásához ismét indukciót használunk. Látható, hogy bármely szabályt alkalmazva bármely L-beli szóra, ismét L-beli szót kapunk. Az $L \subseteq S(G_0)$ bizonyítása mindössze annyiban tér el az 1. esettől, hogy az $Y \rightarrow YY$ szabályt is használjuk.

Ebben az esetben csak az $\{S \rightarrow XS,\ S \rightarrow SX,\ S \rightarrow XY,\ S \rightarrow YX,\ Y \rightarrow YY\}$ szabályok alkalmazása szükséges $S(G_0)$ előállításához.

$S(G_0) \subset Q$ következik abból, hogy minden $W \in S(G_0)$ szó $W = X^m S X^n$ vagy pedig $W = X^m Y^k X^n$ alakú, valamint $SY \notin S(G_0)$.

□

3. eset

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

$$S(G_0) = (X^* \cdot (\{S\} \cup X^+) \cdot Y^*) \setminus X^+ \subset Q.$$

Ismét az 1. esetben látott bízonyítást használjuk. $L \subseteq S(G_0)$ igazolható az $S \overset{n}{\Rightarrow} SY^n \overset{m}{\Rightarrow} X^m SY^n$ és az $S \overset{m-1}{\Rightarrow} X^{m-1}S \overset{n}{\Rightarrow} X^{m-1}SY^n \overset{1}{\Rightarrow} X^m Y^n$ ($m > 0$) levezetésekkel, ezért $SY^n, X^m SY^n, X^m Y^n \in L$.

Csak az $\{S \rightarrow XS,\ S \rightarrow SY,\ S \rightarrow XY\}$ szabályok szükségesek.

$S(G_0) \subseteq Q$ triviális, hiszen minden $W \in S(G_0)$ szó $W = X^m SY^n$ vagy pedig $W = X^m Y^n$ alakú, és $SX \notin S(G_0)$ miatt $S(G_0) \subset Q$.

□

4. eset

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

$$S(G_0) = \{X, Y\}^* \cdot \{S\} \cdot \{X, Y\}^* \subset Q.$$

Ismét hasonló az 1. esethez. $L \subseteq S(G_0)$ következik abból, hogy bármely USV szó, ahol $U, V \in \{X, Y\}^*$ levezethető az $\{S \rightarrow XS,\ S \rightarrow YS,\ S \rightarrow SX,\ S \rightarrow SY\}$ szabályokkal.

Mivel $W \in S(G_0)$ pontosan egy S betűt tartalmaz, és $XY \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

5. eset

$$P_0 = \{S \rightarrow XS,\ X \rightarrow SY,\ X \rightarrow XX,\ X \rightarrow YY\}.$$

$$S(G_0) = \{X, SY, YY\}^* \cdot \{S\}.$$

Az $S(G_0) = L$ igazolása ismét hasonló az 1. esethez. $L \subseteq S(G_0)$ következik az $S \overset{1}{\Rightarrow} XS,\ S \overset{1}{\Rightarrow} XS \overset{1}{\Rightarrow} YYS,\ S \overset{1}{\Rightarrow} XS \overset{1}{\Rightarrow} SYS$ levezetésekből.

A szükséges szabályaink $\{S \to XS,\ X \to YY,\ X \to SY\}$.

Tegyük fel, hogy $W = U^k \in S(G_0),\ k > 1$. Ekkor $U = YU'S$, ahol az U szó elején szereplő Y-ok száma páros, mivel a W szó elejére, az első S elé csak az $X \to YY$ alakú szabály alkalmazásával kerülhet Y. Ugyanakkor az U szó elején szereplő Y-ok száma páratlan, mert a W szóban minden S után páratlan számú Y áll, kivéve az utólsó S-t. Ez ellentmondás.

$SX \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

6. eset

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

$S(G_0) = \{X, SY\} \cdot \{X, Y, SY\}^* \cdot \{S\} \cup \{S\} \subset Q$.

$S(G_0) = L$ igazolása ismét az 1. esethez hasonló. $L \subseteq S(G_0)$ következik az $S \overset{1}{\Rightarrow} XS,\ X \overset{1}{\Rightarrow} XY$, és $X \overset{1}{\Rightarrow} SY \overset{1}{\Rightarrow} XSY$ levezetésekből.

Csak az $\{S \to XS,\ X \to XY,\ X \to SY\}$ szabályok alkalmazása szükséges.

$W = U^k \in S(G_0),\ k > 1$ esetén egyrészt $U = XU'S$ vagy $U = SYU'S$, másrészt viszont $U = YU'S$ ellentmondás.

Mivel $SX \notin S(G_0)$, így $S(G_0) \subset Q$.

□

7. eset

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

$S(G_0) = \{S, X\} \cdot \{Y, XS, XX\}^* \cdot \{Y, XS\} \cup \{S\} \subset Q$.

(a) $S(G_0) \subseteq L$ igazolása megegyezik az 1. esetben látottal.

(b) $L \subseteq S(G_0)$ a következő levezetések következménye: $S \overset{1}{\Rightarrow} SY,\ S \overset{1}{\Rightarrow} XY,\ S \overset{n}{\Rightarrow} SY^n,\ Y \overset{1}{\Rightarrow} XS \overset{1}{\Rightarrow} XSY,\ Y \overset{1}{\Rightarrow} XS \overset{1}{\Rightarrow} XXY$. Az $Y \to YY$ szabály kivételével a többi szabályra szükség van.

(c) Az $S(G_0) \subset Q$ igazolására 4 lehetőséget kell megvizsgálnunk. Ehhez legyen $W \in S(G_0) \setminus Q$, azaz $W = U^k,\ k > 1$.

(ca) $W = SVS$. Ekkor $U = SU'XS$ ellentmondás, hiszen SS nem részszava egyetlen $W \in S(G_0)$ szónak sem.

(cb) $W = SVY$. Ekkor $U = SU'Y$. Ekkor minden egymást követő két S között páratlan számú X áll, míg az utólsó S után páros számú. Ez ellentmondás.

(cc) $W = XVS$. Ekkor $U = XU'XS$. Ekkor minden egymást követő két S között páratlan számú X áll, míg az első S előtt páros számú. Ez ellentmondás.

(cd) $W = XVY$. Ekkor $U = XU'Y$. Ekkor az X-ek és S-ek számának az összege páros bármely két Y között, és páratlan az első Y előtt. Ez ellentmondás.

Így $S(G_0) \subset Q$, mivel $SX \notin S(G_0)$.

□

8. eset

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

Jelölje $n(S), n(X), n(Y)$ az S, X, Y betűk számát a generált szóban. Elindulva az $n(S) = 1,\ n(X) = 0,\ n(Y) = 0$ értékekből indukcióval bizonyítható, hogy az $n(Y) = 2n(X) + n(S) - 1$ egyenlőség mindig teljesül. Legyen $W = U^k,\ k > 1$, és jelölje $n_U(S), n_U(X), n_U(Y)$ rendre az U szóban található S-ek, X-ek és Y-ok számát. Ekkor $n(S) = kn_U(S),\ n(X) = kn_U(X),\ n(Y) = kn_U(Y)$ ellentmondás.

Ezek után $S(G_0) \subset Q$, mivel $SX \notin S(G_0)$.

□

9. eset

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

A szabályok felépítéséből következik, hogy bármely $W \in S(G_0)$ esetén $n(X) = n(Y)$, ahol $n(X)$ jelöli a W szóban szereplő X betűk számát, és $n(Y)$ az Y betűk számát. Továbbá minden $W \in S(G_0),\ W \neq S$ szó $W = XVY$ alakú.

Indukcióval bizonyítható, hogy minden Z szóra, ahol Z a $W \in S(G_0),\ W \neq S$ szó valódi kezdőszelete fenáll, hogy $n_Z(X) > n_Z(Y)$, ahol $n_Z(X), n_Z(Y)$ jelöli a Z szóban található X-ek és Y-ok számát. Ez $W = XY$ esetén triviális. Az $X \to XS$ és az $Y \to SY$ szabály alkalmazása esetén vagy csökken az $n_Z(Y)$ eggyel, vagy pedig az $n_Z(X)$ és az $n_Z(Y)$ értéke is változatlan marad.

Az $S \to XY$ szabály alkalmazásával pedig vagy megnő az $n_Z(X)$ eggyel, vagy az $n_Z(X)$ és az $n_Z(Y)$ is megnő eggyel.

Most legyen $W = U^k$, $k > 1$. Jelölje $n_U(X), n_U(Y)$ az U szóban található X-ek és Y-ok számát. Ekkor $n(X) = kn_U(X)$, $n(Y) = kn_U(Y)$, ezért $n_U(X) = n_U(Y)$. Ugyanakkor az U szó kezdőszelete a W szónak, ez ellentmondás.

