DISS. B 1194

Marton László

EGYSZERŰ KONTEXTUSMENTES NYELVEK

Egyetemi doktori értekezés

József Attila Tudományegyetem Természettudományi Kar
Számítástudományi Tanszék

Szeged 1973

E L Ő S Z Ó

A formális grammatikák és nyelvek tanulmányozása a számitástudomány fontos részét alkotja. Az ilyen jellegü vizsgálatok elinditása Bar-Hillel, Lambek és Chomsky nevéhez füződik. Chomsky az ötvenes évek második felében egyszerü modelleket adott meg a természetes nyelvek számára és ezeket a modelleket matematikai szempontból is tanulmányozta pl. [1, 2, 3]. Hamarosan kiderült, hogy a megadott modellek az elektronikus számológépekhez készitett mesterséges nyelvek, a programnyelvek leirására is nagyon jól használhatók.

Ezekből a vizsgálatokból azóta egy önálló tudományág alakult ki, saját terminológiával, vizsgálati módszerekkel és fontos gyakorlati alkalmazásokkal /pl. a szintaxisorientált fordító/. Számos publikáció jelent meg e témakörből a világ minden táján.

Dolgozatom a kontextusmentes grammatikák két speciális osztályát vizsgálja. A dolgozat négy fejezetre és egy függelékre oszlik. Az első fejezet a tudományág legalapvetőbb fogalmait tartalmazza. A második fejezetben a kontextusmentes nyelvek elméletének azon alapvető fogalmai és tételei szerepelnek, amelyek a további vizsgálatoknál felhasználásra kerülnek. A harmadik és negyedik fejezet tartalmazza a két speciális nyelvosztályra vonatkozó eredményeket, a függelékben pedig táblázatok formájában bizonyos szempontok szerint csoportositva összefoglalom a legfontosabb eredményeket.

Ez uton fejezem ki köszönetemet Dr. Kalmár László professzornak. Az ő előadásai nyomán ismerkedtem meg ezzel a témakörrel és a dolgozat megirásához is tőle kaptam támogatást és biztatást.

TARTALOMJEGYZÉK

		Oldal
I.	Bevezetés	1
	1.§ A nyelv fogalma	1
	2.§ A grammatika fogalma, grammatikatipusok és nyelvosztályok	4
	3.§ Algoritmusok és megoldhatóság	11
II.	Kontextusmentes grammatikák és nyelvek	16
	1.§ Bevezetés	16
	2.§ Alapfogalmak és tételek	16
	3.§ A veremautomata, a determinisztikus kontextusmentes nyelvek	23
	4.§ A reguláris nyelvek és a véges automaták	32
	5.§ Müveletek kontextusmentes nyelvekkel	37
	6.§ Megoldhatóság	40
	7.§ Az egyszerü determinisztikus nyelvek	43
III.	Baloldali levezetés nyelvek	48
	1.§ Bevezetés	48
	2.§ Definiciók és alaptulajdonságok	52
	3.§ A baloldali levezetés nyelvek és a verem-automaták	63
	4.§ Müveletek	67
	5.§ Megoldhatóság	72
	5.1. Néhány általánosan nem megoldható probléma	72
	5.2. A szorzatképzés problémája	77
	5.3. A tartalmazás és az ekvivalencia problémája	80
IV.	Lengyel jelölésmód nyelvek	94
	1.§ Bevezetés	94
	2.§ Definiciók és alaptulajdonságok	99
	3.§ A lengyel jelölésmód nyelvek és a verem-automaták	102

Oldal

4.§ Müveletek 105
5.§ Megoldhatóság 110
5.1. Egy általánosan nem megoldható probléma 110
5.2. Az üresség, végesség, regularitás, tartalmazás és ekvivalencia problémája 110
5.3. Az egyesités, szorzatképzés, reguláris nyelvvel való metszetképzés és megforditás problémája 112
5.4. A homomorf leképezés problémája . . . 115
5.5. A metszetképzés problémája 119

Függelék
1. A tárgyalt nyelvosztályok közti tartalmazási viszonyok összefoglalása.
2. A nyelvosztályok és a müveletek közti összefüggések összefoglaló táblázata
3. A megoldhatósági kérdések összefoglaló táblázata

Irodalomjegyzék

I. fejezet

BEVEZETÉS

1.§ A nyelv fogalma

A Magyar Nyelv Értelmező Szótára a következő meghatározást adja a nyelv fogalmára: "Az emberi társadalom, illetve valamely nép története folyamán kialakult olyan, beszédhangokból összetevődő sajátos jelrendszer, amely meghatározott szabályok szerint /mondatokká/ kapcsolódó szavakból és e szavaknak egymáshoz, valamint a mondanivalóhoz való viszonyát kifejező egyéb elemekből áll, s a gondolatok megformálásának, kicserélésének eszköze."

Természetesen ez a definició nem eléggé pontos ahhoz, hogy egy matematikai elmélet alapja lehessen. A nyelv matematikai meghatározásának szükségképpen valamilyen formális rendszert kell leirnia, hiszen csak akkor tudunk matematikai eszközöket alkalmazni a nyelv vizsgálatára, ha a modelljének tekintett objektum egy matematikailag kielégitő pontossággal definiált rendszer. Nem kevésbé fontos az, hogy egy nyelvmodell mennyire közeliti meg az általa leirni kivánt nyelvi valóságot, de ez a kérdés /a modell adekvátsága/ tulajdonképpen csak a természetes nyelvek esetében vetődik fel, és jóval tulmegy egy formális elmélet határain. Bizonyos egyszerüsitéssel és általánositással egy természetes nyelv felfogható ugy, mint viszonylagosan önálló nyelvi egységek /mondatok/ halmaza, egy mondat pedig ugy, mint elemi összetevők /morfémák/ lánca. Ebből a felismerésből kiindulva adjuk meg a nyelv formális definicióját.

Legyen V egy nemüres halmaz, /ábécé, szótár/ amely-

nek valamilyen szimbólumok az elemei /pl. latin betük, görög betük, cirill betük, szóköz, vagy esetleg szavak /szóalakok/, morfémák, számjegyek, irásjelek stb./. Az ezekből az elemekből a konkatenáció /egymás mellé irás/ müveletével képzett objektumokat l á n c o k n a k nevezzük. Jelölje V^* a V elemeiből képezhető ö s z s z e s v é g e s l á n c o k halmazát, beleértve az ü r e s l á n c o t is, amit a továbbiakban mindig az ε szimbólummal jelölünk. A V^*egy elemét m o n d a t n a k is nevezhetjük. Egy, a V feletti nyelvet ezután már ugy is definiálhatunk, mint a V^* halmaz egy részhalmazát.

Természetesen a V^* halmaznak, lévén végtelen halmaz, végtelen részhalmazai is vannak, tehát végtelen, azaz végtelen sok mondatból álló nyelvek is vannak. Ezzel a jelenséggel egyébként a természetes nyelveknél is találkozunk. Egy természetes nyelv mondatainak a száma potenciálisan végtelen. A természetes nyelvekben ugyanis számos olyan nyelvi jelenség van, amelyek segitségével akármilyen nagyszámu mondatot képezhetünk, tehát a képezhető mondatok száma elvileg végtelen. Például a magyar nyelvben az "és" mellérendelő kötőszó használatával bármely két mondatból egy uj mondat keletkezik, vagy akármilyen hosszu mondatot készithetünk ugy, hogy egy-egy mondatban szereplő főnév elé több jelzőt teszünk, illetve egy jelzőt többször megismétlünk.

A végtelen nyelvekkel kapcsolatban lép fel a v é g e s r e p r e z e n t á l h a t ó s á g problémája, vagyis az, hogy hogyan lehet v é g e s l e i r á s s a l megadni mondatok egy végtelen halmazát. /Ez a probléma igen nagy elemszámu véges rendszereknél is előjön, hiszen gyakorlatilag lehetetlen felsorolni pl. egymilliárd elemet./

Felvetődhet az a kérdés is, hogy vajon minden nyelv meg-

adható-e véges leirással. Egyszerüen belátható, hogy a válasz "nem", hiszen az egy ábécé feletti nyelvek halmaza kontinuum számosságu, a véges leirások halmaza pedig megszámlálható. A véges leirásokat aszerint, hogy milyen módon definiálják az illető nyelvet alapjában véve két csoportra oszthatjuk: f e l i s m e r ő rendszerek és g e n e r a t i v rendszerek.
Egy felismerő rendszer egy olyan eljárást ad meg, amelynek segitségével bármely véges láncról /a V^* bármelyik eleméről/ véges lépésben eldönthetjük, hogy az illető nyelvhez tartozó mondat-e vagy sem. Egy generativ rendszerre viszont az a jellemző, hogy elő tudja állitani az illető nyelv minden mondatát és nem állit elő egyetlen olyan mondatot sem, amely nem tartozik ehhez a nyelvhez. Ha egy nyelvről van már felismerő jellegü véges leirásunk, akkor ennek segitségével készithetünk generativ jellegü véges leirást is. A nyelvek véges leirással való megadásának alapvető módja a nyelv g r a m m a t i k á j á n a k megadása, amellyel a következőkben foglalkozunk részletesebben.

2.§ A grammatika fogalma, grammatikatipusok és nyelvosztályok

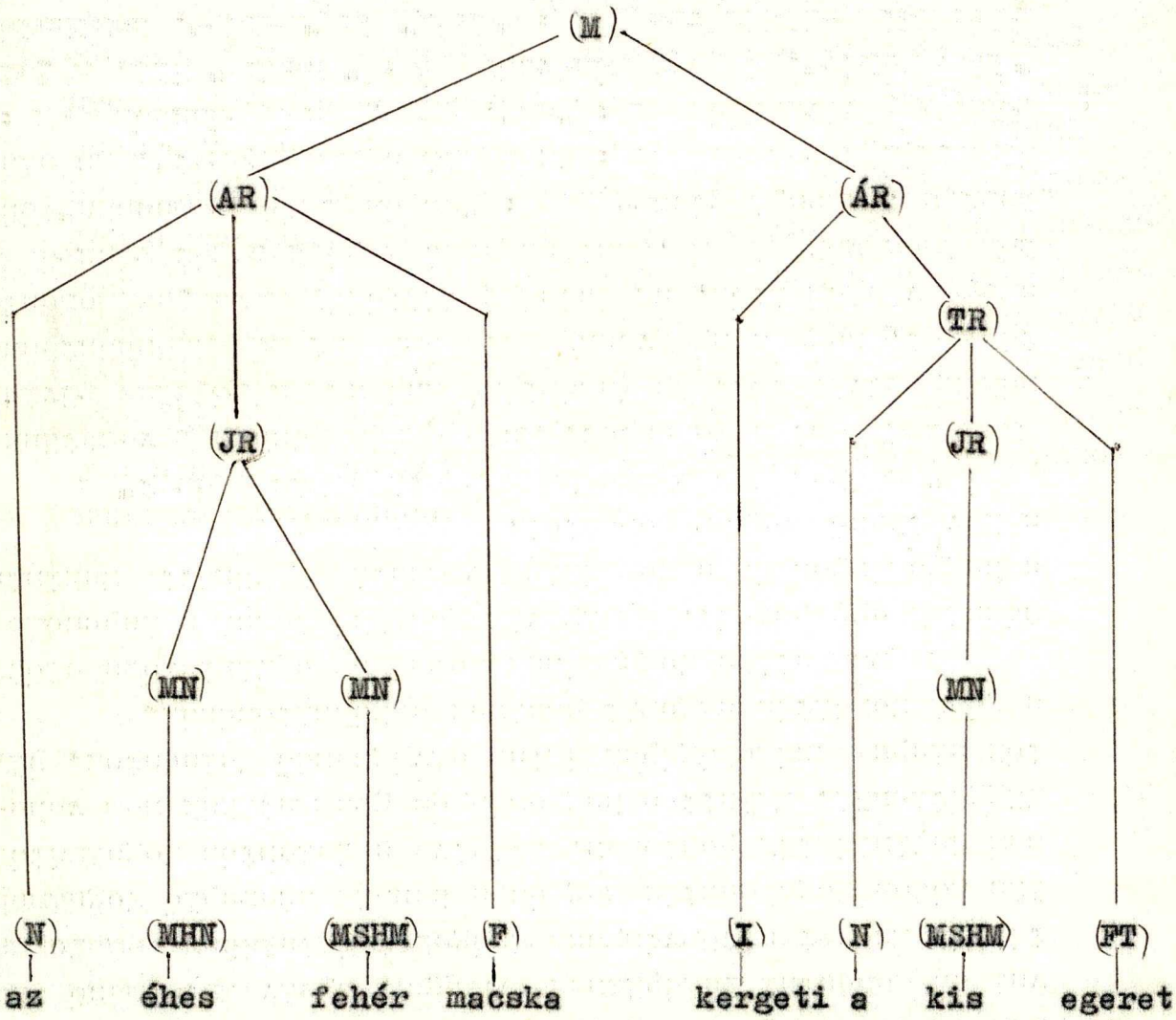
A g r a m m a t i k a fogalma, a természetes nyelvekre vonatkoztatva, történetileg alakult ki a nyelvtudomány fejlődésének folyamán. Általában egy természetes nyelv grammatikáján "azon törvényszerüségek összességét, rendszerét értjük, amelyek szerint az illető nyelv szavai és mondatai felépülnek" /Magyar Nyelv Értelmező Szótára/. Ezeket a törvényszerüségeket általában nem formális módon fogalmazták meg és a bennük szereplő fogalmakat sem definiálták pontosan, illetve a jelentés fogalmától függetlenül. Emiatt a hagyományos grammatikák nem vehetők alapul számológépek által elvégzendő feladatok, pl. a gépi forditás algoritmusainak elkészitésénél.

A grammatika fogalmának formális definiciója előtt nézzünk egy példát, amely világosabbá teszi a kapcsolatot a hagyományos és a formális grammatikák között. Tekintsük a következő magyar mondatot:

"Az éhes fehér macska kergeti a kis egeret"

Elemezzük ezt a mondatot. Az elemzés közben bevezetünk néhány röviditést is /azért használunk zárójeles röviditéseket, hogy a nyelvtani kategóriákat megkülönböztessük a nyelv elemeitől, az ábécé betüitől/. A teljes mondatból indulunk ki (M), ezt az első lépésben két részre bontjuk, alanyi részre (AR): "az éhes fehér macska" és állitmányi részre (ÁR): "kergeti a kis egeret". Az alanyi rész egy névelőből (N): "az", egy jelzői részből (JR): "éhes fehér", és egy főnévből (F): "macska" áll. A jelzői rész két melléknévre (MN): "éhes" és "fehér", bontható. Mivel az első melléknév magánhangzóval kezdődik (MHM), az előtte álló névelő: "az". Az állitmányi rész egy jelentő mód jelen idő egyes szám harmadik személyben álló igéből (I): "kergeti", és egy tárgyi részből (TR): "a kis

egeret", épül fel. Ez egy névelőre: "a", egy jelzői részre: "kis" és egy tárgyesetben álló főnévre (FT): "egeret", bomlik. A jelzői rész egy mássalhangzóval kezdődő melléknév (MSHM). Az elemzést a következő diagrammal szemléltethetjük:



Az elemzésnél használt szabályokat a következő formában is irhatjuk: /itt gondolunk arra a lehetőségre is, hogy egy főnév előtt kettőnél több jelző is állhat és arra is, hogy sok esetben a tárgyi rész hiányzik/.

(M) ⟶ (AR) (ÁR)
(AR) ⟶ (N) (JR) (F)
(ÁR) ⟶ (I) (TR)
(ÁR) ⟶ (I)

(N) (MHM) ⟶ az (MHM)
(N) (MSHM) ⟶ a (MSHM)
(MHM) ⟶ éhes

(TR) ⟶ (N) (JR) (FT)
(JR) ⟶ (MN) (JR)
(JR) ⟶ (MN)
(MN) ⟶ (MHM)
(MN) ⟶ (MSHM)

(MSHM) ⟶ fehér
(MSHM) ⟶ kis
(I) ⟶ kergeti
(F) ⟶ macska
(FT) ⟶ egeret

Ha a szabályokban szereplő nyilaknak azt a jelentést tulajdonitjuk, hogy a baloldali szimbólumot vagy szimbólumláncot a jobboldali lánccal helyettesithetjük, erre irhatjuk át, akkor egy generativ rendszert kapunk, amely a példamondatot a következő módon állitja elő:

(M) ⇒ (AR) (ÁR) ⇒ (N) (JR) (F) (ÁR) ⇒ (N) (MN) (JR) (F) (ÁR) ⇒ (N) (MN) (MN) (F) (ÁR) ⇒ (N) (MHM) (MN) (F) (ÁR) ⇒ az (MHM) (MN) (F) (ÁR) ⇒ az éhes (MN) (F) (ÁR) ⇒ az éhes (MSHM) (F) (ÁR) ⇒ az éhes fehér (F) (ÁR) ⇒ az éhes fehér macska (ÁR) ⇒ az éhes fehér macska (I) (TR) ⇒ az éhes fehér macska kergeti (TR) ⇒ az éhes fehér macska kergeti (N) (JR) (FT) ⇒ az éhes fehér macsak kergeti (N) (MSHM) (FT) ⇒ az éhes fehér macska kergeti a (MSHM) (FT) ⇒ az éhes fehér macska kergeti a kis (FT) ⇒ az éhes fehér macska kergeti a kis egeret

A "⇒" a levezetés egy lépését jelzi, vagyis a szabályrendszer egy szabályának alkalmazását. Észrevehetjük, hogy a megadott szabályrendszer sok más mondatot is tud generálni, bizonyos szabályok többszöri, illetve más sorrend szerinti alkalmazásával. Például: "A kis macska kergeti az éhes fehér egeret". A szótárunk kibővitésével pl. ujabb főnevek felvételével is növekedhet a generálható mondatok száma.

A fenti példából levonhatunk néhány következtetést. Generativ rendszerünkben tulajdonképpen n é g y o b j e k t u m szerepel. A n y e l v t a n i k a t e g ó r i á k halmaza, a n y e l v s z ó a l a k j a i n a k halmaza, az á t i r á s i s z a b á l y o k halmaza, és egy k i t ü n t e t e t t k a t e g ó r i a, a

"mondat" kategóriája, amiből kiindulva alkalmazzuk a szabályokat. A "k a t e g ó r i a" helyett szokásos még a "n e m t e r m i n á l i s j e l" vagy "v á l t o z ó", a "s z ó a l a k" helyett a "t e r m i n á l i s j e l", az "á t i r á s i s z a b á l y" helyett a "p r o d u k c i ó s s z a b á l y", a "k i t ü n t e t e t t k a t e g ó r i a" helyett az "a x i ó m a" vagy "s t a r t s z i m b ó l u m" elnevezés is.

Formálisan egy G m o n d a t s z e r k e z e t g r a m m a t i k á n vagy f o r m á l i s g r a m m a t i k á n /röviden: grammatikán/ egy rendezett négyest értünk:

$$G = (N, T, P, S)$$

ahol az N, a T és a P véges halmazok, az N a kategóriák, a T terminális jelek, a P a produkciós szabályok halmaza, az S pedig a start szimbólum, az axióma. Természetesen $S \in N$. Magától értetődő követelmény, hogy az $N \cap T = \emptyset$ /üres halmaz/ legyen. A P halmaz $\alpha \to \beta$ alaku szabályokból áll, ahol a $V = N \cup T$ és a $V^+ = V^* - \varepsilon$ jelöléssel élve, $\alpha \in V^+$ és $\beta \in V^*$.

Bevezetünk még néhány a továbbiakban általánosan használt jelölést. Ha H szimbólumok egy véges halmaza, akkor H^* jelöli a H elemeiből képezhető összes v é g e s l á n c o k halmazát, H^+ pedig az összes n e m ü r e s v é g e s l á n c o k halmazát, tehát a $H^* - \varepsilon$ halmazt. Ha a $G = (N, T, P, S)$ egy grammatika, akkor a V mindig az $N \cup T$ halmazt jelöli. Legyen $\alpha \in V^*$ és $x \in V$. Az $|\alpha|_x$ jelöli az x szimbólum α láncbeli e l ő f o r d u l á s a i n a k s z á m á t, $|\alpha|$ pedig az α lánc /szimbólumokban mért/ h o s s z á t. Minden $x \in V$ esetén $|\varepsilon|_x = 0$ és $|\varepsilon| = 0$. Ha n egy nemnegativ egész szám, akkor α^n jelöli az α n-s z e r e s e g y m á s u t á n i r á s á v a l keletkező láncot. Minden $\alpha \in V^*$ és $n \geq 0$ egész szám esetén $\alpha^0 = \varepsilon$ és $\varepsilon^n = \varepsilon$. Az α^R jelöli az α m e g f o r d i t á s á v a l kapott láncot, vagyis ha $\alpha = x_1 \dots x_k$ ahol $k \geq 1$ és minden $1 \leq i \leq k$ indexre $x_i \in V$ akkor $\alpha^R = x_k \dots x_1$, $\varepsilon^R = \varepsilon$. Azt mondjuk, hogy a β lánc az α lánc k e z d ő s z e l e t e, ha $\alpha = \beta\gamma$,

és **valódi kezdőszelete**, ha $\beta \neq \varepsilon$ és $\gamma \neq \varepsilon$. Azt mondjuk, hogy az $\alpha_1, \ldots, \alpha_k$ láncok az α lánc egy **felbontását** adják, ha $\alpha = \alpha_1 \ldots \alpha_k$. Egy $\alpha = \alpha_1 \ldots \alpha_k$ felbontást **teljes felbontásnak** nevezzünk, ha minden $1 \leq i \leq k$ indexre $|\alpha_i| = 1$, vagyis $\alpha_i \in V$. Ha H egy véges halmaz, akkor $\#H$ jelöli a H **elemeinek számát**. A továbbiakban általában a nagy latin betüket a kategóriák, a kis latin betüket a terminális jelek és az ezekből képzett láncok, a terminális láncok, a kis görög betüket pedig a vegyes láncok jelölésére használjuk. E technikai jellegü kitérő után, a következőkben a grammatika által generált nyelv formális definiciójával foglalkozunk.

Legyen a $G = (N, T, P, S)$ egy grammatika és $\gamma, \delta \in V^*$. Azt mondjuk, hogy a δ **közvetlenül levezethető** a γ láncból a G grammatikában, ha vannak olyan $\alpha, \beta, \gamma_1, \gamma_2 \in V^*$ láncok, hogy $\alpha \to \beta \in P$ és $\gamma = \gamma_1 \alpha \gamma_2$ valamint $\delta = \gamma_1 \beta \gamma_2$. A közvetlen levezethetőség jelölése:

$$\gamma \underset{G}{\Rightarrow} \delta$$

Ha magától értetődő, hogy melyik grammatikáról van szó, akkor a jelölésből a „G"-t elhagyhatjuk. Azt mondjuk, hogy a δ levezethető a γ láncból a G grammatikában, ha vannak olyan $\alpha_1, \ldots, \alpha_m$, V^* halmazbeli láncok, hogy $m \geq 2$ és minden $1 \leq i \leq m-1$ indexre $\alpha_i \underset{G}{\Rightarrow} \alpha_{i+1}$, valamint $\alpha_1 = \gamma$ és $\alpha_m = \delta$. A levezethetőség jelölése:

$$\gamma \overset{*}{\underset{G}{\Rightarrow}} \delta$$

Ha magától értetődő, hogy melyik grammatikáról van szó, itt is elhagyhatjuk a jelölésből a „G"-t. A levezethetőség fogalmának segitségével definiáljuk a G által generált nyelvet, amit $L(G)$-vel jelölünk.

A G grammatika által **generált nyelven** értjük azon **terminális láncok** összességét, amelyek a grammatika **axiomájából** levezethetők. Formulával:

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

Két grammatikát akkor nevezünk e k v i v a l e n s n e k, ha az általuk generált nyelvek megegyeznek.

A bevezetett fogalmak illusztrálására tekintsük a következő példát:

$G_1 = (N_1, T_1, P_1, S)$ **ahol** $N_1 = \{S, A, B\}$, $T_1 = \{a, b, c\}$ **és**

$$P_1 = \{ S \to AB,\ A \to aAb,\ A \to \varepsilon,\ B \to cB,\ B \to \varepsilon \}$$

$G_2 = (N_2, T_2, P_2, S)$ **ahol** $N_2 = \{S, A, B, C, D\}$, $T_2 = \{a, b, c\}$ **és**

$$P_2 = \{ S \to ABS,\ S \to DC,\ C \to cC,\ C \to \varepsilon,\ BD \to Db,$$
$$AD \to aD,\ Aa \to aA,\ Ba \to aB,\ D \to \varepsilon \}$$

A két grammatika ekvivalens, $L(G_1) = L(G_2) = \{ a^n b^n c^m \mid n, m \geq 0 \}$
Példaképpen vezessük le az a^2b^2c **mondatot:**

$$S \underset{G_1}{\Rightarrow} AB \underset{G_1}{\Rightarrow} aAbB \underset{G_1}{\Rightarrow} aaAbbB \underset{G_1}{\Rightarrow} aabbB \underset{G_1}{\Rightarrow} aabbcB$$
$$\underset{G_1}{\Rightarrow} aabbc$$

$$S \underset{G_2}{\Rightarrow} ABS \underset{G_2}{\Rightarrow} ABABS \underset{G_2}{\Rightarrow} ABABDC \underset{G_2}{\Rightarrow} ABADbC \underset{G_2}{\Rightarrow} ABaDbC$$
$$\underset{G_2}{\Rightarrow} AaBDbC \underset{G_2}{\Rightarrow} AaDbbC \underset{G_2}{\Rightarrow} aADbbC \underset{G_2}{\Rightarrow} aaDbbC$$
$$\underset{G_2}{\Rightarrow} aabbC \underset{G_2}{\Rightarrow} aabbcC \underset{G_2}{\Rightarrow} aabbc$$

Végül tekintsük a formális grammatikák CHOMSKY által megadott o s z t á l y o z á s á t.

A fenti általános definiciónak megfelelő grammatikákat 0 - t i p u s u grammatikáknak nevezzük.

Ha egy 0-tipusu grammatika bármely szabályának a baloldala rövidebb a jobboldalánál vagy egyenlő hosszuságu vele, akkor ezt a grammatikát 1 - t i p u s u vagy másnéven k o n t e x t u s é r z é k e n y grammatikának nevezzük. Ide soroljuk még azokat a grammatikákat is, amelyekben szerepel egy $S \to \varepsilon$ szabály /ahol az S az axioma/, a többi szabályok pedig eleget tesznek a fenti megkötésnek és az S szimbólum egyetlen szabály jobboldalán sem fordul elő.

Azokat a grammatikákat, amelyeknél minden szabály baloldalát egy egyedülálló kategória alkotja, 2 - t i p u s u vagy k o n t e x t u s m e n t e s grammatikáknak nevezzük.

Ha egy grammatika minden szabálya $A \to aB$ vagy $A \to a$ vagy $A \to \varepsilon$ alaku, ahol az A és B kategóriák, az a pedig terminális jel, akkor a grammatikát 3 - t i p u s u, vagy r e g u l á r i s grammatikának nevezzük.

Egy n y e l v e t akkor mondunk 0, 1, 2 vagy 3 - t i p u s u n y e l v n e k /illetve a megfelelő esetekben a kontextusérzékeny, a kontextusmentes, vagy a reguláris nyelvek osztályába tartozónak/, ha van olyan grammatika, amely az illető tipushoz tartozik és ezt a nyelvet generálja.

A definiciókból közvetlenül következik, hogy minden reguláris nyelv egyben kontextusmentes nyelv is, és minden kontextusérzékeny nyelv egyben 0-tipusu nyelv is. Az is egyszerüen bizonyitható, hogy minden kontextusmentes grammatika ekvivalens egy olyan kontextusmentes grammatikával, amely egyben kontextusérzékeny is. Tehát a kontextusérzékeny nyelvek osztálya tartalmazza a kontextusmentes nyelveket.
Alkalmasan megválasztott példákkal bebizonyitható, hogy mindegyik tartalmazás valódi tartalmazás.

3.§ Algoritmus és megoldhatóság

A formális grammatikák és nyelvek elméletének számos alkalmazási területén lényeges a következő tipusu probléma: ha tekintünk egy speciális grammatikaosztályt /a formális grammatikák halmazának egy részhalmazát/, akkor ezen az osztályon belül b á r m e l y k é t adott grammatika esetén e l t u d j uk- e d ö n t e n i, hogy pl.:

a, ekvivalens-e a két grammatika
b, van-e olyan mondat, amelyet mindkét grammatika generál

illetve b á r m e l y grammatika esetén el t u d j u k - e d ö n t e n i, hogy pl.:

a, van-e olyan mondat, amelyet ez a grammatika generál
b, véges vagy végtelen számu mondatot képes generálni az illető grammatika
c, egy adott mondat levezethető-e ebben a grammatikában

Ezek általában nem könnyü problémák, megoldásuk sok esetben bonyolult és hosszadalmas. Hogy ilyen tipusu kérdéseket vizsgálhassunk, természetesen először tisztázni kell, mit értünk azon, hogy "e l t u d j u k d ö n t e n i" illetve "n e m t u d j u k e l d ö n t e n i". Ez a kérdés szoros kapcsolatban van az a l g o r i t m u s fogalmával.

Egy a l g o r i t m u s tulajdonképpen egy olyan szabályrendszer, amely pontosan leir egy eljárást, lépésről lépésre megmondja, hogy mit kell tenni. Az eljárást, a lépések egymásrakövetkezését olyan pontosan kell leirni, hogy egy olyan mechanizmus /vagy ember/ is képes legyen végrehajtani, ami /vagy aki/ csak az egyes lépések önálló elvégzésére és bizonyos feltételek teljesülésének ellenőrzésére képes. /Például két természetes szám legnagyobb közös osztóját az euklideszi algoritmus segitségével az

is ki tudja számitani, aki csak osztani tud és azt tudja ellenőrizni, hogy a maradék nulla-e vagy sem./

Hogy mi tekinthető pontos, egyértelmű leirásnak, az természetesen az algoritmus végrehajtójának "képességeitől" is függ, vagyis attól, hogy mik azok az e l e m i m ü v e l e t e k, amelyeket önállóan el tud végezni.

Végsősoron minden /elméleti jellegü/ algoritmus végrehajtása j e l s o r o z a t o k több lépésben való á t a l a k i t á s a. Ebből az alapelvből indulnak ki azok a javaslatok, amelyek az algoritmusok formális leirási módszerére /tehát tulajdonképpen az algoritmus fogalmának formális definiciójára/ születtek. Ezek közül a számitástechnikai szemlélethez legközelebb áll, és nagyon szemléletes a T u r i n g g é p fogalma.

Egy T u r i n g g é p egy iró-olvasó fejjel felszerelt, véges számu belső állapottal rendelkező vezérlőegységből és egy végtelen hosszu szalagból álló rendszer. A s z a l a g, amely a memória szerepét tölti be, véges hosszuságu részekre, mezőkre van osztva. Minden mező egy jel tárolására képes. Egy Turing gépnek véges számu jelből álló j e l k é s z l e t e van, ennek elemeiből felépülő véges hosszuságu jellláncokon végez átalakitásokat a gép. Müködése elemi lépések sorozatából áll, egy-egy elemi lépésben a következőt hajtja végre: beolvassa az éppen az i r ó - o l v a s ó f e j alatt lévő mező tartalmát és a beolvasott jeltől, valamint az aktuális b e l s ő á l l a p o t t ó l függően a beolvasott jel helyére kiir egy jelet utána pedig áttér egy másik mező vizsgálatára /a szalagot jobbra vagy balra huzza az iró-olvasó fej alatt/. A Turing gép pontos definiciója megtalálható pl.: [8, 10, 14].

A Turing gépekre vonatkozóan ismeretes a T u r i n g - t é z i s, amely szerint minden olyan eljáráshoz, amelyet intuitive pontosan megadottnak tekintünk, konstruálható olyan Turing gép, amely ezt az eljárást hajtja végre /ha szükséges, megfelelő kódolással/.

Tehát ezen tézis szerint az algoritmus fogalma identikus a Turing gép fogalmával. Az algoritmus fogalmát a továbbiakban ilyen értelemben használjuk

Azt mondjuk, hogy egy algoritmus e l d ö n t egy kérdést, ha az algoritmust végrehajtva, v é g e s s z á m u l é p é s elvégzése után a v a l ó s á g n a k m e g f e l e l ő választ kapunk a kérdésre.

Ugyanez a Turing gépek nyelvén megfogalmazva: Azt mondjuk, hogy egy Turing gép eldönt egy kérdést, ha a Turing gép kiindulva a szalagjára /megfelelően kódolva/ felirt kérdésből, véges számu lépés megtétele után megáll és a szalag tartalmazza a /kódolt/ választ.

Egy p r o b l é m a általában v é g t e l e n sok k o n k r é t k é r d é s b ő l áll, vagy másszóval egy problémának végtelen sok konkrét esete van. Például egy adott grammatikaosztályra vonatkozóan az ekvivalencia problémája, mint fentebb kifejtettük, azt jelenti, hogy bármely két /az adott osztályba tartozó/ grammatika esetén eldönthető-e az ekvivalencia kérdése. Mivel egy "valamirevaló" grammatikaosztály végtelen sok elemből áll, a problémának végtelen sok konkrét esete van.

Á l t a l á n o s a n m e g o l d h a t ó n a k tekintünk egy problémát, ha van olyan algoritmus /Turing gép/, amely a probléma minden konkrét esetében eldönti a kérdést. Á l t a l á n o s a n n e m m e g o l d h a t ó a probléma, ha nincs ilyen algoritmus.

Ennek megfelelően egy probléma általános megoldhatóságát általában ugy bizonyitjuk, hogy megadjuk a megfelelő algoritmust /a konstrukció a gyakorlatban legtöbbször nem megy le a Turing gép megadásának szintjére/. Azt, hogy egy probléma általánosan nem megoldható, nagyon bonyolult és hosszadalmas lenne a Turing gépek szintjén bebizonyitani. Az ilyen bizonyitásokhoz hasznos segédeszköz a következő negativ eredmény, a P o s t-f é l e k o r r e s p o n d e n c i a p r o b l é m a /röviden PCP/ általánosan nem megoldhatósága:

Legyen a V véges ábécé, jelölje X és Y a V^+ halmaz két egyenlő elemszámu részhalmazát:

$$X = \{x_1, \ldots, x_k\}, \quad Y = \{y_1, \ldots, y_k\}, \quad k \geq 1$$

Igaz a következő: Ha V legalább kételemü, akkor nincs olyan algoritmus, amely bármely adott X, Y pár esetén eldöntené, hogy van-e olyan $i_1, \ldots, i_m$, $m \geq 1$ indexsorozat, amelyre teljesül, hogy

$$x_{i_1} \ldots x_{i_m} = y_{i_1} \ldots y_{i_m}$$

/A bizonyitásra vonatkozóan ld. pl.: [8, 10, 14]/.

A probléma általánosan nem megoldható volta természetesen nem zárja ki, hogy bizonyos konkrét esetekben eldönthető legyen a kérdés. Legyen például $V = \{a, b\}$ és $Y_2 = \{a, b\}$ $X_1 = \{a, aba, aaa, ab\}$, $Y_1 = \{aaa, abi, b\}$, $X_2 = \{ab, aaa\}$ Az (X_1, Y_1) pár esetén a válasz "igen" ($i_1 = 2, i_2 = 1, i_3 = 1, i_4 = 3$, aba^6b) az (X_2, Y_2) pár esetén a válasz "nem" /mert az X_2 minden eleme hosszabb az Y_2 minden eleménél/.

Általánosan megoldható és nem megoldható problémákra sok példát fogunk látni a következő fejezetekben. Elöljáróban megadunk egy, a formális grammatikákra vonatkozó általánosan nem megoldható problémát:
Nincs olyan algoritmus, amely tetszőlegesen megadott G grammatika és u mondat esetén eldöntené, hogy az hozzátartozik-e a G által generált nyelvhez, vagy sem. Érdemes megjegyezni, hogy ugyanez a probléma már általánosan megoldható, ha a kontextusérzékeny grammatikák osztályára szoritkozunk.

Legyenek a H_1 és a H_2 olyan halmazok, amelyeknek az elemei formális nyelvek és $H_1 \subseteq H_2$. A definiciókból közvetlenül következik:

a, Ha egy probléma a H_2 halmaz esetén általánosan megoldható, akkor ez a probléma a H_1 halmaz esetén is általánosan megoldható.

b, Ha egy probléma a H_1 halmaz esetén általánosan nem oldható meg, akkor ez a probléma a H_2 halmaz esetén

sem oldható meg általánosan.

Ezt a két tényt fel fogjuk használni a továbbiakban. /Ezzel kapcsolatban meg kell jegyeznünk, hogy a fentiek természetesen csak akkor érvényesek, ha a szóbanforgó probléma egyáltalában megfogalmazható. Ugyanis ha pl. H_2 megszámlálható, akkor kontinuum számosságu H_1 részhalmaza van, de csak megszámlálható sok probléma vehető fel egyáltalában./ Ha valamely nyelvosztályra és problémára vonatkozóan az általános megoldhatóság vagy meg nem oldhatóság a fenti állitások valamelyikéből és egy olyan tételből következik, amely a dolgozatban szerepel, akkor erre a megoldhatósági eredményre külön nem utalunk általában. A dolgozat végén a Függelékben t á b l á z a t b a n f o g l a l j u k ö s z s z e a dolgozatban tárgyalt megoldhatósági eredményeket. Itt feltüntetjük azokat is, amelyekre a fent emlitett okból, a dolgozaton belül nem utaltunk.

Végül megjegyezzük: Ha azt mondjuk hogy egy matematikai nyelvészeti objektum /pl. egy grammatika, egy mondat/ "m e g a d h a t ó", "l é t e z i k", "f e l é p i t h e t ő", "k o n s t r u á l h a t ó", ezen végsősoron azt értjük, hogy v a n o l y a n a l g o r i t m u s, amely bizonyos bemenő adatokból kiindulva, véges számu lépés után eredményként megadja a szóbanforgó objektumot.

II. fejezet

KONTEXTUSMENTES GRAMMATIKÁK ÉS NYELVEK

1.§ Bevezetés

A kontextusmentes grammatikák a formális grammatikák egy fontos speciális osztályát alkotják. A matematikai nyelvészetben széleskörüen alkalmazzák ezeket a grammatikákat. Bár egy természetes nyelv szintaxisának leirására nem elegendő egy kontextusmentes grammatika, de kiegészitve bizonyos transzformációs szabályokkal /amelyek a kontextusmentes rész által generált, un. magmondatokat alakitják át/ a transzformációs grammatikákhoz jutunk, amelyek már jól alkalmazhatók a nyelvtudományban. A magasabbszintü programozási nyelvek formális definiálásánál is fontos eszköz a kontextusmentes grammatika. Ebben a fejezetben a kontextusmentes grammatikák és nyelvek elméletének alapjait tárgyaljuk olyan részletességgel, amilyenre a következő fejezetekhez szükség lesz. A tételek részletes bizonyitására nem térünk ki, eseténként példákkal illusztráljuk az állitást. A tételek bizonyitásaival kapcsolatban a következő munkákra utalunk: [4, 8, 10, 12]. A továbbiakban a "kontextusmentes" röviditésére a "CF", a "kontextusérzékeny" röviditésére pedig a "CS" jelölést használjuk /context-free, illetve context-sensitive/.

2.§ Alapfogalmak és tételek

Definició: Legyen a $G=(N,T,P,S)$ egy formális grammatika. A G grammatikát akkor nevezzük CF - grammatikának, ha minden P -beli szabály $A\to\alpha$ alaku, ahol $A\in N$ és $\alpha\in V^*$.

Definició: Egy L nyelvet akkor nevezünk CF - n y e l v - n e k, ha van olyan G CF-grammatika, amely ezt a nyelvet generálja, vagyis $L = L(G)$.

Definició: Legyen a $G=(N,T,P,S)$ egy CF-grammatika. Az $A \to \alpha \in P$ szabályt t e r m i n á l i s s z a b á l y n a k nevezzük, ha $\alpha \in T^*$.

A kontextusmentes grammatika fenti definiciója nem zárja ki, hogy a grammatikában "felesleges" jelek és szabályok is előforduljanak. F e l e s l e g e s j e l e k e n értjük azokat a V halmazbeli elemeket, amelyek nem szerepelnek egyetlen, az S axiómából kiinduló levezetésben sem, és azokat a kategóriákat, amelyekből kiindulva nem vezethető le egyetlen terminális lánc sem. Igaz a következő:

2.2.1. tétel: Legyen a $G=(N,T,P,S)$ CF-grammatika.

a, Bármely $X \in V$ esetén eldönthető, hogy van-e olyan $S \overset{*}{\underset{G}{\Rightarrow}} \alpha$ levezetés, hogy $\alpha = \beta X \gamma$ ahol $\beta, \gamma \in V^*$.

b, Bármely $A \in N$ kategória esetén eldönthető, hogy van-e olyan $u \in T^*$ mondat, amelyre létezik az $A \overset{*}{\underset{G}{\Rightarrow}} u$ levezetés.

Tehát van olyan algoritmus, amely bármely adott CF-grammatika esetén megkeresi a felesleges jeleket. A felesleges jelek nyilván nem vesznek részt a nyelv generálásában, tehát, ha elhagyjuk őket, azokkal a szabályokkal együtt, amelyekben szerepel felesleges jel, a maradék grammatika ugyanazt a nyelvet generálja mint az eredeti. Nyilvánvalóan elhagyhatjuk még az $A \to A$ alaku szabályokat is. Az elhagyások nem változtatják meg a grammatika kontextusmentes jellegét, kivéve azt az esetet, amikor az $L(G) = \emptyset$, ez esetben ugyanis mindent el kellene hagyni, tehát a $T = \emptyset, P = \emptyset$ lenne. Hogy az $L(G) = \emptyset$ fennáll-e az a 2.2.1 tétel b, pontja szerint eldönthető. Ilyen esetben nincs értelme az elhagyásoknak.

Definició: Legyen a $G=(N,T,P,S)$ CF-grammatika és az $L(G)\neq\emptyset$. Azt mondjuk, hogy a G **egyszerüsitett CF-grammatika**, ha teljesül a következő három feltétel:

a, Bármely $x\in V$ esetén van olyan $\alpha\in V^*$, hogy $S\overset{x}{\underset{G}{\Rightarrow}}\alpha$ és $\alpha=\beta x\gamma$.

b, Bármely $A\in N$ kategóriához van olyan $\mu\in T^*$ mondat, hogy $A\overset{*}{\underset{G}{\Rightarrow}}\mu$.

c, A P nem tartalmaz $A\to A$ alaku szabályt.

A fentiek szerint bármely, nemüres nyelvet generáló CF-grammatikához megadható egy vele ekvivalens egyszerüsitett CF-grammatika. Ennek alapján a továbbiakban, anélkül, hogy ezt külön feltennénk, mindig ugy tekintjük a szóbanforgó CF-grammatikát, hogy ha nemüres nyelvet generál, akkor egyszerüsitett grammatika.

Definició: Legyen a $G=(N,T,P,S)$ CF-grammatika. Egy G-beli levezetést **baloldali levezetésnek** nevezünk, ha a levezetés minden $\alpha\Rightarrow\beta$ lépése esetén teljesül a következő: ha $\alpha=\mu A\gamma$, ahol $\mu\in T^*$, $A\in N$ és $\gamma\in V^*$, akkor $\beta=\mu\delta\gamma$, ahol $A\to\delta\in P$. /tehát a levezetés minden lépésében a balszélső kategóriát helyettesitjük/.

2.2.2. lemma: Legyen a G CF-grammatika. Minden $\upsilon\in L(G)$ mondat levezethető baloldali levezetéssel is az axiómából.

Definició: Egy G CF-grammatikát **többértelmünek** nevezünk, ha van olyan $\upsilon\in L(G)$ mondat, amelyet legalább két különböző baloldali levezetéssel le lehet vezetni az axiómából. Egy CF-grammatikát **egyértelmünek** nevezünk, ha nem többértelmü.

A fenti fogalmak illusztrációjaként nézzünk egy példát:

2.2.1. példa: $G_1 = (\{S\}, \{+, \times, a, b, c\}, P_1, S)$

$G_2 = (\{S, A\}, \{+, \times, a, b, c\}, P_2, S)$

P_1 : $S \to S \times S$, $S \to S + S$, $S \to a$, $S \to b$, $S \to c$

P_2 : $S \to aA$, $S \to bA$, $S \to cA$, $A \to +S$, $A \to \times S$, $S \to a$, $S \to b$, $S \to c$

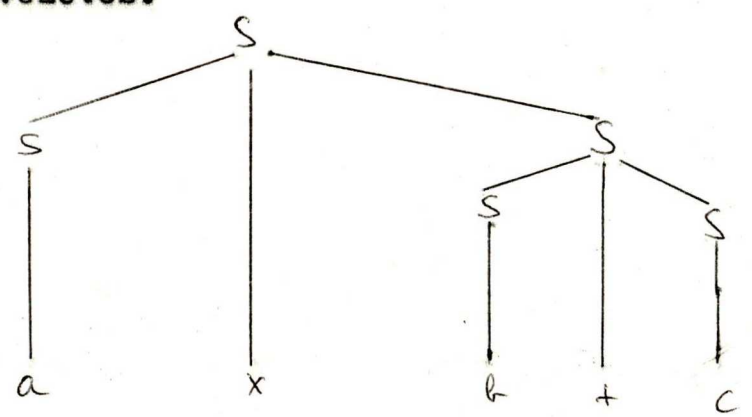
$$S \underset{G_1}{\Rightarrow} S \times S \underset{G_1}{\Rightarrow} a \times S \underset{G_1}{\Rightarrow} a \times S + S \underset{G_1}{\Rightarrow} a \times b + S \underset{G_1}{\Rightarrow} a \times b + c$$

$$S \underset{G_1}{\Rightarrow} S + S \underset{G_1}{\Rightarrow} S \times S + S \underset{G_1}{\Rightarrow} a \times S + S \underset{G_1}{\Rightarrow} a \times b + S \underset{G_1}{\Rightarrow} a \times b + c$$

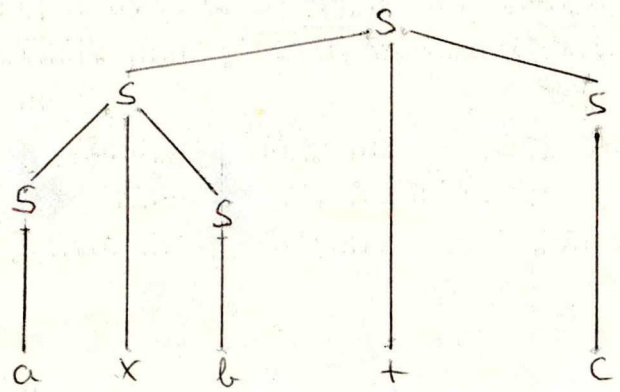
$$S \underset{G_2}{\Rightarrow} aA \underset{G_2}{\Rightarrow} a \times S \underset{G_2}{\Rightarrow} a \times bA \underset{G_2}{\Rightarrow} a \times b + S \underset{G_2}{\Rightarrow} a \times b + c$$

Tehát az $a \times b + c$ mondatnak a G_1 -ben két különböző baloldali levezetése van, a G_1 többértelmü. Ebben a konkrét esetben megadhatjuk a többértelmüség egy interpretációját is. Értelmezzük a $\times$ jelet szorzásjelnek, a $+$ jelet az összeadás jelének, az a , b és c jeleket pedig számértékeket jelölő azonositóknak. Tekintsük a két levezetés diagrammját:

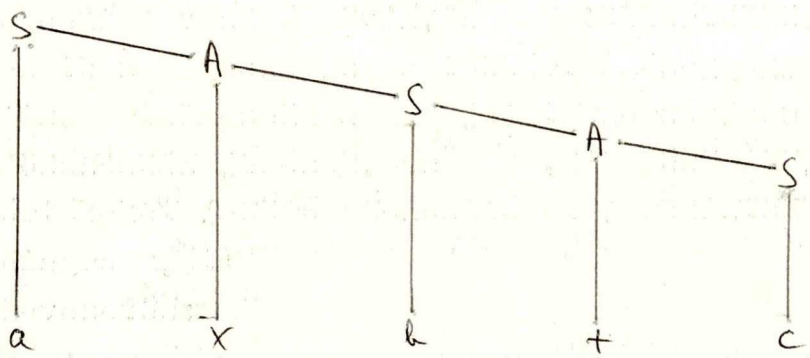
Az első levezetés:



A fentiek alapján ez ugy értelmezhető, hogy $a \times (b+c)$:
A második levezetés:



Ennek értelmezése: $(a \times b)+c$.
Tehát a két különböző baloldali levezetés két különböző "jelentést" tulajdonit a mondatnak /ha $a \neq 1$ és $c \neq 0$ akkor a két érték különböző/.
/A természetes nyelveknél sürün találkozunk ezzel a jelenséggel, pl.: "a fiu és a lány szülei"./
A G_2 grammatika ugyanazt a nyelvet generálja mint a G_1 de egyértelmüen /ez könnyen belátható/. A harmadik levezetés diagrammja:



Ennek a levezetésnek az $a \times (b+c)$ jelentést tulajdonithatjuk.

A példából látható, hogy a többértelmüség általában nem egy nyelv, hanem egy grammatika tulajdonsága. De vannak olyan nyelvek is, amelyek "önmaguknál fogva" többértelmüek, értve ezen azt, hogy bármely CF-grammatika, amely egy ilyen nyelvet generál, szükségszerüen többértelmü. Ilyen nyelv például a következő [10]:

$L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$

Definició: Egy CF - n y e l v e t e g y é r t e l m ü n e k nevezünk, ha van olyan egyértelmü CF-grammatika, amely ezt a nyelvet generálja. Egy CF - n y e l v e t t ö b b é r t e l m ü n e k nevezünk, ha nem egyértelmü.

Definició: Legyen a $G=(N,T,P,S)$ CF-grammatika. Azt mondjuk, hogy G Greibach-féle normál alaku grammatika, vagy másnéven s t a n d a r d g r a m m a t i k a, ha minden $A \to \alpha \in P$ szabály esetén $\alpha = a\beta$, ahol $a \in T$ és $\beta \in N^*$.

2.2.3. tétel: Bármely CF-grammatikához megadható egy vele ekvivalens standard CF-grammatika. [7]

Definició: Legyenek a $G_1=(N_1,T_1,P_1,S_1)$ és a $G_2=(N_2,T_2,P_2,S_2)$ CF-grammatikák. Azt mondjuk, hogy a G_2 a G_1 h o m o m o r f k é p e, ha $T_1 = T_2$ és van a V_1^* halmaznak a V_2^* halmazba való olyan leképezése, amelyre igaz:

a, $f(\varepsilon) = \varepsilon$ és minden $a \in T_1$ esetén $f(a) = a$.

b, $f(N_1) = N_2$ és $f(S_1) = S_2$

c, minden $X_1 \dots X_k \in V_1^*$ esetén $f(X_1 \dots X_k) = f(X_1) \dots f(X_k)$

d, minden $A \to \alpha \in P_1$ esetén $f(A) \to f(\alpha) \in P_2$ és a P_2 halmaznak más eleme nincs, vagyis:

$$P_2 = \{ f(A) \to f(\alpha) \mid A \to \alpha \in P_1 \}$$

Azt mondjuk, hogy a két g r a m m a t i k a i z o m o r f, ha a fent leirt f leképezés kölcsönösen egyértelmü.

2.2.4. tétel: Legyenek a G_1 és a G_2 CF-grammatikák. Ha a G_2 a G_1 homomorf képe, akkor az $L(G_1) \subseteq L(G_2)$. Ha a két grammatika izomorf, akkor ekvivalensek, vagyis $L(G_1) = L(G_2)$.

Definició: Legyenek a $G_1=(N_1,T_1,P_1,S_1)$ és a $G_2=\{N_2,T_2,P_2,S_2\}$ CF-grammatikák. Azt mondjuk, hogy

a G_2 grammatika a G_1 **grammatika részgrammatikája**, ha teljesülnek a következő feltételek:

a, $N_2 \subseteq N_1$ és $T_2 \subseteq T_1$

b, $P_2 \subseteq P_1$

c, $S_2 = S_1$

2.2.5. tétel: Legyenek a G_1 és a G_2 CF-grammatikák. Ha a G_2 a G_1 részgrammatikája, akkor $L(G_2) \subseteq L(G_1)$.

2.2.2. példa:

$G_1 = (\{S, A, B\}, \{a, b\}, P_1, S)$

$G_2 = (\{S, A, B\}, \{a, b\}, P_2, S)$

$G_3 = (\{S\}, \{a, b\}, P_3, S)$

P_1: $S \to bA$, $S \to aB$, $A \to aS$, $B \to bS$, $A \to bAA$, $B \to aBB$, $A \to a$, $B \to b$

P_2: $S \to bA$, $A \to aS$, $A \to bAA$, $A \to a$, $S \to aB$, $B \to bS$, $B \to b$

P_3: $S \to bS$, $S \to bSS$, $S \to aS$, $S \to a$, $S \to b$

$L(G_2) \subseteq L(G_1)$, $L(G_2) \subseteq L(G_3)$

A G_2 a G_1 **részgrammatikája a** G_3 **pedig a** G_2 **homomorf képe.**

Végül egy, a későbbiekben felhasználandó negatív eredmény:

2.2.6. lemma: Az $L = \{a^n b^n c^n \mid n \geq 1\}$ **nyelv nem CF-nyelv.**

3.§ A veremautomata, a determinisztikus kontextusmentes nyelvek

Az előző paragrafusban a kontextusmentes nyelveket generativ jellegü véges leirásokkal, a CF-grammatikákkal definiáltuk. A következőkben olyan, felismerő jellegü, véges leirásokkal foglalkozunk, amelyek ugyancsak a kontextusmentes nyelveket definiálják. Ezek alapvető fogalma a v e r e m m e m ó r i a, ezzel foglalkozunk először.

A veremmemória egy végtelen kapacitásu absztrakt memória, amelyben a memóriaegységek, a "memóriarekeszek" lineárisan vannak rendezve. Tehát beszélhetünk "legfelső", "alattalévő" stb. memóriarekeszekről. Minden memóriarekesz egy jel tárolására képes.

A veremmemóriába csak a "verem tetején" keresztül tudunk uj információt bevinni. Ez azt jelenti, hogy ha be akarunk irni a memóriába egy $k \geq 1$ hosszuságu jelláncot, akkor a beirási parancs hatására a memóriában lévő jellánc k hellyel automatikusan visszalép, vagyis az eddig a legfelső rekeszben lévő jel a $k+1$. rekeszbe, az alatta lévő jel a $k+2$. rekeszbe, stb. kerül. A szabaddá vált felső k helyet a beirandó lánc ugy tölti fel, hogy a lánc első /balszélső/ eleme kerül a legfelső rekeszbe, a rákövetkező a következő rekeszbe és igy tovább. Ha a beirási parancs az üres (ε) láncra vonatkozik, akkor a veremmemóriában semmi sem változik.

A kiolvasás ennek a forditottja. Egyszerre csak egy jelet lehet kiolvasni, a legfelső rekesz tartalmát. Ha ezt kiolvastuk, a memóriában lévő jellánc automatikusan "eggyel feljebb lép", vagyis az eddig második helyen lévő jel átkerül a legfelső rekeszbe a harmadik rekesz tartalma a második rekeszbe megy és igy tovább. Tehát a teljes tárolt információhoz hozzájuthatunk, de csak a forditott beirási sorrendben. A veremmemória technika jól használható programozási nyelvek forditóprogramjainak megszerkesztéséhez.

Ha a veremmemóriát kiegészitjük egy, a bemenő adatok

/szintén jelláncok/ tárolására szolgáló "szalaggal" /röviden: bemenő szalag/, és egy, véges számu belső állapottal rendelkező, valamint egy olvasó és egy iró-olvasó fejjel felszerelt vezérlő egységgel, akkor egy v e r e m a u t o m a t á t kapunk.

A v e r e m a u t o m a t a /push-down automata, PDA/ egyes alkotórészeinek állapota a müködés megkezdése előtt a következő:

a, A b e m e n ő s z a l a g o n egy véges jellánc van.

b, A vezérlőegység egy kitüntetett állapotban, a k e z d ő á l l a p o t b a n van, az o l v a s ó f e j a bemenő szalagon lévő jellánc első /balszélső/ jele felett áll olvasásra készen, az i r ó - o l v a s ó f e j készen áll a veremmemória kiolvasására.

c, A veremmemória üres, kivéve a legfelső rekeszt, ebben egy kitüntetett jel, a kezdőjel van.

Az automata müködése lépések sorozatából áll. Egy müködési lépés kétféle lehet:

1, a, Az olvasófej leolvassa a bemenő szalag megfelelő jelét és az olvasófej egy pozicióval jobbra lép /a következő jel kerül az olvasófej alá/.

b, Az iró-olvasó fej kiolvassa a memória legfelső rekeszének tartalmát.

c, Az iró-olvasó fej beir egy jelláncot a veremmemóriába és a vezérlőegység átmegy egy uj belső állapotba /ami lehet ugyanaz mint a régi/. Hogy mi lesz az uj belső állapot és a beirt lánc, az a két beolvasott jeltől, a régi belső állapottól és általános esetben a véletlentől is függ, ugyanis egy (belső állapot, bemenő jel, veremmemória jel) rendezett hármashoz az általános esetben az automatának több /de véges számu/ lehetősége van az uj belső állapot és a beirt jellánc "megválasztására". Előfordulhat az is, hogy egyetlen választási lehetőség

sincs, ilyenkor az automata nem tud tovább müködni, "blokkol".

2. Ebben az esetben az a, ütem elmarad. Minden más megegyezik az 1, esettel, de ugy, hogy itt a bemenő jel alatt az üres láncot értjük.

Egy ilyen tipusu lépésnél az automata megszakitja a beolvasást és a bemenő adatoktól függetlenül hajt végre egy lépést. Vagyis az olvasófej nem lép tovább és az automata attól függetlenül hajt végre egy lépést, hogy milyen jel van az olvasófej alatt illetve hogy egyáltalában van-e valami az olvasófej alatt. Azon, hogy "nincs semmi az olvasófej alatt" azt értjük, hogy az olvasófej, mozgása közben már /jobbra/ kilépett a bemenő szalagon lévő jellánchól. Ez esetben csak a 2. tipusu lépés lehetséges. Ha van valami az olvasó fej alatt /még nem lépett ki/ akkor mind az 1. mind a 2. tipusu lépés lehetséges.

A veremautomata formális definiciója a következő:

Definició: Egy M **v e r e m a u t o m a t a** /röviden PDA/ egy 7 objektumból álló rendszer,

$M=(K,T,V,\delta,q_0,Z_0,F)$, ahol:

a, A K egy véges halmaz, a belső állapotok halmaza.
b, A T egy véges ábécé, a bemenő ábécé.
c, A V egy véges ábécé a veremmemória ábécé.
d, A $q_0 \in K$, a kezdő állapot.
e, A $Z_0 \in V$, a kezdőjel.
f, Az $F \subseteq K$, a végállapotok halmaza.
g, A δ egy leképezés, a $K\times(T\cup\{\varepsilon\})\times V$ halmazt képezi le a $K\times V^*$ halmaz véges részhalmazainak halmazába.

Legyen $q, p_1, \ldots, p_m \in K$, $a \in T$, $Z \in V$, $m \geq 1$, $\gamma_1, \ldots, \gamma_m \in V^*$

A $\delta(q,a,Z)=\{(p_1,\gamma_1),\ldots,(p_m,\gamma_m)\}$ és a $\delta(q,a,Z)=\emptyset$ kifejezéseket a fenti 1, pontban leirt müködési lépés,

a $\delta(q, \varepsilon, Z) = \{(p_1, \gamma_1), \ldots, (p_m, \gamma_m)\}$ és a $\delta(q, \varepsilon, Z) = \emptyset$ kifejezéseket pedig a fenti 2, pontban leirt müködési lépés formális definiciójaként értelmezzük /a $(p_1, \gamma_1), \ldots, (p_m, \gamma_m)$ a választási lehetőségek, az $\emptyset$ jelzi, hogy nincs választási lehetőség/.

A következőkben azzal foglalkozunk, hogy milyen módon ad meg egy PDA egy nyelvet.

Definició: Egy $M = (K, T, V, \delta, q_0, Z_0, F)$ PDA egy **konfigurációján** értünk egy (q, γ) párt, ahol a $q \in K$ és a $\gamma \in V^*$. A (q_0, Z_0) konfigurációt **kezdő konfigurációnak** nevezzük. Ha a $q \in F$ akkor a (q, γ) párt **végkonfigurációnak** nevezzük. Azt mondjuk, hogy az M a (q, γ) **konfigurációban van**, ha belső állapota q, a veremmemória tartalma pedig γ /felülről lefelé olvasva/.

Definició: Legyen az $M = (K, T, V, \delta, q_0, Z_0, F)$ egy PDA és $a, a_1, \ldots, a_n \in T \cup \{\varepsilon\}$, a $p, q, q_1, \ldots, q_{n+1} \in K$ a $\beta, \gamma, \gamma_1, \ldots, \gamma_{n+1} \in V^*$, $n \geq 1$, $Z \in V$. Ha $(p, \beta) \in \delta(q, a, Z)$ akkor azt mondjuk, hogy a **átviszi** az automatát a $(q, Z\gamma)$ konfigurációból a $(p, \beta\gamma)$ konfigurációba. Jelöléssel:

$$a : (q, Z\gamma) \vdash_M (p, \beta\gamma)$$

Ha minden $1 \leq i \leq n$ indexre $a_i : (q_i, \gamma_i) \vdash_M (q_{i+1}, \gamma_{i+1})$ akkor azt mondjuk, hogy az $a_1 \ldots a_n$ **mondat átviszi** az M automatát a (q_1, γ_1) konfigurációból a (q_{n+1}, γ_{n+1}) konfigurációba. Jelöléssel:

$$a_1 \ldots a_n : (q_1, \gamma_1) \vdash_M^* (q_{n+1}, \gamma_{n+1})$$

Ha az egyértelmüséget nem zavarja, az M elhagyható a jelölésekből.

Definició: Azt mondjuk, hogy az M PDA f e l i s m e r i az u mondatot, ha az u mondat átviszi az M automatát a kezdő konfigurációból valamely végkonfigurációba.

Definició: Legyen az $M=(K,T,V,\delta,q_0,Z_0,F)$ egy PDA. Az M által meghatározott nyelvnek nevezzük és $T(M)$-el jelöljük a T^* halmaz azon részhalmazát, amely pontosan azokból a mondatokból áll, amelyeket az M felismer. Jelöléssel:

$$T(M)=\{u\in T^* \mid u:(q_0,Z_0)\overset{*}{\underset{M}{\vdash}}(q,\gamma),\ q\in F\}$$

2.3.1. példa:

$$M_0=(\{p,q,r\},\{a,b\},\{R,A,B\},\delta,p,R,\{r\})$$

$$\delta(p,a,R)=(p,AR) \qquad \delta(q,a,A)=(q,\varepsilon)$$

$$\delta(p,b,R)=(p,BR) \qquad \delta(q,b,B)=(q,\varepsilon)$$

$$\delta(p,a,A)=\{(p,AA),(q,\varepsilon)\} \qquad \delta(q,\varepsilon,R)=(r,R$$

$$\delta(p,a,B)=(p,AB)$$

$$\delta(p,b,A)=(p,BA)$$

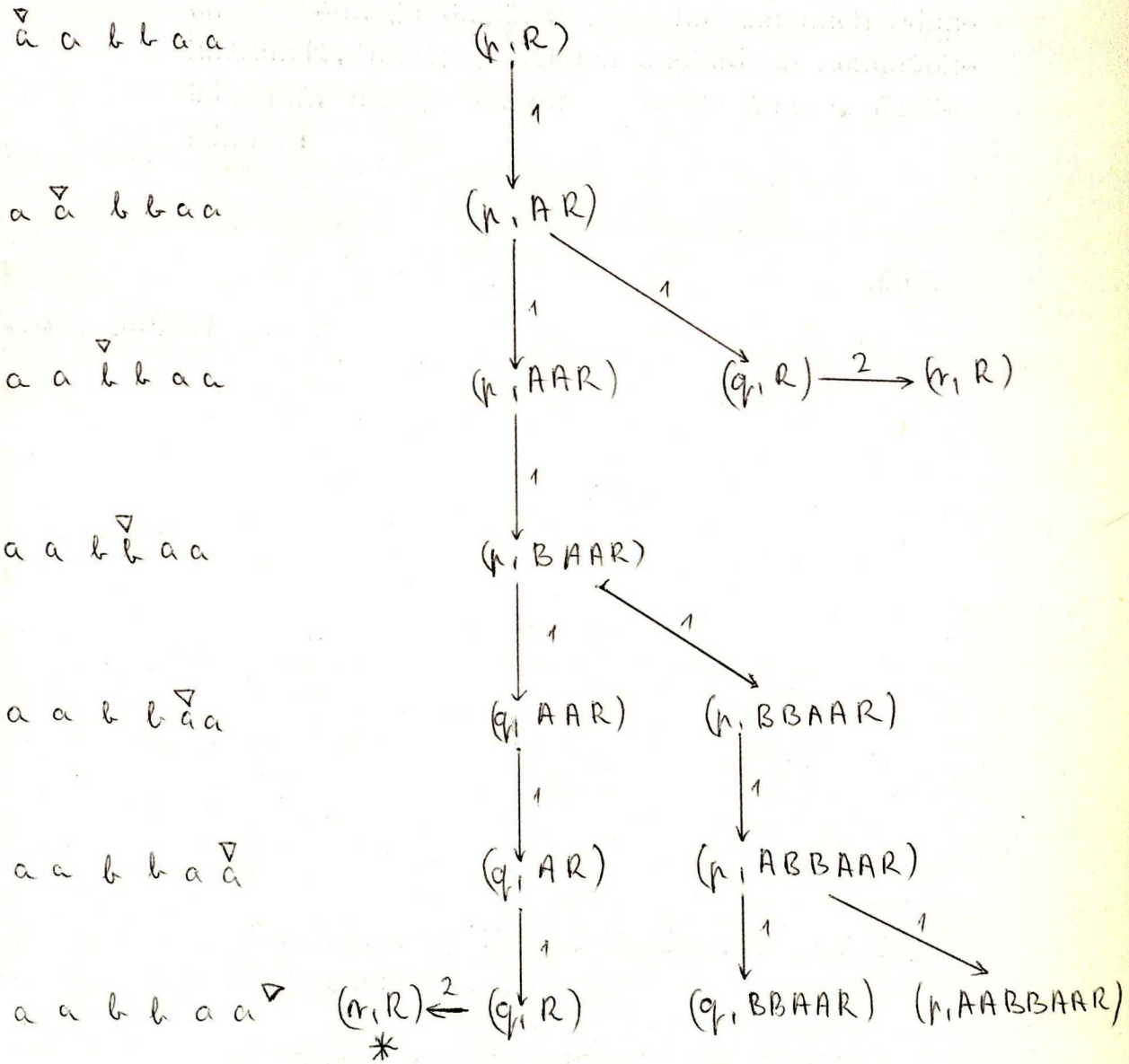
$$\delta(p,b,B)=\{(p,BB),(q,\varepsilon)\}$$

Minden más esetben a leképezés eredménye legyen az üres halmaz.

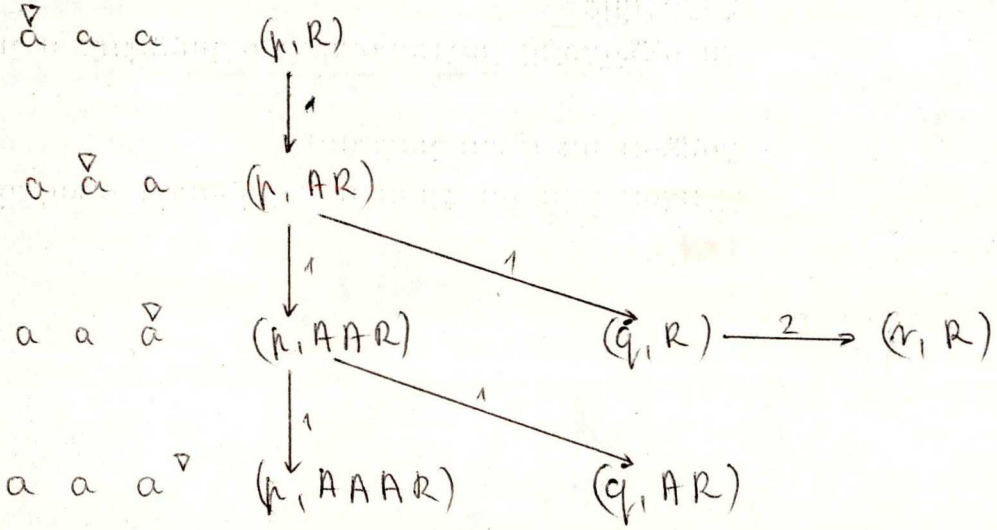
pl.: $\delta(q,a,R)=\delta(p,\varepsilon,A)=\phi$

Az egyelemü halmazoknál elhagytuk a halmazjelölést.

Legyen $u_1 = aabbaa$, és tegyük fel, hogy az M_o bemenő szalagján az u_1 szerepel. A következő ábrán megadjuk az összes lehetséges működését az u_1 bemenetre. A ∇ az olvasófej pozicióját, a konfigurációkat összekötő nyilakra irt szám pedig a működési lépés tipusát jelzi.



Nézzük ugyanezt az $u_2 = aaa$ mondattal.



Mivel az M_0 -nak van olyan működése, amellyel az u_1 átviszi az automatát egy végkonfigurációba, / * eset/, tehát az $u_1 \in T(M)$.
/A példából az is kitünik, hogy $aa \in T(M)$/
Az u_2 esetén ilyen nincs, tehát az $u_2 \notin T(M)$.
Megjegyezzük még /a példában ez nem igy van/, hogy a végkonfiguráció nem szükségképpen "blokkoló" konfiguráció.

A következő tétel megmutatja, hogy milyen szoros összefüggés van a CF-grammatikák és a veremautomaták között.

2.3.1. tétel: Bármely M PDA esetén a $T(M)$ nyelv CF-nyelv, és bármely L CF-nyelvhez megadható olyan M PDA, hogy $L = T(M)$.

Tehát a CF grammatikák és a veremautomaták ugyanazon nyelvosztályt irják le, a grammatikák generativ módon, a veremautomaták pedig felismerő módon.

2.3.2. példa: $G_0 = (\{S\}, \{a, b\}, P, S)$

$P: S \to aSa$

$S \to bSb$

$S \to aa$

$S \to bb$

Könnyen belátható, hogy $L(G_0) = T(M_0) = \{vv^R \mid v \in \{a,b\}^+\}$ ahol az M_0 a 2.3.1. példa automatája a G_0 pedig a fenti grammatika.

A CF-nyelvek egy fontos osztályát, a determinisztikus kontextusmentes nyelveket is a PDA fogalmának a segitségével definiáljuk. /Ismeretes a megfelelő grammatika-tipus, az $LR(k)$ grammatikatipus is, de nekünk a dolgozatban elegendő az automatákkal való leirás./

Definició: Egy $M=(K,T,V,\delta,q_0,Z_0,F)$ veremautomatát **determinisztikus veremautomatának** /röviditve DPDA/ nevezzük, ha minden $q \in K$, $a \in T$ és $Z \in V$ esetén teljesülnek a következő feltételek:

a, A $\delta(q,a,Z)$ legfeljebb egy elemet tartalmaz.

b, A $\delta(q,\varepsilon,Z)$ legfeljebb egy elemet tartalmaz.

c, Ha $\delta(q,\varepsilon,Z) \neq \emptyset$ akkor $\delta(q,a,Z) = \emptyset$.