Mivel $SX \notin S(G_0)$ ismét $S(G_0) \subset Q$.

$\square$

A fenti esetben $S(G_0)$ nem reguláris. Ezt a következőképpen igazoljuk:

$$(S(G_0) \cap \{X,Y\}^*) \cup \{\lambda\} = h(S(G_0)) = \{X\} \cdot D(X,Y) \cdot \{Y\} \cup \{\lambda\},$$

ahol a $h : \{S,X,Y\} \to \{X,Y\}$ homomorfizmus $h(S) = \lambda$, $h(X) = X$, $h(Y) = Y$, valamint $D(X,Y)$ az $\{X,Y\}$ ábécé feletti Dyck nyelv, azaz $\lambda \in D(X,Y)$ és minden $W \in D(X,Y)$, $W \neq \lambda$ egyedi módon felírható

$$W = \prod_{i=1}^{k} (X \cdot U_i \cdot Y)$$

alakban, ahol $U_i \in D(X,Y)$, $1 \leq i \leq k$.

Mivel a Dyck nyelvek nem generálhatóak reguláris nyelvtannal, ezért elegendő a fenti állítást bizonyítani.

$XY \in S(G_0)$, és tegyük fel, hogy $XUY \in S(G_0)$ minden $U \in D(X,Y)$, $|U| < |W|$ szóra. Ekkor

$$S \Rightarrow XY \overset{k}{\Rightarrow} XS^kY \overset{*}{\Rightarrow} X \cdot (\prod_{i=1}^{k} XU_iY) \cdot Y$$

miatt $XWY \in S(G_0) \cap \{X,Y\}^*$ minden $W \in D(X,Y)$ szóra.
Ebből következik, hogy $\{X\} \cdot D(X,Y) \cdot \{Y\} \subseteq S(G_0) \cap \{X,Y\}^*$.

A másik irányhoz tekintsük az alábbi állítást:

$$h(S(G_0)) \subseteq \{X\} \cdot D(X,Y) \cdot \{Y\} \cup \{\lambda\}.$$

Ezt is indukcioval igazolhatjuk, a következő módon:
$h(S) = \lambda$ és $h(XY) = XY \in D(X,Y)$. Tegyük fel, hogy $h(W) \in D(X,Y)$.
Ha $W = USV \overset{1}{\Rightarrow} UXYV = W'$ akkor $h(W') \in D(X,Y)$,

ha $W = UXV \overset{1}{\Rightarrow} UXSV = W'$ akkor $h(W') = h(W) \in D(X,Y)$, és ha $W = UYV \overset{1}{\Rightarrow} USYV = W'$ akkor $h(W') = h(W) \in D(X,Y)$. Mindezek miatt $\{X\} \cdot D(X,Y) \cdot \{Y\} \subseteq S(G_0) \cap \{X,Y\}^* \subseteq h(S(G_0)) \subseteq \{X\} \cdot D(X,Y) \cdot \{Y\} \cup \{\lambda\}$, így

$$S(G_0) \cap \{X,Y\}^* \cup \{\lambda\} = h(S(G_0)) = \{X\} \cdot D(X,Y) \cdot \{Y\} \cup \{\lambda\}.$$

Ezért $S(G_0)$ nem reguláris nyelv.

□

10. eset

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

Jelölje $n(S)$ és $n(Y)$ az S és Y betűk számát a generált szóban. A szabályok felépítéséből következik, hogy az $n(Y) = n(S) - 1$ egyenlőség minden $W \in S(G_0)$ szóra teljesül. Továbbá minden $W \in S(G_0)$ legalább egy S betűt tartalmaz. Legyen $W = U^k$, $k > 1$, és jelölje $n_U(S)$ és $n_U(Y)$ az U szóban található S-ek és Y-ok számát. Ekkor $n(S) = kn_U(S)$, $n(Y) = kn_U(Y)$ ellentmond a fenti egyenlőségnek.

Mivel $SY \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

11. eset

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

(a) Legyen $P_1 = \{S \to SYY,\ Y \to SYS\}$, $G_1 = (N \setminus \{X\}, S, P_1)$, és a $h : \{X,S,Y\} \to \{SY,S,Y\}$ homomorfizmus $h(X) = SY$, $h(S) = S$, $h(Y) = Y$.

(b) $S(G_1) = S(G_0) \cap \{S,Y\}^* = h(S(G_0)).$

$S(G_1) \subseteq S(G_0) \cap \{S,Y\}^*$ következik abból, hogy mindkét G_1-beli szabály levezethető a G_0 szabályaiból: $S \overset{1}{\Rightarrow} XY \overset{1}{\Rightarrow} SYY$, $Y \overset{1}{\Rightarrow} XS \overset{1}{\Rightarrow} SYS$.

A másik irány bizonyításához tekintsük egy tetszőleges $W \in S(G_0) \cap \{S,Y\}^*$ szó levezetési fáját. Az X egy csúcsba csak az $S \to XY$ vagy pedig az $Y \to XS$ szabály alkalmazásával kerülhet. Mivel nincs X-el jelölt levél a levezetési fában, ezért minden X-el jelölt csúcsra alkalmaztuk az $X \to SY$ szabályt. Ebből következik, hogy minden ilyen alkalommal G_1-ben vagy az $S \to SYY$ vagy pedig az $Y \to SYS$ szabálynak megfelelő szabályokat alkalmaztunk, két lépésben. Ezért $S(G_0) \cap \{S, Y\}^* \subseteq S(G_1)$.

$h(S(G_0)) \subseteq S(G_0) \cap \{S, Y\}^*$, hiszen egy tetszőleges X-re a $h(X) = SY$ homomorfizmus alkalmazása pontosan azt a hatást éri el, mintha az $X \to SY$ szabályt alkalmaztuk volna.

$S(G_1) = S(G_0) \cap \{S, Y\}^* \subseteq h(S(G_0))$, mivel $S(G_1) \subseteq S(G_0)$ és ezért $S(G_1) = h(S(G_1)) \subseteq h(S(G_0))$.

Mindezek következménye, hogy $S(G_0) \subseteq Q \Leftrightarrow S(G_1) \subseteq Q_{\{S,Y\}}$.

(c) Minden $W \in S(G_1)$ felírható az alábbi alakban:

$$W = (\prod_{i=0}^{m-1} S^{t_i} \cdot Y) \cdot S^{t_m}$$

ahol $m \geq 0$ és $t_i \geq 0,\ 0 \leq i \leq m$.

Jelölje $n(S)$ és $n(Y)$ az S és Y betűk számát a W szóban. Ekkor $n(Y) = m = 2n_Y$ és $n(S) = 2n_S + 1$ valamely n_Y, n_S egészekre, mivel az $S \to SYY$ szabály alkalmazása növeli $n(Y)$ értékét kettővel, az $Y \to SYS$ alkalmazása pedig az $n(S)$ értékét növeli kettővel. W levezetésében az $S \to SYY$ szabályt n_Y alkalommal használjuk.

Legyen $t_{0,0} = 1$, és jelölje $t_{i,j}$ az S-ek számát egy S-blokkban, ahol i az Y indexe és j a levezetés lépésszáma. Egy levezetésben az $Y \to SYS$ szabály alkalmazásakor a megfelelő $i \geq 0$ esetén $t_{i,j+1} = t_{i,j} + 1$ és $t_{i+1,j+1} = t_{i+1,j} + 1$. Ennek következtében egy páros indexű S-blokk és egy páratlan indexű S-blokk száma nő eggyel. Az $S \to SYY$ szabály alkalmazásakor $t_{k,j+1} = t_{k,j}$ ha $k < i$, továbbá $t_{i,j+1} + t_{i+2,j+1} = t_{i,j}$, $t_{i+1,j+1} = 0$, és végül $t_{k+2,j+1} = t_{k,j}$ ha $k > i$, a megfelelő i esetén. Így a páros indexű S-blokkok és a páratlan indexű S-blokkok számossága is változatlan marad.

Mindezek miatt igaz az alábbi egyenlőség:

$$\sum_{i=0}^{n_Y} t_{2i} = 1 + \sum_{i=0}^{n_Y - 1} t_{2i+1}.$$

(d) Legyen $W = U^k,\ k > 1$ valamely $U \in \{S, Y\}^+$ szóra. Ekkor

$$U = (\prod_{i=0}^{n-1} S^{t_i} \cdot Y) \cdot S^{t_m}$$

ahol $m = kn \equiv 0 \ (mod\ 2)$.