A három megkötés megszünteti a véletlen szerepét az automata müködésében. Az első kettő miatt minden lépésben legfeljebb egy (uj állapot, kiirandó jelláncz) pár áll rendelkezésre, tehát nincs választási lehetőség. A harmadik megakadályozza, hogy az automata véletlenszerüen dönthessen arról, hogy egy lépésben felhasználja-e a bemenő információt vagy sem.

Definició: Egy L nyelvet **determinisztikus kontextusmentes** /röviden DCF/ nyelvnek nevezünk, ha van olyan M DPDA, amelyre igaz, hogy $T(M)=L$.

A DCF nyelvek a CF-nyelvek valódi részhalmazát alkotják. Bebizonyitható pl., hogy a 2.3.1. példában megadott $L_1 = \{vv^R \mid v \in \{a,b\}^+\}$ nyelvhez nincs DPDA. De ha "megjelöljük" a mondatok közepét, akkor már DCF-nyelvet kapunk, vagyis az $L_2 = \{vcv^R \mid v \in \{a,b\}^+\}$ nyelv DCF-nyelv, feltéve, hogy $c \neq a$ és $c \neq b$.

2.3.3. példa:

$$M = (\{p,q\}, \{a,b\}, \{Z_0, A, B\}, \delta, p, Z_0, \{p\})$$

$\delta(p,a,Z_0)=(q,AZ_0)$ $\quad$ $\delta(q,a,B)=(q,\varepsilon)$

$\delta(p,b,Z_0)=(q,BZ_0)$ $\quad$ $\delta(q,a,A)=(q,AA)$

$\delta(q,\varepsilon,Z_0)=(p,Z_0)$ $\quad$ $\delta(q,b,A)=(q,\varepsilon)$

$\delta(q,b,B)=(q,BB)$

$$T(M)=\{u\in\{a,b\}^* \mid |u|_a=|u|_b\}$$

Az M nyilván DPDA.

2.3.2. tétel: Minden DCF-nyelv egyértelmű.

2.3.3. tétel: Legyen a T egy véges ábécé, az $L\subseteq T^*$ és a $d\notin T$. Az $L_1=\{ud \mid u\in L\}$ nyelv akkor és csak akkor DCF-nyelv, ha az L DCF-nyelv.

Megemlitjük még a DCF-nyelvek egy fontos tulajdonságát. A 2.3.1. példából láthatjuk, hogy egy nem determinisztikus PDA felesleges lépéseket is végez, értve ezen azt, hogy egy olyan mondatnál, amelyet felismer "tévutra" is mehet, nem biztos, hogy már az első menetben felismeri. Hogy egy mondatról azt mondhassuk, hogy nem tartozik az automata nyelvéhez, minden lehetséges esetet végig kell próbálni. Ez konkrét alkalmazásoknál, pl. egy program szintaktikai vizsgálatánál időveszteséggel jár. Ez a veszteség nem lép fel a DPDA esetén. Egy menetben eldönti, hogy az illető mondat eleme-e a nyelvnek, vagy sem.

4.§ A reguláris nyelvek és a véges automaták

A reguláris nyelvek halmaza a formális nyelvek legszükebb olyan osztályát alkotja, amelyet még fel lehet használni a természetes nyelvekben előforduló bizonyos általános jelenségek modellezésére. A formális nyelvészet fejlődésének folyamán először ezeket a nyelveket tanulmányozták részletesen.

Definició: Egy $G=(N, T, P, S)$ kontextusmentes grammatikát **reguláris grammatikának** nevezünk, ha minden produkciós szabálya $A \to aB$ vagy $A \to a$ vagy $A \to \Sigma$ alaku, ahol $A, B \in N$ és $a \in T$.

Definició: Egy L nyelvet **reguláris nyelvnek** nevezünk, ha van olyan G reguláris grammatika, hogy $L = L(G)$.

A reguláris nyelvek osztályának felismerő jellegü véges leirását a **véges automaták** adják. Egy véges automata egy olyan rendszer modellje, amelynek véges számu, egymástól jól elkülönithető, stabil belső állapota van. A rendszert egy állapotából egy környezeti hatás /bemenőjel/ viszi át egy másik állapotba és eközben a rendszer kimenőjelet adhat a környezet számára. Nekünk itt elegendő olyan véges automatákkal foglalkozni, amelyek nem adnak kimenőjelet. A formális definició a következő:

Definició: Egy A **véges automata** egy 5 objektumból álló rendszer, $A=(K, T, \delta, q_0, F)$
a, A K egy véges halmaz, a belső állapotok halmaza.
b, A T egy véges halmaz, a bemenő ábécé.
c, A $q_0 \in K$ a kezdő állapot.
d, Az $F \subseteq K$, a végállapotok halmaza
e, A δ a $K \times T$ halmaz leképezése a K halmaz részhalmazainak halmazába.

Legyen $q, p_1, \ldots, p_k \in K$, $k \geq 1$ és $a \in T$. A

$\delta(q,a)=\{p_1,\dots,p_k\}$ kifejezést ugy értelmezzük, hogy ha az automata a q állapotban az a bemenő jelet kapja, akkor átmegy a $p_1,\dots,p_k$ állapotok valamelyikébe. A $\delta(q,a)=\emptyset$ formulát ugy értelmezzük, hogy ha az automata a q állapotban az a bemenő jelet kapja, akkor nem tud tovább müködni, "blokkol". A δ leképezést kiterjesztjük a $K\times T^*$ halmaz elemeire is, a következő módon: Minden $q\in K$, $a\in T$ és $u\in T^*$ esetén legyen:

$$\delta(q,\varepsilon)=q \quad \text{és} \quad \delta(q,ua)=\bigcup_{p\in\delta(q,u)}\delta(p,a)$$

Definició: Azt mondjuk, hogy az $A=(K,T,\delta,q_0,F)$ véges automata **f e l i s m e r i** az $u\in T^*$ mondatot, ha a $\delta(q_0,u)\cap F\neq\emptyset$.

A fenti definició részletesebben kifejtve a következőt jelenti: ha az $u=x_1\dots x_k$ ahol $k\geq 1$ és minden $1\leq i\leq k$ esetén $x_i\in T$, akkor van olyan $p_1,\dots,p_{k+1}$ állapotsorozat, hogy teljesül: $\delta(p_i,x_i)=p_{i+1}$, $1\leq i\leq k$, $p_1=q_0$ és $p_{k+1}\in F$.

Definició: Legyen az $A=(K,T,\delta,q_0,F)$ egy véges automata. Az A **által m e g h a t á r o z o t t n y e l v n e k** nevezzük, és $T(A)$ -val jelöljük a T^* halmaz azon részhalmazát, amely pontosan azokból a mondatokból áll, amelyeket az A véges automata felismer. Jelöléssel:

$$T(A)=\{u\in T^*\mid \delta(q_0,u)\cap F\neq\emptyset\}$$

Definició: Az $A=(K,T,\delta,q_0,F)$ véges automatát **d e t e r m i n i s z t i k u s v é g e s a u t o m a t á n a k** nevezzük, ha minden $q\in K$ és $a\in T$ esetén a $\delta(q,a)$ legfeljebb egy elemet tartalmaz.

A következő tétel kimondja, hogy a nyelvek meghatározása szempontjából a véges automaták és a determinisztikus véges automaták osztálya ekvivalens.

2.4.1. tétel: Minden A_1 véges automatához megadható olyan A_2 determinisztikus véges automata, hogy $T(A_1) = T(A_2)$.

2.4.2. tétel: Minden G reguláris grammatikához van olyan A véges automata, hogy $L(G) = T(A)$ és minden A véges automatához van olyan G reguláris grammatika, hogy $T(A) = L(G)$.

2.4.1. példa:

$$G = (\{S, A\}, \{a, b\}, P, S)$$

P: $S \to aS$ $\quad A \to bS$

$S \to bA$ $\quad A \to b$

$S \to aA$ $\quad S \to a$

$A \to bA$ $\quad S \to b$

$$A_1 = (\{p, q\}, \{a, b\}, \delta_1, p, \{q\})$$

$\delta_1(p, a) = \{p, q\}$ $\quad \delta_1(q, b) = \{p, q\}$

$\delta_1(p, b) = q$ $\quad \delta_1(q, a) = \emptyset$

$$A_2 = (\{p, q, r\}, \{a, b\}, \delta_2, p, \{q, r\})$$

$\delta_2(p, a) = r$ $\quad \delta_2(q, b) = r$

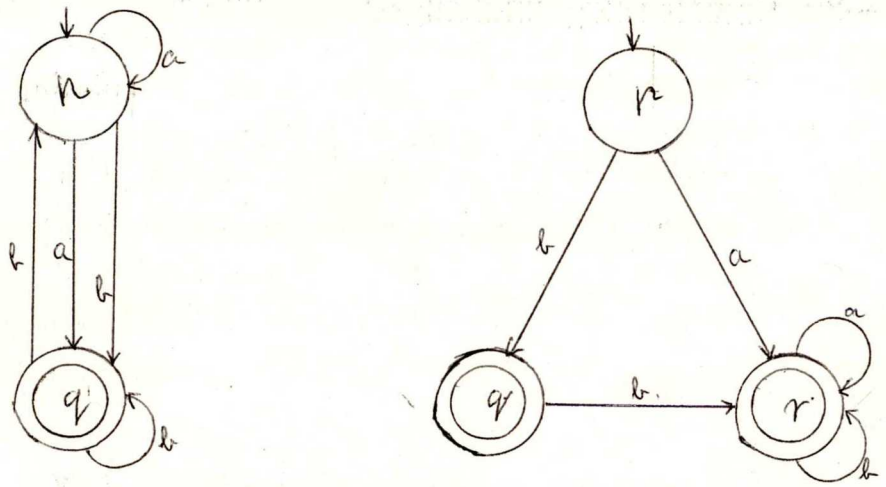
$\delta_2(p, b) = q$ $\quad \delta_2(r, a) = r$

$\delta_2(r, b) = r$ $\quad \delta_2(q, a) = \emptyset$

$$L(G) = \{b\} \cup \{bbu \mid u \in \{a, b\}^*\} \cup \{av \mid v \in \{a, b\}^*\}$$

$$L(G) = T(A_1) = T(A_2)$$

A véges automaták megadásának könnyebben áttekinthető formája a diagrammokkal való megadás, ahol az állapotokat körökkel, a végállapotokat kettős körökkel, az átmenetek irányát nyilakkal szemléltetjük, minden nyil mellé odairva, hogy milyen bemenő jel hatására következik be az átmenet. A kezdő állapotot nyillal jelöljük meg. Ezzel a jelölésmóddal a fenti példa automatái:



2.4.3. tétel: Minden reguláris nyelv-DCF nyelv.

A tétel közvetlenül adódik abból a tényből, hogy minden determinisztikus véges automatát felfoghatunk egy olyan veremautomataként, amelynél a $V = \{Z_0\}$ és minden $q \in K$ és $a \in T$ esetén $\delta(q, a, Z_0) = (p, Z_0)$ tehát a veremmemória tartalma nem változik.

A DCF nyelvek halmaza valódi részhalmazként tartalmazza a reguláris nyelvek halmazát. Bebizonyitható, hogy az $L = \{a^n b^n \mid n \geq 1\}$ nyelv nem reguláris DCF-nyelv.

Nemcsak reguláris CF-grammatikák generálhatnak reguláris nyelveket.

Definició: Legyen a $G = (N, T, P, S)$ CF grammatika. Azt mondjuk, hogy a G **önbeágyazó grammatika**, ha van olyan $A \in N$ és $\alpha, \beta \in V^*$ hogy $\alpha \neq \Sigma$, $\beta \neq \Sigma$ és $A \overset{*}{\underset{G}{\Rightarrow}} \alpha A \beta$.

2.4.4. tétel: Ha a G nem önbeágyazó CF-grammatika, akkor az $L(G)$ nyelv reguláris nyelv.

Ennek közvetlen következménye:

2.4.5. tétel: Minden véges nyelv reguláris nyelv.

A fenti állitások nem fordithatók meg. Ezt egy példával igazoljuk.

2.4.2. példa: $G_1 = (\{S, A\}, \{a, b\}, P_1, S)$

$G_2 = (\{S, A\}, \{a, b\}, P_2, S)$

P_1: $S \to aAb$ $\quad$ P_2: $S \to aS$

$S \to Ab$ $\quad$ $S \to bA$

$A \to aA$ $\quad$ $A \to bA$

$A \to Ab$ $\quad$ $A \to \varepsilon$

$A \to \varepsilon$

A G_1 grammatika önbeágyazó, mert $A \overset{*}{\underset{G_1}{\Rightarrow}} aAb$ a G_2 grammatika reguláris. A két grammatika ekvivalens, ugyanis: $L(G_1) = L(G_2) = \{a^n b^m \mid n \geq 0, m \geq 1\}$

Végül még két, a későbbiekben felhasználandó példát adunk nem reguláris CF-nyelvekre. Könnyen bebizonyitható, hogy ezek nem reguláris nyelvek.

2.4.3. példa: $L_1 = \{a^n c b^n \mid n \geq 1\}$

2.4.4. példa: $L_2 = \{a^n b^{n+1} \mid n \geq 1\}$

5.§ Müveletek kontextusmentes nyelvekkel

A következőkben azt vizsgáljuk, hogy a definiált nyelvosztályok bizonyos müveletekre nézve zártak-e. Az ilyen vizsgálatok fontos szerepet játszanak a formális nyelvek elméletében. Például eszközül szolgálhatnak egy nyelvosztály algebrai jellemzéséhez, vagy annak eldöntéséhez, hogy egy nem grammatikával megadott nyelv generálható-e egy adott tipusu grammatikával. Először is definiáljuk azokat a müveleteket, amelyekkel foglalkozunk.

Definició: Legyenek a V_1 és V_2 véges ábécék és az f a V_1^* halmaz egy leképezése a V_2^* halmazba.
Azt mondjuk, hogy az f leképezés h o m o m o r f l e k é p e z é s, ha teljesülnek a következő feltételek:
a, $f(\varepsilon) = \varepsilon$
b, minden $x_1 \dots x_k \in V_1^+$ esetén $f(x_1 \dots x_k) = f(x_1) \dots f(x_k)$

Definició: Legyen a T egy véges ábécé és $L_1 L_{11} L_2 \subseteq T^*$.
Az L_1 és L_2 nyelv e g y e s i t é s é n a következő nyelvet értjük:

$$L_1 \cup L_2 = \{u \mid u \in L_1 \text{ vagy } u \in L_2\}$$

Az L_1 és L_2 nyelv m e t s z e t é n a következő nyelvet értjük:

$$L_1 \cap L_2 = \{u \mid u \in L_1 \text{ és } u \in L_2\}$$

Az L_1 nyelv L_2 -vel való s z o r z a t á n értjük a következő nyelvet:

$$L_1 L_2 = \{uv \mid u \in L_1 \text{ és } v \in L_2\}$$

Az L nyelv m e g f o r d i t á s á n értjük a következő nyelvet:

$$L^R = \{u^R \mid u \in L\}$$

Az L nyelv komplementerén értjük a következő nyelvet:

$$\overline{L} = \{ u \in T^* \mid u \notin L \}$$

Legyen az f egy homomorf leképezés. Az L nyelv h o m o m o r f k é p é n értjük a következő nyelvet: $f(L) = \{ f(u) \mid u \in L \}$.

2.5.1. tétel: Ha az L_1 L_1 és L_2 CF-nyelvek, akkor az $\overline{L}$ és az $L_1 \cap L_2$ nyelvek CS-nyelvek.

2.5.2. tétel: Ha az L_1 L_1 és L_2 CF-nyelvek, az f homomorf leképezés és az R reguláris nyelv, akkor az $L_1 \cup L_2$, $L_1 L_2$, L^R, $f(L)$, $L \cap R$ nyelvek is CF-nyelvek.

2.5.3. tétel: A metszetképzés és a komplementerképzés kivezetnek a CF-nyelvek halmazából.

2.5.4. tétel: Ha az L DCF-nyelv és az R reguláris nyelv, akkor az $\overline{L}$ és az $L \cap R$ nyelvek is DCF-nyelvek.

2.5.5. tétel: A DCF-nyelvek halmazából kivezetnek a következő müveletek: egyesités, metszetképzés, szorzatképzés, megforditás, homomorf leképezés.

A reguláris nyelvek halmaza abban is "szabályos", hogy a felsorolt müveletekre nézve zárt.

2.5.6. tétel: A reguláris nyelvek halmaza zárt a következő müveletekre: egyesités, szorzatképzés, metszetképzés, megforditás, komplementerképzés, homomorf leképezés.

A későbbiekben felhasználjuk a homomorf leképezésre vonatkozó következő állitást:

2.5.7. tétel: Legyen a $G_1 = (N_1, T_1, P_1, S_1)$ CF-grammatika és az f a T_1^* halmaz egy homomorf leképezése a T_2^* halmazba, ahol T_2 egy véges ábécé. Legyen a g az $(N_1 \cup T_1)^*$ halmaznak az $(N_1 \cup T_2)^*$

halmazba való, a következő formulákkal megadott, homomorf leképezése:

a, $g(\varepsilon) = \varepsilon$, $g(a) = f(a)$ minden $a \in T_1$ esetén

b, $g(A) = A$ minden $A \in N_1$ esetén

Legyen a $P_2 = \{ A \to g(\alpha) \mid A \to \alpha \in P_1 \}$ és a $G_2 = (N_1, T_2, P_2, S_1)$

Állitás:

$$f(L(G_1)) = L(G_2)$$

2.5.1. példa: $G_1 = (\{S\}, \{a, b, c\}, P_1, S)$

$P_1 : S \to aSb$

$S \to bSb$

$S \to c$

$f(\varepsilon) = \varepsilon$, $f(a) = abd$, $f(b) = c$, $f(c) = \varepsilon$

$G_2 = (\{S\}, \{a, b, c, d\}, P_2, S)$

$P_2 : S \to abdSc$

$S \to cSc$

$S \to \varepsilon$

6.§ Megoldhatóság

A megoldhatóság fogalmával az I. fejezetben részletesen foglalkoztunk. Itt felsorolunk néhány olyan problémát, amelyekkel a már definiált és a később definiálandó nyelvosztályok esetében foglalkozunk.
Először a CF-nyelvek halmazára vonatkozó néhány tételt tekintünk.

2.6.1. tétel: Van olyan algoritmus, amely bármely G CF-grammatika és u mondat esetén eldönti, hogy az u az $L(G)$ nyelvhez tartozik-e vagy sem.

2.6.2. tétel: Van olyan algoritmus, amely bármely G CF-grammatika esetén eldönti, hogy az $L(G)$ nyelv üres-e vagy sem és ha nem üres, akkor véges-e vagy sem.

2.6.3. tétel: Legyen T egy véges ábécé. Nincs olyan algoritmus, amely bármely L CF-nyelv esetén eldöntené, hogy $L=T^*$ teljesül-e vagy sem.

2.6.4. tétel: Legyen R reguláris nyelv. Nincs olyan algoritmus, amely bármely L CF-nyelv esetén eldöntené, hogy az $L=R$ teljesül-e vagy sem.

2.6.5. tétel: Nincs olyan algoritmus, amely bármely G CF-grammatika esetén eldöntené, hogy az $L(G)$ nyelv reguláris-e vagy sem.

2.6.6. tétel: Nincs olyan algoritmus, amely bármely L CF-nyelv esetén eldöntené, hogy az $\overline{L}$ nyelv CF-nyelv-e vagy sem.

Az utóbbi négy probléma a DCF-nyelvek halmazára szoritkozva már általánosan megoldható.

2.6.7. tétel: Legyen R reguláris nyelv. Van olyan algoritmus, amely bármely L DCF-nyelv esetén eldönti, hogy $L=R$ teljesül-e vagy sem.

2.6.8. tétel: Van olyan algoritmus, amely bármely L DCF-nyelv esetén eldönti, hogy az L reguláris nyelv-e vagy sem.

A 2.6.3. tételnek megfelelő probléma általános megoldhatósága a 2.6.7. tétel következménye, hiszen a T^* nyelv reguláris. A komplementerképzés problémája triviálisan általánosan megoldható a 2.5.5. tétel szerint.

2.6.9. tétel: Nincs olyan algoritmus, amely bármely L_1 és L_2 DCF-nyelvek esetén eldöntené, hogy az $L_1 \cap L_2$ nyelv üres-e vagy sem, vagy véges-e vagy sem.

2.6.10. tétel: Nincs olyan algoritmus, amely bármely L_1 és L_2 DCF-nyelvek esetén eldöntené, hogy az $L_1 \subseteq L_2$ teljesül-e vagy sem.

2.6.11. tétel: Nincs olyan algoritmus, amely bármely L_1 és L_2 CF-nyelvek esetén eldöntené, hogy az $L_1 = L_2$ teljesül-e vagy sem.

Az ekvivalenciaproblémára vonatkozóan a DCF-nyelvek esetén nem ismeretes általános tétel.

2.6.12. tétel: Nincs olyan algoritmus, amely bármely L_1 és L_2 DCF-nyelvek esetén eldöntené, hogy az $L_1 \cup L_2$ nyelv DCF-nyelv-e vagy sem.

2.6.13. tétel: Nincs olyan algoritmus, amely bármely L_1 és L_2 DCF-nyelvek esetén eldöntené, hogy az $L_1 L_2$ nyelv DCF nyelv-e vagy sem.

Ez a két probléma a CF-nyelvek esetén triviálisan általánosan megoldható /2.5.2./.

2.6.14. tétel: Nincs olyan algoritmus, amely bármely L_1 és L_2 CF-nyelvek esetén eldöntené, hogy az $L_1 \cap L_2$ nyelv CF-nyelv-e vagy sem.

2.6.15. tétel: Nincs olyan algoritmus, amely bármely L_1 és L_2 DCF-nyelvek esetén eldöntené, hogy az $L_1 \cap L_2$ nyelv DCF-nyelv-e vagy sem.

2.6.16. tétel: Nincs olyan algoritmus, amely bármely G CF-grammatika esetén eldöntené, hogy az $L(G)$ nyelv DCF-nyelv-e vagy sem.

Az eddigiekben említett problémák /kiegészítve még néhány más problémával/ szerepelnek az összefoglaló táblázatban, az ezután definiálandó nyelvosztályokra vonatkozóan is /Függelék/.

A következő állitásokat a későbbiekben más tételekhez használjuk fel.

2.6.17. lemma: Van olyan algoritmus, amely bármely CF-grammatika esetén eldönti, hogy önbeágyazó-e vagy sem.

2.6.18. lemma: Legyen T egy véges ábécé. Van olyan algoritmus, amely bármely $a \in T$ és G CF-grammatika esetén megadja az $L(G)$ nyelv legrövidebb olyan u mondatait, amelyekre $|u|_a \geq 1$ teljesül.

2.6.19. lemma: Van olyan algoritmus, amely bármely G CF-grammatika esetén megadja az $L(G)$ nyelv legrövidebb mondatát.

7.§ Az egyszerű determinisztikus nyelvek

Az "egyszerű determinisztikus nyelveket" A.J. Korenjak és J.E. Hopcroft definiálták és vizsgálták először a "Simple deterministic languages" című munkájukban [12]. Ezt a cikket vesszük alapul ezen nyelvosztály ismertetésénél.

Definició: Legyen $G=(N,T,P,S)$ egy standard CF-grammatika. Ha bármely (A,a) pár esetén, ahol $A\in N$ és $a\in T$, legfeljebb egy olyan $\alpha\in N^*$ lánc van, amelyre az $A\to a\alpha\in P$ teljesül, akkor a G grammatikát egyszerű determinisztikus grammatikának, vagy röviden S-grammatikának nevezzük.

Az S-grammatika definiciójából közvetlenül következik:

2.7.1. tétel: Minden S-grammatika egyértelmű.

Definició: Egy nyelvet egyszerű determinisztikus nyelvnek, vagy röviden S-nyelvnek nevezünk, ha van olyan S-grammatika, amely ezt a nyelvet generálja.

Ezzel megadtuk az S-nyelvek generativ leirását, a megfelelő felismerő jellegü leirás a következő:

Definició: Legyen az $M=(K,T,V,\delta,q_0,Z_0,F)$ egy DPDA. Ha teljesül a következő két feltétel:

1, $F=K=\{q_0\}$

2, Minden $Z\in V$ esetén $\delta(q_0,\varepsilon,Z)=\phi$

akkor azt mondjuk, hogy az M egy egyszerű determinisztikus automata, vagy röviden SA.

Mint a definicióból kiderül, egy SA nem szakitja félbe a bemenő szalag beolvasását, mindig a bemenet figyelembevételével müködik. Ebből következik, hogy a DPDA-nál "gazdaságosabb", hiszen pontosan annyi lépésben ellenőriz egy mondatot, amennyi a mondat hossza /egy DPDA-nál a

beolvasás megszakitható és két beolvasás között akárhány lépés végrehajtható/. Mivel csak egy belső állapot van, ami egyben a kezdő és a végállapot is, a "felismerés" definiciója a PDA-hoz képest módosul.

Definició: Legyen az $M=(\{q_0\},T,V,\delta,q_0,Z_0,\{q_0\})$ egy SA. Azt mondjuk, hogy az $u\in T^*$ mondatot f e l i s m e r i az M S-automata, ha az u mondat az automatát átviszi a (q_0,Z_0) konfigurációból a (q_0,ε) konfigurációba /kiüriti a veremmemóriát/. Jelöléssel:

$$u:(q_0,Z_0)\overset{*}{\underset{M}{\vdash}}(q_0,\varepsilon)$$

Definició: Legyen az $M=(\{q_0\},T,V,\delta,q_0,Z_0,\{q_0\})$ egy SA. Azt a nyelvet, amely pontosan azokból a mondatokból áll, amelyeket az M felismer, az M által m e g h a t á r o z o t t n y e l v n e k nevezzük és $T(M)$ -el jelöljük. Jelöléssel:

$$T(M)=\{u\in T^*\mid u:(q_0,Z_0)\overset{*}{\underset{M}{\vdash}}(q_0,\varepsilon)\}$$

2.7.2. tétel: Minden G S-grammatikához van olyan M SA, hogy $L(G)=T(M)$ és minden M S-automatához van olyan G S-grammatika, hogy $L(G)=T(M)$.

Tehát az S-automaták és az S-grammatikák a nyelvleirás szempontjából ekvivalensek. Ennek közvetlen következménye:

2.7.3. tétel: Minden S-nyelv DCF-nyelv.

2.7.1. példa:

$$L=\{a^m b^n a c^n \mid m\geq 0,\ n\geq 1\}$$

$$G=(\{S,A,B\},\{a,b,c\},P,S)$$

$$P:\ S\to aS \qquad A\to a$$

$$S\to bAB \qquad B\to c$$

$$A\to bAB$$

$M = (\{p\}, \{a, b, c\}, \{A, B, S\}, \delta, p, S, \{p\})$

$\delta(p, a, S) = (p, S)$

$\delta(p, b, S) = (p, AB)$

$\delta(p, b, A) = (p, AB)$

$\delta(p, a, A) = (p, \varepsilon)$

$\delta(p, c, B) = (p, \varepsilon)$

Minden más esetben a jobboldal az üres halmaz.

$L(G) = T(M) = L$

2.7.2. példa:

$$L = \{a^n b a^n b \mid n > 0\} \cup \{a^m c a^m c \mid m > 0\}$$

A 2.7.2. példában szereplő nyelv nem S-nyelv, de DCF-nyelv. Ebből következően az S-nyelvek halmaza a DCF-nyelvek halmazának valódi részét alkotja.

Definició: Legyen T egy véges ábécé és $L \subseteq T^*$. Azt mondjuk, hogy az L nyelvnek megvan a p r e f i x t u l a j d o n s á g a, ha bármely $u \in L$ és $v \in T^*$ esetén, ha $v \neq \varepsilon$, akkor $uv \notin L$.

2.7.4. tétel: Minden S-nyelv rendelkezik a prefix tulajdonsággal.

Megjegyezzük, hogy a DCF-nyelvekre ez már nem áll, hiszen pl. az $L = \{a^n b^m \mid n \geq 0, m \geq 0\}$ nyelv DCF nyelv /mivel reguláris/, de nem rendelkezik a prefix tulajdonsággal. Ebből az is következik, hogy az S-nyelvek halmaza nem tartalmazza a reguláris nyelvek halmazát. Természetesen, az S nyelvek között vannak reguláris nyelvek is.

2.7.5. tétel: Legyen a T egy véges ábécé és az $c \notin T$ egy további jel. Minden $R \subseteq T^*$ reguláris nyelv esetén az $L = Rc$ nyelv /ami reguláris nyelv/ S-nyelv.

2.7.6. tétel: Legyen G egy S-grammatika. Az $L(G)$ nyelv akkor és csak akkor reguláris, ha a G nem önbeágyazó grammatika.

A bevezetett müveletekre vonatkozóan a következő tételek érvényesek:

2.7.7. tétel: Ha az L_1 és L_2 nyelvek S-nyelvek, akkor az $L_1 L_2$ nyelv is S-nyelv.

2.7.8. tétel: Az S nyelvek halmazából kivezetnek a következő müveletek: egyesités, metszetképzés, reguláris nyelvvel való metszetképzés, megforditás és homomorf leképezés. Ha az L S-nyelv, akkor az $\overline{L}$ nem S-nyelv.

Mivel az S-nyelvek a DCF-nyelvek részhalmazát alkotják, azok a problémák, amelyek a DCF-nyelvekre vonatkozóan általánosan megoldhatók, az S-nyelvekre vonatkozóan is általánosan megoldhatók. Ezekkel nem foglalkozunk külön, kivéve a következő problémát, amelynél a megoldó algoritmus triviálissá egyszerüsödik:

2.7.9. tétel: Legyen a T egy véges ábécé. A T^* nyelv nem S-nyelv.

A tétel a 2.7.4. tétel következménye.

Az ekvivalenciaprobléma itt általánosan megoldható:

2.7.10. tétel: Van olyan algoritmus, amely bármely G_1 és G_2 S-grammatikák esetén eldönti, hogy az $L(G_1) = L(G_2)$ teljesül-e vagy sem.

Bizonyos problémák ebben a nyelvosztályban sem oldhatók meg általánosan.

2.7.11. tétel: Nincs olyan algoritmus, amely bármely G_1 és G_2 S-grammatikák esetén eldöntené, hogy az $L(G_1) \cap L(G_2)$ nyelv üres-e vagy sem, vagy véges-e vagy sem.

2.7.12. tétel: Nincs olyan algoritmus, amely tetszőleges G CF-grammatika esetén eldöntené, hogy a G S-nyelvet generál-e vagy sem.

Bizonyos problémákra nem ismeretes általános tétel. Ilyen pl. a tartalmazás problémája: Van-e olyan algoritmus, amely bármely G_1 és G_2 S-grammatikák esetén eldönti, hogy az $L(G_1) \subseteq L(G_2)$ teljesül-e vagy sem.

III. fejezet

BALOLDALI LEVEZETÉS NYELVEK

1.§ Bevezetés

Legyen a $G=(N,T,P,S)$ egy kontextusmentes grammatika. A grammatika minden, az axiomából kiinduló, baloldali levezetéshez hozzárendelünk egy jellláncot a következő módon: A grammatika szabályainak, a P elemeinek, adjunk egyedi "neveket", c i m k é z z ü k meg őket valamilyen k ö l c s ö n ö s e n e g y é r t e l m ü módon. Legyen a P szabályainak száma $m \geq 1$ és legyenek a szabályok nevei $X_1, \ldots, X_m$ páronként különböző jelek. Tekintsünk egy G -beli baloldali levezetést:

$$\alpha_0 \Rightarrow u_1\alpha_1 \Rightarrow \ldots \Rightarrow u_n\alpha_n$$

ahol $\alpha_0 = S$ és $n \geq 1$. A baloldali levezetés definiciójának megfelelően minden $1 \leq i \leq n$ indexre $u_i \in T^*$, $\alpha_i \in V^*$ valamint minden $0 \leq i \leq n-1$ indexre $\alpha_i = A_i\beta_i$, $\alpha_{i+1} = \gamma_i\beta_i$ és $A_i \to \gamma_i \in P$. A levezetés m i n d e n l é p é s é h e z rendeljük hozzá az ezen lépésben f e l h a s z n á l t s z a b á l y n e v é t, a l e v e z e t é s h e z magához pedig a szabályok neveiből a levezetés lépéseinek megfelelő sorrendben alkotott l á n c o t. Tehát, ha X_{j_i} -vel jelöljük az $A_i \to \gamma_i$ szabály nevét, akkor a fenti levezetéshez hozzárendelt lánc a következő:

$$X_{j_0} \ldots X_{j_{n-1}}$$

ahol $1 \leq j_0, \ldots, j_{n-1} \leq m$.

Szoritkozzunk azokra a baloldali levezetésekre, amelyek az $L(G)$ nyelv egy m o n d a t á t adják /vagyis $\alpha_n = \varepsilon$/. Az ezekhez hozzárendelt $\{X_1, \ldots, X_m\}^*$ halmazbeli láncok

összességét a G grammatikához rendelt **b a l o l d a l i l e v e z e t é s - n y e l v n e k** nevezzük.

3.1.1. példa:

$$G=(\{S,A\},\{a,b\},P,S)$$

$P:\ x: S\to aSb \qquad v: A\to a$

$y: S\to A \qquad r: S\to \varepsilon$

$z: A\to bAa \qquad q: S\to a$

$S\Rightarrow a$

q

$S\Rightarrow aSb\Rightarrow aaSbb\Rightarrow aaAbb\Rightarrow aaabb$

$x\ x\ y\ v$

$S\Rightarrow A\Rightarrow a$

$y\ v$

$S\Rightarrow \varepsilon$

r

Különböző baloldali levezetésekhez különböző hozzárendelt láncok tartoznak. Ez könnyen belátható, hiszen két ilyen levezetés csak akkor lehet különböző, ha az egyik több lépésből áll mint a másik /ez esetben a hozzárendelt láncok hossza különböző/ vagy az egyik levezetés valamely lépésben más szabályt alkalmaztunk mint a másik levezetés ugyanannyiadik lépésében /ez esetben a szabályok neveinek egyedi volta miatt lesz a két jellánc különböző/. Igy a hozzárendelés, mivel a hozzárendelési eljárás egyértelmü, kölcsönösen egyértelmü a levezetések és a láncok között. Mivel minden baloldali levezetéshez egyértelmüen hozzátartozik a nyelv egy mondata, a fentiek és a CF-grammatikák egyértelmüsége definiciójának

alapján igaz a következő: A fent leirt hozzárendelési eljárás minden G CF-grammatika esetén, egyben a hozzárendelt láncok halmazának az $L(G)$ nyelvre való leképezését is definiálja és ez a leképezés akkor és csak akkor kölcsönösen egyértelmü, ha a G grammatika egyértelmü.

A 3.1.1. példa grammatikája nem egyértelmü ennek megfelelően két lánc, az yv és a q képe a.

Minden G CF-grammatikához egyszerü módon megadható egy olyan G' CF-grammatika, amely pontosan a G baloldali levezetéseihez rendelt láncok halmazát generálja:

Legyen a $G=(N,T,P,S)$ CF-grammatika. Definiáljuk a V^* halmaz N^* halmazba való f leképezését a következőképpen:

a, $f(a)=\varepsilon$ ha $a\in T\cup\{\varepsilon\}$

b, $f(A)=A$ ha $A\in N$

c, $f(\alpha)=f(y_1)\dots f(y_k)$ ha $\alpha\in V^*$ és $\alpha=y_1\dots y_k$

A G grammatikában a T halmazt helyettesitsük az $X=\{x_1,\dots,x_m\}$ halmazzal és a P minden $A\to\alpha$ szabályát az $A\to x\,f(\alpha)$ szabállyal, ahol az $x\in X$ az $A\to\alpha$ szabály cimkéje. Tehát a

$G'=(N,X,P',S)$ ahol $P'=\{A\to x f(\alpha)\mid A\to f(\alpha)\in P\}$

3.1.2. példa: $G'=(\{S,A\},\{x,y,z,v,q,r\},P',S)$

P': $S\to xS$

$S\to yA$

$A\to zA$

$A\to v$

$S\to q$

$S\to r$

A fenti példa a 3.1.1. példa grammatikájának megfelelő grammatikát adja meg.

A fenti, általános definicióban megadott G' grammatika rendelkezik a következő tulajdonságokkal:

a, standard grammatika

b, a szabályok és a terminális jelek közt kölcsönösen egyértelmü megfeleltetés van.

Az is könnyen belátható, hogy minden ilyen tipusu G' grammatikához van /végtelen sok/ olyan G grammatika, amelyből kiindulva a fenti konstrukció a G' grammatikát eredményezi /ilyen maga a G' is/.

Tehát a fenti a, és b, tulajdonságok a CF-grammatikák halmazának egy olyan részhalmazát jellemzik, amelynek lényeges matematikai nyelvészeti jelentés is tulajdonitható.

A dolgozat hátralévő részében ezen grammatikákkal és az általuk generált nyelvekkel foglalkozunk, ez a dolgozatunk fő célja.

2.§ Definiciók és alaptulajdonságok

Definició: Legyen a $G=(N,T,P,S)$ egy CF-grammatika. Azt mondjuk, hogy a G egy **baloldali levezetés-grammatika**, ha teljesül a következő két feltétel:

a, A G standard grammatika.

b, A T halmaz minden eleméhez pontosan egy szabály van a P halmazban, amelyben ez az elem szerepel.

Definició: Egy nyelvet **baloldali levezetés-nyelvnek** nevezünk, ha van olyan baloldali levezetés-grammatika, amely ezt a nyelvet generálja.

A "baloldali levezetés" rövidebb jelölésére a következőkben az "LD" röviditést használjuk /leftmost derivation/. Az LD-grammatika definiciójával kapcsolatban megjegyezzük, hogy a benne szereplő két feltétel egyike sem hagyható el. A CF-nyelvek közül az LD-nyelveket csak a két feltétel együtt tudja kiválasztani. Az a, egymagában hatástalan megszoritás a generált nyelvre vonatkozóan /2.23./. Ha a grammatikánk a b, feltételt teljesiti, standard alakra hozva általában nem lesz LD-grammatika, mert elveszitheti a b, tulajdonságát [7].

3.2.1. példa:

$$G_1=(\{S,B\},\{a,b,c\},P,S)$$

$$G_2=(\{S\},\{a,b,c\},P_2,S)$$

$$G_3=(\{S,C,D\},\{a,b,c,d,e\},P_3,S)$$

$$G_4=(\{S,A,B\},\{a,b,c,d,e\},P_4,S)$$

$$P_1=\{S\to aSB,\ B\to b,\ S\to c\}$$

$$P_2=\{S\to aS,\ S\to bS,\ S\to c\}$$

$$P_3=\{S\to aSC,\ S\to bSD,\ S\to e,$$

$$C\to c,\ D\to d\}$$

$P_4 = \{ S \to aS,\ S \to bA,\ A \to cB,\ B \to dA,\ B \to e \}$

$L_1 = L(G_1) = \{ a^n c b^n \mid n \geq 0 \}$

$L_2 = L(G_2) = \{ \mu c \mid \mu \in \{a,b\}^* \}$

$L_3 = L(G_3) = \{ \mu c \nu \mid \mu = f(\nu^R),\ \nu \in \{c,d\}^* \}$

$L_4 = \{ a^n b c (dc)^m e \mid n, m \geq 0 \}$

ahol az f a következő leképezés:

$f(e) = \varepsilon$, $f(c) = a$, $f(d) = b$

$f(x_1 \dots x_n) = f(x_1) \dots f(x_n)$ $x_i \in \{c,d\}$ $i = 1, \dots, n$

3.2.1. lemma: Legyen G_1 LD-grammatika és $L(G_1) \neq \emptyset$. Ha a G_2 a G_1 egyszerüsitésével kapott grammatika, akkor G_2 is LD-grammatika.

Bizonyitás: Ez az állitás közvetlenül adódik az egyszerüsitési eljárás definiciójából /II.2.§/, ugyanis egyszerüsitéssel csak elhagyhatunk a kategóriák, a terminális jelek és a szabályok halmazából. Ha egy terminális jelet elhagyunk, a hozzátartozó szabály is feleslegessé válik, illetve ha egy szabályt elhagyunk, akkor a hozzátartozó terminális jel is felesleges lesz.

A továbbiakban, ha csak mást nem mondunk, külön feltevés nélkül ugy tekintünk egy G LD-grammatikát, hogy az egyszerüsitett és az általa generált nyelv nem üres.

3.2.2. lemma: Ha a $G = (N, T, P, S)$ egy LD-grammatika, akkor $\#N \leq \#T$.

Bizonyitás: Mivel a G egyszerüsitett, minden kategóriának meg kell jelennie legalább egy szabály baloldalán.

Tehát $\#N \leq \#P$. Mivel a G LD-grammatika $\#T = \#P$ tehát $\#N \leq \#T$.

3.2.3. lemma: Minden LD-grammatika S-grammatika.
Bizonyitás: Az állitás közvetlenül következik a definiciókból.

3.2.4. lemma: Az $L_1 = \{(ba)^m a \mid m \geq 0\}$ és $L_2 = \{a^m b^{m+1} \mid m \geq 0\}$ nyelvek S-nyelvek de nem LD-nyelvek.
Bizonyitás: Tekintsük a $G = (\{A, S\}, \{a, b\}, P, S)$ grammatikát, ahol $P = \{S \to bA, A \to aS, S \to a\}$. Ez S-grammatika és $L(G) = L_1$.
Tegyük fel, hogy van olyan G_1 LD-grammatika, hogy az $L_1 = L(G_1)$. A G_1 -nek kettő szabálya és a 3.2.2. lemma szerint legfeljebb kettő kategóriája van. Egyik az axióma, jelölje ezt X . Mivel $a \in L_1$ és $baa \in L_1$ mindkét szabály baloldalán az axiomának kell állni, tehát csak egy kategória van, az X . Mivel $a \in L_1$ az egyik szabály az $X \to a$ szabály, a másik pedig, mivel a $baa \in L_1$, az $X \to bXX$ szabály. De akkor a $bbaaa \in L(G_1)$. Tehát ellentmondásra jutottunk, a G_1 nem lehet LD-grammatika. Az L_2 nyelvre hasonlóan bizonyítható az állitás.
A 3.2.3. és a 3.2.4. lemmákból következik:
3.2.5. tétel: Az LD-nyelvek halmaza az S-nyelvek halmazának valódi részhalmaza.

Az LD-nyelvek és a dolgozatban tárgyalt nyelvosztályok viszonyát a tartalmazás és közös részek szempontjából a Függelék 1. tárgyalja.

Felhasználva az S-nyelvekre vonatkozó ismereteinket /2.7.1. és 2.7.4./ adódik:

3.2.6. tétel: Minden LD-grammatika egyértelmü.
3.2.7. lemma: Minden LD-nyelv rendelkezik a prefix tulajdonsággal.

A következő lemmák az LD-grammatikák néhány fontos tulajdonságát adják meg.

3.2.8. lemma: Legyen a $G=(N,T,P,S)$ egy LD-grammatika, $A,B\in N$ és $A\overset{*}{\underset{G}{\Rightarrow}} u\alpha$, $B\overset{*}{\underset{G}{\Rightarrow}} v\beta$, ahol $u,v\in T^+$ és $\alpha,\beta\in N^*$. Ha $u=v$ akkor $A=B$ és $\alpha=\beta$.

Bizonyitás: Legyen az $u=v=a_1\dots a_k$ ahol $k\geq 1$ és $a_i\in T$ minden $1\leq i\leq k$ indexre. Tekintsük a két baloldali levezetést:

$$A=\alpha_0\Rightarrow a_1\alpha_1\Rightarrow\dots a_1\dots a_{i-1}\alpha_{i-1}\Rightarrow a_1\dots a_i\alpha_i\Rightarrow\dots a_1\dots a_k\alpha_k=u\alpha$$

$$B=\beta_0\Rightarrow a_1\beta_1\Rightarrow\dots a_1\dots a_{i-1}\beta_{i-1}\Rightarrow a_1\dots a_i\beta_i\Rightarrow\dots a_1\dots a_k\beta_k=v\beta$$

A levezetések első lépésében alkalmazott szabály $A\to a_1\alpha_1$ illetve $B\to a_1\beta_1$. Mivel a G LD-grammatika, $A=B$ és $\alpha_1=\beta_1$. Tegyük fel, hogy $\alpha\neq\beta$ és legyen i_0 a legkisebb olyan index, amelyre az $\alpha_{i_0}\neq\beta_{i_0}$. Az $1<i_0\leq k$ és $\alpha_{i_0-1}=\beta_{i_0-1}$. Mivel $0<i_0-1<k$, az $\alpha_{i_0-1}=\beta_{i_0-1}\neq\varepsilon$ tehát van olyan, $C\in N$ hogy az $\alpha_{i_0-1}=\beta_{i_0-1}=C\delta$ és $C\to a_{i_0}\delta\in P$. Mivel a G LD-grammatika, mindkét levezetésénél ezt a szabályt kell alkalmaznunk az i_0. lépésben, tehát $\alpha_{i_0}=\beta_{i_0}$. Ellentmondásra jutottunk, amely bizonyitja az állitás igazságát.

Megjegyezzük, hogy mint a bizonyitásból kiderül, ha feltesszük az állitás első felét ($A=B$) akkor az $\alpha=\beta$ állitás akkor is bebizonyitható, ha a G -ről csak azt tesszük fel, hogy S-grammatika. A teljes állitás nem teljesül S-grammatikákra. Pl. a $G=(\{S,A\},\{a,b\},\{S\to aSA, S\to b, A\to b\},S)$ grammatika esetén $S\overset{*}{\Rightarrow} b$, $A\overset{*}{\Rightarrow} b$.

3.2.9. lemma: Legyen a $G=(N,T,P,S)$ egy LD-grammatika, $\alpha\in N^+$ és $u,v\in T^*$. Ha $\alpha\overset{*}{\underset{G}{\Rightarrow}} u$ és $\alpha\overset{*}{\underset{G}{\Rightarrow}} uv$, akkor $v=\varepsilon$.

Bizonyitás: Legyen az $\alpha=A_1\dots A_k$ az α teljes felbontása. Tekintsük az $\alpha\overset{*}{\Rightarrow} u$ baloldali levezetést. Ennek

alapján megadhatók azok az $u_1, \ldots, u_k$ mondatok, amelyekre igaz, hogy az $u = u_1 \ldots u_k$ és minden $1 \leq i \leq k$ indexre $A_i \overset{*}{\Rightarrow} u_i$. Ugyanigy kapjuk a $v_1, \ldots, v_k$ mondatokat, amelyekre igaz, hogy az $uv = v_1 \ldots v_k$ és minden $1 \leq i \leq k$ indexre $A_i \overset{*}{\Rightarrow} v_i$. Azt állitjuk, hogy $u_i = v_i$ minden $1 \leq i \leq k$ indexre. Tegyük fel az ellenkezőjét, és legyen az $1 \leq i_0 \leq k$ a legkisebb olyan index, amelyre $u_{i_0} \neq v_{i_0}$. Akkor $u = t u_{i_0} \ldots u_k$, $uv = t v_{i_0} \ldots v_k$ és $A_{i_0} \ldots A_k \overset{*}{\Rightarrow} u_{i_0} \ldots u_k$, $A_{i_0} \ldots A_k \overset{*}{\Rightarrow} v_{i_0} \ldots v_k$. Két eset lehetséges:

a, $u_{i_0} = v_{i_0} \gamma$ ahol $\gamma \neq \varepsilon$

b, $v_{i_0} = u_{i_0} \rho$ ahol $\rho \neq \varepsilon$

Elegendő az egyik esetet vizsgálni. Nézzük az a, esetet. Az $A_{i_0} \overset{*}{\Rightarrow} u_{i_0}$ és az $A_{i_0} \overset{*}{\Rightarrow} u_{i_0}\beta$, $\beta \overset{*}{\Rightarrow} \gamma$. A 3.2.8. lemma szerint $\beta = \varepsilon$. Ez ellentmond annak, hogy $\gamma \neq \varepsilon$. Tehát minden $1 \leq i \leq k$ indexre $u_i = v_i$ következésképpen $u = uv$ és igy $v = \varepsilon$.

Megjegyezzük, hogy ez a lemma igaz S-grammatikák esetén is.

3.2.10. lemma: Legyen a $G = (N, T, P, S)$ egy LD-grammatika, $\alpha, \beta \in N^+$ és $u \in T^+$. Ha $\alpha\beta \overset{*}{\underset{G}{\Rightarrow}} u$ akkor pontosan egy olyan $u = u_1 u_2$ felbontása van az u mondatnak, amelyre teljesül, hogy $\alpha \overset{*}{\underset{G}{\Rightarrow}} u_1$ és $\beta \overset{*}{\underset{G}{\Rightarrow}} u_2$.

Bizonyitás: Legyenek a v_1 és v_2 olyan mondatok, amelyek teljesitik a lemma állitását. Mivel $u = u_1 u_2 = v_1 v_2$, két eset lehetséges:

a, $u_1 = v_1 \gamma$

b, $v_1 = u_1 \rho$

Elegendő csak az egyik esetet vizsgálni, nézzük az a, esetet: $\alpha \overset{*}{\Rightarrow} v_1$ és $\alpha \overset{*}{\Rightarrow} v_1 \gamma$ A 3.2.9. lemma szerint $\gamma = \varepsilon$ tehát $u_1 = v_1$ és $u_2 = v_2$

3.2.11. lemma: Legyen a $G = (N, T, P, S)$ egy LD-grammatika, $\alpha, \beta, \gamma \in N^*$ és $u \in T^+$. Ha az $\alpha \overset{*}{\underset{G}{\Rightarrow}} u$ és $\beta \overset{*}{\underset{G}{\Rightarrow}} u\gamma$, akkor a $\beta = \alpha\gamma$.

Bizonyitás: Legyen a v egy olyan mondat, amely a γ -ból

levezethető. A feltételekből következően az $\alpha, \beta \in N^+$. Legyen a $\delta \in N^*$ az α és β láncok leghosszabb közös kezdőszelete. Mivel az $u \neq \varepsilon$, a $\delta \neq \varepsilon$. Tegyük fel, hogy a lemma állitásával ellentétben a $\delta \neq \alpha$. Akkor van olyan $\gamma_1 \in N^+$, hogy az $\alpha = \delta\gamma_1$. Legyen az $u = u_1u_2$ ugy hogy $\delta \overset{*}{\Rightarrow} u_1$ és $\gamma_1 \overset{*}{\Rightarrow} u_2$ /3.2.10/. Két eset lehetséges:

a, $\beta = \delta$. Ez esetben a $\delta \overset{*}{\Rightarrow} u_1$ és a $\delta \overset{*}{\Rightarrow} u_1\gamma_2$ teljesül, ahol a $\gamma_2 \overset{*}{\Rightarrow} u_2\gamma \overset{*}{\Rightarrow} u_2v$, tehát $\delta \overset{*}{\Rightarrow} u_1u_2v$. A 3.2.9. lemma szerint az $u_2v = \varepsilon$, ellentmondásban a $\gamma_1 \in N^+$ feltétellel.

b, $\beta = \delta\gamma_3$ és $\gamma_3 \in N^+$. Legyen a $\gamma_1 = A\gamma_4$ és $\gamma_3 = B\gamma_5$, ahol az $A, B \in N$. Mivel $A\gamma_4 \overset{*}{\Rightarrow} u_2$ és $B\gamma_5 \overset{*}{\Rightarrow} u_2\gamma$, valamint az $u_2 \neq \varepsilon$, igy az $A = B$. Ez pedig ellentmond annak, hogy a δ a leghosszabb közös kezdőszelet.

Mindkét esetben ellentmondásra jutottunk, ezzel a lemma igazolt.

3.2.12. lemma. Legyen a $G = (N, T, P, S)$ egy LD-grammatika, az $\alpha, \beta \in N^*$ és az $u \in T^+$. Ha az $\alpha \overset{*}{\underset{G}{\Rightarrow}} u$ és a $\beta \overset{*}{\underset{G}{\Rightarrow}} u$, akkor az $\alpha = \beta$.

Bizonyitás: A lemma a 3.2.11. lemma $\gamma = \varepsilon$ esetre adódó következménye.

Definició: Legyen a $G = (N, T, P, S)$ egy LD-grammatika. A G grammatika **sulyfüggvényének** nevezzük és a φ_G -vel jelöljük a V^* halmaznak az egész számok halmazába való következő leképezését:

a, $\varphi_G(\varepsilon) = 0$

b, $\varphi_G(A) = -1$ ha $A \in N$

c, $\varphi_G(a) = |\alpha| - 1$ ha $a \in T$ és $A \to a\alpha \in P$

d, $\varphi_G(x_1 \dots x_k) = \varphi_G(x_1) + \dots + \varphi_G(x_k)$ ha az $x_i \in V$ minden $1 \leq i \leq k$ indexre.

Ha magától értetődik, hogy melyik grammatikáról van szó, akkor a jelölésből a „G" elhagyható.

A sulyfüggvény fogalma nagyon jól használható az LD-nyelvek vizsgálatánál. A következőkben néhány, erre vonatkozó állitást bizonyitunk be.

3.2.13. lemma: Legyen a $G = (N, T, P, S)$ egy LD-grammatika és φ_G a sulyfüggvénye, valamint $\alpha, \beta \in V^*$ és $u \in T^*$.

a, Ha $\alpha \overset{*}{\underset{G}{\Rightarrow}} \beta$, akkor $\varphi_G(\beta) = \varphi_G(\alpha)$

b, Ha $u \in L(G)$, akkor a $\varphi_G(u) = -1$ és bármely $u = u_1 u_2$ felbontás esetén, ha az $u_2 \neq \varepsilon$, akkor a $\varphi_G(u_1) \geq 0$.

Bizonyitás: Először az a, állitást bizonyitjuk. Nyilvánvalóan elegendő bizonyitani, hogy egylépéses levezetés esetén igaz az állitás, vagyis ha $\alpha \Rightarrow \beta$, akkor $\varphi(\beta) = \varphi(\alpha)$. Tegyük fel, hogy $\alpha \Rightarrow \beta$. Akkor van olyan $\gamma_1, \gamma_2, \delta \in V^*$ és $A \in N$ valamint $a \in T$, hogy az $\alpha = \gamma_1 A \gamma_2 \Rightarrow \gamma_1 a \delta \gamma_2 = \beta . A \varphi$ definiciójából, mivel a $\delta \in N^*$, következik:

$$\varphi(\beta) = \varphi(\gamma_1) + \varphi(a) + \varphi(\delta) + \varphi(\gamma_2) = \varphi(\gamma_1 \gamma_2) + |\delta| - 1 - |\delta| =$$

$$\varphi(\gamma_1 \gamma_2) - 1 = \varphi(\gamma_1) + \varphi(A) + \varphi(\gamma_2) = \varphi(\alpha)$$

Áttérünk a b, állitás bizonyitására. Legyen az $u = u_1 u_2$. Tekintsük az $S \overset{*}{\Rightarrow} u_1 \alpha$ levezetést, ahol az $\alpha \in N^*$. Az a, állitásból adódik, hogy a $\varphi(u_1 \alpha) = \varphi(S) = -1$. Ha az $u_2 = \varepsilon$ akkor az $\alpha = \varepsilon$ tehát $u = u_1$ és $\varphi(u) = -1$. Ha az $u_2 \neq \varepsilon$, akkor mivel $\alpha \overset{*}{\Rightarrow} u_2$, az $\alpha \neq \varepsilon$, az $|\alpha| \geq 1$. Mivel a $\varphi(u_1 \alpha) = \varphi(u_1) - |\alpha| = -1$, igy a $\varphi(u_1) \geq 0$.

Ha csak egy kategória van, akkor a sulyfüggvény egyértelmüen meghatározza a nyelvet. Ezt mondja ki a következő tétel.

3.2.14. tétel: Legyen a $G = (N, T, P, S)$ egy LD-grammatika. Ha az $N = \{S\}$ akkor bármely $u \in T^*$ mondat esetén igaz a következő: Az $u \in L(G)$ akkor és csak akkor teljesül, ha a $\varphi_G(u) = -1$, és bármely $u = u_1 u_2$, $u_2 \neq \varepsilon$ felbontás esetén a $\varphi_G(u_1) \geq 0$.

Bizonyitás: Az állitás egyik felét a 3.2.13. lemma bizonyit-

ja, csak a másik felét kell bizonyitanunk. Tegyük fel, hogy az $u \in T^*$ mondatra teljesülnek a feltételek. Nyilvánvalóan az $u \neq \varepsilon$. Legyen az $u = a_1 \dots a_k$ az u teljes felbontása. Legyen i_0 a legnagyobb index, amelyre létezik az $S \overset{*}{\Rightarrow} a_1 \dots a_{i_0} \alpha_{i_0}$ baloldali levezetés. Mivel van olyan $\alpha \in N^*$, hogy az $S \to a_1 \alpha \in P$, az $i_0 \geq 1$. Tegyük fel, hogy $i_0 < k$ és legyen az $u_1 = a_1 \dots a_{i_0}$ az $u_2 = a_{i_0+1} \dots a_k$. Mivel az $u_2 \neq \varepsilon$, a $\varphi(u_1) \geq 0$, tehát az $S \overset{*}{\Rightarrow} a_1 \dots a_{i_0} \alpha_{i_0}$ levezetés alapján felirt $\varphi(u_1 \alpha_{i_0}) = \varphi(u_1) + \varphi(\alpha_{i_0}) = -1$ egyenlőség miatt a $\varphi(\alpha_{i_0}) \leq -1$. Ebből viszont az $\alpha_{i_0} \neq \varepsilon$ következik, tehát az $\alpha_{i_0} = S\beta$. Ez esetben viszont a levezetés folytatható, mert van olyan $\gamma \in N^*$, hogy az $S \to a_{i_0+1} \gamma \in P$. Tehát az $i_0 < k$ nem teljesülhet. Igy az $i_0 = k$ és az $S \overset{*}{\Rightarrow} a_1 \dots a_k \alpha_k = u \alpha_k$. A $\varphi(u) = -1$, valamint a $\varphi(u \alpha_k) = -1$ miatt a $\varphi(\alpha_k) = 0$, tehát az $\alpha_k = \varepsilon$ vagyis az $u \in L(G)$.

A következő állitás az LD-nyelvek egy nagyon jellemző tulajdonságát mondja ki.

3.2.15. lemma: Legyen a $T_0 = \{a_1, \dots, a_n\}$ egy véges ábécé, az $N_0 = \{A_1, \dots, A_n\}$ kategóriák egy véges halmaza, és az $L_0 = \{u_1, \dots, u_m\} \subset T^*$ egy olyan véges nyelv, amelyre teljesül, hogy minden $1 \leq i \leq n$ indexre az $|u_i|_{a_i} > 0$.
Megadható egy LD-grammatikákból álló olyan véges H halmaz, amely tartalmaz minden olyan $G = (N, T, P, S)$ LD-grammatikát, amely rendelkezik a következő két tulajdonsággal:
1, $N \subseteq N_0$ és $T = T_0$
2, $L_0 \subseteq L(G)$

Bizonyitás: Legyen a $G = (N, T, P, S)$ egy olyan LD-grammatika, amely teljesiti a két feltételt. Legyen a φ a G sulyfüggvénye és vezessük be a következő jelöléseket:

$$d_{ij} = |u_i|_{a_j} \qquad 1 \leq i, j \leq n$$

$$d_k = d_{1k} + \dots + d_{nk} \qquad 1 \leq k \leq n$$

$d = \min(d_1, \ldots, d_n)$ $\quad c = d_1 + \ldots + d_n$

$x_k = \varphi(a_k) + 2 \quad 1 \leq k \leq n$

Az L_0 definiciójából következik, hogy $d_{ij} \geq 0$ és $d_{ii} > 0$ ha $1 \leq i, j \leq n$ és igy $d_k > 0$ minden $1 \leq k \leq n$ indexre, tehát a $d > 0$ és a $c > 0$. Felhasználva azt, hogy az $L_0 \subseteq L(G)$ kapjuk:

$$\varphi(u_i) = d_{i_1} \varphi(a_1) + \ldots + d_{i_n} \varphi(a_n) = -1 \qquad 1 \leq i \leq n$$

Az egyenleteket összeadva:

$$d_1 \varphi(a_1) + \ldots + d_n \varphi(a_n) = -n$$

Ebből:

$$d_1 x_1 + \ldots + d_n x_n = 2c - n$$

Leosztva a d -vel, és felhasználva azt, hogy az $x_1, \ldots, x_n \geq 1$ kapjuk:

$$x_1 + \ldots + x_n \leq (2c - n) / d$$

$$\max(x_1, \ldots, x_n) \leq (2c - n) / d < 2c/d$$

Jelöljük a $2c/d$ mennyiséget K -val. Mivel minden $1 \leq i \leq n$ indexre az x_i éppen az a_i terminális jelhez tartozó szabály jobboldalának a hosszát adja meg, azt kaptuk, hogy a G minden szabálya jobboldalának a hossza kisebb a K értékénél. A K érték csak az L_0 nyelvtől függ. Olyan $G' = (N', T', P', S')$ LD-grammatika, amelyre igaz, hogy $N' \subseteq N_0$, $T' = T_0$ és minden szabály jobboldala rövidebb mint a K szám, nyilvánvalóan csak véges sok van $(\#H \leq n!\, n^{nK})$ és ezek a grammatikák egyszerü módon felsorolhatók. Tehát a halmaz megadható.

A következő tétel az LD-nyelvek és a CF-nyelvek közt bizonyos kapcsolatot ad meg.

3.2.16. tétel: Legyen a G egy tetszőleges CF-grammatika. Megadható olyan f homomorf leképezés, és G' LD-grammatika, hogy az $L(G) = f(L(G'))$ teljesül.

Bizonyítás: Legyen a $G=(N, T, P, S)$ egy CF-grammatika. A 2.2.3. tétel szerint megadható egy olyan $G_1=(N_1, T, P_1, S_1)$ standard CF-grammatika, amelyre az $L(G_1)=L(G)$ teljesül. Tekintsük ezt a G_1 grammatikát. Számozzuk meg valamilyen módon a szabályokat. Ha a $\#P_1 = M$, akkor legyen a $T'=\{a_1, \ldots, a_M\}$ ahol az $a_1, \ldots, a_M$ páronként különböző /egyébként tetszőleges/ szimbólumok. Legyen az f a következő egyenlőségekkel definiált homomorf leképezés:

a, $f(\varepsilon) = \varepsilon$

b, $f(a_i) = a$ ha $a \in T$ és az a a P_1 i-edik szabályában szerepel.

c, $f(A) = A$ ha $A \in N_1$

A P_1 minden szabályában cseréljük ki a terminális jelet a T' halmaz megfelelő elemére, vagyis ha az $A \to a\alpha$ a P_1 i-edik eleme, akkor ezt cseréljük ki az $A \to a_i\alpha$ szabályra. Az így kapott szabályhalmazt nevezzük P'-nek és tekintsük a $G'=(N_1, T', P', S_1)$ grammatikát. A konstrukcióból következik, hogy a G' LD-grammatika, és a 2.5.7. tétel alkalmazva az is, hogy $f(L(G')) = L(G_1) = L(G)$.

3.2.2. példa: $G = (\{S, A\}, \{a, b, c\}, P, S)$

$G_1 = (\{S, A, B, C\}, \{a, b, c\}, P_1, S)$

$G' = (\{S, A, B, C\}, \{a, b, c, d, e, x, y\}, P', S)$

P: $S \to SS$, $S \to aASb$, $A \to abc$, $S \to c$

P_1: $S \to cS$, $S \to aASBS$, $A \to aBC$, $B \to b$, $C \to c$, $S \to c$

P': $S \to a\,S$ $\qquad$ $B \to e$

$S \to b\,A\,S\,B\,S$ $\qquad$ $C \to x$

$S \to c\,A\,S\,B$ $\qquad$ $S \to y$

$A \to d\,B\,C$

$f(a) = f(x) = f(y) = c$

$f(b) = f(c) = f(d) = a$

$f(e) = b$

$f(L(G')) = L(G_1) = L(G)$

3.§ A baloldali levezetés nyelvek és a veremautomaták

Bár az állitásainkat és bizonyitásainkat, ettől a paragrafustól eltekintve, a grammatikák nyelvén fogalmazzuk meg, a teljesség kedvéért megadjuk az LD-nyelvek felismerő jellegü leirását szolgáltató veremautomatatipust.

Definició: Legyen az $M=(\{q_0\}, T, V, \delta, q_0, Z_0, \{q_0\})$ egy S-automata. Azt mondjuk, hogy az M egy LD-a u t o m a t a /röviden LDA/, ha bármely $a \in T$ esetén pontosan egy olyan $Z \in V$ jel van, amelyre a $\delta(q_0, a, Z) \neq \emptyset$.

Definició: Legyen az $M=(\{q_0\}, T, V, \delta, q_0, Z_0, \{q_0\})$ egy LDA. Azt mondjuk, hogy az M f e l i s m e r i az $u \in T^*$ m o n d a t o t, ha az u mondat átviszi az automatát a (q_0, Z_0) konfigurációból a (q_0, ε) konfigurációba. Jelöléssel:

$$u : (q_0, Z_0) \overset{*}{\underset{M}{\vdash}} (q_0, \varepsilon)$$

Definició: Legyen az $M=(\{q_0\}, T, V, \delta, q_0, Z_0, \{q_0\})$ egy LDA. Az M által m e g h a t á r o z o t t n y e l v n e k nevezzük és $T(M)$ -el jelöljük a T^* azon részhalmazát, amely pontosan az M által felismert mondatokból áll. Jelöléssel:

$$T(M) = \{u \in T^* \mid u : (q_0, Z_0) \overset{*}{\underset{M}{\vdash}} (q_0, \varepsilon)\}$$

3.3.1. tétel: Minden G LD-grammatikához van olyan M LDA, hogy az $L(G) = T(M)$ és minden M LD-automatához van olyan G LD-grammatika, hogy a $T(M) = L(G)$.

Bizonyitás: Jelöljeˆ H_1 az LD-automaták halmazának azt a részhalmazát, amely pontosan azokból az LD-automatákból áll, amelyeknél a belső állapotot a „q_0" szimbólum jelöli. A H_1 halmaz és az LD-grammatikák halmaza közt kölcsönösen egyértelmü leképezés létesithető a következő módon:

Legyen a $G=(N, T_1, P, S)$ LD-grammatika és az $M=(\{q_0\}, T_2, V, \delta, q_0, z_0, \{q_0\})$ egy LDA. A G és az M akkor és csak akkor feleljenek meg egymásnak, ha teljesülnek a következők:

a, $T_1=T_2$ és $V=N$ valamint $S=z_0$

b, Minden $a\in T_1$, $A\in N$ és $\alpha\in N^*$ esetén $\delta(q_0,a,A)=(q_0,\alpha)$ akkor és csak akkor teljesül, ha $A\to a\alpha\in P$.

Az LD-grammatikák és az LD-automaták definiciója miatt a megfeleltetés kölcsönösen egyértelmü. Állitjuk, hogy ha a G LD-grammatika és az M LDA ebben a leképezésben megfelelnek egymásnak, akkor az $L(G)=T(M)$.

1. Legyen az $u\in L(G)$ és az $u=a_1\dots a_k$ az u teljes felbontása. Tekintsük az u G-beli baloldali levezetését.

$$S=\alpha_0\Rightarrow a_1\alpha_1\Rightarrow\dots a_1\dots a_{i-1}\alpha_{i-1}\Rightarrow a_1\dots a_i\alpha_i\Rightarrow\dots a_1\dots a_k\alpha_k=u$$

Állitjuk, hogy minden $1\leq i\leq k$ indexre:

$$a_1\dots a_i : (q_0,S)\vdash^* (q_0,\alpha_i)$$

Tegyük fel, hogy ez nem igaz és legyen az i_0 a legkisebb olyan index, amelyre a fenti állitás nem igaz. Mivel az $S\to a_1\alpha_1\in P$ igy a $\delta(q_0,a_1,S)=(q_0,\alpha_1)$ tehát az $i_0>1$. Tehát $a_1\dots a_{i_0-1}:(q_0,S)\vdash^*(q_0,\alpha_{i_0-1})$. Mivel az $i_0-1<k$ az $\alpha_{i_0-1}\neq\varepsilon$, van olyan $A\in N$ és $\gamma\in N^*$, hogy az $\alpha_{i_0-1}=A\gamma$. A fenti baloldali levezetés miatt van olyan $\gamma_1\in N^*$, hogy az $A\to a_{i_0}\gamma_1\in P$ és $\alpha_{i_0}=\gamma_1\gamma$, tehát $\delta(q_0,a_{i_0},A)=(q_0,\gamma_1)$ és igy $a_{i_0}:(q_0,\alpha_{i_0-1})\vdash(q_0,\gamma_1\gamma)=(q_0,\alpha_{i_0})$, ennélfogva $a_1\dots a_{i_0}:(q_0,S)\vdash^*(q_0,\alpha_{i_0})$ ellentmondásban az indirekt feltevéssel. Ekkor viszont $a_1\dots a_k:(q_0,S)\vdash^*(q_0,\alpha_k)=(q_0,\varepsilon)$, tehát az $a_1\dots a_k=u\in T(M)$

2, Legyen a $v\in T(M)$ és a $v=b_1\dots b_l$ a v teljes felbontása. Minden $1\leq i\leq l$ indexre:

$$b_1\dots b_i : (q_0,S)\vdash^*(q_0,\beta_i)$$

és a $\beta_i=\varepsilon$ akkor és csak akkor teljesül, ha $i=l$. Legyen az i_0 a legnagyobb olyan index, amelyre az

$S \overset{*}{\underset{G}{\Rightarrow}} b_1 \dots b_{i_0} \beta_{i_0}$

baloldali levezetés létezik. Mivel $\delta(q_0, b_1, S) = (q_0, \beta_1)$, $S \to b_1 \beta_1 \in P$, tehát az $i_0 \geq 1$. Tegyük fel, hogy az $i_0 < l$. Tehát $S \overset{*}{\Rightarrow} b_1 \dots b_{i_0} \beta_{i_0}$. Mivel a $\beta_{i_0} \neq \varepsilon$ van olyan $B \in N$ és $\varrho \in N^*$, hogy a $\beta_{i_0} = B\varrho$. Mivel $b_{i_0+1} : (q_0, \beta_{i_0}) \vdash (q_0, \beta_{i_0+1})$, van olyan $\varrho_1 \in N^*$ hogy $\delta(q_0, b_{i_0+1}, B) = (q_0, \varrho_1)$ és $\beta_{i_0+1} = \varrho_1 \varrho$. De akkor a $B \to b_{i_0+1} \varrho_1 \in P$ és igy $S \overset{*}{\Rightarrow} b_1 \dots b_{i_0} \beta_{i_0} \Rightarrow b_1 \dots b_{i_0+1} \varrho_1 \varrho =$ $= b_1 \dots b_{i_0+1} \beta_{i_0+1}$ ellentmondásban az i_0 -ra tett feltevéssel. Tehát az $i_0 = l$, és $S \overset{*}{\Rightarrow} b_1 \dots b_l \beta_l = v$ vagyis $v \in L(G)$.

Ezzel bebizonyitottuk, hogy $T(M) = L(G)$. Ebből viszont már következik a tétel, hiszen nyilvánvaló, hogy az LD-automaták belső állapotának jelölésére tett megkötés lényegtelen, a H_1 halmaz elemei által meghatározott nyelvek halmaza ugyanaz mint az összes LD-automaták által meghatározott nyelvek halmaza.