Mivel $n_W(S) = kn_U(S) \equiv 1 \ (mod\ 2)$ és $n_W(Y) = kn_U(Y) \equiv 0 \ (mod\ 2)$, ezért $k \equiv 1 \ (mod\ 2)$, $n_U(S) \equiv 1 \ (mod\ 2)$ és $n = n_U(Y) \equiv 0 \ (mod\ 2)$. Így $n_U(Y) = 2n'$.

Továbbá minden $0 < i < n$, $0 \leq j < k$ esetén $t_i = t_{i+jn}$, és minden $0 < j < k$ esetén $t_{jn} = t_0 + t_m$, hiszen $W = U^k$.

Mindezek következménye, hogy

$$\sum_{i=0}^{n_Y-1} t_{2i+1} = k \cdot \sum_{i=0}^{n'-1} t_{2i+1}$$

és

$$\sum_{i=0}^{n_Y} t_{2i} = k \cdot (\ (t_0 + t_m) + \sum_{i=1}^{n'-1} t_{2i})$$

ezért

$$\sum_{i=0}^{n_Y} t_{2i} - \sum_{i=0}^{n_Y-1} t_{2i+1} \equiv 0 \ (mod\ k),$$

ami ellentmond annak, hogy

$$\sum_{i=0}^{n_Y} t_{2i} - \sum_{i=0}^{n_Y-1} t_{2i+1} \equiv 1 \ (mod\ k).$$

$SX \notin S(G_0)$, és mivel igazoltuk, hogy $S(G_1) \subseteq Q_{\{S,Y\}}$ ezért $S(G_0) \subset Q$.

□

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.

5.5. 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. Ezt az állítást indirekten bizonyítjuk. Tegyük fel, hogy az ASB szót generálja a nyelvtan, ahol S a mondatszimbólum, A, B pedig megfelelő szavak. Ekkor létezik olyan primitív szó, hogy X^nBAX^n valamely X betűre és n pozitív egészre, tehát az X^nBAX^n szó levezethető a mondatszimbólumból. Ekkor viszont az AX^nBAX^nB szó is levezethető, ami nemprimitív szó, ezzel ellentmondáshoz jutottunk.

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

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

$$S(G_0) = (X^+ \cdot Y^+) \cup (Y^+ \cdot X^+) \cup (X^+ \cdot Z^+) \cup (Z^+ \cdot X^+) \cup (Y^+ \cdot Z^+) \cup$$
$$(Z^+ \cdot Y^+) \subset Q.$$

A bizonyítás hasonló az 5.4. fejezet 1. esetének bizonyításához. $S(G_0) \subseteq L$ igazolásához ismét indukciót használunk. Látható, hogy bármely szabályt alkalmazva bármely L-beli szóra, ismét L-beli szót kapunk. Az $L \subseteq S(G_0)$ bizonyítása mindössze annyiban tér el az 5.4. fejezet 1. esetének bizonyításától, hogy minden szabályt felhasználunk $S(G_0)$ előállításához.

$S(G_0) \subset Q$ következik $S(G_0)$ felépítéséből, valamint abból, hogy $XYX \notin S(G_0)$.

$\square$

2. eset

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

$$S(G_0) = (X^+ \cdot Y^+ \cdot X^*) \cup (Y^+ \cdot X^+ \cdot Y^*) \cup (X^+ \cdot \{Z\}) \cup (\{Z\} \cdot X^+) \cup\\ (Y^+ \cdot \{Z\}) \cup (\{Z\} \cdot Y^+) \subset Q.$$

A bizonyítás gyakorlatilag megegyezik az 1. eset bizonyításával. $S(G_0) \subseteq L$ indukcióval igazolható, $L \subseteq S(G_0)$ bizonyításához ismét minden szabály alkalmazása szükséges.

$S(G_0) \subset Q$, mivel $XYZ \notin S(G_0)$.

$\square$

3. eset

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

$$S(G_0) = (Y^* \cdot X^* \cdot \{Z, \lambda\} \cdot Y^*) \cup (X^* \cdot \{Z, \lambda\} \cdot Y^* \cdot X^*) \setminus \{\lambda, X^+, Y^+, Z\} \subset Q.$$

A bizonyítás ismét megegyezik az 1. eset bizonyításával. $S(G_0) \subseteq L$, és az $L \subseteq S(G_0)$ bizonyításához minden szabály alkalmazása szükséges.

$XYZ \notin S(G_0)$, így $S(G_0) \subset Q$.

$\square$

4. eset

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

$$S(G_0) = (X^+ \cdot Y^+) \cup (Y^+ \cdot X^+) \cup (\{X, Y\}^* \cdot \{Z\} \cdot \{X, Y\}^*) \setminus \{Z\} \subset Q.$$

Az $L \subseteq S(G_0)$ bizonyításához minden szabály alkalmazása szükséges.

$S(G_0) \subseteq Q$, mivel minden $W \in S(G_0)$ szó vagy pontosan egy Z betűt tartalmaz, vagy X^+Y^+ illetve Y^+X^+ alakú, és $S(G_0) \subset Q$, mivel $XYX \notin S(G_0)$.

$\square$

5. eset

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

Jelölje $n(Y), n(Z)$ az Y, Z betűk számát a generált szóban. Indukcióval bizonyítható, hogy az $n(Y) \pm 1 = n(Z)$ egyenlőség teljesül minden $W \in S(G_0)$, $|W| \geq 3$ szóra. Legyen $W = U^k$, $k > 1$, és jelölje $n_U(Y), n_U(Z)$ rendre az U szóban található Y-ok és Z-k számát. Ekkor $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, így $kn_U(Y) \pm 1 = kn_U(Z)$, $k > 1$, ami ellentmondás.

Ezek után $S(G_0) \subset Q$, mivel $XYX \notin S(G_0)$.

□

6. eset

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

(a) Ha a generálás az $\{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX\}$ szabályok valamelyikéből indul, akkor az előző esetben bizonyított $n(Y) \pm 1 = n(Z)$ egyenlőség fennáll.

(b) Ha a generálás az $S \to YZ$ szabályból indul, akkor az 5.4. fejezet 9. esetében bizonyított vázzal betű-izomorf vázat kapunk. (Az $X \to XX$ szabály felesleges.)

(c) Végül vizsgáljuk meg azt az esetet, hogy a generálás az $S \to ZY$ szabályból indul. Tegyük fel, hogy létezik W nemprimitív szó, melyet az $S \to ZY$ szabályból indulva generál a nyelvtan. Ekkor, mivel a nemprimitív szavak zártak a ciklikus permutációra nézve, létezik W_2 nemprimitív szó is, melyet az $S \to YZ$ szabályból kiindulva kapunk. Ez viszont ellentmond a (b) esetben bizonyított eredménynek.

$XYY \notin S(G_0)$, így $S(G_0) \subset Q$.

□

7. eset

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

Ebben az esetben minden $S(G_0)$-beli W szóban

(a) pontosan egy Y betű található, ha az $\{S \to XY,\ S \to YX,\ S \to YZ,\ S \to ZY\}$ szabályok valamelyikéből indulunk ki,

(b) vagy pedig pontosan egy Z betű, ha az $\{S \to XZ,\ S \to ZX\}$ szabályokból indulunk ki.

$XYY \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

8. eset

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

(a) Az $\{S \to XZ,\ S \to ZX,\ S \to YZ,\ S \to ZY\}$ szabályok közül bármelyikből kiindulva, a generált szóban mindig pontosan egy X vagy pedig pontosan egy Y lesz.

(b) Az $\{S \to XY,\ S \to YX\}$ szabályokból indulva pedig az $n(X)+n(Y) = 2$ egyenlőség fennáll minden szóban, és amennyiben $n(X) = 2$ vagy pedig $n(Y) = 2$, akkor $n(Z)$ páratlan.

$XYX \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

9. eset

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

$S(G_0) = (X^+ \cdot \{Y, Z\}^+) \cup (\{Y, Z\}^+ \cdot X^+) \subset Q.$

A bizonyítás ismét megegyezik az 1. eset bizonyításával. Az $L \subseteq S(G_0)$ bizonyításához csak az $\{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\ Y \to YY,\ Z \to ZZ,\ Y \to YZ,\ Y \to ZY\}$ szabályok alkalmazása szükséges.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

10. eset

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

$$S(G_0) = (X^* \cdot \{Z, Y^+\} \cdot X^*) \setminus \{Z, Y^+\} \subset Q.$$

A bizonyítás ismét az 1. eset bizonyításához hasonló. Az $L \subseteq S(G_0)$ bizonyításához csak az $\{S \to XY,\ S \to YX,\ S \to XZ,\ S \to ZX,\ X \to XX,\ Y \to YY,\ Z \to XY,\ Z \to XZ\}$ szabályok alkalmazása szükséges.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

11. eset

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

$$S(G_0) = (X^* \cdot \{Z, \lambda\} \cdot Y^* \cdot X^*) \setminus \{\lambda, X^+, Y^+, ZY^*\} \subset Q.$$

A bizonyítás ismét az 1. eset bizonyításához hasonló. Az $L \subseteq S(G_0)$ bizonyításához a $Z \to YY$ szabály kivételével minden szabály szükséges.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

12. eset

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

(a) Az $\{S \to XZ,\ S \to ZX\}$ szabályok közül bármelyikből kiindulva a generált szóban mindig pontosan egy Z betű lesz.

(b) Az $S \to YX$ szabályból indulva a generált W szó mindig $Y\{Y, Z, X^+Z\}^*X^+$ alakú lesz. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = YU'X$ alakú, másrészt viszont $U = ZU'X$ vagy $U = XU'X$ alakú, ami ellentmondás.

(c) Mivel az $S \to YX$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve, így az $S \to YX$ szabályból indulva is csak primitív szavakat kapunk.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

13. eset

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

Jelölje $n(Y), n(Z)$ az Y, Z betűk számát a generált szóban. Indukcióval bizonyítható, hogy az $n(Y) \pm 1 = n(Z)$ egyenlőség teljesül minden $W \in S(G_0)$ szóra. Legyen $W = U^k$, $k > 1$, és jelölje $n_U(Y), n_U(Z)$ rendre az U szóban található Y-ok és Z-k számát. Ekkor $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, így $kn_U(Y) \pm 1 = kn_U(Z)$, $k > 1$, ami ellentmondás.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

14. eset

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

(a) Az $S \to XY$ szabályból kiindulva a generált W szó mindig $\{X, Y\}V\{Y, XZ^i, YZ^j\}$ alakú lesz a megfelelő V szóra, ahol $i \geq 1$, i páratlan, $j \geq 2$, j páros. Ugyanakkor W-ben minden Y betűt, minden XZ^i, $i \geq 1$, i páratlan szót és minden YZ^j, $j \geq 2$, j páros szót is Z követ. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = \{X, Y\}U'$ alakú, másrészt viszont $U = ZU'$ alakú, ami ellentmondás.

(b) Az $S \to YX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XY$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

(c) Az $S \to XZ$ szabályból kiindulva a generált W szó vagy XZ^+ vagy YZ^+ vagy pedig $\{X, Y\}V\{XZ^i, YZ^j\}$ alakú lesz a megfelelő V szóra, ahol $i \geq 1$, i páratlan, $j \geq 2$, j páros. Ugyanakkor W-ben minden XZ^i, $i \geq 1$, i páratlan szót és minden YZ^j, $j \geq 2$, j páros szót is Z követ. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = \{X, Y\}U'$ alakú, másrészt viszont $U = ZU'$ alakú, ami ellentmondás.

(d) Az $S \to ZX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XZ$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

15. eset

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

Indukcióval bizonyítható, hogy a generált szó mindig pontosan egy Y, vagy pedig pontosan egy Z betűt tartalmaz.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

16. eset

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

Indukcióval bizonyítható, hogy a generált szó mindig pontosan egy X betűt tartalmaz.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

17. eset

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

(a) Az $\{S \to XZ,\ S \to ZX\}$ szabályokból kiindulva az $\{XZ, ZX\}$ véges nyelvet kapjuk.

(b) Az $S \to YX$ szabályból kiindulva a generált W szó mindig $\{X, Y, Z^i\}V\{X\}$ alakú lesz a megfelelő V szóra, ahol $i \geq 2$, i páros. Ugyanakkor W-ben minden X betűt páratlan számú Z követ. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = \{X, Y, Z^i\}U'$ alakú, ahol i páros, másrészt viszont $U = Z^iU'$ alakú, ahol i páratlan. Ez ellentmondás.

(c) Az $S \to XY$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to YX$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

18. eset

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

(a) Az $\{S \to XZ,\ S \to ZX\}$ szabályokból indulva az $\{XZ^+, Z^+X\}$ csak primitív szavakból álló nyelvet kapjuk.

(b) Az $S \to YX$ szabályból kiindulva a generált W szó mindig $\{X,Y\}V\{X\}$ alakú lesz a megfelelő V szóra. Ugyanakkor W-ben minden X betűt Z betű követ. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = \{X,Y\}U'$ alakú, másrészt viszont $U = ZU'$ alakú. Ez ellentmondás.

(c) Az $S \to XY$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to YX$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

19. eset

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

Jelölje $n(X), n(Y), n(Z)$ az X, Y, Z betűk számát a generált szóban.

(a) Először vizsgáljuk meg azokat a szavakat, melyeket az $\{S \to XY,\ S \to YX\}$ szabályokból indulva generál a váz. Indukcióval bizonyítható, hogy ebben az esetben a $2n(Y) + n(Z) - n(X) = 1$ egyenlőség teljesül minden $W \in S(G_0)$ szóra. Legyen $W = U^k$, $k > 1$, és jelölje $n_U(X), n_U(Y), n_U(Z)$ rendre az U szóban található X-ek, Y-ok és Z-k számát. Ekkor $n(X) = kn_U(X)$, $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, így $k(2n_U(Y) + n_U(Z) - n_U(X)) = 1$, ami ellentmondás.

(b) Most tekintsük azt az esetet, mikor az $S \to XZ$ szabályból indulunk el. Ekkor a $2n(Y) + n(Z) = n(X)$ egyenlőség teljesül minden $W \in S(G_0)$ szóra. Ugyanakkor minden $W \in S(G_0)$ szó valódi kezdőszeletében igaz a $2n(Y) + n(Z) < n(X)$ egyenlőtlenség. Legyen $W = U^k$, $k > 1$, és jelölje $n_U(X), n_U(Y), n_U(Z)$ rendre az U szóban található X-ek, Y-ok és Z-k számát. Ekkor $n(X) = kn_U(X)$, $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, és mivel az U szó a W szó valódi kezdőszelete, így teljesül a $k(2n_U(Y) + n_U(Z)) < kn_U(X)$ egyenlőtlenség, ami ellentmond annak, hogy $2n(Y) + n(Z) = n(X)$.

(c) Az $S \to ZX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XZ$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$YZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

20. eset

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

(a) Az $S \to XY$ szabályból kiindulva a generált W szó mindig $\{X, Y, Z^i\}V\{Y\}$ alakú lesz a megfelelő V szóra, ahol $i \geq 2$, i páros. Ugyanakkor W-ben minden Y betűt páratlan számú Z követ. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = \{X, Y, Z^i\}U'$ alakú, ahol i páros, másrészt viszont $U = Z^iU'$ alakú, ahol i páratlan. Ez ellentmondás.

(b) Az $S \to YX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XY$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

21. eset

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

$$S(G_0) = (\{X, Z\}^* \cdot X \cdot Z^* \cdot \{Y, Z\}^* \cdot Y \cdot Z^*) \cup\\ (\{Y, Z\}^* \cdot Y \cdot Z^* \cdot \{X, Z\}^* \cdot X \cdot Z^*) \subset Q.$$

Ebben az esetben az $L \subseteq S(G_0)$ bizonyításához a $Z \to ZZ$ szabály kivételével minden szabályra szükség van.

Legyen L_1 az a nyelv, melyet az $S(G_0)$ nyelvből úgy kapunk, hogy az $S(G_0)$ szavaiból az összes Z betűt elhagyjuk. $L_1 = \{X^+Y^+, Y^+X^+\}$. Tegyük fel, hogy létezik a váz által generált W nemprimitív szó. Ekkor létezik $W' \in L_1$

nemprimitív szó, melyet úgy kapunk, hogy a W szóból elhagyjuk az összes Z betűt. Ez ellentmondás.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

22. eset

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

(a) Az $S \to XY$ szabályból kiindulva a generált W szó mindig $\{X\}V\{Y\}$ alakú lesz a megfelelő V szóra. Ugyanakkor W-ben minden Y betűt Y vagy Z követ. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = XU'$ alakú, másrészt viszont $U = \{Y, Z\}U'$ alakú. Ez ellentmondás.