3.3.1. példa: $G = (\{S, A, B, C\}, \{a, b, c, d, e\}, P, S)$

$P: S \to aASBS$

$B \to bAA$

$A \to cSC$

$C \to dS$

$S \to e$

$M = [\{q_0\}, \{a, b, c, d, e\}, \{S, A, B, C\}, \delta, q_0, S, \{q_0\}]$

$\delta(q_0, a, S) = (q_0, ASBS)$

$\delta(q_0, b, B) = (q_0, AA)$

$\delta(q_0, c, A) = (q_0, SC)$

$\delta(q_0, d, C) = (q_0, S)$

$\delta(q_0, e, S) = (q_0, \varepsilon)$

Minden más esetben a leképezés eredménye az üres halmaz.

$L(G) = T(M)$.

4.§ Müveletek

Ebben a paragrafusban az LD-nyelvek halmazának az előző fejezetekben bevezetett müveletekre való zártságát vizsgáljuk. Bizonyos esetekben megadjuk azt a legszükebb nyelvosztályt is /a CS, CF, DCF és LD-nyelveket figyelembevéve/, amelyből már nem vezet ki a müvelet LD-nyelveken való elvégzése.

3.4.1. lemma: A $G_1=(N, T\cup\{d\}, P_1, S)$ és a $G_2=(N, T\cup\{e\}, P_2, S)$ CF-grammatika, ahol $T=\{a,b,c\}$, $N=\{S\}$, $P_1=\{S\to aS, S\to bSS, S\to c, S\to d\}$ és $P_2=\{S\to aSS, S\to b, S\to cS, S\to e\}$

Az $L(G_1)\cup L(G_2)$ nyelv nem DCF-nyelv.

Bizonyitás: Tegyük fel, hogy az $L(G_1)\cup L(G_2)$ nyelv DCF nyelv. Akkor az $R=\{a^ib^jc^kd \mid i,j,k\geq 1\}\cup\{a^ib^jc^ke \mid i,j,k\geq 1\}$ reguláris nyelvvel való metszete is DCF-nyelv /2.5.4./. Legyen az $L'=\{a^ib^ic^jd \mid i,j\geq 1\}\cup\{a^ib^ic^je \mid i,j\geq 1\}$. Ha belátjuk, hogy $L\cap R=L'$ akkor készen vagyunk, mert az L' nyelv nem DCF-nyelv [5].

1. Legyen az $u\in L\cap R$. Mivel az $u\in R$, vannak olyan $i_0, j_0, k_0\geq 1$ számok, hogy az $u=a^{i_0}b^{j_0}c^{k_0}x$, ahol az $x=d$ vagy $x=e$. Tegyük fel először, hogy $x=d$. Mivel $u\in L$, igy $u\in L(G_1)$. De akkor $\varphi_{G_1}(u)=-1$. Részletesen kiirva: $\varphi_{G_1}(u)=i_0\cdot 0+j_0\cdot 1+k_0\cdot(-1)+(-1)=-1$. Ebből $j_0=k_0$ tehát $u\in L'$. Legyen az $x=e$. Ez esetben $u\in L(G_2)$, tehát $\varphi_{G_2}(u)=i_0\cdot 1+j_0(-1)+k_0\cdot 0+(-1)=-1$, tehát $i_0=j_0$ vagyis $u\in L'$.
2. Legyen a $v\in L'$. Mivel az $L'\subset R$ triviálisan teljesül, elegendő belátni, hogy $v\in L$. Két eset van:
   a, $v=a^{i_0}b^{j_0}c^{j_0}d$ ahol $i_0, j_0\geq 1$.
   A $\varphi_{G_1}(v)=j_0-j_0-1=-1$. Legyen a $v_2\neq\varepsilon$ és $v=v_1v_2$. Ha a v_1 nem tartalmaz c jelet, akkor nyilván $\varphi_{G_1}(v_1)>0$. Tegyük fel, hogy a $v_1=a^{i_0}b^{j_0}c^{k_0}$. Mivel a $v_2\neq\varepsilon$, $k_0\leq j_0$. Ebből viszont $\varphi_{G_1}(v_1)=j_0-k_0\geq 0$.

A 3.2.14. tétel szerint a $v \in L(G_1)$, tehát a $v \in L$.

b, $v = a^{i_0} b^{i_0} c^{i_0} e$, ahol $i_0, j_0 \geq 1$.

A $\varphi_{G_2}(v) = i_0 - i_0 - 1 = -1$. Legyen a $v_2 \neq \varepsilon$ és $v = v_1 v_2$. Ha a v_1 tartalmaz c jelet, akkor $\varphi_{G_2}(v_1) = 0$. Ha $v_1 = a^{i_0} b^{k_0}$, ahol $i_0 \geq k_0$, akkor $\varphi_{G_2}(v_1) = i_0 - k_0 \geq 0$. A 3.2.14. tétel szerint $v \in L(G_2)$, tehát $v \in L$

Ezzel a tételt bebizonyitottuk.

3.4.2. lemma: A $G_1 = (N, T, P_1, S)$ és a $G_2 = (N, T, P_2, S)$ LD grammatikák, ahol $N = \{S\}$, $T = \{a, b, c, d\}$,
$P_1 = \{S \to aSS, S \to bS, S \to c, S \to d\}$
$P_2 = \{S \to aS, S \to bSS, S \to c, S \to d\}$
Az $L(G_1) \cap L(G_2)$ nyelv nem CF-nyelv.

Bizonyitás: Legyen az $L = L(G_1) \cap L(G_2)$ és $R = \{a^i b^j c^k d \mid i, j, k \geq 1\}$ valamint $H = \{a^n b^n c^n d \mid n \geq 1\}$. Azt állitjuk, hogy $L \cap R = H$.

1, Legyen $u \in H$. Mivel $H \subset R$, elegendő belátni, hogy $u \in L$. Van olyan $n_0 \geq 1$ egész szám, hogy $u = a^{n_0} b^{n_0} c^{n_0} d$. A $\varphi_{G_1}(u) = n_0 - n_0 - 1 = -1$, a $\varphi_{G_2}(u) = \varphi_{G_1}(u) = n_0 - n_0 - 1 = -1$. Legyen az $u_2 \neq \varepsilon$ és $u = u_1 u_2$. Ha az u_2 nem tartalmaz c jelet, akkor $\varphi_{G_1}(u) \geq 0$, és a $\varphi_{G_2}(u) \geq 0$ triviálisan teljesül. Legyen $u_1 = a^{n_0} b^{n_0} c^{k_0}$. Mivel az $u_2 \neq \varepsilon$, igy a $k_0 \leq n_0$. De akkor $\varphi_{G_1}(u_1) = \varphi_{G_2}(u_1) = n_0 - k_0 \geq 0$. Tehát $u \in L(G_1)$ és $u \in L(G_2)$,igy $u \in L$. /3.2.14./.

2. Legyen az $u \in L \cap R$. Mivel $u \in L$, vannak olyan $i_0, j_0, k_0 \geq 1$ egész számok, hogy $u = a^{i_0} b^{j_0} c^{k_0} d$. Mivel $u \in L$, $\varphi_{G_1}(u) = \varphi_{G_2}(u) = -1$,
$\varphi_{G_1}(u) = i_0 - k_0 - 1 = \varphi_{G_2}(u) = j_0 - k_0 - 1 = -1$.
Ebből $i_0 = j_0 = k_0$, tehát $u \in H$.

Ezzel bebizonyitottuk, hogy az $L \cap R = H$. Ebből már következik a lemma állitása. Ugyanis, ha az L CF-nyelv, akkor az $L \cap R$ nyelv is CF-nyelv és ennek az $f(a) = a$, $f(c) = c$, $f(b) = b$ és $f(d) = \varepsilon$ leképezéssel adódó homomorf képe, az $\{a^n b^n c^n \mid n \geq 1\}$ halmaz is CF-nyelv. Erről viszont tud-

juk, hogy nem CF-nyelv /2.2.6./.

3.4.3. lemma: A $G=(N,T,P,S)$ LD-grammatika, ahol $N=\{S\}$, $T=\{a,b,c,d\}$ és $P=\{S\to aSS, S\to b, S\to c, S\to d\}$ Az $R=\{a^ib^jd \mid i,j\geq 1\}\cup\{a^icjd \mid i,j\geq 1\}$ reguláris nyelv. Az $L(G)\cap R$ nem S-nyelv.

Bizonyitás: Legyen a $H=\{a^nb^nd \mid n\geq 1\}\cup\{a^mc^md \mid m\geq 1\}$. Állitjuk, hogy az $L(G)\cap R=H$.

1, Legyen az $u\in H$. Mivel $H\subset R$ elegendő belátni, hogy az $u\in L(G)$. Az $u=a^{i_0}x^{i_0}d$ ahol az $i_0\geq 1$ és az $x=b$ vagy az $x=c$. Mindkét esetben $\varphi(u)=i_0-i_0-1=-1$. Legyen az $u_2\neq\varepsilon$ és $u=u_1u_2$. Ha az u_1 nem tartalmaz b vagy c jelet, akkor nyilvánvalóan $\varphi(u_1)>0$. Ha az $u_1=a^{i_0}x^{k_0}$ akkor mivel $u_2\neq\varepsilon$, $i_0\geq k_0$ tehát a $\varphi(u_1)=i_0-k_0\geq 0$ következésképpen $u\in L(G)$ /3.2.14./.

2. Legyen az $u\in L(G)\cap R$. Mivel az $u\in R$, vannak olyan $i_0,j_0\geq 1$ egész számok, hogy $u=a^{i_0}x^{j_0}d$, ahol az $x=b$ vagy $x=c$. Mivel az $u\in L(G)$, $\varphi(u)=i_0-j_0-1=-1$, tehát $i_0=j_0$, ebből pedig $u\in H$. Ezzel a lemmát bebizonyitottuk, ugyanis a H nem S-nyelv [12].

3.4.4. lemma: A $G=(N,T,P,S)$ LD-grammatika, ahol $N=\{S,C,D\}$, $T=\{a,b,c,d,e\}$ és $P=\{S\to aSC, S\to bSD, C\to c, D\to d, S\to e\}$. Van olyan f homomorf leképezés, amelyre az $f(L(G))$ nem DCF-nyelv.

Bizonyitás: Legyen az f a következőképpen definiálva:

$f(e)=f(\varepsilon)=\varepsilon$, $f(a)=f(c)=a$, $f(b)=f(d)=b$.

Az $f(L(G))=\{u\,u^R \mid u\in\{a,b\}^*\}$. Ez pedig nem DCF-nyelv [10].

3.4.5. lemma: A $G=(N,T,P,S)$ LD-grammatika, ahol $N=\{S\}$, $T=\{a,b\}$ és $P=\{S\to aS, S\to b\}$.

a, Az $(L(G))^R$ nyelv nem S-nyelv

b, Az $L(G)L(G)$ nyelv nem LD-nyelv.

Bizonyitás: Az $L=L(G)=\{a^nb \mid n\geq 0\}$ ennek megfelelően az $L^R=\{ba^n \mid n\geq 0\}$. Az L^R nyelv nem rendelkezik a prefix

tulajdonsággal, tehát nem S-nyelv /2.7.4./.
Az $L_1 = L(G_1)L(G_1) = \{a^u b a^m b \mid u, m \geq 0\}$. Tegyük fel, hogy van olyan $G_1 = (N_1, \{a, b\}, P_1, S_1)$ LD-grammatika, amelyre $L(G_1) = L_1$. Mivel $bab \in L_1$ és $abb \in L_1$ vannak olyan $\alpha_1, \alpha_2 \in N_1^*$ láncok, hogy az $S_1 \to a\alpha_1, S_1 \to b\alpha_2 \in P$. Mivel a G_1 LD-grammatika, a P_1-nek nincs több szabálya. Az $L_1 \neq \emptyset$ miatt $\alpha_1 = \varepsilon$ vagy $\alpha_2 = \varepsilon$. Egyik esetben $a \in L_1$ a másik esetben $b \in L_1$. Mindkettő ellentmond az L_1 definiciójának. Tehát az L_1 nem LD-nyelv.

A következő tételekben szereplő "legszükebb nyelvosztály" kifejezés ugy értendő, hogy a következő nyelvosztályok között a legszükebb: CS, CF, DCF, S, LD.

3.4.6. tétel: Az egyesités müvelete kivezet az LD-nyelvek halmazából. A legszükebb nyelvosztály, amely bármely két LD-nyelv egyesitését tartalmazza, a CF-nyelvek halmaza.

Bizonyitás: A tétel a 3.4.1. lemmából és abból következik, hogy bármely két CF-nyelv egyesitése CF-nyelv. /2.5.2./.

3.4.7. tétel: A metszetképzés müvelete kivezet az LD-nyelvek halmazából és a legszükebb nyelvosztály, amely bármely két LD-nyelv metszetét tartalmazza, a CS-nyelvek halmaza.

Bizonyitás: A tétel a 3.4.2. lemmából és abból következik, hogy a CS-nyelvek halmaza a metszetképzésre zárt /2.5.1./.

3.4.8. tétel: A reguláris nyelvvel való metszetképzés müvelete kivezet a LD-nyelvek halmazából és a legszükebb nyelvosztály, amely bármely LD-nyelv és bármely reguláris nyelv metszetét tartalmazza, a DCF-nyelvek halmaza.

Bizonyitás: A tétel a 3.4.3. lemmából és abból következik, hogy a DCF-nyelvek halmaza zárt erre a müveletre nézve /2.5.4./.

3.4.9. tétel: A homomorf leképezés müvelete kivezet az LD-nyelvek halmazából és a legszükebb nyelvosztály, amely bármely LD-nyelv bármely homomorf képét tartalmazza, a CF-nyelvek halmaza.

Bizonyitás: A tétel a 3.4.4. lemmából és abból következik, hogy a CF-nyelvek halmaza zárt erre a müveletre nézve /2.5.2./.

3.4.10. tétel: Ha az L nyelv LD-nyelv, akkor az $\overline{L}$ nyelv nem LD-nyelv és nem is S-nyelv, de DCF-nyelv.

Bizonyitás: A tétel abból következik, hogy egyetlen S-nyelv komplementere sem S-nyelv /2.7.8./, de minden DCF-nyelv komplementere DCF-nyelv /2.5.4./.

3.4.11. tétel: A szorzatképzés müvelete kivezet az LD-nyelvek halmazából és a legszükebb olyan nyelvosztály, amely bármely két LD-nyelv szorzatát tartalmazza, az S-nyelvek halmaza.

Bizonyitás: A tétel a 3.4.5. lemmából és abból következik, hogy az S-nyelvek halmaza a szorzatképzésre zárt /2.7.7./.

3.4.12. tétel: A megforditás müvelete kivezet az LD-nyelvek halmazából.

Bizonyitás: A tétel a 3.4.5. lemma következménye.

5.§ Megoldhatóság

Ebben a paragrafusban a megoldhatósági problémák közül vizsgálunk meg néhányat. Bizonyos problémák általános megoldhatóságát az LD-nyelvek halmaza "készen kapja" az őt tartalmazó CF, DCF és S-nyelvosztályoktól, ezekre nem térünk ki külön. Ilyen pl. az ekvivalencia probléma általános megoldhatósága.

5.1. Néhány általánosan nem megoldható probléma

Először két olyan problémát vizsgálunk meg, amelyek az S-nyelvek halmazán nem voltak általánosan megoldhatók /2.7.12. és 2.7.11./ és mint látni fogjuk az LD-nyelvekre szoritkozva is megmarad ez a tulajdonságuk.

3.5.1.1. lemma: Legyen a T véges ábécé és a $c \notin T$ egy további jel. Tekintsük a $G=(N,T,P,S)$ CF-grammatikát, ahol $N=\{S\}$, $P=\{S\to aS \mid a\in T\}\cup\{S\to c\}$. A G LD-grammatika és az $L(G)=T^*c$.

Bizonyitás: Az állitás a G definiciójából közvetlenül következik.

3.5.1.2. tétel: Nincs olyan algoritmus, amely bármely CF-grammatika esetén eldöntené, hogy a grammatika által generált nyelv LD-nyelv-e vagy sem.

Bizonyitás: Legyen $X=\{x_1,\dots,x_k\}$, $Y=\{y_1,\dots,y_k\}$, ahol a $k\geq 1$ és minden $1\leq i\leq k$ indexre $x_i, y_i \in \{a,b\}^+$. Rendeljük hozzá az (X,Y) párhoz a $G_{XY}=(N,T,P,S)$ grammatikát, ahol az $N=\{S,A,B\}$, $T=\{a,b\}\cup\{d_1,\dots,d_k\}\cup\{c\}$. Az $a, b, d_1, \dots, d_k, c$ szimbólumok legyenek páronként különbözők. A P halmaz elemei legyenek a következő szabályok:

$$S \to AcB$$

$$A \to d_i A x_i^R$$

$$A \to d_i x_i^R \qquad \text{minden } 1\leq i\leq k \text{ indexre}$$

$$\left.\begin{array}{l} B \to y_i B d_i \\ B \to y_i d_i \end{array}\right\} \quad \text{minden } 1 \leq i \leq k \quad \text{indexre}$$

A G_{XY} nyilván CF-grammatika, és az általa generált nyelv a következő

$$L_{XY} = \{d_{i_1} \dots d_{i_u} x_{i_u}^R \dots x_{i_1}^R c\, y_{i_1} \dots y_{j_m} d_{j_m} \dots d_{j_1} \mid u, m \geq 1\},$$

ahol $i_1, \dots, i_u, j_1, \dots, j_m \in \{1, \dots, k\}$. Tekintsük a következő CF-grammatikát: $G_1 = (N_1, T_1, P_1, S_1)$, ahol az $N_1 = \{S_1\}$ a $T_1 = T$, a P_1 szabályai a következőek:

$$S \to a S a$$

$$S \to b S b$$

$$S \to c$$

$$S \to d_i S d_i \qquad \text{minden } 1 \leq i \leq k \text{ indexre.}$$

A G_1 definíciójából közvetlenül adódik, hogy az általa generált nyelv:

$$L_{XY}^{(1)} = \{u c u^R \mid u \in T^*\}$$

Legyen az $L_{XY}^{(2)} = L_{XY} \cap L_{XY}^{(1)}$ és az $L_{XY}^{(3)} = \overline{L_{XY}^{(2)}} d$ ahol a $d \notin T$ egy uj jel /$\overline{\ }$ komplementálás a T^* -ra vonatkozik/. A szóbanforgó nyelvek definícióiból közvetlenül adódik, hogy az $L_{XY}^{(2)}$ akkor és csak akkor nem üres, ha van olyan $i_1, \dots, i_\ell$ $\ell \geq 1$ indexsorozat, amelyre az $x_{i_1} \dots x_{i_\ell} = y_{i_1} \dots y_{i_\ell}$ teljesül. Másrészt, ha az $L_{XY}^{(2)}$ üres, akkor az $L_{XY}^{(3)} = T^* d$ tehát LD-nyelv /3.5.11./, ha pedig az $L_{XY}^{(2)}$ nem üres, akkor nem CF-nyelv [13], tehát az $L_{XY}^{(3)}$ nem LD-nyelv. /Ha az $L_{XY}^{(3)}$ LD-nyelv, akkor DCF-nyelv is, viszont akkor $\overline{L_{XY}^{(2)}}$ DCF-nyelv /2.3.3./, következésképpen az $L_{XY}^{(2)}$ is DCF /2.5.4./, tehát CF-nyelv lenne./ Tehát az

$L^{(2)}_{XY}$ akkor és csak akkor üres, ha az $L^{(3)}_{XY}$ LD-nyelv. Az $L^{(3)}_{XY}$ nyelv CF-nyelv [13], jelöljük az ezt a nyelvet generáló CF-grammatikát G'_{XY} -al.

Tegyük fel, hogy a tétel állitásával ellentétben van olyan algoritmus, amely bármely G CF-grammatika esetén eldönti, hogy az $L(G)$ nyelv LD nyelv-e vagy sem. Akkor ez az algoritmus bármely (X,Y) pár esetén eldönti, hogy az $L(G'_{XY}) = L^{(3)}_{XY}$ nyelv LD-nyelv-e vagy sem, vagy ami ezzel ekvivalens, hogy az $L^{(2)}_{XY}$ üres-e vagy sem. Ez viszont azt jelenti, hogy ez az algoritmus bármely konkrét (X,Y) pár esetén eldönti a Post-féle korrespondencia problémát, tehát a PCP általánosan megoldható. Ezzel ellentmondásra jutottunk, ami bizonyitja a tételt.

3.5.1.3. tétel: Nincs olyan algoritmus, amely bármely G_1 és G_2 LD grammatikák esetén eldöntené, hogy az $L(G_1) \cap L(G_2)$ nyelv üres-e vagy sem, vagy véges-e vagy sem.

Bizonyitás: Legyen $X = \{x_1, \dots, x_k\}, Y = \{y_1, \dots, y_k\}$, ahol $k \geq 1$ és minden $1 \leq i \leq k$ indexre $x_i, y_i \in \{a,b\}^+$. Legyen az A és B két különböző szimbólum és az f az $\{a,b\}^*$ halmaznak az $\{A,B\}^*$ halmazra való, a következő egyenlőségekkel definiált homomorf leképezése:

a, $f(\varepsilon) = \varepsilon$

b, $f(a) = A$ és $f(b) = B$

Tekintsük a következő grammatikákat: $G'_{XY} = (N,T,P_1,S)$, $G''_{XY} = (N,T,P_2,S)$ ahol az $N = \{S,A,B\}$, $T = \{a,b\} \cup \{d_1, \dots, d_k\} \cup \{c_1, \dots, c_k\}$. Az $a, b, d_1, \dots, d_k, c_1, \dots, c_k$ legyenek páronként különböző jelek.

$$P_1: \left.\begin{array}{l} S \to d_i S\,(f(x_i))^R \\ S \to c_i\,(f(x_i))^R \end{array}\right\} \text{ minden } 1 \leq i \leq k \text{ indexre.}$$

$$A \to a$$

$$B \to b$$

$$P_2 : \left.\begin{array}{l} S \to d_i S\,(f(y_i))^R \\ S \to c_i\,(f(y_i))^R \end{array}\right\} \text{ minden } 1 \leqq i \leqq k \text{ indexre}$$

$$A \to a \;,\; B \to b$$

A definiciók alapján mindkét grammatika LD-grammatika és

$$L'_{XY} = L(G'_{XY}) = \{ d_{i_1} \ldots d_{i_m} c_\ell x_\ell^R x_{i_m}^R \ldots x_{i_1}^R \mid m \geqq 0,$$
$$i_1, \ldots, i_m, \ell \in \{1, \ldots, k\}\}$$

$$L''_{XY} = L(G''_{XY}) = \{ d_{i_1} \ldots d_{i_m} c_\ell y_\ell^R y_{i_m}^R \ldots y_{i_1}^R \mid m \geqq 0,$$
$$i_1, \ldots, i_m, \ell \in \{1, \ldots, k\}\}$$

Legyen az $L_{XY} = L'_{XY} \cap L''_{XY}$ és $u \in L_{XY}$. Akkor van olyan $i_1, \ldots, i_r, \ell$ indexsorozat, hogy $u = d_{i_1} \ldots d_{i_r} c_\ell v$ ahol $v = x_\ell^R x_{i_r}^R \ldots x_{i_1}^R = y_\ell^R y_{i_r}^R \ldots y_{i_1}^R$. Képezve a megforditást: $v^R = x_{i_1} \ldots x_{i_r} x_\ell = y_{i_1} \ldots y_{i_r} y_\ell$. A levezetés megfordítható, tehát, ha van olyan $j_1, \ldots, j_m, m \geqq 1$ indexsorozat, hogy $x_{j_1} \ldots x_{j_m} = y_{j_1} \ldots y_{j_m}$, akkor bevezetve a $v_1 = x_{j_1} \ldots x_{j_m}$, $v_2 = y_{j_1} \ldots y_{j_m}$ és $\ell = j_m$ jelöléseket, azt kapjuk, hogy az $u_1 = d_{j_1} \ldots d_{j_{m-1}} c_\ell v_1^R \in L'_{XY}$ az $u_2 = d_{j_1} \ldots d_{j_{m-1}} c_\ell v_2^R \in L''_{XY}$ és az $u_1 = u_2 \in L_{XY}$. Tehát ha lenne olyan algoritmus, amely minden (X, Y) pár esetén eldönti, hogy az $L_{XY} = \emptyset$ teljesül-e vagy sem, akkor a Post-féle korrespondencia probléma általánosan megoldható lenne. Ezzel az első állitást bebizonyitottuk.

A második állitás bebizonyitásához elegendő megmutatni, hogy az L_{XY} nyelv akkor és csak akkor végtelen, ha nem üres. Nyilvánvalóan, hogyha végtelen, akkor nem üres. Tegyük fel, hogy az L_{XY} nem üres, legyen az $u = d_{i_1} \ldots d_{i_r} c_\ell v \in L_{XY}$. Akkor $S \overset{*}{\underset{G'_{XY}}{\Rightarrow}} d_{i_1} \ldots d_{i_r} d_\ell S v$ és $S \overset{*}{\underset{G''_{XY}}{\Rightarrow}} d_{i_1} \ldots d_{i_r} d_\ell S v$ tehát minden $n \geqq 1$ természetes számra az $u_n = (d_{i_1} \ldots d_{i_r} d_\ell)^n d_{i_1} \ldots d_{i_r} c_\ell v v^n$ mondat eleme az L'_{XY} és az L''_{XY} nyelveknek, tehát az L_{XY} nyelvnek is. Mivel az $u \neq \varepsilon$ a $v \neq \varepsilon$, tehát az L_{XY} nyelv végtelen, ha nem üres.

A következő probléma általános meg nem oldhatósága szorosan összefügg az LD-nyelvek azon tulajdonságával, hogy bármely CF-nyelv előállitható ugy mint egy LD-nyelv homomorf képe.

3.5.1.4. tétel: Ha a H halmaz a CF-nyelvek halmazának olyan részhalmaza, amelyre teljesül a következő két feltétel:

a, nincs olyan algoritmus, amely tetszőleges G CF-grammatika esetén eldöntené, hogy az $L(G)$ nyelv eleme-e a H halmaznak vagy sem,

b, a H halmaz tartalmazza az LD-nyelvek halmazát,

akkor nincs olyan algoritmus, amely tetszőleges L, H halmazbeli nyelv és f homomorf leképezés esetén eldöntené, hogy az $f(L)$ nyelv is H halmazbeli-e vagy sem.

Bizonyitás: Van olyan A algoritmus, amely bármely adott G_1 CF-grammatika esetén megad egy olyan L_1 LD-nyelvet és f homomorf leképezést, amelyekre teljesül, hogy $f(L_1)=L(G_1)$ /3.2.16./. Tegyük fel, hogy ellentétben a tétel állitásával van olyan B algoritmus, amely megfelel a tételben leirt követelménynek. Legyen a G egy tetszőleges CF-grammatika, az A algoritmus megad egy olyan L LD-nyelvet és egy olyan f homomorf leképezést, amelyekre igaz, hogy $f(L)=L(G)$. Mivel az L H halmazbeli nyelv, a B algoritmus eldönti, hogy az $f(L)$ is eleme-e a H halmaznak vagy sem. De ezzel azt is eldönti, hogy az $L(G)$ nyelv H halmazbeli nyelv-e vagy sem. Az A és B összekapcsolásával kapott algoritmus tehát tetszőleges G CF-grammatika esetén eldönti, hogy az $L(G) \in H$ teljesül-e vagy sem. Ez viszont ellemtmond az a, feltételnek, tehát a tételt bebizonyitottuk.

Ennek a tételnek közvetlen következményei, felhasználva rendre a 2.6.16., 2.7.12, és a 3.5.1.2. tételeket a következő állitások:

3.5.1.5. tétel: Nincs olyan algoritmus, amely bármely L DCF-nyelv és f homomorf leképezés esetén eldöntené, hogy az $f(L)$ nyelv is DCF-nyelv-e vagy sem.

3.5.1.6. tétel: Nincs olyan algoritmus, amely bármely G S-grammatika és f homomorf leképezés eldöntené, hogy $f(L(G))$ nyelv is S-nyelv-e vagy sem.

3.5.1.7. tétel: Nincs olyan algoritmus, amely bármely G LD-grammatika és f homomorf leképezés esetén eldöntené, hogy az $f(L(G))$ nyelv is LD-nyelv-e vagy sem.

Megjegyezzük, hogy bár a tételekben a homomorf leképezésről nem kötöttünk ki semmit, az állitások akkor is igazak, ha olyan homomorf leképezésekre szoritkozunk, amelyek hossztartó leképezést adnak, vagyis minden u mondatra $|u| = |f(u)|$. Ez a tény következik a 3.2.16. tétel bizonyitásából.

5.2. A szorzatképzés problémája

Tudjuk, hogy a szorzatképzés kivezet az LD nyelvek halmazából. Felhasználva azt a tényt, hogy bármely két S-nyelv szorzata is S-nyelv, valamint azt, hogy az S nyelvek halmazán az ekvivalenciaprobléma megoldható, bármely két LD nyelv szorzatáról el tudjuk dönteni, hogy LD nyelv-e vagy sem. Ennek részletes bizonyitását adjuk meg az alábbiakban.

3.5.2.1. lemma: Legyenek a G_1 és a G_2 LD-grammatikák. Megadható egy olyan, LD-grammatikákból álló véges halmaz, amely tartalmazza az összes olyan G LD-grammatikát /izomorfizmustól eltekintve/, amelyre az $L(G) = L(G_1)L(G_2)$ teljesül.

Bizonyitás: Legyen a $T_1 = \{a_1, \ldots, a_k\}$ a $T_2 = \{b_1, \ldots, b_l\}$ ahol $k, l \geq 1$. Minden $1 \leq i \leq k$ indexre jelölje az u_i az $L(G_1)$ legrövidebb olyan mondatát, amelyben az a_i legalább egyszer előfordul. Ugyanigy minden $1 \leq j \leq l$

indexre jelölje v_j az $L(G_2)$ nyelv legrövidebb olyan mondatát, amelyben a b_j legalább egyszer előfordul. Legyen az $L(G_1)$ nyelv legrövidebb mondata az u , az $L(G_2)$ nyelv legrövidebb mondata a v . Legyen az $n = k+\ell$, a $T_0 = \{a_1, \ldots, a_k, b_1, \ldots, b_\ell\}$, az $L_0 = \{u, v, \ldots, u^k v, uv, \ldots, uv^\ell\}$ és $N_0 = \{A_1, \ldots, A_n\}$. Alkalmazzuk a 3.2.15. lemmát, a kapott grammatikahalmazt jelöljük H -val. Legyen a $G = (N, T, P, S)$ olyan grammatika, amelyre teljesül, hogy $L(G) = L(G_1)L(G_2)$. Közvetlenül adódik, hogy $T = T_0$ ebből viszont az, hogy $\#N \leq n$ tehát a G kategóriáit kölcsönösen egyértelműen átjelölhetjük ugy, hogy az $N' \subseteq N_0$ legyen, ahol N' az N elemeinek átjelölésével keletkezett halmaz. Az átjelöléseket megfelelően végrehajtva a P halmazon is, egy $G' = (N', T, P', S')$ grammatikát kapunk, amely a G izomorf képe, tehát $L(G) = L(G')$. Mivel az $L_0 \subseteq L(G')$ teljesül, azt kapjuk, hogy a $G' \in H$. Tehát a H halmaz eleget tesz a lemma követelményeinek.

Az N_0 halmazt rögzitve a H halmaz csak az n számtól, a T_0 és L_0 halmazoktól függ, ezek viszont a G_1 és G_2 grammatika ismeretében egyértelműen megadhatók /2.6.18. és 2.6.19./. Az N_0 , T_0 és L_0 ismeretében viszont a H is megadható /3.2.15./.

3.5.2.2. tétel: Van olyan algoritmus, amely bármely G_1 és G_2 LD-grammatikák esetén eldönti, hogy az $L(G_1)L(G_2)$ nyelv LD-nyelv-e vagy sem, és ha a nyelv LD-nyelv, akkor megad egy olyan G LD-grammatikát, amely ezt a nyelvet generálja.

Bizonyitás: Tudjuk, hogy az $L = L(G_1)L(G_2)$ nyelv S-nyelv és megadható az a G' S-grammatika, amelyre teljesül, hogy $L(G') = L$ [12]. A 3.5.2.1. lemmát alkalmazva megkapjuk azt a $H = \{G_1, \ldots, G_m\}$ véges grammatikahalmazt, amely tartalmazza az összes olyan LD-grammatikát, amely az L nyelvet generálja. Minden (G', G_i) $i = 1, \ldots, m$ grammatikapár ese-

tén végezzük el az ekvivalenciavizsgálatot /2.7.10./. Ha a H valamely G_{i_0} eleme ekvivalens a G' grammatikával, akkor az L nyelv LD-nyelv és $L = L(G_{i_0})$. Ha a H nem tartalmaz a G' grammatikával ekvivalens elemet, akkor a H halmaz tulajdonságai miatt az L nyelv nem lehet LD-nyelv.

Mint a bizonyitásokból látható, a szorzatképzés problémája ugyan megoldható, a megoldó algoritmus bonyolult és hosszadalmas általában. Bizonyos speciális esetekben azonban egyszerűbben is eldönthető a probléma. Néhány ilyen esetet adunk meg az alábbiakban.

3.5.2.3. tétel: Legyenek a $G_1 = (N_1, T_1, P_1, S_1)$ és a $G_2 = (N_2, T_2, P_2, S_2)$ LD grammatikák.

a, Ha az S_1 egyetlen P_1 -beli szabály jobboldalán sem fordul elő és $T_1 \cap T_2 = \emptyset$, akkor megadható olyan G LD-grammatika, hogy $L(G) = L(G_1)\,L(G_2)$.

b, Ha az $L(G_1) \cap L(G_2) \neq \emptyset$ akkor az $L(G_1)\,L(G_2)$ nem LD-nyelv.

Bizonyitás: Feltehető, hogy az N_1, T_1, N_2, T_2 halmazok páronként idegenek. Legyen a $P_1 = P_1' \cup P_1''$ ahol a P_1' pontosan azokból a szabályokból áll, amelyek baloldalán az S_1 axioma van. Tekintsük a következő grammatikát: $G = (N, T, P, S)$ ahol az $N = N_1 \cup N_2$, $T = T_1 \cup T_2$, $S = S_1$ valamint $P = \{ S_1 \to a\alpha S_2 \mid S_1 \to a\alpha \in P_1' \} \cup P_1'' \cup P_2$. A G grammatika nyilván LD grammatika és figyelembevéve a feltételeket, az is könnyen belátható, hogy $L(G) = L(G_1)\,L(G_2)$. Ezzel az a, állitást igazoltuk.

Vezessük be a következő jelölést: $L = L(G_1)\,L(G_2)$. Tegyük fel, hogy van olyan $G = (N, T, P, S)$ LD grammatika, amelyre $L(G) = L$. Legyen $u \in L(G_1) \cap L(G_2)$ akkor az $uu \in L$. Tehát van olyan $\alpha \in N^*$ lánc, hogy $S \overset{*}{\Rightarrow} u\alpha$ és $\alpha \overset{*}{\Rightarrow} u$. Mivel az $u \neq \varepsilon$ alkalmazható a 3.2.11. lemma, eszerint $S = \alpha\alpha$. Ez viszont nyilvánvalóan lehetetlen. Ezzel igazoltuk a b, állitást is.

A fenti tételhez érdemes megjegyezni, hogy az a, feltétel második része /$T_1 \cap T_2 \neq \emptyset$/ nem szükséges ahhoz, hogy a szorzat LD-nyelv legyen. Ezt mutatja a következő példa.