(b) Az $S \to YX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XY$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

23. eset

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

(a) Az $S \to XY$ szabályból kiindulva a generált W szó mindig $\{X, Z, Y^i\}V\{Y^j\}$ alakú lesz a megfelelő V szóra, ahol $i \geq 2$, i páros, $j \geq 1$, j páratlan. Ugyanakkor W-ben minden páratlan számú Y betűt páratlan számú Y követ. (Eltekintve a szó végén található Y-blokktól.) Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = \{X, Z, Y^i\}U'$ alakú, ahol $i \geq 2$, i páros, másrészt viszont $U = Y^jU'$ alakú, ahol $j \geq 1$, j páratlan. Ez ellentmondás.

(b) Az $S \to YX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XY$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

24. eset

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

(a) Legyen L_1 az a nyelv, melyet az $S(G_0)$ nyelvből úgy kapunk, hogy az $S(G_0)$ szavaiból az összes Z betűt elhagyjuk. Az $S \to XY$ szabályból kiindulva a generált $W \in S(G_0)$ szóhoz tartozó $W' \in L_1$ szó mindig $\{X, Y^i\}V\{Y^j\}$ alakú lesz a megfelelő V szóra, ahol $i \geq 2$, i páros, $j \geq 1$, j páratlan. Ugyanakkor W'-ben minden páratlan számú Y betűt páratlan számú Y követ. (Eltekintve a szó végén található Y-blokktól.) Ekkor ha $W' = U^k$, $k \geq 2$, akkor egyrészt $U = \{X, Y^i\}U'$ alakú, ahol $i \geq 2$, i páros, másrészt viszont $U = Y^jU'$ alakú, ahol $j \geq 1$, j páratlan. Ezzel bizonyítottuk, hogy minden, az $S \to XY$ szabályból induló levezetéssel kapható $W \in S(G_0)$ szóhoz tartozó $W' \in L_1$ szó primitív szó. Tegyük fel, hogy a G_0 váz generál az $S \to XY$ szabályból indulva $W \in S(G_0)$ nemprimitív szót. Ekkor az a $W' \in L_1$ szó is nemprimitív lesz, melyet úgy kapunk, hogy a W szóból elhagyjuk az összes Z betűt. Ez ellentmondás.

(b) Az $S \to YX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XY$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

25. eset

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

(a) Az $S \to XY$ szabályból kiindulva a generált W szó mindig $\{X, Y\}V\{Y\}$ alakú lesz a megfelelő V szóra. Ugyanakkor W-ben minden Y betűt Z követ. Ekkor ha $W = U^k$, $k \geq 2$, akkor egyrészt $U = \{X, Y\}U'$ alakú, másrészt viszont $U = ZU'$ alakú, ami ellentmondás.

(*b*) Az $S \to YX$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to XY$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

26. eset

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

Jelölje $n(X), n(Y), n(Z)$ az X, Y, Z betűk számát a generált szóban. Indukcióval bizonyítható, hogy a $2n(X) + n(Z) - n(Y) = 1$ egyenlőség teljesül minden $W \in S(G_0)$ szóra. Legyen $W = U^k$, $k > 1$, és jelölje $n_U(X), n_U(Y), n_U(Z)$ rendre az U szóban található X-ek, Y-ok és Z-k számát. Ekkor $n(X) = kn_U(X)$, $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, így $k(2n_U(X) + n_U(Z) - n_U(Y)) = 1$, ami ellentmondás.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

27. eset

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

Legyen L_1 az a nyelv, melyet úgy kapunk, hogy az $S(G_0)$ szavaiból az összes Z betűt elhagyjuk. Mivel minden $W \in S(G_0)$ tartalmaz X és Y betűt is, ezért ha W nemprimitív szó, akkor az összes Z betűt elhagyva W-ből $W' \in L_1$ nemprimitív szót kapunk. Ezért annak bizonyításához, hogy $S(G_0)$ nem tartalmaz nemprimitív szót, elegendő igazolni, hogy az L_1 nyelv nem tartalmaz nemprimitív szót.

Az L_1 nyelv szavainak felépítése a következő: Az XY vagy pedig az YX szóból indulunk, és bármely betű elé vagy után beszúrhatunk XX vagy YY szót.

Legyen $V = U^k$, $k > 1$ tetszőleges nemprimitív szó az $\{X, Y\}$ ábécé felett. Hagyjuk el U-ból az összes XX szót, majd az összes YY szót, és ismételjük ezt az eljárást addig, amíg létezik XX vagy YY szó. Az így kapott U' szó eleme lesz az $\{\{XY\}^i \cdot \{X, \lambda\}, \{YX\}^i \cdot \{Y, \lambda\} \mid i \geq 0\}$ halmaznak. Legyen V' az a szó, melyet úgy kapunk, hogy az U'^k szóra végrehajtjuk a fentebbi eljárást, azaz az U'^k szóból elhagyjuk az összes XX szót, majd az összes YY szót. Ekkor ha az U' szó hossza páros volt, akkor V' eleme lesz az $\{\{XY\}^{i*k}, \{YX\}^{i*k} \mid i \geq 0,\ k > 1\}$ halmaznak, ha pedig az U' szó hossza páratlan volt, akkor V' eleme lesz az $\{U', \lambda\}$ halmaznak.

Ezek után tegyük fel, hogy létezik $W' \in L_1,\ W' = U^k,\ k > 1$ nemprimitív szó. Ekkor mivel a W' szóban páratlan számú X és páratlan számú Y található, ezért k páratlan, és az U szó hossza páros. Hajtsuk végre a W' szón a fentebbi eljárást. A kapott W'' szó ekkor eleme lesz az $\{\{XY\}^{i*k}, \{YX\}^{i*k} \mid i \geq 0,\ k > 1\}$ halmaznak, mivel U' páros. Ugyanakkor a generálás az XY vagy pedig az YX szóból indult, ami nem felel meg a fentebbi alaknak, ez ellentmondás.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

28. eset

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

(a) Ha a generálás az $S \to YX$ szabályból indul, akkor az 5.4. fejezet 11. esetében bizonyított vázzal betű-izomorf vázat kapunk.

(b) Az $S \to XY$ szabályból indulva a generált W szó mindig primitív lesz, hisz az $S \to YX$ szabályból indulva csak primitív szavakat kapunk, és a primitív szavak halmaza zárt a ciklikus permutációra nézve.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

□

Végezetül tekintsük a nem redukált eseteket:

29. eset

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

$S(G_0) = (X^+ \cdot Y^+) \cup (Y^+ \cdot X^+) \subset Q$.

Nem redukált nyelvtan, Z nem szerepel egyetlen $W \in S(G_0)$ szóban sem. Az $L \subseteq S(G_0)$ bizonyításához csak az $\{S \to XY,\ S \to YX,\ X \to XX,\ Y \to YY\}$ szabályok szükségesek.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

30. eset

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

$S(G_0) = (X^* \cdot Y \cdot X^*) \setminus \{Y\} \subset Q$.

Nem redukált nyelvtan, Z nem szerepel egyetlen $W \in S(G_0)$ szóban sem. Az $L \subseteq S(G_0)$ bizonyításához csak az $\{S \to XY,\ S \to YX,\ Y \to XY,\ Y \to YX\}$ szabályok szükségesek.

$XZ \notin S(G_0)$, ezért $S(G_0) \subset Q$.

$\square$

Összefoglaló

A disszertáció öt fejezetből áll. Az első fejezetben rövid áttekintést adunk a formális nyelvek elméletének alapvető fogalmairól, melyeket a disszertáció további részeiben használunk. A második és a harmadik fejezetet a kutatási téma ismertetésére és az irodalmi előzmények bemutatására szánjuk. A negyedik és az ötödik fejezetben a saját eredmények bemutatására kerül sor.

Az **első fejezet**ben bevezetjük az ábécé, a szó, a nyelv és a nyelvtan fogalmát, ismertetjük a Chomsky-hierarchiát, végül pedig a Chomsky-féle normál forma és a környezetfüggetlen nyelvtanok kapcsolatáról szóló tételt ismertetjük, melyet használni fogunk a későbbiekben.

A **második fejezet**ben különböző nyelveknek a Chomsky-féle hierarchiában elfoglalt helyéről esik szó. Bevezetjük a palindróma, a repetitív-, a nemrepetitív-, a primitív- és a nemprimitív szó fogalmát, és megadjuk az általuk alkotott nyelvek Chomsky-féle hierarchiában elfoglalt helyét. Megmutatjuk, hogy a legalább kétbetűs ábécé feletti összes palindrómák halmaza környezetfüggetlen, a legalább hárombetűs ábécé feletti összes nemrepetitív szavak halmaza környezetfüggő és a legalább hárombetűs ábécé feletti összes repetitív szavak halmaza is környezetfüggő nyelvet alkot. A legalább kétbetűs ábécé feletti összes nemprimitív szavak halmaza szintén környezetfüggő.

A fentebbi állítások igazolásában a Bar-Hillel Lemma - közismert nevén pumpálós lemma környezetfüggetlen nyelvekre - igen fontos szerepet játszik, mivel azok a nyelvek melyek nem elégítik ki a Bar-Hillel feltételt nem generálhatóak környezetfüggetlen nyelvtannal. Q, egy legalább kétbetűs ábécé feletti összes primitív szavak halmaza kielégíti a Bar-Hillel feltételt. Ezek után számos, erősebb iterációs feltételt ismertetünk, (Ogden feltétel, Bader-Moura feltétel, erős Bader-Moura feltétel valamint Sokolowsky feltétel,) és megmutatjuk, hogy mindezen feltételeket kielégíti a Q nyelv, ezért ezzel a módszerrel nem lehet igazolni a Q nyelv környezetfüggőségét. Továbbra is megoldatlan probléma marad, hogy Q valódi környezetfüggő nyelv-e, vagy pedig generálható környezetfüggetlen nyelvtannal is.

A **harmadik fejezet** elején új fogalmakat vezetünk be: konjugált szavak, sűrű nyelvek és diszjunktív nyelvek. Ezek után felvetünk számos kérdést a nemprimitív szavakat generáló környezetfüggetlen nyelvek, sűrű nyelvek és diszjunktív nyelvek témaköréből. Ezek egy része könnyen megválaszolható, míg más kérdések megválaszolásához szükség van az $X^+ \setminus Q$ és a $Q^{(2)}$ elemeiből álló környezetfüggetlen nyelvek felépítéseinek meghatározására. A következő részben megadjuk az $X^+ \setminus Q$ elemeiből álló környezetfüggetlen nyelvek karakterizációját, mely után már meg tudjuk válaszolni az egyik kérdést: Ha $i \geq 3$, akkor nem létezik $L \subseteq X^*$ végtelen környezetfüggetlen nyelv úgy, hogy $L \subseteq Q^{(i)}$. A következő rész fő eredménye, hogy ismertetjük a $Q^{(2)}$ elemeiből álló környezetfüggetlen nyelvek felépítését, majd ennek ismeretében újabb két megállapítást tehetünk: Nem létezik olyan $L \subseteq X^*$ sűrű környezetfüggetlen nyelv, melyre $L \subseteq Q^{(2)}$. ; Legyen $L \subseteq X^*$ sűrű környezetfüggetlen nyelv. Ekkor $L \cap Q$ sűrű nyelv. Végezetül egy megoldatlan problémát ismertetünk: Ha $L \subseteq X^*$ diszjunktív környezetfüggetlen nyelv, akkor $L \cap Q$ diszjunktív nyelv?

A harmadik fejezetben ismertettük Ito és Katsura eredményét, akik megadták az $X^+ \setminus Q$ elemeiből álló környezetfüggetlen nyelvek karakterizációját. A **negyedik fejezet**ben ezt az eredményt terjesztjük ki lineáris és reguláris nyelvekre, teljessé téve ezzel a Chomsky-féle hierarchia különböző szintjein elhelyezkedő, csak nemprimitív szavakból álló nyelvek felépítéseinek meghatározását. Elsőként ismertetünk néhány tételt, melyeket a bizonyításainkban használni fogunk: megadjuk Lyndon és Schützenberger egyik tételét, valamint az iterációs lemmákat lineáris és reguláris nyelvekre. Ezek után megadjuk az $X^+ \setminus Q$ elemeiből álló lineáris és reguláris nyelvek karakterizációját, majd összefoglaljuk az ismert eredményeket, megadva a nemprimitív szavakból álló nyelvek hierarchiáját.

Az **ötödik fejezet**ben az összes, legfeljebb három nemterminálist tartalmazó, Chomsky-féle normál formájú, csak primitív szavakat generáló nyelvtan karakterizációját adjuk meg. Ezek a nyelvtanok kicsik és maximálisak. Kicsik abban az értelemben, hogy legfeljebb három nemterminálist tartalmaznak, és maximálisak, mivel bármely újabb szabályt hozzájuk véve generálnak nemprimitív szót is. 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, valamint a vázakhoz tartozó mondatformák halmazát. 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 elején bevezetjük a betű-izomorf nyelvtan, a redukált nyelvtan, a váz és a váz által generált mondatformájú nyelv fogalmát. Ezek után megadjuk az egy és két nemterminálisból álló, csak primitív szavakat generáló vázakat és az általuk generált nyelvek karakterizációját. Három nemterminális esetén a lehetséges vázak nagy száma miatt szükség volt számítógép és megfelelő program használatára, mellyel meg lehetett határozni a lehetséges maximális vázakat. A program segítségével 12 különböző maximális vázat kaptunk, eltekintve a mondatszimbólumtól különböző két nemterminális felcserélésével vagy pedig az összes $X \to YZ$ alakú szabály $X \to ZY$ alakú szabályra cserélésével kapható szimmetrikus vázaktól. A program feladata az volt, hogy kiválogassa azokat a maximális vázakat, melyek 12 hosszúságig csak primitív szavakat generálnak. Ezek után matematikai módszerekkel igazoltuk, hogy a fentebbi 11 redukált és 1 nem redukált váz mindegyike csak primitív szavakat generál. 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.

Summary

"Context-free languages have played an important role as bases of programming languages from the practical point of view. On the other hand, the algebraic studies of languages have been indispensable not only for the theory of formal languages but also for the theory of semigroups from the theoretical point of view. There have been many tools for classifying languages. Among them, the study on relationships between the free semigroup X^+ generated by an alphabet X and the set of all primitive words Q over X has received special interest."[18] In this dissertation we investigate Q, the language of all primitive words, and $X^+ \setminus Q$, the language of all nonprimitive words over an alphabet.

The dissertation consists of five chapters. In the first chapter we introduce the basic notions and notations of formal language theory. The second and third chapters are the preliminaries of our new results, which are presented in the fourth and fifth chapters.

In the **first chapter** we introduce the notions of alphabet, words, languages and grammar. We present the Chomsky-hierarchy and we show the theorem about the connection of the context-free grammars and the Chomsky normal form, which is used in the following chapters.

In the **second chapter** we investigate various combinatorial properties of languages and their position in the Chomsky-hierarchy. We introduce the notions of palindromicity, repetitivity, nonrepetitivity, primitivity and nonprimitivity. We show that the set of all palindromes over the alphabet contains at least two letters is properly context-free. The set of all nonrepetitive words over a three or more letters alphabet is not context-free, and the set of all repetitive words over an alphabet having three or more letters is not context-free. The set of all nonprimitive words over an alphabet having at least two letters is not context-free.

In the proofs of these theorems, the Bar-Hillel lemma - a well-known iteration lemma for context-free languages - plays an important role. Q, the language of all primitive words over an alphabet having two or more letters satisfies the Bar-Hillel condition. We investigate various iteration conditions, such as Ogden's, Bader and Moura's, BM strong and Sokolowsky's condition, and we show that Q satisfies the conditions of all above-mentioned iteration lemmas. It remains an unsolved problem whether or not Q is properly context-sensitive.

At the beginning of the **third chapter** we introduce the notions of conjugate words, dense languages, principal congruence, and disjunctive languages. After them, we present various problems related to context-free languages consisting of nonprimitive words, dense languages, and disjunctive languages. Some of these problems are quite simple, but to answer the other part of questions we have to determine the structures of context-free languages contained in $X^+ \setminus Q$ and in $Q^{(2)}$. In the next section we give the characterization of context-free languages consisting of $X^+ \setminus Q$, and then we can answer one of the previous questions: For any $i \geq 3$, there is no infinite context-free language $L \subseteq X^*$, such that $L \subseteq Q^{(i)}$. The main result of the following section is the characterization of the context-free languages contained in $Q^{(2)}$. The solving of this problem answers some more questions: There is no dense context-free language $L \subseteq X^*$, such that $L \subseteq Q^{(2)}$. ; Let $L \subseteq X^*$ be a dense context-free language. Then $L \cap Q$ is a dense language. Finally we present an open problem: Is $L \cap Q$ a disjunctive language when $L \subseteq X^*$ is a disjunctive context-free language?

In the third chapter we present the result of Ito and Katsura, determining the structure of the context-free languages contained in $X^+ \setminus Q$. In the **fourth chapter** we extend the characterization for linear and regular languages. First we introduce Lyndon and Schützenberger's theorem, and the iteration lemmas for linear and regular languages. These theorems play an important role in our following proofs. In the following section we give the characterization of the linear languages contained in $X^+ \setminus Q$, and the characterization of regular languages contained in $X^+ \setminus Q$. Finally we summarize our results, and present a hierarchy of languages consisting of non-primitive words. It remained a further problem to characterize some important special classes of context-sensitive languages containing non-primitive words.