3.5.2.1. példa: $G_1 = (\{S, C, B\}, \{a, b, c, d\}, P_1, S)$

$G_2 = (\{S, B\}, \{b, c, d\}, P_2, S)$

$G_3 = (\{S, C, B\}, \{a, b, c, d\}, P_3, S)$

P_1: $S \to aCB$, $B \to bB$, $C \to c$, $B \to d$

P_2: $S \to bS$, $S \to dB$, $B \to c$

P_3: $S \to aCBBC$, $B \to bB$, $C \to c$, $B \to d$

$$L(G_1) = \{a c b^n d \mid n \geq 0\}$$
$$L(G_2) = \{b^m d c \mid m \geq 0\}$$
$$L(G_3) = \{a c b^n d b^m d c \mid n, m \geq 0\}$$
$$L(G_3) = L(G_1)\,L(G_2)$$

5.3. A tartalmazás és az ekvivalencia problémája

Ismeretes, hogy az S-nyelvek halmazán megoldható az ekvivalencia problémája, emiatt az LD-nyelvek esetén is megoldható ez a probléma. Olyan algoritmus viszont nem ismeretes, amely bármely két G_1 és G_2 S-grammatika esetén eldöntené, hogy az $L(G_1) \subseteq L(G_2)$ teljesül-e vagy sem. Az sincs bizonyitva, hogy ilyen algoritmus nem létezik. A következőkben az LD-nyelvekre vonatkozóan vizsgáljuk meg ezt a problémát és megadunk néhány nem triviális esetet, amelyekben a probléma eldönthető.

Legyenek a $G_1 = (N_1, T_1, P_1, S_1)$ és a $G_2 = (N_2, T_2, P_2, S_2)$ LD-grammatikák. A két grammatika nyilván csak akkor lehet ekvivalens, ha a $T_1 = T_2$ Az $L(G_1) \subset L(G_2)$ is csak akkor teljesülhet, ha $T_1 \subseteq T_2$. Sőt ebben az esetben is elegendő a G_2 azon szabályait figyelembevenni, amelyekben a T_1 -beli terminális jelek szerepelnek, hiszen egy olyan szabály, amelynek a terminális jele a $T_2 - T_1$ halmazba tartozik, nem szerepelhet egyetlen $L(G_1)$ -beli mondat generálásánál sem. A fentiekből következik, hogy a tartalmazás és az

ekvivalencia vizsgálatánál elegendő csak az olyan grammatikákat vizsgálni, amelyeknél ugyanaz a terminális jelkészlet. Az általános eset erre egyszerüen visszavezethető.

Definició: Legyenek a $G_1=(N_1,T,P_1,S_1)$ és a $G_2=(N_2,T,P_2,S_2)$ közös terminális jelkészletü LD-grammatikák. Ha van az N_1 halmaznak az N_2 halmazra való olyan f leképezése, amely teljesiti a következő feltételeket:

a, minden $A\in N_1$ és $B\in N_2$ esetén az $f(A)=B$ akkor és csak akkor teljesül, ha van olyan $a\in T$, $\alpha\in N_1^*$, $\beta\in N_2^*$, hogy az $A\to a\alpha\in P_1$ és $B\to a\beta\in P_2$.

b, $f(S_1)=S_2$

akkor azt mondjuk, hogy az N_1 **t-leképezhető** az N_2 -re, és a fent leirt leképezést az N_1 halmaz N_2 -re való **t-leképezésének** nevezzük.

3.5.3.1. lemma: Legyenek a $G_1=(N_1,T,P_1,S_1)$ és a $G_2=(N_2,T,P_2,S_2)$ LD-grammatikák. Ha az $L(G_1)\subseteq L(G_2)$ akkor az N_1 az N_2 halmazra t-leképezhető.

Bizonyitás: Legyen az A az N_1 halmaz egy tetszőleges eleme. Egy $B\in N_2$ kategóriát akkor és csak akkor rendelünk hozzá az A kategóriához, ha van olyan $a\in T, \alpha'\in N_1^*$ és $\beta'\in N_2^*$, hogy az $A\to a\alpha'\in P_1$ és a $B\to a\beta'\in P_2$. Állitjuk, hogy ez a hozzárendelés leképezés, vagyis az A -hoz rendelt kategóriák halmaza egyelemü. Legyen a B_1 és a B_2 két, az A -hoz rendelt kategória. Akkor van olyan $a_1,a_2\in T$, $\alpha_1,\alpha_2\in N_1^*$, $\beta_1,\beta_2\in N_2^*$, hogy az $A\to a_1\alpha_1\in P_1$ és $B_1\to a_1\beta_1\in P_2$ valamint az $A\to a_2\alpha_2\in P_1$ és a $B_2\to a_2\beta_2\in P_2$. Van olyan $u,v,u_1,u_2\in T^*$ és $\alpha\in N_1^*$ hogy $S_1\overset{*}{\underset{G_1}{\Rightarrow}}uA\alpha$, $\alpha\overset{*}{\underset{G_1}{\Rightarrow}}v$, $\alpha_1\overset{*}{\underset{G_1}{\Rightarrow}}u_1$ és $\alpha_2\overset{*}{\underset{G_1}{\Rightarrow}}u_2$. Tehát $ua_1u_1v\in L(G_1)$ és $ua_2u_2v\in L(G_1)$. Mivel $L(G_1)\subseteq L(G_2)$, ua_1u_1v, $ua_2u_2v\in L(G_2)$. Tehát van olyan $\beta\in N_2^*$ hogy $S_2\overset{*}{\underset{G_2}{\Rightarrow}}u\beta$ és $\beta\overset{*}{\underset{G_2}{\Rightarrow}}a_1u_1v$, $\beta\overset{*}{\underset{G_2}{\Rightarrow}}a_2u_2v$. Mivel a G_2 LD-grammatika, a $\beta=B_1\gamma$

és $\beta = B_2\gamma$. Ebből viszont $B_1 = B_2$. Tehát a definiált hozzárendelés leképezés. Mivel minden N_2 -beli elem előfordul valamely P_2 -beli szabály baloldalán a leképezés az N_2 -re való leképezés. Jelöljük a leképezést f - el. Ha belátjuk, hogy az $f(S_1) = S_2$ akkor igazolva lesz, hogy az f t-leképezés. Legyen az $S_1 \to c\gamma_1 \in P_1$ és $B \to c\gamma_2 \in P_2$ valamint az $u \in T^*$ olyan, hogy $\gamma_1 \overset{*}{\underset{G_1}{\Rightarrow}} u$. Akkor $cu \in L(G_1)$ és a tartalmazás miatt $cu \in L(G_2)$. Ebből viszont a $B = S_2$ következik.

3.5.3.2. lemma: Legyen a $G_2 = (N_2, T, P_2, S_2)$ LD-grammatika a $G_1 = (N_1, T, P_1, S_1)$ LD-grammatika homomorf képe, a leképezést jelölje f . Minden $A \in N_1, u \in T^+$ és $\alpha \in N_1^*$ esetére igaz, hogy ha $A \overset{*}{\underset{G_1}{\Rightarrow}} u\alpha$ akkor $f(A) \overset{*}{\underset{G_2}{\Rightarrow}} u f(\alpha)$.

Bizonyitás: Tekintsük az u G_1 -beli baloldali levezetését:

$$A = \alpha_0 \Rightarrow a_1\alpha_1 \Rightarrow \dots a_1 \dots a_i\alpha_i \Rightarrow a_1 \dots a_{i+1}\alpha_{i+1} \Rightarrow \dots a_1 \dots a_k\alpha_k = u\alpha_k$$

ahol az $\alpha_k = \alpha$ és $\alpha_i \neq \varepsilon$ ha $1 \leq i < k$ .Legyen i_0 a legnagyobb index, amelyre $f(A) \overset{*}{\underset{G_2}{\Rightarrow}} a_1 \dots a_{i_0} f(\alpha_{i_0})$. Mivel az $A \to a_1\alpha_1 \in P_1$ ezért az $f(A) \to a_1 f(\alpha_1) \in P_2$, tehát $i_0 \geq 1$. Tegyük fel, hogy $i_0 < k$. Ha $f(\alpha_{i_0}) = \varepsilon$, akkor $\alpha_{i_0} = \varepsilon = \alpha$, tehát $i_0 = k$ lenne. Tehát van olyan $C \in N_2$ és $\beta \in N_2^*$, hogy $f(\alpha_{i_0}) = C\beta$. Mivel $\alpha_{i_0} \neq \varepsilon$, $\alpha_{i_0} = B\gamma$ ahol $B \in N_1$. Tehát $f(B) = C$. Mivel van olyan $\gamma_1 \in N_1^*$, hogy a $B \to a_{i_0+1}\gamma_1 \in P_1$ és $\alpha_{i_0+1} = \gamma_1\gamma$, igy az $f(B) = C$, $C \to a_{i_0+1} f(\gamma_1) \in P_2$, tehát $f(\alpha_{i_0}) \underset{G_2}{\Rightarrow} a_{i_0+1} f(\gamma_1)\beta = a_{i_0+1} f(\gamma_1) f(\gamma) = a_{i_0+1} f(\alpha_{i_0+1})$. Ellentmondásra jutottunk, hiszen az $i_0+1 > i_0$ és $f(A) \overset{*}{\underset{G_2}{\Rightarrow}} a_1 \dots a_{i_0+1} f(\alpha_{i_0+1})$. Tehát az $i_0 = k$, $f(A) \overset{*}{\underset{G_2}{\Rightarrow}} u f(\alpha)$.

3.5.3.3. lemma: Legyen a $G_1 = (N_1, T, P_1, S_1)$ és a $G_2 = (N_2, T, P_2, S_2)$ LD-grammatikák és az $\# N_1 = \# N_2$. Az $L(G_1) \subseteq L(G_2)$ akkor és csak akkor teljesül, ha ha az $L(G_1) = L(G_2)$ teljesül.

Bizonyitás: Ha az $L(G_1) = L(G_2)$ akkor nyilvánvalóan $L(G_1) \subseteq L(G_2)$. Tegyük fel, hogy $L(G_1) \leq L(G_2)$ és legyen az f az N_1 halmaz N_2 -re való t-leképezése /3.5.3.1./. Mivel az $\#N_1 = \#N_2$ az f szükségképpen kölcsönösen egyértelmü. Tegyük fel, hogy van olyan u mondat, hogy $u \in L(G_2)$ de $u \notin L(G_1)$. Legyen az u_1 az u leghosszabb olyan kezdőszelete, amelyre létezik az $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} u_1\beta$ levezetés, ahol a $\beta \in N_1^*$. Mivel az $u \neq \varepsilon$, van olyan $a' \in T$ és $u' \in T^*$, hogy $u = a'u'$, tehát van olyan $\alpha' \in N_2^*$, hogy $S_2 \to a'\alpha' \in P_2$. De akkor van olyan $\beta' \in N_1^*$ is, hogy az $S_1 \to a'\beta' \in P_1$, hiszen $f(S_1) = S_2$ és az f kölcsönösen egyértelmü t-leképezés. Kaptuk, hogy $|u_1| \geq 1$.

a, Tegyük fel, hogy az $u_1 = u$. Ez esetben a $\beta \neq \varepsilon$/ha $\beta = \varepsilon$, akkor $u \in L(G_1)$/, tehát van olyan $v \in T^+$ mondat, hogy a $\beta \overset{*}{\underset{G_1}{\Rightarrow}} v$. Ekkor viszont az $uv \in L(G_1)$ és $v \neq \varepsilon$. A prefix tulajdonság miatt az $uv \notin L(G_2)$ ellentmondásban az $L(G_1) \subseteq L(G_2)$ feltétellel.

b, Tegyük fel, hogy $u = u_1u_2$ és $u_2 \neq \varepsilon$, $u_2 = au_3$, ahol $a \in T$. Ha a $\beta = \varepsilon$, akkor $u_1 \in L(G_1)$, viszont $u_1 \notin L(G_2)$ /prefix tulajdonság/, ellentétben az $L(G_1) \subseteq L(G_2)$ feltétellel. Ha a $\beta \neq \varepsilon$ akkor legyen $\beta = B\gamma$ és $B \to b\delta \in P_1$. Az u_1 definiciója miatt $a \neq b$. Legyen a $v \in T^*$ olyan mondat, amelyre $\delta\gamma \overset{*}{\underset{G_1}{\Rightarrow}} v$, akkor $u_1bv \in L(G_1)$. Azt állitjuk, hogy $u_1bv \notin L(G_2)$. Mivel $u \in L(G_2)$, van olyan $\alpha \in N_2^+$, hogy $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} u_1K$. Legyen $\alpha = A\alpha_2$, ahol $A \in N_2$. Mivel $u = u_1au_3$, van olyan $\delta_1 \in N_2^*$, hogy $A \to a\delta_1 \in P_2$. Ha $u_1bv \in L(G_2)$, akkor léteznie kell egy $A \to b\delta_2 \in P_2$ alaku szabálynak is. Ekkor viszont $f(B) = A$. Legyen $C \in N_1$ az a kategória, amelyre $C \to a\gamma_1 \in P_1$. Az f definiciója szerint $f(C) = A$. Mivel az f kölcsönösen egyértelmü, a $B = C$. Ebből pedig az következik, hogy $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} u_1B\gamma \underset{G_1}{\Rightarrow} u_1a\gamma_1\gamma$ ellentétben az u_1 definiciójával. Tehát az $u_1bv \notin L(G_2)$, és az $u_1bv \in L(G_1)$. Ellentmondásra jutottunk az $L(G_1) \subseteq L(G_2)$ feltétellel.

Mivel minden esetben ellentmondásra jutottunk, nincs olyan

u mondat, amelyre az $u \in L(G_2)$ és az $u \notin L(G_1)$ teljesülne, tehát $L(G_1) = L(G_2)$.

Definició: Legyen a $G_1 = (N_1, T_1, P_1, S_1)$ és a $G_2 = (N_2, T_2, P_2, S_2)$ két tetszőleges LD-grammatika.
Jelölje E_0 a következőképpen definiált halmazt:
a, Az E_0 elemei LD-grammatikából alkotott rendezett párok
b, A $(G_1, G_2) \in E_0$ akkor és csak akkor áll fenn, ha $T_1 = T_2$ és $\# N_1 \leq \# N_2$.

3.5.3.4. tétel: Legyenek a G_1 és a G_2 LD-grammatikák. Ha a $(G_1, G_2) \in E_0$, akkor eldönthető, hogy az $L(G_1) \subseteq L(G_2)$ teljesül-e vagy sem.

Bizonyitás: Legyen a $(G_1, G_2) \in E_0$. Ha $\# N_1 < \# N_2$ akkor az $L(G_1) \subseteq L(G_2)$ nem állhat fenn, mivel az N_1 nem t-leképezhető az N_2 -re /az N_1 semmilyen módon sem képezhető le az N_2 -re/, viszont a t-leképezhetőség a tartalmazás szükséges feltétele /3.5.3.1./. Ha $\# N_1 = \# N_2$, akkor vizsgáljuk meg, hogy a két grammatika ekvivalens-e. Ha igen, akkor az $L(G_1) \subseteq L(G_2)$ is igaz, ha nem, akkor az $L(G_1) \subseteq L(G_2)$ sem teljesülhet /3.5.3.3./.

Definició: Legyen a $G = (N, T, P, S)$ LD-grammatika, $a \in T$ és $A \to a\alpha \in P$. Bevezetjük a következő jelöléseket:
1, $L_G(a) = \{u \in T^* \mid a\alpha \overset{*}{\underset{G}{\Rightarrow}} u\}$
2, $\mu_G(a)$ jelöli az $L_G(a)$ legrövidebb elemeinek a halmazát.

Definició: H_0 jelöli azt a halmazt, amelyet a következő két tulajdonság határoz meg:
a, A H_0 elemei LD-grammatikákból alkot rendezett párok.
b, Legyenek a $G_1 = (N_1, T_1, P_1, S_1)$ és a $G_2 = (N_2, T_2, P_2, S_2)$ LD-grammatikák. A (G_1, G_2) pár akkor és csak akkor eleme a H_0 halmaznak, ha teljesülnek a következő feltételek:

1, $T_1 = T_2$

2, Ha az $L(G_1) \subseteq L(G_2)$ akkor minden $a \in T$ esetén az $\mu_{G_1}(a) \subseteq L_{G_2}(a)$.

3.5.3.5. tétel: Legyenek a G_1 és a G_2 LD grammatikák. Ha a $(G_1, G_2) \in H_0$, akkor eldönthető, hogy az $L(G_1) \subseteq L(G_2)$ teljesül-e vagy sem. Igaz a következő állitás: Az $L(G_1) \subseteq L(G_2)$ akkor és csak akkor teljesül, ha a G_2 a G_1 homomorf képe.

Bizonyitás: Legyen a $G_1 = (N_1, T, P_1, S_1)$ a $G_2 = (N_2, T, P_2, S_2)$ és a $(G_1, G_2) \in H_0$. Ha a G_2 a G_1 homomorf képe akkor az $L(G_1) \subseteq L(G_2)$ /2.2.4./. Tegyük fel, hogy az $L(G_1) \subseteq L(G_2)$ és vezessük be a következő röviditéseket: Minden $a \in T$ esetén az $\mu(a)$ jelölje az $\mu_{G_1}(a)$ egy elemét és az $L_2(a)$ az $L_{G_2}(a)$ nyelvet. Legyen az f az N_1 N_2-re való t-leképezése /3.5.3.1./. Azt állitjuk, hogy ezzel a leképezéssel a G_2 előáll a G_1 homomorf képeként. Tegyük fel az ellenkezőjét. Ez a feltevés azt jelenti, hogy van olyan $a \in T$, $A \in N_1$ és $B \in N_2$ valamint $\alpha \in N_1^*$ és $\beta \in N_2^*$ hogy $A \to a\alpha \in P_1$, $B \to a\beta \in P_2$ és $f(A) = B$ de $f(\alpha) \neq \beta$.

A következő esetek lehetnek:

a, Ha az $\alpha = \varepsilon$, akkor $\mu(a) = a$. Mivel $f(\alpha) \neq \beta$, a $\beta \neq \varepsilon$, tehát $a \notin L_2(a)$ ellentmondást kaptunk.

b, Ha $\beta = \varepsilon$, akkor $L_2(a) = \{a\}$, és mivel $f(\alpha) \neq \beta$, igy $\alpha \neq \varepsilon$, tehát az $|\mu(a)| > |a|$ ezért $\mu(a) \notin L_2(a)$, ellentmondást kaptunk.

c, Ha $\alpha \neq \varepsilon$ és $\beta \neq \varepsilon$ akkor legyen a $\delta \in N_1^*$ az α leghosszabb olyan kezdőszelete, amelyre $f(\delta)$ a β kezdőszelete. Tehát $\alpha = \delta\alpha_1$ és $\beta = f(\delta)\beta_1$. Ha $\delta = \varepsilon = f(\delta)$, akkor legyen $\alpha_1 = A_1'\alpha_2$, $\beta_1 = B_1'\beta_2$, ahol $A_1' \in N_1$, $B_1' \in N_2$ és $f(A_1') \neq B_1'$. Ha $\mu(a) = a b v_1$, ahol $b \in T$, akkor $A_1' \to b\delta_1 \in P_1$. De mivel $f(A_1') \neq B_1'$, nincs olyan P_2-beli szabály, amelynek a baloldalán a B_1' állna a jobboldala pedig a b-vel kezdődne, tehát $\mu(a) \notin L_2(a)$ ami ellentmondás. Ha a $\delta \neq \varepsilon$ akkor legyen a $\delta = A_1 \ldots A_k$ a δ teljes felbontása. Mivel $a\alpha \overset{*}{\underset{G_1}{\Rightarrow}} \mu(a)$, megadható az $\mu(a)$ következő felbontása: $\mu(a) = a u_1 \ldots u_k v$ ahol $\alpha_1 \overset{*}{\underset{G_1}{\Rightarrow}} v$

és minden $1 \leq i \leq k$ indexre $A_i \overset{*}{\Rightarrow}_{G_1} u_i = a_i v_i$, ahol $a_i \in T$ és az $u(a)$ minimumtulajdonsága miatt $u_i \in u_{a_1}(a_i)$. Vezessük be az $u_i = u(a_i)$, $1 \leq i \leq k$ jelöléseket. Legyen az $f(\sigma) = B_1 \dots B_k$ az $f(\sigma)$ teljes felbontása. Minden $1 \leq i \leq k$ indexre $f(A_i) = B_i$ tehát van olyan $\gamma_i \in N_2^*$, hogy $B_i \to a_i \gamma_i \in P_2$. Mivel minden $1 \leq i \leq k$ indexre $u(a_i) \in L_2(a_i)$, tehát $B_i \overset{*}{\Rightarrow}_{G_2} u(a_i)$. Ebből következően $f(\sigma) \overset{*}{\Rightarrow}_{G_2} a u_1 \dots u_k$. Tehát bevezetve az $u' = a u_1 \dots u_k$ jelölést, kaptuk: $u(a) = u' v$, $a\alpha \overset{*}{\Rightarrow}_{G_1} u' \alpha_1$ és $a\beta \overset{*}{\Rightarrow}_{G_2} u' \beta_1$, ahol az $f(\alpha_1) \neq f(\beta_1)$.

Három eset van:

1, Ha az $\alpha_1 = \varepsilon$, akkor $u(a) = u'$, de mivel $\beta_1 \neq \varepsilon$, az $u' \notin L_2(a)$ ami ellentmondás.

2, Ha a $\beta_1 = \varepsilon$, akkor mivel $u' \in L_2(a)$ és $u(a) = u'v$, ahol a $v \neq \varepsilon$ /mivel $\alpha_1 \neq \varepsilon$ / igy $u(a) = u'v \in L_2(a)$ ami ellentmondás.

3. Ha az $\alpha_1 = A_0 \alpha_3$ és a $\beta_1 = B_0 \beta_3$, ahol $A_0 \in N_1$ és $B_0 \in N_2$ valamint $f(A_0) \neq B_0$, akkor mivel a $v \neq \varepsilon$, van olyan $b \in T$, hogy $v = bv_1$. Ekkor viszont van olyan $\delta_2 \in N_1^*$, amelyre $A_0 \to b\delta_2 \in P_1$. De mivel $f(A_0) \neq B_0$, bármely $C \in T$ esetén, ha $B_0 \to C\varrho \in P_2$, akkor $b \neq C$ tehát $u'bv_1 \notin L_2(a)$, ami szintén ellentmondás.

Mivel minden lehetséges esetben ellentmondásra jutottunk, bebizonyitottuk azt, hogy a G_2 a G_1 homomorf képe. Ezután a döntési eljárás már egyszerü. Két véges halmaz közt csak véges sok leképezés létesithető, ezek a leképezések egyszerü módon felsorolhatók. Ha van olyan leképezés, amelyre nézve a G_2 a G_1 homomorf képe, akkor az $L(G_1) \subseteq L(G_2)$ teljesül, ha nincs ilyen leképezés, akkor nem teljesül. Ezzel a tételt bebizonyitottuk.

Az eddigiekben láttuk, hogy a probléma két speciális esetben, az E_0 és a H_0 halmazok esetén általánosan megoldható. Az E_0 halmaz esetén azon alapszik a módszerünk, hogy a tartalmazási viszony ugyanaz mint az /általánosan megoldható/ ek-

vivalencia viszony. A H_o halmazánál pedig azt használjuk fel, hogy a két nyelv közti tartalmazási viszony ekvivalens a két grammatika közti homomorfizmus meglétével. Az alábbiakban megmutatjuk azt is, hogy ezekkel a módszerekkel tovább nem mehetünk, minden olyan esetet, amelyre ezek alkalmazhatók az E_o illetve a H_o halmaz eleve magábafoglal.

3.5.3.6. lemma: Legyen az E halmaz LD-grammatikákból álló rendezett párok olyan halmaza, amely rendelkezik a következő tulajdonsággal: Minden $(G_1, G_2) \in E$ esetén az $L(G_1) \subseteq L(G_2)$ akkor és csak akkor teljesül, ha $L(G_1) = L(G_2)$. Az E halmaz részhalmaza az E_o halmaznak.

Bizonyitás: Legyenek a $G_1 = (N_1, T_1, P_1, S_1)$ és a $G_2 = (N_2, T_2, P_2, S_2)$ LD grammatikák és a $(G_1, G_2) \in E$. Tegyük fel, hogy $L(G_1) \subseteq L(G_2)$. Ez esetben nyilván $T_1 \subseteq T_2$ és az $\# N_1 \geq \# N_2$. De mivel a $(G_1, G_2) \in E$, a feltevésből az is következik, hogy $L(G_2) \subseteq L(G_1)$, tehát $T_2 \subseteq T_1$ és $\# N_2 \geq \# N_1$ Összevetve kapjuk, hogy a $T_1 = T_2$ és $\# N_1 = \# N_2$. Tehát a $(G_1, G_2) \in E_o$.

3.5.3.7. lemma. Legyen a H halmaz LD-grammatikákból álló rendezett párok olyan halmaza, amely rendelkezik a következő tulajdonsággal: Minden $(G_1, G_2) \in H$ esetén az $L(G_1) \subseteq L(G_2)$ akkor és csak akkor teljesül, ha a G_2 grammatika a G_1 homomorf képe. A H halmaz részhalmaza a H_o halmaznak.

Bizonyitás: Legyenek a $G_1 = (N_1, T_1, P_1, S_1)$ és a $G_2 = (N_2, T_2, P_2, S_2)$ LD grammatikák és a $(G_1, G_2) \in H$. Tegyük fel, hogy $L(G_1) \subseteq L(G_2)$. Mivel a $(G_1, G_2) \in H$ van olyan f homomorf leképezés, amellyel a G_2 a G_1 homomorf képe. Ebből rögtön adódik, hogy $T_1 = T_2 = T$. Legyen $a \in T$ és legyen $u \in u_{G_1}(a)$. Ekkor van olyan $A \in N_1$ és $\alpha \in N_1^*$, hogy $A \to a\alpha \in P_1$, $a\alpha \overset{*}{\underset{G_1}{\Rightarrow}} u$, tehát $A \overset{*}{\underset{G_1}{\Rightarrow}} u$. De akkor a 3.5.3.2. lemma szerint $f(A) \overset{*}{\underset{G_2}{\Rightarrow}} u$, ebből viszont az adódik, hogy van olyan $\beta \in N_2^*$ hogy $f(A) \to a\beta \in P_2$ és $a\beta \overset{*}{\underset{G_2}{\Rightarrow}} u$. Következésképpen $u_{G_1}(a) \subset L_{G_2}(a)$, tehát a $(G_1, G_2) \in H_o$.

Természetesen ahhoz, hogy az említett két módszer segítségével egy megoldó algoritmust definiálhassunk, feltétlenül szükséges az is, hogy bármely két grammatika esetén el tudjuk dönteni, hogy használható-e a módszer vagy sem. Erre vonatkozik a következő állitás.

3.5.3.8. tétel: Legyen a $D_0 = E_0 \cup H_0$. Van olyan algoritmus, amely bármely G_1 és G_2 LD-grammatikák esetén eldönti, hogy a $(G_1, G_2) \in D_0$ teljesül-e vagy sem.

Bizonyitás: Legyenek a $G_1=(N_1, T_1, P_1, S_1)$ és a $G_2=(N_2, T_2, P_2, S_2)$ LD grammatikák. A $T_1=T_2$ feltétel egyszerüen ellenőrizhető. Ha nem teljesül, akkor a $(G_1, G_2) \notin D_0$. Ha teljesül, akkor vizsgáljuk meg, hogy az $\#N_1 \leqq \#N_2$ teljesül-e. Ha igen, akkor a $(G_1, G_2) \in D_0$, ha nem, akkor minden $a \in T_1$ esetén adjuk meg az $\mu_{G_1}(a)$ halmazt és vizsgáljuk meg, hogy ez része-e az $L_{G_2}(a)$ nyelvnek. Az $\mu_{G_1}(a)$ megadható és a tartalmazási viszony teljesülése eldönthető /2.6.1. és 2.6.18./. Ha minden $a \in T$ esetén az $\mu_{G_1}(a) \subset L_{G_2}(a)$, akkor a $(G_1, G_2) \in H_0 \subset D_0$ ellenkező esetben a $(G_1, G_2) \notin D_0$.

Az nyilvánvaló, hogy az $E_0 \neq \emptyset$. A következőkben bebizonyitjuk, hogy a H_0 sem üres és megadjuk egy részhalmazát.

3.5.3.9. lemma: Legyenek a $G_1=(N_1, T, P_1, S_1)$ és a $G_2=(N_2, T, P_2, S_2)$ LD-grammatikák. Ha minden $a \in T$ esetén az $A \to a\alpha \in P_1$ és $B \to a\beta \in P_2$ szabályokra teljesül, hogy az $|\alpha| \leqq |\beta|$ akkor a $(G_1, G_2) \in H_0$.

Bizonyitás: Könnyen belátható, hogy bármely $\gamma, \gamma_1 \in N_1^*$, $\delta, \delta_1 \in N_2^*$ és $u \in T^+$ esetén, ha a $\gamma \overset{*}{\underset{G_1}{\Rightarrow}} u\gamma_1$, $\delta \overset{*}{\underset{G_2}{\Rightarrow}} u\delta_1$ akkor, ha a $|\gamma| = |\delta|$ akkor $|\gamma_1| \leqq |\delta_1|$ és ha a $|\gamma| < |\delta|$ akkor $|\gamma_1| < |\delta_1|$. Legyen $a \in T$, $A \Rightarrow a\alpha \in P_1$, valamint $B \to a\beta \in P_2$. Tegyük fel, hogy $L(G_1) \leqq L(G_2)$. Legyen $u \in \mu_{G_1}(a)$, akkor $A \overset{*}{\underset{G_1}{\Rightarrow}} u$. Legyen $u = u_1 u_2$ és u_1 a leghosszabb olyan kezdőszelete az u mondatnak, amelyre a $B \overset{*}{\underset{G_2}{\Rightarrow}} u_1\beta_1$ levezetés létezik. Mivel $A \to a\alpha \in P_1$ és

$B \to a\beta \in P_2$, igy az $u_1 \neq \varepsilon$. Tegyük fel, hogy $u_2 \neq \varepsilon$. Mivel $A \overset{*}{\underset{G_1}{\Rightarrow}} u_1\alpha_1$ és $\alpha_1 \overset{*}{\underset{G_1}{\Rightarrow}} u_2$ valamint $u_2 \neq \varepsilon$ igy $\alpha_1 \neq \varepsilon$, tehát az $|\alpha_1| \leq |\beta_1|$ miatt $\beta_1 \neq \varepsilon$. Legyen $\alpha_1 = C\alpha_2$ és $\beta_1 = D\beta_2$ ahol $C \in N_1$ és $D \in N_2$. Van olyan $v_1 \in T^*$ és $\alpha_3 \in N_1^*$, hogy $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} v_1 A\alpha_3 \overset{*}{\underset{G_1}{\Rightarrow}} v_1 u_1 C\alpha_2\alpha_3$. De mivel az $L(G_1) \subseteq L(G_2)$ van olyan $\beta_3 \in N_2^*$, hogy $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} v_1 B\beta_3 \overset{*}{\underset{G_2}{\Rightarrow}} v_1 u_1 D\beta_2\beta_3$. Legyen a $C \to b\gamma_2 \in P_1$. Ha nincs olyan $\delta_2 \in N_2^*$ amelyre a $D \to b\delta_2 \in P_2$ teljesülne, akkor ellentmondásra jutunk az $L(G_1) \subseteq L(G_2)$ állitással, ha pedig van olyan δ_2 hogy $D \to b\delta_2 \in P_2$ akkor ellentmondásra jutunk az u_1 leghosszabb voltával. Az $u_2 \neq \varepsilon$ feltevés mindenképpen ellentmondásra vezet, tehát $u_2 = \varepsilon$ vagyis $u_1 = u$. Azt kaptuk, hogy $A \overset{*}{\underset{G_1}{\Rightarrow}} u$ és $B \overset{*}{\underset{G_2}{\Rightarrow}} u\beta_1$. Ha $\beta_1 = \varepsilon$ akkor készen vagyunk. Tegyük fel, hogy $\beta_1 \neq \varepsilon$. Ekkor, mivel $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} v_1 A\alpha_3$ és $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} v_1 B\beta_3$, $|\alpha_3| \leq |\beta_3|$, és mivel $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} v_1 u\alpha_3$ és $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} v_1 u\beta_1\beta_2$. Igy $|\alpha_3| < |\beta_1\beta_3|$. Legyen a $v_2 \in T^*$ és $\alpha_3 \overset{*}{\underset{G_1}{\Rightarrow}} v_2$. Mivel $L(G_1) \subseteq L(G_2)$, $\beta_1\beta_3 \overset{*}{\underset{G_2}{\Rightarrow}} v_2$. Ez viszont ellentmond az $|\alpha_3| < |\beta_1\beta_3|$ állitásnak, tehát $\beta_1 = \varepsilon$, igy az $u \in L_{G_2}(a)$. Ezzel bebizonyitottuk, hogy a $(G_1, G_2) \in H_0$.

Bizonyitás nélkül megjegyezzük még, hogy a [8]-ban definiált B-nyelvek esetére is alkalmazható a módszerünk, ugyanis, ha a G_1 és G_2 B-grammatikák, akkor a (G_1, G_2) pár eleme a H_0 halmaznak.

3.5.3.10. lemma: Legyen a G egy tetszőleges LD-grammatika. Izomorfiától eltekintve, csak véges számu olyan LD-grammatika van, amely a G -vel ekvivalens, és van olyan algoritmus, amely megadja ezeket a grammatikákat.

Bizonyitás: Legyen a $G = (N, T, P, S)$ és $T = \{a_1, \ldots, a_k\}$. Minden $1 \leq i \leq k$ indexre jelölje u_i a legrövidebb olyan mondatot, amely legalább egy a_i jelet tartalmaz. Legyen $L_0 = \{u_1, \ldots, u_k\}$, $T_0 = \{a_1, \ldots, a_k\}$ és $N_0 = \{A_1, \ldots, A_k\}$, ahol az $A_1, \ldots, A_k$ páronként idegen jelek. Alkalmazva a 3.2.15. lemmát, kapunk egy $H = \{G_1, \ldots, G_m\}$ grammatikahalmazt. Le-

gyen a G' LD-grammatika ekvivalens a G grammatikával. Jelöljük át a G' kategóriát kölcsönösen egyértelmü módon az N_0 elemeivel. A kapott grammatikát jelöljük G'' -vel. Mivel a G'' és a G' izomorfok, az $L(G'') = L(G') = L(G)$. Tehát $L_0 \subseteq L(G'')$ igy a H halmaz definiciója szerint $G'' \in H$. Tehát a H tartalmazza izomorfizmustól eltekintve az összes, a G -vel ekvivalens LD grammatikát. Minden $1 \leq j \leq m$ indexre vizsgáljuk meg, hogy a G és a G_j grammatikák ekvivalensek-e. Ezen a módon megkapjuk az összes G -vel ekvivalens H -beli grammatikát.

Definició: Jelölje R_0 a következő tulajdonságokkal megadott halmazt:

a, A halmaz elemei LD-grammatikákból álló rendezett párok.

b, Ha a $(G_1, G_2) \in R_0$ akkor az $L(G_1)$ vagy az $L(G_2)$ reguláris nyelv.

3.5.3.11. tétel: Bármely G_1 és G_2 LD-grammatikák esetén eldönthető, hogy a $(G_1, G_2) \in R_0$ teljesül-e vagy sem. Ha a $(G_1, G_2) \in R_0$, akkor eldönthető, hogy az $L(G_1) \subseteq L(G_2)$ állitás teljesül-e vagy sem.

Bizonyitás: Legyenek a $G_1 = (N_1, T, P_1, S_1)$ és a $G_2 = (N_2, T, P_2, S_2)$ LD-grammatikák. A tétel első állitása közvetlenül adódik a 2.6.17. és a 2.7.6. lemmákból. Az $L(G_1) \subseteq L(G_2)$ állitás ekvivalens a következővel: $\overline{L(G_2)} \cap L(G_1) = \emptyset$. Ha az $L(G_1)$ reguláris, akkor, mivel az $L(G_2)$ DCF-nyelv, az $\overline{L(G_2)}$ is DCF-nyelv /2.5.4./, igy az $\overline{L(G_2)} \cap L(G_1)$ nyelv is DCF-nyelv /2.5.4./, tehát üres vagy nem üres volta eldönthető /2.6.2./. Ha az $L(G_2)$ reguláris, akkor az $\overline{L(G_2)}$ is reguláris /2.5.6./ és mivel az $L(G_1)$ DCF nyelv, az $\overline{L(G_2)} \cap L(G_1)$ ebben az esetben is DCF-nyelv, igy eldönthető, hogy üres-e vagy sem.