In the **fifth chapter** we characterize all context-free grammars having Chomsky normal form with not more than three nonterminals generating only primitive words. These grammars are small with respect to nonterminals and maximal with respect to productions. Since a necessary condition for the generated language to contain only primitive words (over terminal symbols) is that all sentential forms are also primitive words (over nonterminals) it suffices to consider only the sentential form languages. It was the hope to deduce from the structure of such grammars some insight for a proof of the conjecture that the entire set of primitive words is not context-free by showing that there are always certain primitive words missing in the language generated by the grammar. Another conjecture was that any such grammar (and also all non-maximal ones) generate only regular sets of primitive words. This conjecture does not hold since we show that there exists a maximal grammar with 3 nonterminals generating a nonregular infinite set of primitive words. It turned out that all such grammars generate infinite sets of primitive words.

At the beginning of this chapter we introduce the notions of letter-isomorphic grammars, reduced grammars, skeletons and sentential form languages generated by a skeleton. After that we present the maximal skeletons with 1 and 2 nonterminal symbols. For the 3 nonterminal cases a computer program is created. Using the computer program we found 12 different maximal skeleton candidates, (one of them is not reduced) up to symmetries, with 3 nonterminals (S, X, Y). These symmetries are σ defined by $\sigma(X) = Y$, $\sigma(Y) = X$, and π defined by $\pi(A \to BC) = A \to CB$, with the properties $\sigma^2 = \pi^2 = 1$, $\pi\sigma = \sigma\pi$. The computer program in question checked that none of these 12 skeletons generates a nonprimitive word W of nonterminals with length $|W| \leq 12$. We present the exact mathematical proof for all of these skeletons generating only primitive words with no length limitation. The program was run for the case of 4 nonterminals, starting with length 6, and repeating the procedure for length 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, and 27. The program produced 413 candidates for maximal skeletons which number is too big to prove for all of them to be maximal skeletons indeed. However, a grammar which generates all primitive words over an (at least two letter) alphabet and does not generate any non-primitive word can not contain the start symbol on the right part of any rule. So we limited our examination for skeletons, the right part of whose rules do not contain the start symbol. Using the computer program we found 30 different maximal, above mentioned skeleton candidates, (2 of them are not reduced) and we present the exact mathematical proof for all of these skeletons generating only primitive words.

Irodalomjegyzék

[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.

Melléklet

Az 5. fejezet mellékleteként megadjuk a 3 és 4 nemterminális esetében használt C programot.

```
/*********************************************************************/
/*                                                                   */
/*      primitiv.cc                                                  */
/*                                                                   */
/*********************************************************************/

/* Looking for maximum grammars with three symbols and rules in CNF,    */
/* which do not generate non-primitive words of length >= 'limit'.      */

#include <stdlib.h>
#include <string.h>
#include <iomanip.h>
#include <fstream.h>

const rl_num  =  24,
      gr_num  = 500,
      max_len =  15,
      trace   =   0;

struct string
{
   int   len;
   char  val[max_len+1];

                    string      ()                  {};
                    string      (const string &s);
                    string      (const char   *s);
          void      operator =  (const string &s);
                   ~string      ()                  {};

   const  char     &operator [] (int      i) const     { return val[i]; }
```

```
          char     &operator () (int      i)          { return val[i]; }
   friend ostream  &operator << (ostream &s, const string &r);
};

inline string::string (const char *s) : len (strlen (s))
{
   strcpy (val, s);
}

inline ostream &operator << (ostream &s, const string &r)
{
//**    return s << "(" << r.len << ", '" << r.val << "')";
   return s << r.val;
}

struct rule
{
   char  l, r[3];
   int   db;
};

struct grammar
{
   int bits;
                 grammar     ()                  : bits (0     )   { }
                 grammar     (      int      i)  : bits (i     )   { }
                 grammar     (const grammar &g)  : bits (g.bits)   { }
          void   operator =  (const grammar &g)  { bits = g.bits;    }
                ~grammar     ()                  {                   }

          void   set         (int i)       { bits |=    1 << i;      }
          void   clear       (int i)       { bits &= ~ (1 << i);     }
          void   rotate      ();

          bool   operator <= (const grammar g) const;
          bool   operator [] (int i) const { return (bits >> i) & 1; }

   friend ostream  &operator << (ostream &s, const grammar g);
};

inline bool grammar::operator <= (const grammar g) const
{
   return (bits & ~g.bits) == 0;
}

/* List of all three symbol rules except S=>SS; S=>XX; S=>YY           */
```

```
const string start ("S");

const rule rl[rl_num] =
      {{'S', "XY", 0}, {'S', "YX", 0},
       {'S', "SX", 0}, {'S', "SY", 0}, {'S', "XS", 0}, {'S', "YS", 0},
       {'X', "SS", 0}, {'Y', "SS", 0},
       {'X', "SX", 0}, {'Y', "SY", 0}, {'X', "SY", 0}, {'Y', "SX", 0},
       {'X', "XS", 0}, {'Y', "YS", 0}, {'X', "YS", 0}, {'Y', "XS", 0},
       {'X', "XX", 1}, {'Y', "YY", 1}, {'X', "YY", 0}, {'Y', "XX", 0},
       {'X', "XY", 0}, {'Y', "YX", 0}, {'X', "YX", 0}, {'Y', "XY", 0}};

const grammar even = 0xAAAAAA, odd = 0x555555;

int      p_num = 0,     /* number of primitive grammars found so far    */
         n_num = 0,     /* number of nonprimitive grammars found so far */
         cp    = 0,     /* number of rules in the actual grammar        */
         limit = 0;     /* max. length of primitiv words to generate    */

grammar  g,             /* the actual grammar                           */
         p_list[500],   /* list of maximal primitive grammars           */
         n_list[500];   /* list of minimal nonprimitive grammars        */

ifstream fin;           /* known primitive grammars                     */
ofstream fout;          /* output medium                                */

inline void clear ()               { cout << "\033[H\033[2J"; }
inline void home  ()               { cout << "\033[\005H"; }
inline void clreol  ()             { cout << "\033[K"; flush (cout); }
inline void goto_xy (int x, int y)
   { cout << "\033[" << y << ';' << x << 'H'; }

void write (int row, int col, int val)
{
   goto_xy  (3 * col + 7, row);
   cout << setw (3) << val;
   clreol ();
   home ();
}

/* This function decides whether or not the actual word contains        */
/* a repetitive part of length n.                                        */

bool rep (const string &w, int n)
{
   for (int i = 0  ; i <   n;   i++   )
```

```
   for (int j = i+n; j < w.len; j += n)
      if (w[i] != w[j])  return false;
   return true;
}

/* Application of 'r' on position 'pos' in 'src' results in 'dest'.     */

string &apply (string &dest, const string &src, int pos, int ri)
{
   dest.len = src.len + 1;
   strncpy (&dest(0), &src[0], pos);
   dest(pos  ) = rl[ri].r[0];
   dest(pos+1) = rl[ri].r[1];
   strcpy (&dest(pos+2), &src[pos+1]);

   if (trace) write (3, src.len, ri);
   return dest;
}

/* This function decides whether the actual grammar generates a         */
/* non-primitive word of length up to 'len'.                             */
/* It returns true if this is not the case.                              */
/* Rule applications on positions before 'pos' are assumed to be        */
/* already checked.                                                      */

bool primitive (const string &word, int pos, int len)  return r
{
   int    wi, ri;
   string ww;

   if (trace)  { goto_xy (12, 4);  cout << word;  clreol; }

   switch (word.len)
   {
      case  1:  r =  true;                            break;
      case  4:  r = !rep (word, 2);                   break;
      case  6:  r = !rep (word, 2) && !rep (word, 3); break;
      case  8:  r = !rep (word, 4);                   break;
      case  9:  r = !rep (word, 3);                   break;
      case 10:  r = !rep (word, 5) && !rep (word, 2); break;
      case 12:  r = !rep (word, 6) && !rep (word, 4); break;
      case 14:  r = !rep (word, 7);                   break;
      case 15:  r = !rep (word, 3) && !rep (word, 5); break;
      default:  r = !rep (word, 1);
   }
```

```
   if (r && word.len < len)
   {
      ri = rl_num;
      for (wi = pos; wi < word.len && ri == rl_num; wi++)
      {
         if (trace)  { write (2, word.len, wi);
    write (3, word.len, 0); }

         ri = 0;
         while (        ri < rl_num
                 && (   !g[ri]
                     || word[wi] != rl[ri].l
                     || (rl[ri].db && wi && word[wi] == word[wi-1])
                     || primitive (apply (ww, word, wi, ri), wi, len)))
    ri++;
      }
      return ri == rl_num;
   }
}

/* Equivalent to exchanging all 'X' and 'Y'-symbols                    */

void grammar::rotate ()
{
   bits = ((bits << 1) & even.bits) + ((bits >> 1) & odd.bits);
}

/* Write the actual rule to a file                                     */

ostream &operator << (ostream &s, const grammar g)
{
   int i = 0;
   while (i < rl_num && !g[i]) i++;
   if (i < rl_num)
   {
      s << rl[i].l << "=>" << rl[i].r;
      while (++i < rl_num)
         if (g[i])  s << ", " << rl[i].l << "=>" << rl[i].r;
      s << '.' << endl;
   }
   return s;
}

/* Is the given grammar g or its X<-->Y permutation covered by the set  */
/* of primitive grammars?                                               */
```

```
bool covered (grammar g)
{
   int i;
   for (i = 0; i < p_num && !(g <= p_list[i]); i++);
   if (i < p_num) return true;

   g.rotate ();
   for (i = 0; i < p_num && !(g <= p_list[i]); i++);
   g.rotate ();
   return  i < p_num;
}

/* Is the given grammar g or its X<-->Y permutation covering an element */
/* of the set of nonprimitive grammars?                                 */

bool covering (grammar g)
{
   int i;
   for (i = 0; i < n_num && !(n_list[i] <= g); i++);
   if (i < n_num) return true;

   g.rotate ();
   for (i = 0; i < n_num && !(n_list[i] <= g); i++);
   g.rotate ();
   return  i < n_num;
}

/* append the actual grammar to the set of primitive grammars if        */
/* neither the grammar itself nor its X<-->Y permutation is already     */
/* covered by the set.                                                  */

void p_append (grammar g)
{
   if (!covered (g))
   {
      if (p_num == gr_num)
      {
         cerr << "Too many primitive grammars" << endl;
         exit (1);
      }
      p_list[p_num++] = g;
      fout << g;
   }
}

/* append the actual grammar to the set of nonprimitive grammars if     */
```

```
/* neither the grammar itself nor its X<-->Y permutation is already     */
/* covering an element of the set.                                       */

void n_insert (grammar g)
{
   if (!covering (g))
   {
      grammar h = g;
      h.rotate ();

      int i = 0, j;
      for (j = 0; j < n_num; j++)
         if (!(g <= n_list[j]) && !(h <= n_list[j]))
    n_list[i++] = n_list[j];
      if (i == gr_num)
      {
         cerr << "Too many nonprimitive grammars" << endl;
         fout << "Too many grammars" << endl;
         for (i = 0; i < gr_num; i++) fout << n_list[i];
         exit (1);
      }
      n_list[i] = g;
      n_num = i + 1;
   }
}

/* This procedure partitions the set of grammars consisting of some     */
/* of the 24 main rules into a set N of grammars that do generate       */
/* non-primitive words of length up to 'len' and the set P of           */
/* so-called 'primitive grammars'.                                       */
/*                                                                       */
/* N is represented by the set of its minimal elements in 'n_list'.     */
/* Accordingly, the maximal elements of P are stored in 'p_list'.       */

bool test (int act)  return b (0)
{
   const int bounds[] = { 4, 6, 8, 9, 10, 12, 14, 15, 0x7FFF};
   g.set (act);  cp++;

   if (trace)  { write (1, cp, act); }

   if (!covering (g))
   {
      if (!covered (g))
      {
         int i;
```

```
         for (i = 0; bounds[i] <= limit &&
    primitive (start, 0, bounds[i]); i++);
         b = bounds[i] > limit;
         if (!b) n_insert (g);
      }
      else b = 1;

      if (b)
      {
         int longer = 0;
         for (int i = act + 1; i < rl_num; i++) longer += test (i);
         if (longer == 0) p_append (g);
      }
   }

   g.clear (act);  cp--;
};

void read_it (istream &s, int &num, grammar *list)
{
   char l, r[3], x[256];
   int  i;

   cerr << "Reading ...";
   while (fin.peek () != EOF)
   {
      list[num] = 0;
      do {
         if (!fin.scan ("%c=>%2s", &l, &r))
            { cerr << "Invalid input!" << endl; exit (1); }

         for (i = 0; i < rl_num && (l != rl[i].l ||
    strcmp (r, rl[i].r)); i++);
         if  (i == rl_num)
         {
            cerr << "Invalid Rule: " << l << "=>" << r << '!' << endl;
            exit (1);
         }
         list[num].set (i);
         fin >> l;
      }
      while (l == ',');
      if (l != '.') { cerr << "Invalid input!" << endl; exit (1); }
      fin.get ();
      num++;
   }
```

```
   cerr << "Done ..." << endl;
}

main (int argc, char **argv)
{
   if (argc >= 2) limit = atoi (argv[1]);

   if (argc >= 3) fin.open (argv[2]);
   else           fin.open ("known.txt");
   read_it (fin, p_num, p_list);
   fin.close ();

   if (argc >= 4) fin.open (argv[3]);
   else           fin.open ("down.txt");
   read_it (fin, n_num, n_list);
   fin.close ();

   if (argc >= 5) fout.open (argv[4]);
   else           fout.open ("output.prt");

   if (trace)
   {
      clear ();
      cout << "Grammar:\n" << "Position:\n" << "Rule:\n" << "Word:\n";
      home ();
   }

   for (int i = 0; i < p_num;      i++) fout << p_list[i];
   for (int i = 0; rl[i].l == 'S'; i++) test (i);

   fout << endl;
   for (int i = 0; i < n_num;      i++) fout << n_list[i];

   fin.close ();
   fout.close ();
}
```