Összegezve a tartalmazási problémával kapcsolatos eddigi állitásainkat, kimondhatjuk a következő tételt:

3.5.3. tétel: Van olyan algoritmus, amely bármely G_1 és G_2 LD-grammatikák esetén eldönti, hogy van-e olyan, a G_1 grammatikával ekvivalens G_1' és a G_2 grammatikával ekvivalens G_2' grammatika, amelyekre a (G_1', G_2') pár eleme az $M = E_0 \cup H_0 \cup R_0$ halmaznak, és ha a $(G_1', G_2') \in M_0$, akkor eldönti, hogy az $L(G_1) \subseteq L(G_2)$ állítás teljesül-e vagy sem.

A fenti tétel széleskörüen használható, de van olyan eset is, amikor nem alkalmazható.

3.5.3.1. példa: $G_1 = (\{S, B, C, E\}, \{a, b, c, d, e\}, P_1, S)$

$G_2 = (\{S\}, \{a, b, c, d, e\}, P_2, S)$

P_1: $S \to aBE$ $\quad$ P_2: $S \to aS$

$B \to bBC$ $\quad$ $S \to bS$

$C \to cE$ $\quad$ $S \to cSS$

$B \to dE$ $\quad$ $S \to dSS$

$E \to e$ $\quad$ $S \to e$

$L(G_1) = \{a b^n d e (ce)^n e \mid n \geq 0\}$

A példát elemezve megállapithatjuk, hogy az $L(G_1) \subset L(G_2)$ teljesül. Ez abból adódik, hogy a $\varphi_{G_2}(a b^n d e (ce)^n e) = -1$ minden $n \geq 0$ természetes számra. Nézzük meg, hogy vannak-e olyan G_1' és G_2' LD-grammatikák, hogy $L(G_1) = L(G_1')$, $L(G_2) = L(G_2')$ és a $(G_1', G_2') \in H_0$. A következő fejezetben bebizonyitjuk azt a tényt, hogy a G_2 izomorfiától eltekintve csak saját magával ekvivalens. Tegyük fel, hogy van olyan G_1' amely teljesiti a feltételeket. Mivel a $(G_1', G_2) \notin R_0 \cup E_0$ triviálisan teljesül, csak a $(G_1', G_2) \in H_0$ fordulhat elő. A G_1' első két szabálya /mivel a G_2 a G_1' homomorf képe/ $S \to aX$ és $Y \to bZ$ alaku, ahol az S az axioma. Az $X = S$ nem teljesülhet, de az $X = Y = Z$ feltétlenül teljesül, ez viszont ellemtmond a G_1 és G_1' ekvivalenciájának.

Végezetül arra is hozunk egy példát, hogy a 3.5.3.10. lemma S-nyelvekre nem vihető át, vagyis izomorfiától eltekintve is végtelen sok olyan S-grammatika van, amelyek egymással ekvivalensek, de különbözők.

3.5.3.2. példa: $G_0 = (N_0, \{a, b\}, P_0, Z_0)$

$N_0 = \{Z_0\}$, $P_0 = \{Z_0 \to a Z_0 Z_0, Z_0 \to b\}$

$G_n = \{N_n, \{a, b\}, P_n, Z_n\}$

$N_n = \{Z_n\} \cup N_{n-1}$, $Z_n \notin N_{n-1}$

$P_n = \{Z_n \to a Z_{n-1} Z_n, Z_n \to b\} \cup P_{n-1}$

$n = 1, 2, \ldots$

A definícióból látható, hogy a G_n minden n esetén S-grammatika és ha $i \neq j$ akkor a G_i nem izomorf a G_j -vel. Könnyen belátható, hogy

$$L(G_n) = \{(a u)^m b \mid u \in L(G_{n-1}),\ m \geq 0\}$$

$n = 1, 2, \ldots$

Teljes indukcióval igazoljuk, hogy $L(G_n) = L(G_0)$, $n = 1, 2, \ldots$
Legyen $n \geq 1$.
$L(G_1) = \{(a u)^m b \mid u \in L(G_0),\ m \geq 0\}$. Legyen a $v = (a u)^m b \in L(G_1)$. $\varphi_{G_0}(v) = m(\varphi_{G_0}(a) + \varphi_{G_0}(u)) + \varphi_{G_0}(b) = -1$, $v \in L(G_0)$.
Legyen az $u \in L(G_0)$, ha $u = b$, akkor $u \in L(G_1)$ nyilvánvalóan. Ha $u \neq b$ akkor $u = a u_1 b$, ahol $Z_0 \overset{*}{\underset{G_0}{\Rightarrow}} u_1$ tehát $u \in L(G_1)$.
Legyen $n > 1$. Tegyük fel, hogy $L(G_{n-1}) = L(G_0)$. Akkor
$L(G_n) = \{(a u)^m b \mid u \in L(G_{n-1}),\ m \geq 0\} = \{(a u)^m b \mid u \in L(G_0), m \geq 0\}$
$= L(G_1) = L(G_0)$.

IV. fejezet

LENGYEL JELÖLÉSMÓD NYELVEK

1.§ Bevezetés

Ebben a fejezetben a Lukasiewicz-féle zárójelmentes jelölés fogalmának megfelelő nyelvekkel foglalkozunk. Ezeket nevezzük l e n g y e l j e l ö l é s m ó d n y e l v e k n e k. Elöljáróban röviden áttekintjük, hogy mit értünk a lengyel jelölésmódon. Egy programnyelven felirt aritmetikai vagy logikai kifejezés egy, müveleti jelekből és azonositókból álló véges lánc. Természetesen egy ilyen láncnak értéket is akarunk tulajdonitani, mégpedig valamilyen egyértelmü módon. Ha nem teszünk pótlólagos kikötéseket, akkor bizonyos kifejezésekhez több értéket is rendelhetünk, például az $a \times b + c$ kifejezés értéke, az azonositók $a=2, b=3, c=1$ értékei mellett, lehet 7 vagy 8, attól függően, hogy a szorzást vagy összeadást végezzük el először. Ezt a többértelmüséget többféle módon megszüntethetjük.

Ilyen módszer pl. az, hogy a müveleti jelek közt prioritási szabályokat adunk meg, az egyenlő prioritásu müveleti jelekhez pedig egy pótlólagos szabályt adunk /balról jobbra szabály/. Egy másik módszer uj terminális jelek, a zárójelek bevezetése és a kifejezés teljes zárójelezése. /E két módszer kombinációja az általában használt kifejezésfogalom./ A lengyel jelölésmód egy, kifejezések zárójelmentes felirására szolgáló módszer. Definiciója a következő: Legyen k -féle müveletünk, a megfelelő müveleti jelek legyenek $a_1, \ldots, a_k$. Minden $1 \leq i \leq k$ indexre az a_i jelü müvelet legyen l_i változós müvelet $(l_i \geq 1)$. Egy, az $a_1, \ldots, a_k$ jelekből és az azonositókból álló véges láncot akkor és csak akkor tekintünk lengyel jelölésmóddal felirt kifejezésnek, ha teljesülnek a következők:

a, Nincsenek felesleges azonositók és müveletek, tehát a

kifejezés egyáltalán kiértékelhető. Ezt a tulajdonságot számszerüen is ki lehet fejezni. Legyen minden $1 \leq i \leq k$ esetén az a_i előfordulásainak száma $m_i \geq 0$ és a kifejezésben szereplő azonositók száma s. Teljesülni kell a következő összefüggésnek: $\sum_{i=1}^{k} m_i(l_i-1)+1=s$. A képlet a következőképpen magyarázható. Ha a kiértékelés egy lépésében az a_i jelü müveletet hajtjuk végre, akkor ehhez l_i számu érték kell. Egy értéket szolgáltat az eddigi lépések eredménye, tehát szükség van még l_i-1 db azonositóra. Ebből jön az egyenlet baloldalán szereplő összeg. A plusz egy azonositó a legelső lépéshez kell, mivel ennél nincs még "előző müvelet eredménye".

b, A kifejezés bármely valódi kezdőszelete esetén a

$\sum_{i=1}^{k} m'_i(l_i-1)+1>s'$, ahol minden $1 \leq i \leq k$ indexre m'_i jelöli az a_i előfordulásainak számát a kezdőszeletben, s' pedig a kezdőszeletben szereplő azonositók száma.

Az a, követelmény magától értetődő. A b, követelmény definiálja tulajdonképpen a lengyel jelölésmódot. Az ezzel a módszerrel felirt kifejezések egyértelmü kiértékelése /vagy lefordítása pl. gépi nyelvre/ szekvenciálisan történhet. Az ezt megvalósitó algoritmusokra nem térünk ki, csak néhány egyszerü példát adunk. Legyen a $+$ az összeadás, a $\times$ a szorzás jele az a, b, c és d pedig azonositók.

lengyel jelölésmódu kifejezés	zárójeles alak.
a	a
$+ \times a b c$	$(a \times b) + c$
$\times + a b c$	$(a + b) \times c$
$\times + a b + c d$	$(a+b) \times (c+d)$
$+ + a \times b c d$	$(a + (b \times c)) + d$

Két példa nem lengyel jelölésmódu kifejezésre:
$++abbc$ /az a, nem teljesül/, $a+b\times c$ /a b, nem teljesül/.

A lengyel jelölésmód és az LD-grammatikák egy speciális osztálya közt szoros kapcsolat van. Ez az osztály az egy kategóriával rendelkező LD-grammatikák halmaza. A kapcsolat az LD-grammatikákra vonatkozó eddigi ismereteink és a lengyel jelölésmód definiciója alapján könnyen levezethető. Legyenek az $a_1, \ldots, a_k$ müveleti jelek és a hozzájuk tartozó müveletek legyenek rendre $l_1, \ldots, l_k$ változósak. A $b_1, \ldots, b_r$ jelek jelöljék az azonositókat. Tekintsük a következő grammatikát:

$$G=(\{S\},\{a_1,\ldots,a_k\}\cup\{b_1,\ldots,b_r\},P,S)$$

$$P:\ S\to a_1S^{l_1}$$
$$\vdots$$
$$S\to a_kS^{l_k}$$
$$S\to b_1$$
$$\vdots$$
$$S\to b_r$$

A G egy LD-grammatika. Azt állitjuk, hogy az $L(G)$ nyelv minden mondata egy lengyel jelölésmódban felirt kifejezés és minden lengyel jelölésmódu kifejezés előállitható a G grammatikával. Ez az állitás a sulyfüggvénynyel könnyen bizonyitható. Legyen a G sulyfüggvénye a φ, 3.2.14. tétel szerint az $L(G)$ pontosan azokból az u mondatokból áll, amelyekre:

1, $\varphi(u)=-1$

2, ha $u=u_1u_2$ és $u_2\neq\varepsilon$ akkor $\varphi(u_1)\geqq 0$

De mivel bármely v mondat esetén a

$$\varphi(v)=\sum_{i=1}^{k}|v|_{a_i}(l_i-1)-\sum_{j=1}^{r}|v|_{b_j}$$

az 1, és 2, követelmény pontosan megegyezik a lengyel jelö-

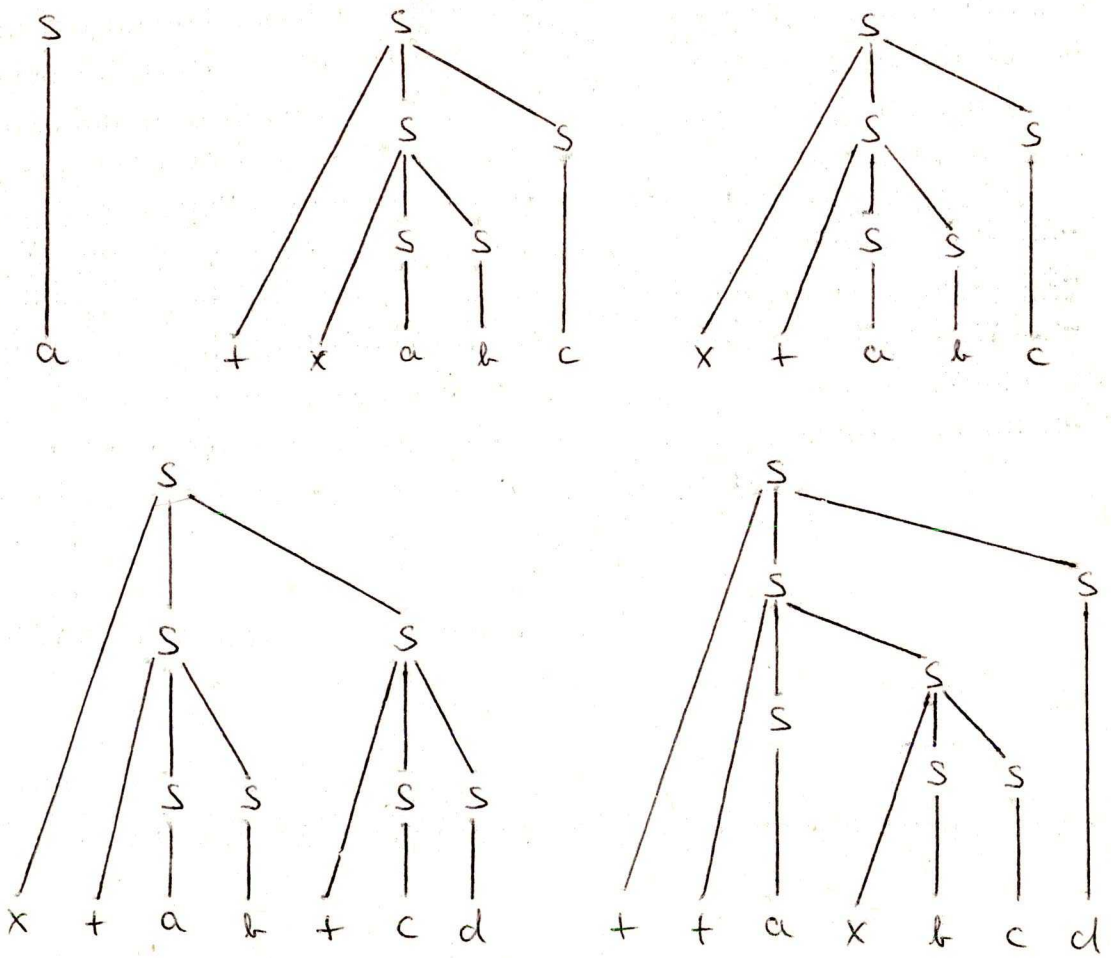
lésmódu kifejezések definiciójában szereplő a, és b, követelményekkel.

Nézzünk egy konkrét példát. A + és × valamint az a, b, c és d jelek jelentsék ugyanazt mint a fentebbi példában. Mivel a + és × müveletek kétváltozósak, a megfelelő grammatika a következő:

$$G_1 = (\{S\}, \{+, \times, a, b, c, d\}, P_1, S)$$

$$P_1: \quad S \to +SS \qquad S \to c$$
$$S \to \times SS \qquad S \to d$$
$$S \to a$$
$$S \to b$$

A példában szereplő kifejezések levezetéseinek diagrammjai:



A diagrammokból a zárójelezés /kiértékelés/ módja is könnyen leolvasható. Az LD-grammatikákra vonatkozó egyértelműségi tétel/3.2.6./ biztositja a zárójelezés /kiértékelés/ egyértelműségét. A következő részben ezekkel a nyelvekkel, mint speciális LD-nyelvekkel foglalkozunk matematikai nyelvészeti szempontból.

2.§ Definiciók és alaptulajdonságok

A "lengyel jelölésmód" helyett a továbbiakban a "PN" rövidítést használjuk /polnish notation/.

Definició: Egy $G=(N,T,P,S)$ LD-grammatikát **PN - g r a m - m a t i k á n a k** nevezünk, ha az $N=\{S\}$.

Definició: Egy L nyelvet **PN - n y e l v n e k** nevezünk, ha van olyan G PN-grammatika, amelyre az $L=L(G)$ teljesül.

A teljesség kedvéért itt is kimondjuk a PN-nyelvek sulyfüggvénnyel való jellemzését megadó, az ezen nyelvek vizsgálatánál leghasznosabb eszközül szolgáló 3.2.14. tételt.

4.2.1. tétel: Legyen a $G=(N,T,P,S)$ egy PN-grammatika és az $u\in T^*$. Az $u\in L(G)$ akkor és csak akkor teljesül, ha $\varphi_G(u)=-1$ és bármely $u=u_1u_2$ felbontás esetén, ha $u_2\neq\varepsilon$ akkor $\varphi_G(u_1)\geq 0$.

Definició: Legyen a $G=(N,T,P,S)$ PN-grammatika. Vezessük be a következő elnevezéseket:

a, Minden $a\in T$ esetén $l_G(a)=\varphi_G(a)+1$

b, $A_G=\{a\in T \mid l_G(a)>1\}$

c, $B_G=\{a\in T \mid l_G(a)=1\}$

d, $C_G=\{a\in T \mid l_G(a)=0\}$

4.2.2. lemma: Legyen a $G=(N,T,P,S)$ PN-grammatika. Igazak a következő állitások:

a, Az $L(G)=\phi$ akkor és csak akkor ha $C_G=\phi$.

b, Az $L(G)$ akkor és csak akkor véges, ha $A_G\cup B_G=\phi$.

c, Az $L(G)$ akkor és csak akkor reguláris, ha $A_G=\phi$.

d, Ha az $L(G)$ véges, akkor $L(G)=C_G$.

e, Ha az $L(G)$ reguláris akkor $L(G)=B_G^*C_G$.

Bizonyitás: Az a, és b, állitás közvetlenül adódik a defi-

nicíókból, a d, állitás a definiciókból és az a, állitásból. Igazoljuk a c, állitást. Ha az $A_G = \phi$ akkor a definiciókból adódik, hogy a G egy reguláris grammatika, tehát az $L(G)$ reguláris. Ha az $A_G \neq \phi$ akkor a G önbeágyazó, tehát az $L(G)$ nem lehet reguláris /2.7.6./.

A következőkben, külön feltevés nélkül is, ugy tekintjük, hogy a szóbanforgó PN-grammatikák nemüres nyelvet generálnak. Az $L(G) = \phi$ eset teljesen triviális ezért a továbbiakban nem foglalkozunk vele.

4.2.3. tétel: Egy R reguláris nyelv akkor és csak akkor PN-nyelv, ha vannak olyan T_1 és T_2 véges ábécék, hogy az $R = T_1^* T_2$ és a $T_1 \cap T_2 = \phi$.

Bizonyitás: Az állitás egyik fele a 4.2.2. e, következménye, hiszen bármely G PN-grammatika esetén a $B_G \cap C_G = \phi$. Legyen a T_1 és a T_2 véges ábécé és tekintsük a következő grammatikát: $G = (\{S\}, T_1 \cup T_2, P, S)$, $P = \{S \to aS \mid a \in T_1\} \cup \{S \to b \mid b \in T_2\}$. Nyilvánvaló, hogy $L(G) = T_1^* T_2$. Mivel a $T_1 \cap T_2 = \phi$ a G PN-grammatika, tehát az $L(G)$ PN-nyelv.

4.2.4. lemma: Egy LD-grammatika akkor és csak akkor generál PN-nyelvet, ha maga is PN-grammatika.

Bizonyitás: Legyen a $G = (N, T, P, S)$ LD-grammatika. Tegyük fel, hogy $L(G) = L(G_1)$ ahol a $G_1 = (N_1, T, P_1, S_1)$ PN-grammatika. A 3.5.3.1. lemma szerint $\#N_1 \geq \#N$ tehát $\#N = 1$ vagyis a G PN-grammatika.

4.2.5. tétel: A PN-nyelvek halmaza az LD-nyelvek halmazának valódi részhalmaza.

Bizonyitás: A tétel közvetlenül következik a PN-grammatikák definiciójából és a 4.2.4. lemmából.

4.2.6. lemma: Legyen a $G = (N, T, P, S)$ PN-grammatika és $u \in T^*$. A következő három állitás közül egy feltétlenül teljesül:

a, $u \in L(G)$

b, Van olyan $v \in T^+$, hogy $\mu v \in L(G)$.
c, Van olyan $\mu_2 \in T^+$, hogy $\mu = \mu_1 \mu_2$ és $\mu_1 \in L(G)$.

Bizonyitás: Legyen az μ_1 az μ leghosszabb olyan kezdőszelete, amelyre az $S \overset{*}{\Rightarrow} \mu_1 S^k$ levezetés létezik, ahol a $k \geq 0$. Mivel az $\mu \in T^*$ az $\mu_1 \neq \varepsilon$. Két eset van:

1, Tegyük fel, hogy az $\mu_1 = \mu$. Ha $k = 0$, akkor az a, állitás teljesül. Ha a $k > 0$, akkor a $v = b^k$ mondattal, ahol a $b \in C_G$, a b, állitás teljesül.

2. Tegyük fel, hogy $\mu = \mu_1 \mu_2$ és $\mu_2 \neq \varepsilon$. Ha $k = 0$, akkor a c, állitás teljesül. A $k > 0$ eset itt nem fordulhat elő az μ_1 definiciója miatt /mivel minden $a \in T$ esetén van $S \to a\alpha$ alaku szabály/.

4.2.7. lemma: Legyen az L PN-nyelv. Ha $\mu \in L$ és $|\mu| \geq 2$, akkor van az μ mondatnak olyan $\mu = \mu_1 \mu_2$ felbontása, hogy $\mu_1 \neq \varepsilon$, $\mu_2 \neq \varepsilon$ és minden $n \geq 0$ egész számra az $\mu_1^n \mu_2 \in L$. Ha az μ_1 a leghosszabb ilyen tulajdonságu mondat, akkor az $|\mu_2| = 1$.

Bizonyitás: Legyen az $L = L(G)$, ahol a G PN-grammatika. Jelölje a G sulyfüggvényét φ. Legyen az μ_1 az μ egy olyan kezdőszelete, amelyre $\varphi(\mu_1) = 0$. Ilyen valódi kezdőszelet van, mert az $|\mu| > 1$, a $\varphi(\mu) = -1$ és minden μ_1' kezdőszelet esetén $\varphi(\mu_1') \geq 0$. Legyen $\mu = \mu_1 \mu_2$ és $v_n = \mu_1^n \mu_2$ Mivel $\varphi(\mu_2) = -1$, az $\mu_2 \neq \varepsilon$. A $\varphi(v_n) = n\,\varphi(\mu_1) + \varphi(\mu_2) = \varphi(\mu_2) = -1$. Legyen $v_n = v_n' v_n''$ és $v_n'' \neq \varepsilon$. Ha $v_n' = \mu_1^k \mu_3$, ahol $k \leq n$ és az μ_3 az μ_1 kezdőszelete, akkor $\varphi(v_n') = k\,\varphi(\mu_1) + \varphi(\mu_3) = \varphi(\mu_3) \geq 0$. Ha $v_n' = \mu_1^n \mu_4$, ahol az μ_4 az μ_2 valódi kezdőszelete, akkor $\varphi(v_n') = n\,\varphi(\mu_1) + \varphi(\mu_4) = \varphi(\mu_4) = \varphi(\mu_1 \mu_4) \geq 0$. Tehát a $v_n \in L$ minden $n \geq 0$ egész számra. Legyen $\mu = va$, ahol $a \in T$; Mivel az $|\mu| > 1$, a $v \neq \varepsilon$ és mivel az $\mu \in L$, a $\varphi(v) = 0$ és a $\varphi(a) = -1$.

Tehát az $u_1 = G$ és $u_2 = a$ választással is helytálló a fenti bizonyitás. Ezzel a lemma mindkét állitását igazoltuk.

3.§ A lengyel jelölésmód-nyelvek és a veremautomaták

Az eddig definiált nyelvosztályokhoz hasonlóan a PN-nyelvekhez is hozzárendelhetjük a veremautomaták egy speciális tipusát.

Definició: Egy $M=(\{q_0\},T,V,\delta,q_0,Z_0,\{q_0\})$ S-automatát PN a u t o m a t á n a k /röviden PNA/ nevezzünk, ha a $V=\{Z_0\}$ teljesül.

4.3.1. tétel: Minden G PN-grammatikához van olyan M PN-automata, hogy $L(G)=T(M)$ és minden M PN-automatához van olyan G PN-grammatika, hogy $T(M)=L(G)$.

Bizonyitás: Ha a belső állapotot minden PNA-nál a „q_0" jelöli, akkor a PN automaták és a PN-grammatikák közt kölcsönösen egyértelmü megfeleltetés definiálható a következő módon:

Legyen a $G_1=(\{S\},T_1,P_1,S)$ PN-grammatika és az $M=(\{q_0\},T,\{Z_0\},\delta,q_0,Z_0,\{q_0\})$ PNA. A G_1 és az M akkor és csak akkor feleljenek meg egymásnak, ha teljesülnek a következők:

a, $S=Z_0$

b, $T_1=T$

c, Minden $a\in T$ esetén $S\to aS^k\in P_1$ és $\delta(q_0,a,Z_0)=(q_0,Z_0^k)$.

A definicióból következik, hogy bármely $u\in T^+$ esetén az $S\overset{*}{\underset{G_1}{\Rightarrow}}uS^l$ akkor és csak akkor teljesül, ha $u:(q_0,Z_0)\overset{*}{\underset{M}{\vdash}}(q_0,Z_0^l)$. Figyelembevéve azt, hogy az M mint S-automata akkor és csak akkor ismer fel egy $u\in T^*$ mondatot, ha $u:(q_0,Z_0)\overset{*}{\underset{M}{\vdash}}(q_0,Z_0^l)$ és $l=0$, és hogy az $u\in L(G_1)$ akkor és csak akkor teljesül, ha az $S\overset{*}{\underset{G_1}{\Rightarrow}}uS^l$ és $l=0$ közvetlenül következik,

hogy az $L(G_1) = T(M)$

4.3.1. példa: $G = (\{S\}, \{a, b, c, d\}, P, S)$

$P: S \to aSS$

$S \to bS$

$S \to c$

$S \to d$

$M = (\{q_0\}, \{a, b, c, d, e\}, \{S\}, \delta, q_0, S, \{q_0\})$

$\delta(q_0, a, S) = (q_0, SS)$

$\delta(q_0, b, S) = (q_0, S)$

$\delta(q_0, c, S) = (q_0, \varepsilon)$

$\delta(q_0, d, S) = (q_0, \varepsilon)$

$T(M) = L(G)$

4.§ Müveletek

Ebben a fejezetben azt vizsgáljuk, hogy a PN-nyelvek halmaza zárt-e a definiált müveletekre, illetve a CS, CF, DCF, S, LD és PN nyelvosztályok közül melyik az a legszükebb nyelvosztály, amelyből már az illető művelet nem vezet ki.

4.4.1. lemma: Van olyan L PN-nyelv és f homomorf leképezés, hogy az $f(L)$ nem S-nyelv.

Bizonyitás: Legyen az $L = \{a^m b \mid m \geq 0\}$. Az L PN-nyelv /4.2.3./. Legyen az f az $\{a,b\}^*$ halmaznak az $\{a\}^*$ halmazba való, a következő egyenletekkel megadott, homomorf leképezése:

a, $f(\varepsilon) = \varepsilon$

b, $f(a) = f(b) = a$

Az $f(L) = \{a^m \mid m \geq 1\}$ ez pedig nem S-nyelv, mert nem rendelkezik a prefix tulajdonsággal /2.7.4./.

4.4.2. tétel: A PN-nyelvek halmazából kivezetnek a következő müveletek: egyesités, metszetképzés, szorzatképzés, komplementerképzés, megforditás, homomorf leképezés, reguláris halmazzal való metszetképzés

Bizonyitás: Mivel az LD-nyelvekre vonatkozó megfelelő részben /III.4.§./ PN-grammatikákat használtunk az egyesitésnél a 3.4.1. lemmára, a metszetképzésnél a 3.4.2. lemmára, a reguláris nyelvvel való metszetképzésnél a 3.4.3. lemmára, a megforditásnál és a szorzatképzésnél a 3.4.5. lemmára hivatkozhatunk. A komplementerképzésnél az állitás abból következik, hogy S-nyelv komplementere nem S-nyelv /2.7.8./. A homomorf leképezésre vonatkozó állitás a 4.4.1. lemmából következik.

4.4.3. lemma: Vannak olyan $G_1, G_2, G_3, G_4, G_5, G_6, G_7$ és G_8 PN-grammatikák, valamint olyan R reguláris nyelv és f homomorf leképezés, hogy

a, az $L(G_1) \cup L(G_2)$ nem DCF-nyelv

b, az $L(G_3) \cap L(G_4)$ nem CF-nyelv

c, az $L(G_5) \cap R$ nem S-nyelv

d, az $L(G_6)L(G_7)$ nem LD-nyelv

e, az $\varphi(L(G_8))$ nem S-nyelv

Bizonyitás: Az a,, b,, c, és d, esetekben az állitás közvetlenül következik a 3.4.1., 3.4.2., 3.4.3. és a 3.4.5. lemmából. Az e, esetben a 4.4.1. lemma adja az állitást.

A megforditásra vonatkozóan általános tétel adható:

4.4.4. tétel: Bármely L PN-nyelv esetén:

a, Az L^R akkor és csak akkor PN-nyelv, ha L véges.

b, Ha az L végtelen, akkor az L^R nyelv nem S-nyelv.

c, Az L^R nyelv DCF-nyelv.

Bizonyitás: Legyen az $L = L(G)$ ahol $G = (\{S\}, T, P, S)$ és $\varphi = \varphi_G$. Tegyük fel, hogy az L véges. Akkor az $L = C_G$ és $L^R = L$ /4.2.2./. Tegyük fel, hogy az L végtelen. Akkor van olyan $u \in L$, hogy az $|u| \geq 2$, tehát a 4.2.7. lemma szerint van olyan $u_1, u_2 \in T^*$ hogy $u_1 \neq \varepsilon$, $u_2 \neq \varepsilon$ és $u_1 u_2 \in L$, $u_1 u_1 u_2 \in L$.

Ekkor viszont $(u_1 u_2)^R = u_2^R u_1^R \in L^R$ és $(u_1 u_1 u_2)^R = u_2^R u_1^R u_1^R \in L^R$. Tehát a prefix tulajdonság nem teljesül, igy az L^R nem S-nyelv /2.7.4./. Ezzel az a, és b, állitásokat igazoltuk. A c, állitás bizonyitásához először megadjuk az L^R nyelv egy sulyfüggvénnyel való jellemzését. Minden $v \in T^*$ mondat esetén legyen az $\eta(v) = -\varphi(v)$. Azt állitjuk, hogy egy $u \in T^*$ mondat akkor és csak akkor eleme az L^R nyelvnek, ha teljesül a következő két állitás:

1, $\eta(u) = 1$

2, Minden $u = u_1 u_2$ felbontás esetén, ha $u_1 \neq \varepsilon$ akkor $\eta(u_1) \geq 1$. Tegyük fel, hogy az $u = u_1 u_2 \in L^R$ és $u_1 \neq \varepsilon$. Akkor az $u^R = u_2^R u_1^R \in L$. Az $\eta(u) = -\varphi(u) = -\varphi(u^R) = 1$. Mivel $u_1 \neq \varepsilon$ a $\varphi(u_2^R) \geq 0$ és $\varphi(u_1^R) \leq -1$, tehát $\eta(u_1) = -\varphi(u_1) = -\varphi(u_1^R) \geq 1$. Tegyük fel, hogy az $u \in T^*$ teljesiti az 1, és 2, tulajdonságokat. Legyen a $v = u^R$. A $\varphi(v) = \varphi(u^R) = \varphi(u) = -\eta(u) = -1$. Legyen $v = v_1 v_2$ és $v_2 \neq \varepsilon$. Az $u = v^R = v_2^R v_1^R$, tehát $\eta(v_2^R) \geq 1$ igy az $\eta(v_1^R) < 1$. Következésképpen a $\varphi(v_1) = \varphi(v_1^R) =$

$= -\eta(v_1^R) > -1$, tehát a $\varphi(v_1) \geq 0$. Ezzel /4.2.1./ igazoltuk, hogy $v \in L$, ebből viszont $v^R = u \in L^R$. Legyen a $z \notin T$ egy uj jel. Ismeretes, hogy az $L^R z$ akkor és csak akkor DCF-nyelv, ha az L^R nyelv DCF-nyelv [5]. Az L^R nyelv η -val való jellemzése alapján megadunk egy olyan M DPDA-t, amelyre az $L^R z = T(M)$ teljesül. Minden $a \in T$ esetén jelölje k_a az a jelhez tartozó szabály jobboldalán szereplő "S" jelek számát / $k_a = \varphi_G(a) + 1$ / és legyen $k = \max_{a \in T} \{k_a\}$. Ha a $k = 0$ akkor az L véges, tehát az L^R PN-nyelv igy DCF-nyelv is. Legyen a $k \geq 1$. Tekintsük a következő M PDA-t:

$$M = (\{q_0, q_1, \ldots, q_k, q_{k+1}\}, T \cup \{z\}, \{Z_0, X\}, \delta, q_0, Z_0, \{q_{k+1}\})$$

$$\delta(q_0, c, Z_0) = (q_1, Z_0) \text{ minden } c \in C_G \text{ esetén}$$

$$\delta(q_1, z, Z_0) = (q_{k+1}, Z_0)$$

$$\left.\begin{aligned} \delta(q_1, c, Z_0) &= (q_1, X Z_0) \\ \delta(q_1, c, X) &= (q_1, XX) \end{aligned}\right\} \text{minden } c \in C_G \text{ esetén}$$

$$\left.\begin{aligned} \delta(q_1, a, Z_0) &= (q_{k_a}, Z_0) \\ \delta(q_1, a, X) &= (q_{k_a}, X) \end{aligned}\right\} \text{minden } a \in A_G \cup B_G \text{ esetén}$$

$$\delta(q_i, \varepsilon, X) = (q_{i-1}, \varepsilon) \text{ minden } 1 < i \leq k \text{ esetén}$$

Bármely olyan (q, X, Z) hármasra, amely a fenti felsorolásban nem fordul elő, legyen a $\delta(q, X, Z) = \emptyset$.
Az M müködése, mint ez a definiciójából kiolvasható, a következő: Minden olyan $c \in T$ beolvasáskor, amelyre az $\eta(c) = 1$ egyel növeli, és minden olyan $a \in T$ beolvasáskor, amelyre az $\eta(a) = -(k_a - 1), k_a - 1$ -el csökkenti a verem-

memória tartalmát. Tehát pontosan azt ellenőrzi, hogy a bemenő mondat teljesiti-e az L^R nyelvhez tartozás feltételeit. Az M DPDA, ez könnyen belátható. A fentiek alapján $T(M)=L^R z$ tehát az L^R DCF-nyelv.

4.4.1. példa: $G=(\{S\},\{a,b,c,d,e\},P,S)$

$P: S \to aSSSS$

$S \to bSS$

$S \to cS$

$S \to d$

$S \to e$

$M=(\{q_0,q_1,q_2,q_3,q_4,q_5\},\{a,b,c,d,e\}\cup\{z\},\{Z_0,X\},\delta,q_0,Z_0,\{q_5\})$

$\delta(q_0,d,Z_0)=(q_1,Z_0)$ $\quad$ $\delta(q_1,a,Z_0)=(q_4,Z_0)$

$\delta(q_0,e,Z_0)=(q_1,Z_0)$ $\quad$ $\delta(q_1,b,Z_0)=(q_2,Z_0)$

$\delta(q_1,z,Z_0)=(q_5,Z_0)$ $\quad$ $\delta(q_1,c,Z_0)=(q_1,Z_0)$

$\delta(q_1,d,Z_0)=(q_1,XZ_0)$ $\quad$ $\delta(q_1,a,X)=(q_4,X)$

$\delta(q_1,e,Z_0)=(q_1,XZ_0)$ $\quad$ $\delta(q_1,b,X)=(q_2,X)$

$\delta(q_1,d,X)=(q_1,XX)$ $\quad$ $\delta(q_1,c,X)=(q_1,X)$

$\delta(q_1,e,X)=(q_1,XX)$ $\quad$ $\delta(q_4,\varepsilon,X)=(q_3,\varepsilon)$

$\delta(q_3,\varepsilon,X)=(q_2,\varepsilon)$

$\delta(q_2,\varepsilon,X)=(q_1,\varepsilon)$

Megjegyezzük még, hogy a fenti tétel tulajdonképpen azt jelenti, hogy a forditott lengyel jelöléssel felirt kifejezések halmaza is DCF-nyelv, tehát ezek is egyértelmüen kiértékelhetőek.