A Függelék

A jelölt publikációs jegyzéke

1. Pál Dömösi, Géza Horváth, Manfred Kudlek, Dirk Hauschildt, 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. Géza Horváth, Zoltán Nagylaki, PDL Robots Represented in VRML Environment, *Proc. Conf. Languages, Algebra and Computer Systems, R.I.M.S. Proceedings* No. 1106, Kyoto University, (1999), 174-185.

3. Géza Horváth, Katsushi Inoue, Akira Ito, Yue Wang, Closure Property of Probabilistic Turing Machines and Alternating Turing Machines with Subalgorithmic Spaces, *International Journal of Foundations of Computer Science*, Vol. 12, No. 3, (2001), 397-409.

4. Mária Demény, Géza Horváth, Zoltán Nagylaki, Csaba Nagylaki, Visualization of Cellular Automata, *Proceedings of The third International Colloquium on Words, Languages and Combinatorics,* Kyoto Sangyo University, (2000), elfogadva.

5. Pál Dömösi, Géza Horváth, Masami Ito, 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.

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

B Függelék

A jelölt konferencia előadásainak jegyzéke

1. PDL Robots Represented in VRML Environment. First International Student Forum-Contest on Multimedia, University of Aizu, Aizu-Wakamatsu, Japán, 1998.
2. Visualization of Cellular Automata. Second International Student Forum-Contest on Multimedia, University of Aizu, Aizu-Wakamatsu, Japán, 1999.
3. Some Results on Small Context-free Grammars Generating Primitive Words. Workshop on Automata and Languages, Kyoto Sangyo University, Kyoto, Japán, 1999.
4. PDL Robots Represented in VRML Environment. Languages, Algebra and Computer Systems, Kyoto University, Kyoto, Japán, 1999.
5. Visualization of Cellular Automata. 9^{th} International Conference on Automata and Formal Languages, Vasszécsény, 1999.
6. Closure Property of Probabilistic Turing Machines and Alternating Turing Machines with Subalgorithmic Spaces. The third International Colloquium on Words, Languages and Combinatorics, Kyoto Sangyo University, Kyoto, Japán, 2000.

A primitív és nemprimitív szavak nyelvei

Értekezés a doktori (PhD) fokozat megszerzése érdekében a matematika tudományában.

Írta: Horváth Géza okleveles programtervező matematikus

Készült a Debreceni Egyetem Matematika doktori programja (Informatika alprogramja) keretében

Témavezető: Dr. Dömösi Pál

A doktori szigorlati bizottság:

elnök:	Dr. ....
tagok:	Dr. ....
	Dr. ....

A doktori szigorlat időpontja: 200... ....

Az értekezés bírálói:

	Dr. ....
	Dr. ....
	Dr. ....

A bírálóbizottság:

elnök:	Dr. ....
tagok:	Dr. ....
	Dr. ....
	Dr. ....
	Dr. ....

Az értekezés védésének időpontja: 200... ....