A müveletekre vonatkozó vizsgálataink összefoglalásképp-

pen kimondhatjuk a következő tételt:

4.4.5. tétel: A legszükebb nyelvosztály, amely bármely L_1, L_1 és L_2 PN-nyelvek, valamint bármely R reguláris nyelv esetén tartalmazza

a, az $L_1 \cup L_2$ nyelvet, a CF-nyelvek halmaza.
b, az $L_1 \cap L_2$ nyelvet, a CS-nyelvek halmaza.
c, az $L_1 L_2$ nyelvet, az S-nyelvek halmaza.
d, az $L \cap R$ nyelvet, a DCF-nyelvek halmaza.
e, az $\overline{L}$ nyelvet, a DCF-nyelvek halmaza.
f, az L^R nyelvet, a DCF-nyelvek halmaza.

Bizonyitás: Az a, esetben a 2.5.2. tételre és a 4.4.3. lemmára, a b, esetben a 2.5.1. tételre és a 4.4.3. lemmára, a c, esetben a 2.7.7. tételre és a 4.4.3. lemmára, a d, esetben a 2.5.4. tételre és a 4.4.3. lemmára, az e, esetben pedig a 4.4.4. tételre való hivatkozással adódik az állitás.

5.§ Megoldhatóság

Ebben a paragrafusban a PN nyelvek halmazán vizsgáljuk az előzőekben felvetett problémák megoldhatóságát. Bár van olyan probléma, amely a PN nyelvekre szoritkozva is általánosan nem megoldható marad, a felvetett problémák többsége itt már általánosan megoldhatónak bizonyul. Az olyan problémákat, amelyek már valamely, a PN nyelveket tartalmazó nyelvosztály esetén általánosan megoldhatónak bizonyultak, itt csak akkor emlitjük meg, ha a PN nyelvek esetén a megoldó algoritmus lényegesen leegyszerüsödik.

5.1. Egy általánosan nem megoldható probléma

4.5.1.1. tétel: Nincs olyan algoritmus, amely bármely G CF-grammatika esetén eldöntené, hogy a G által generált nyelv PN-nyelv-e vagy sem.

Bizonyitás: A tétel bizonyitása szószerint megegyezik az LD nyelvekre vonatkozó hasonló tétel /3.5.1.2./ bizonyitásával. Ugyanis ott az LD nyelvek tulajdonságai közül csak a következő kettőt használtuk fel.

1, Minden LD nyelv DCF nyelv

2, Ha a T egy véges ábécé és $c \notin T$ akkor a $T^* c$ LD-nyelv.

Ez a két tulajdonsága megvan a PN-nyelvek halmazának is /4.2.5. és 4.2.3./.

5.2. Az üresség, végesség, regularitás, tartalmazás és ekvivalencia problémája

A cimben felsorolt problémák, a tartalmazást kivéve, általánosan megoldhatók az LD-nyelvek halmazán is. Itt viszont a megoldó algoritmus nagyon leegyszerüsödik, amint azt a következő állitások és bizonyitások mutatják.

4.5.2.1. tétel: Van olyan algoritmus, amely bármely G PN-grammatika esetén eldönti, hogy az $L(G)$ nyelv üres-e vagy sem, véges-e vagy sem, reguláris-e vagy sem.

Bizonyitás: Legyen a $G=(\{S\},T,P,S)$ PN-grammatika. A 4.2.2. lemma szerint, ha a P -ben van terminális szabály, akkor az $L(G)\neq\phi$, ha nincs terminális szabály, akkor az $L(G)=\phi$, ha csak terminális szabály van, akkor az $L(G)$ véges, ha van nemterminális szabály is, akkor az $L(G)$ végtelen, ha a G reguláris, akkor az $L(G)$ reguláris, ha a G nem reguláris, akkor az $L(G)$ sem reguláris.

4.5.2.2. tétel: Két PN-grammatika akkor és csak akkor ekvivalens, ha izomorf.

Bizonyitás: Legyen a $G_1=(\{S_1\},T_1,P_1,S_1)$ és a $G_2=(\{S_2\},T_2,P_2,S_2)$ Ha a két grammatika izomorf, akkor az $L(G_1)=L(G_2)$ /2.2.4./. Tegyük fel, hogy $L(G_1)=L(G_2)$. Nyilvánvaló, hogy $T_1=T_2$. Legyen a $c\in G_{G_1}$. Ez esetben a $c\in L(G_1)$, tehát a $c\in L(G_2)$ teljesül, tehát $S_1\to c\in P_1$ és $S_2\to c\in P_2$. Ha a két grammatika nem izomorf, akkor van olyan $a\in T$, hogy $S_1\to a S_1^k\in P_1$ és $S_2\to a S_2^l\in P_2$, ahol a $k\neq l$. Feltehető, hogy a $k>l\geq 0$. Akkor az $ac^l c^{k-l}\in L(G_1)$ és az $ac^l\in L(G_2)$. Mivel a $k-l>0$ a prefix tulajdonság /2.7.4./ miatt ellentmondásra jutottunk az $L(G_1)=L(G_2)$ feltevéssel. Tehát a két grammatika izomorf.

4.5.2.3. tétel: Legyenek a G_1 és G_2 PN-grammatikák, Az $L(G_1)\subseteq L(G_2)$ akkor és csak akkor teljesül, ha a G_2 grammatikának van a G_1 grammatikával izomorf részgrammatikája.

Bizonyitás: Ha a G_1 a G_2 egy részgrammatikájával izomorf, akkor az $L(G_1)\subseteq L(G_2)$ nyilvánvalóan teljesül. Tegyük fel, hogy az $L(G_1)\subseteq L(G_2)$. Legyen a $G_1=(\{S_1\},T_1,P_1,S_1)$ a $G_2=(\{S_2\},T_2,P_2,S_2)$. Nyilvánvalóan a $T_1\subseteq T_2$. Tekintsük a G_2 $G_2'=(\{S_2\},T_1,P_2',S_2)$ részgrammatikáját, ahol a $P_2'=\{S\to a\alpha \mid S\to a\alpha\in P_2 \text{ és } a\in T_1\}$. Az $L(G_1)\subseteq L(G_2)$ csak akkor teljesül, ha $L(G_1)\subseteq L(G_2')$. Tegyük fel, hogy $L(G_1)\subseteq L(G_2')$. Mivel $\#\{S_1\}=\#\{S_2\}=1$, a 3.5.3.3. lemma szerint $L(G_1)=L(G_2')$ ebből viszont a 4.5.2.2. tétel szerint a G_1 és a G_2' izomorfiája következik.

A fenti két tétel közvetlen következménye a megoldhatósági tétel:

4.5.2.4. tétel: Van olyan algoritmus, amely bármely G_1 és G_2 PN-grammatikák esetén eldönti, hogy az $L(G_1) = L(G_2)$ vagy az $L(G_1) \subseteq L(G_2)$ teljesül-e vagy sem.

5.3. Az egyesités, szorzatképzés, reguláris nyelvvel való metszetképzés és a megforditás problémája

A cimben emlitett problémák a PN-nyelvek esetén általánosan megoldhatók. Ezt bizonyitjuk a következőkben.

4.5.3.1. lemma: Legyenek a $G_1 = (N_1, T_1, P_1, S_1)$ és a $G_2 = (N_2, T_2, P_2, S_2)$ PN-grammatikák. Az $L(G_1) \cup L(G_2)$ akkor és csak akkor PN nyelv, ha a következő három feltétel közül valamelyik teljesül.

a, A G_1 és G_2 reguláris grammatikák és $B_{G_1} = B_{G_2}$.

b, A G_1 izomorf a G_2 egy részgrammatikájával.

c, A G_2 izomorf a G_1 egy részgrammatikájával.

Bizonyitás: Vezessük be a következő jelöléseket: $L_1 = L(G_1)$, $L_2 = L(G_2)$, $L = L_1 \cup L_2$, $B_1 = B_{G_1}$, $B_2 = B_{G_2}$, $C_1 = C_{G_1}$ és $C_2 = C_{G_2}$. Tegyük fel, hogy valamelyik feltétel teljesül. Ha a b, vagy a c, teljesül, akkor a 4.5.2.3. tétel szerint az $L_1 \subseteq L_2$ illetve az $L_2 \subseteq L_1$ és igy az $L = L_2$ illetve az $L = L_1$ vagyis az L PN-nyelv. Ha az a, teljesül, akkor az $L_1 = B_1^* C_1$ az $L_2 = B_2^* C_2$ /4.2.2./ és mivel a $B_1 = B_2$ az $L = B_1^*(C_1 \cup C_2)$, ez pedig PN-nyelv, hiszen a $(B_1 \cap C_1) \cup (B_2 \cap C_2) = \emptyset$. /4.2.3./. Ezzel az állitás egyik felét bebizonyitottuk. Tegyük fel, hogy az L PN-nyelv, de nem teljesül egyik feltétel sem. Legyen az $L = L(G)$, ahol a G PN-grammatika. Jelölje a G sulyfüggvényét φ. Legyen a $c_1 \in C_1$ és a $c_2 \in C_2$. Mivel $c_1 \in L_1$ és $c_2 \in L_2$, igy $c_1 c_2 \in L$. Vezessük be a $T_0 = T_1 \cap T_2$ jelölést.

1, Tegyük fel, hogy van olyan $a \in T_0$, hogy $S_1 \to a S_1^n \in P_1$

és $S_2 \to a S_2^q \in P_2$, ahol $p, q \geq 0$ egész számok és $p \neq q$. Mivel $a c_1^p \in L_1$ és $a c_2^q \in L_2$, igy $a c_1^p \in L$ és $a c_2^q \in L$, tehát $\varphi(a c_1^p) = \varphi(a c_2^q) = -1$. Ebből $\varphi(a) = p - 1 = q - 1$, tehát $p = q$, ellentmondásra jutottunk.

2, Tegyük fel, hogy minden $a \in T_0$ esetén, ha $S_1 \to a S_1^r \in P_1$ és $S_2 \to a S_2^q \in P_2$, akkor $r = q$. Az indirekt feltevés miatt $T_1 - T_0 \neq \emptyset$ és $T_2 - T_0 \neq \emptyset$. Legyen $b \in T_1$ és $b \notin T_2$ valamint $c \in T_2$ és $c \notin T_1$. Legyenek a megfelelő szabályok $S_1 \to b S_1^r$ és $S_2 \to c S_2^t$. Három eset van:

2.1, Tegyük fel, hogy a G_1 nem reguláris. Akkor van olyan $d \in T_1$, amelyre $S_1 \to d S_1^k \in P_1$ és $k \geq 2$. Mivel $d c_1^k \in L_1$, igy $d c_1^k \in L$, tehát $\varphi(d) = k - 1 \geq 1$.

Tekintsük a következő mondatot: $u = d b c_1^r c c_2^t c_1^{k-2}$. Felhasználva, hogy $b c_1^r, c c_2^t \in L$, a φ segitségével könnyen belátható, hogy $u \in L$. Viszont a b és c definiciója miatt $u \notin L_1$ és $u \notin L_2$, tehát ellentmondásra jutottunk.

2.2, Ha a G_2 nem reguláris, a bizonyitás az előző eset értelemszerü megismétlése.

2.3, Tegyük fel, hogy mindkét grammatika reguláris. Ez esetben $0 \leq r, t \leq 1$ és $T_1 = B_1 \cup C_1$ valamint $T_2 = B_2 \cup C_2$. Ha $r = 1$, akkor mivel $b c_1 \in L_1$ és $c c_2^t \in L_2$, a $b c_1 \in L$ és $c c_2^t \in L$, a $\varphi(b) = 0$. Ebből adódóan $b c c_2^t \in L$, de a b és c definiciója miatt $b c c_2^t \notin L_1$ és $b c c_2^t \notin L_2$. A $t = 1$ feltevésből kiindulva, ugyanigy ellentmondásra jutunk.

Tehát azt kaptuk, hogy $r = t = 0$, vagyis $b \in C_1$, $c \in C_2$. Ebből következik, hogy $T_1 - T_0 \subseteq C_1$ és $T_2 - T_0 \subseteq C_2$, és mivel a $B_1 \cap C_1 = B_2 \cap C_2 = \emptyset$ ezért $B_1 \subseteq T_0$ és $B_2 \subseteq T_0$. Ez viszont azt jelenti, hogy a $B_1 = B_2$, ellentmondásban az indirekt feltevéssel.

Mivel minden esetben ellentmondásra jutottunk, az indirekt feltevés hamis, vagyis az állitást bebizonyitottuk.

4.5.3.2. tétel: Van olyan algoritmus, amely bármely G_1 és G_2 PN-grammatikák esetén eldönti, hogy az $L = L(G_1) \cup L(G_2)$ nyelv PN-nyelv-e és ha az L PN-nyelv, akkor megadja azt a G PN-grammatikát, amelyre az $L = L(G)$ teljesül.

Bizonyitás: A 4.5.3.1. lemma feltételei teljesülésének ellenőrzése nyilvánvalóan végrehajtható. Ha egyik sem teljesül, akkor az L nem PN-nyelv. Ha az a, teljesül, akkor az $L(G_1) = B^*_{G_1} C_{G_1}$ az $L(G_2) = B^*_{G_2} C_{G_2}$ és az $L = B^*_{G_1}(C_{G_1} \cup C_{G_2})$. A $G = (\{S\}, B_{G_1} \cup C_{G_1} \cup C_{G_2}, \{S \to bS \mid b \in B_{G_1}\} \cup \{S \to c \mid c \in C_{G_1} \cup C_{G_2}\}, S)$ generálja az L nyelvet. A G nyilván PN-grammatika. Ha a b, teljesül, akkor $L(G_1) \subseteq L(G_2)$ /4.5.2.3./, tehát $L = L(G_2)$, igy a $G = G_2$. Hasonlóan adódik az, hogy ha a c, teljesül, akkor az $L = L(G_1)$, vagyis a $G = G_1$.

A szorzatképzés problémája triviálisan általánosan megoldható:

4.5.3.3. tétel: Ha az L_1 és L_2 PN-nyelvek, akkor az $L_1 L_2$ nem PN-nyelv.

Bizonyitás: A PN-grammatikák definiciójából következik, hogy minden PN-nyelvnek vannak 1 hosszuságu elemei /a C_G elemei/. Az $L_1 L_2$ nyelv legrövidebb mondatai viszont 2 hosszuságuak.

4.5.3.4. tétel: Van olyan algoritmus, amely bármely G PN-grammatika és R reguláris nyelv esetén eldönti, hogy az $L = L(G) \wedge R$ PN-nyelv-e vagy sem, és ha PN-nyelv, akkor megad egy olyan PN-grammatikát, amely az L nyelvet generálja.

Bizonyitás: Legyen a $G = (N, T, P, S)$ PN-grammatika. Képezzük az $R' = R \wedge T^*$ nyelvet. Ez is reguláris nyelv /2.5.6./ és nyilvánvalóan az $L(G) \wedge R = L(G) \wedge R'$. Megadható egy olyan $G_1 = (N_1, T_1, P_1, S_1)$ egyszerüsitett CF-grammatika /2.5.2./, amely a metszetet generálja, tehát

az $L(G_1) = L$. Mivel a G_1 egyszerüsitett, a $T_1 \subseteq T$ és a T_1 egyetlen eleme sem felesleges. Tekintsük a G $G' = (\{S\}, T_1, P', S)$ részgrammatikáját, ahol a $P' = \{S \to aS^k \mid S \to aS^k \in P_1$ és $a \in T_1\}$. Nyilvánvaló, hogy az $L = L(G) \cap R' = L(G') \cap R'$. Ha az L PN-nyelv, akkor, mivel az $L \subseteq L(G)$ az L nyelvet generáló G_0 PN-grammatikának ekvivalensnek kell lennie a G egy részgrammatikájával /4.5.2.3./. Ez a részgrammatika csak a G' lehet /mivel $L(G_1) = L$ és a G_1 egyszerüsitett/. Ha az $L = L(G')$, akkor az L nyilván PN-nyelv. Azt kaptuk, hogy az L akkor és csak akkor PN-nyelv, ha az $L = L(G')$ teljesül. Ez viszont, mivel az $L = L(G') \cap R'$, ekvivalens az $L(G') \subseteq R'$ teljesülésével. Mivel az R' reguláris és az $L(G')$ PN-nyelv az $L(G') \subseteq R'$ tartalmazás kérdése eldönthető /3.5.3.11./. Ha az $L(G') \subseteq R'$ igaz, akkor az $L = L(G') = L(G) \cap R$. Ha az $L(G') \subseteq R'$ nem teljesül, akkor az L nem PN-nyelv.

4.5.3.5. tétel: Van olyan algoritmus, amely bármely G PN-grammatika esetén eldönti, hogy az $(L(G))^R$ nyelv PN nyelv-e vagy sem.

Bizonyitás: A 4.5.2.1. tétel szerint eldönthető, hogy az $L(G)$ véges- e vagy sem. A 4.4.4. tétel szerint, ha az $L(G)$ véges, akkor az $(L(G))^R$ PN-nyelv, ha az $L(G)$ nem véges, akkor az $(L(G))^R$ nem PN-nyelv.

5.4. A homomorf leképezés problémája

Definició: A l t a l á n o s i t o t t PN - g r a m m a t i k á n a k nevezünk egy $G = (N, T, P, S)$ grammatikát, ha a G teljesiti a következő két feltételt:

a, $N = \{S\}$

b, $P = \{S \to u_i S^{k_i} \mid i = 1, \ldots, n\}$, ahol az $n \geq 1$ és minden $1 \leq i \leq n$ indexre $u_i \in T^+$ és $k_i \geq 0$ egész szám.

Definició: R e d u k á l t á l t a l á n o s i t o t t PN - g r a m m a t i k á n a k nevezünk egy $G = (N, T, P, S)$ általánositott PN grammatikát, ha bármely $S \to u S^k \in P$ esetén igaz a következő: Ha a $k \geq 1$ akkor tetszőleges $u = u_1 u_2$ felbontás

esetén az $u_2 \notin L(G)$.

4.5.4.1. lemma: Ha egy általánositott PN grammatika egy olyan nemüres nyelvet generál, amely rendelkezik a prefix tulajdonsággal, akkor megadhatóegy vele ekvivalens redukált általánositott PN grammatika.

Bizonyitás: Legyen a $G=(N,T,P,S)$ egy általánositott PN grammatika. A redukált alakra hozást szabályonként, az egyes szabályokra pedig lépésenként végezzük el, ugy, hogy minden lépés az előzővel ekvivalens grammatikát ad. Először is megjegyezzük a következőt: Ha az $S \to u_1 u_2 S^k \in P$, $u_2 \in L(G)$, $u_1 \neq \varepsilon$, akkor az $S \overset{*}{\underset{G}{\Rightarrow}} u_2$ levezetésben az $S \to u_1 u_2 S^k$ szabály nem vesz részt, mivel az $S \to \varepsilon \notin P$ és $|u_1 u_2| > |u_2|$. A redukciós eljárás egy lépése a következő: Tekintsük az $S \to u S^k$ szabályt, ahol az $u \in T^+$ és $k \geq 1$. Legyen az $|u| = r$ és minden $0 \leq i \leq r$ indexre jelölje u_i az u i hosszuságu kezdőszeletét a v_i pedig az $u = u_i v_i$ egyenlőséggel definiált mondatot. Az $u_0 = v_r = \varepsilon$, az $u_r = v_0 = u$. Minden i indexre vizsgáljuk meg, hogy a $v_i \in L(G)$ teljesül-e /2.6.1./. Ha az egyetlen indexre sem teljesül, akkor a vizsgált szabály redukált szabály, áttérhetünk egy másik szabályra. Legyen a $v_{i_0} \in L(G)$ ahol a $0 \leq i_0 \leq r$. Az $i_0 = 0$ vagy $i_0 = r$ nem lehet az $L(G)$ prefix tulajdonsága miatt. Tehát $0 < i_0 < r$ és $u_{i_0} \neq \varepsilon$, $v_{i_0} \neq \varepsilon$ és $S \to u_{i_0} v_{i_0} S^k \in P$ és $S \overset{*}{\underset{G}{\Rightarrow}} v_{i_0}$.
A vizsgált szabályt helyettesitsük az $S \to u_{i_0} S^{k+1}$ szabállyal. A helyettesitéssel kapott uj G' grammatika is általánositott PN grammatika, mivel $u_{i_0} \neq \varepsilon$ és ekvivalens az előzővel, mert az $S \overset{*}{\underset{G}{\Rightarrow}} u S^k$ levezetési lépés és az $S \underset{G'}{\Rightarrow} u_{i_0} S^{k+1} \overset{*}{\underset{G'}{\Rightarrow}} u_{i_0} v_{i_0} S^k = u S^k$ levezetés kölcsönösen helyettesithetők egymással. Ezután az $S \to u_{i_0} S^{k+1}$ szabály vizsgálatával folytatjuk az eljárást.
Az eljárást addig folytatjuk, mig minden szabály redukált nem lesz. Ha a szabályok száma m_1 és a leghosszabb szabály hossza m_2, akkor a szükséges lépések száma nem nagyobb mint $m_1 m_2$.

4.5.4.2. lemma: Ha egy G redukált általánositott PN-grammatika ekvivalens egy G_1 PN-grammatikával, akkor van a G grammatikának olyan G' részgrammatikája, amely izomorf a G_1 grammatikával.

Bizonyitás: Legyen a $G=(\{S\}, T, P, S)$ egy redukált általánositott PN-grammatika, a $G_1=(\{S_1\}, T, P_1, S_1)$ egy PN grammatika és $L(G)=L(G_1)$. Legyen a $k \in T$ és $S_1 \to k \in P_1$. Akkor a $k \in L(G_1)$, tehát a $k \in L(G)$, következésképpen az $S \to k \in P$. Legyen az $a \in T$, az $S \to a S^k \in P_1$ és $k \geq 1$. Legyen az $S \underset{G}{\Rightarrow} a\alpha$ az $a k^k$ baloldali levezetésének első lépése. Mivel $\alpha \underset{G}{\overset{*}{\Rightarrow}} k^k$, $\alpha \in \{k, S\}^*$. Mivel a G redukált, az $\alpha \in \{S\}^*$ és mivel az $S \to k \in P$ és /prefix tulajdonság miatt/ nincs olyan $r > 1$ és $t \geq 0$, hogy $S \to k^r S^t \in P$ lenne, az $|\alpha| = k$, tehát az $\alpha = S^k$, az $S \to a S^k \in P$. Ezzel az állitást bebizonyitottuk.

4.5.4.3.lemma: Bármely, nem üres nyelvet generáló általánositott PN-grammatika esetén eldönthető, hogy PN-nyelvet generál-e vagy sem, és ha PN-nyelvet generál, akkor megadható a megfelelő PN-grammatika.

Bizonyitás: Legyen a $G_0=(N, T, P_0, S)$ egy általánositott PN-grammatika. Ha a redukciós eljárás nem hajtható végre a G_0 grammatikán, akkor az $L(G_0)$ nem PN-nyelv /4.5.4.1./. Ha végrehajtható, akkor legyen a $G=(\{S\}, T, P, S)$ olyan általánositott redukált PN-grammatika, amelyre az $L(G_0)=L(G)$ teljesül. Ha van olyan $a \in T$, amelyhez nincs $S \to a S^k$ alaku szabály a G grammatikában, akkor az $L(G)$ nem PN-nyelv /4.5.4.2./. Ellenkező esetben tekintsük a G $G'=(\{S\}, T, P', S)$ részgrammatikáját, ahol $P' = \{S \to u S^k \in P \mid |u| = 1\}$. Ha a G' nem PN-grammatika, akkor van olyan $a \in T$, hogy $S \to a S^{k_1}, S \to a S^{k_2} \in P'$ és $k_1 \neq k_2$. Ez esetben a prefix tulajdonság miatt nem lehet az $L(G)$ PN-nyelv, vagy nincs olyan $k \in T$, amelyre az $S \to k \in P'$ lenne, ez esetben az $L(G)$ szintén nem lehet PN-nyelv

/4.5.4.2./. Tegyük fel, hogy a G' PN-grammatika. Az $L(G)$ akkor és csak akkor PN-nyelv, ha a G ekvivalens a G' -vel. Ugyanis a G' tartalmazza a G bármely PN-tulajdonságu részgrammatikáját, igy ha az $L(G)$ PN-nyelv, akkor a G ekvivalens a G' egy G'' részgrammatikájával /4.5.4.2./, mivel az $L(G') \leq L(G)$ ez a részgrammatika csak maga a G' lehet. Tehát csak azt kell eldönteni, hogy van-e olyan $v \in L(G)$ mondat, amely nem eleme az $L(G')$ nyelvnek. Ha ilyen v van akkor az $L(G)$ nem PN-nyelv, ha nincs, akkor PN-nyelv, Egy ilyen mondat létezése viszont ekvivalens egy olyan $S \to uS^k \in P$ szabály létezésével, amelyhez nincs $S \overset{*}{\underset{G'}{\Rightarrow}} uS^k$ G' -beli levezetés. Az összes $P - P'$ halmazbeli szabályok ebből a szempontból való ellenőrzésével /ami nyilvánvalóan véges számu eset végigpróbálását jelenti/ eldönthető a kérdés.

4.5.4.4. lemma: Legyen a T egy véges ábécé és $u \in T^*$. Nincs olyan L PN-nyelv és $n \geq 2$ természetes szám, amelyre az $u^n \in L$ teljesülne.

Bizonyitás: Legyen a G egy tetszőleges PN grammatika és φ a sulyfüggvénye, valamint az $n \geq 0$ természetes szám. A $\varphi(u^n) = n\varphi(u) = -1$ csak akkor teljesülhet, ha $n = 1$.

Az eddigi segédeszközökkel már megoldható a homomorf leképezés problémája.

4.5.4.5. tétel: Bármely G PN-grammatika és f homomorf leképezés esetén eldönthető, hogy az $f(L(G))$ nyelv PN-nyelv-e vagy sem, és ha PN-nyelv, akkor megadható egy G' PN-grammatika, amelyre az $L(G') = f(L(G))$.

Bizonyitás: Legyen a $G = (\{S\}, T, P, S)$ PN-grammatika és legyen az f a T^* halmaz egy T_1^* halmazba való homomorf leképezése, ahol a T_1 egy véges ábécé. Az $L_1 = f(L(G))$ nyelv CF-nyelv és a következő grammatika generálja: $G_1 = (\{S\}, T_1, P_1, S)$ ahol $P_1 = \{S \to f(a)S^k \mid a \in T, S \to aS^k \in P\}$ /2.5.7./. Két eset van az f leképezéstől függően:

1, Minden $a \in T$ esetén $f(a) \in T_1^+$. Ekkor a G_1 általánositott PN-grammatika, tehát alkalmazva a 4.5.4.3.

lemmát, a tétel bizonyitott.

2, Van a T -nek egy olyan nemüres T' részhalmaza, hogy minden $t \in T'$ esetén az $f(t) = \varepsilon$.

a, Ha van olyan $t \in T'$ hogy az $S \to t \in P$, akkor a $t \in L(G)$ tehát $\varepsilon \in L_1$, következésképpen az L_1 nem PN-nyelv.

b, Ha van olyan $t \in T'$, hogy az $S \to t S^k \in P$ és $k \geq 2$, akkor bármely $u \in L(G)$ esetén a $t(u^k) \in L(G)$, tehát az $f(t u^k) = (f(u))^k \in L_1$. A 4.5.4.4. lemma szerint az L_1 nem PN-nyelv.

c, Ha minden $t \in T'$ esetén az $S \to t S \in P$, akkor a P_1 halmazba az $S \to S$ szabályok kerülnek be, ezek viszont nyilván elhagyhatók. Elhagyásukkal a G_1 egy általánositott PN-grammatika lesz, tehát alkalmazva a 4.5.4.3. lemmát, a tétel bizonyitott.

Ezzel minden esetben eldöntöttük a kérdést.

5.5. A metszetképzés problémája

Tudjuk, hogy nincs olyan algoritmus, amely bármely két G_1 és G_2 CF-grammatika esetén eldöntené, hogy az $L(G_1) \cap L(G_2)$ nyelv üres- e, véges-e, illetve CF-tipusu-e. Láttuk, hogy ezek a problémák a DCF, S és LD-nyelvek halmazára szoritkozva sem oldhatók meg általánosan. A PN-nyelvek halmazán viszont már általánosan megoldhatónak bizonyulnak ezek a problémák. Ezt igazoljuk a következőkben.

Legyenek a $G_1 = (\{S_1\}, T_1, P_1, S_1)$ és a $G_2 = (\{S_2\}, T_2, P_2, S_2)$ PN-grammatikák. Az $L = L(G_1) \cap L(G_2)$ nyelvet vizsgáljuk. Mivel az $L \subset T_1^*$ és az $L \subset T_2^*$, tehát az $L \subset (T_1 \cap T_2)^*$, vizsgálatainknál elegendő a közös terminális jelkészletü grammatikákra szoritkozni, hiszen a megfelelő terminális jelek és szabályok elhagyásával az általános eset erre egyszerüen visszavezethető. Ha a $T_1 \cap T_2 = \emptyset$ akkor az $L = \emptyset$ ez az eset triviális, a továbbiakban külön feltevés nélkül ugy tekintjük, hogy a $T_1 \cap T_2 \neq \emptyset$.

4.5.5.1. lemma: Legyenek a $G_1 = (\{S_1\}, T_1, P_1, S_1)$ és a $G_2 = (\{S_2\}, T_1, P_2, S_2)$ PN-grammatikák. Az $L = L(G_1) \cap L(G_2)$

nyelv akkor és csak akkor nem üres, ha van olyan $C \in T$, hogy $S_1 \to C \in P_1$ és $S_2 \to C \in P_2$.

Bizonyitás: Ha ilyen $C \in T$ van, akkor $C \in L$, tehát $L \neq \phi$. Tegyük fel, hogy van olyan $u \in T^*$, hogy $u \in L$. Az $u \neq \varepsilon$ nyilván, tehát van olyan $a \in T$ és $v \in T^*$, hogy $u = va$. Tekintsük az $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} v\alpha$, $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} v\beta$ baloldali levezetéseket, ahol az $\alpha \in \{S_1\}^*$ és $\beta \in \{S_2\}^*$. Mivel $\alpha \overset{*}{\underset{G_1}{\Rightarrow}} a$ és $\beta \overset{*}{\underset{G_2}{\Rightarrow}} a$ és $|a| = 1$, az $\alpha = S_1$ és a $\beta = S_2$, tehát $S_1 \to a \in P_1$ és $S_2 \to a \in P_2$.

4.5.5.2. lemma: Legyenek a $G_1 = (\{S_1\}, T, P_1, S_1)$ és a $G_2 = (\{S_2\}, T, P_2, S_2)$ PN-grammatikák és az $L = L(G_1) \cap L(G_2)$ egy nemüres PN-nyelv. Legyen $a \in T$ és az $S_1 \to a S_1^{k_1} \in P_1$, az $S_2 \to a S_2^{k_2} \in P_2$. Ha $k_1 \neq k_2$ akkor bármely $u \in L$ esetén az $|u|_a = 0$.

Bizonyitás: Mivel az $L \neq \phi$ van olyan $C \in T$, hogy a $C \in L$ /4.5.5.1./. Legyen az $L = L(G)$ ahol a G PN-grammatika. Legyen a G sulyfüggvénye a φ . Ha van olyan $u \in L$, hogy az $|u|_a \neq 0$, akkor a $\varphi(a)$ értelmezett. /Van $S \to a S^{k_a}$ alaku szabály a G -ben./ Legyen a $\varphi(a) = k \geq -1$, akkor az $a\, C^{k+1} \in L$. De akkor $a\, C^{k+1} \in L(G_1)$ és $a\, C^{k+1} \in L(G_2)$. Mivel az $a\, C^{k_1} \in L(G)$ és az $a\, C^{k_2} \in L(G)$ a prefix tulajdonság miatt $k_1 = k+1 = k_2$, ellentmondásban a feltétellel.

4.5.5.3. lemma: Legyenek a $G_1 = (\{S_1\}, T, P_1, S_1)$ és a $G_2 = (\{S_2\}, T, P_2, S_2)$ PN-grammatikák. Ha az $L = L(G_1) \cap L(G_2)$ nemüres és véges, akkor PN-nyelv és $L \subseteq T$.

Bizonyitás: Tegyük fel, hogy van olyan $u \in L$, hogy $u \notin T$. Akkor az $|u| > 1$. Legyen $u = vta$, ahol $a \in T$. Mivel $u \in L$, $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} v S_1$ és $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} v S_2$, tehát bármely $n \geq 1$ természetes számra $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} v^n S_1$ és $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} v^n S_2$, tehát $v^n a \in L(G_1)$ és $v^n a \in L(G_2)$ igy $v^n a \in L$.

Mivel $|v| \neq 0$ az L végtelen, ellentmondásban a feltevéssel. Azt kaptuk, hogy ha az L nemüres és véges, akkor $L \subseteq T$. Ekkor viszont nyilvánvaló, hogy PN-nyelv.

Definició: Legyenek a $G_1=(\{S_1\}, T, P_1, S_1)$ és a $G_2=(\{S_2\}, T, P_2, S_2)$ PN-grammatikák. Minden $a \in T$ esetén jelölje a $\varphi_{G_1}(a)+1$ értéket l_a a $\varphi_{G_2}(a)+1$ értéket h_a és vezessük be a következő elnevezéseket:

$$C_0 = \{a \in T \mid l_a > h_a > 1\}$$
$$C_1 = \{a \in T \mid l_a > h_a = 1\}$$
$$C_2 = \{a \in T \mid l_a > h_a = 0\}$$

$$D_0 = \{b \in T \mid h_b > l_b > 1\}$$
$$D_1 = \{b \in T \mid h_b > l_b = 1\}$$
$$D_2 = \{b \in T \mid h_b > l_b = 0\}$$

$$E_0 = \{c \in T \mid l_c = h_c > 1\}$$
$$E_1 = \{c \in T \mid l_c = h_c = 1\}$$
$$E_2 = \{c \in T \mid l_c = h_c = 0\}$$

$$\left.\begin{aligned} C_{0i} &= C_0 \cup \ldots C_i \\ D_{0i} &= D_0 \cup \ldots D_i \\ E_{0i} &= E_0 \cup \ldots E_i \end{aligned}\right\} i = 0, 1, 2$$

$$C = C_{02}, \; D = D_{02}, \; E = E_{02}$$

4.5.5.4. lemma: Legyen a $G=(N,T,P,S)$ PN-grammatika, $T_0=\{a\in T \mid \varphi_G(a)\geq 0\}$ és $T_1=\{b\in T \mid \varphi_G(b)=-1\}$, valamint az $u\in T_0^*T_1^*$. Az $u\in L(G)$ akkor és csak akkor teljesül ha a $\varphi_G(u)=-1$.

Bizonyitás: Ha $u\in L(G)$, akkor a $\varphi_G(u)=-1$ /4.2.1./. Tegyük fel, hogy $\varphi_G(u)=-1$ és legyen $u=u_1u_2$, ahol az $u_2\neq\varepsilon$. Ha $u_1\in T_0^*$, akkor a $\varphi_G(u_1)\geq 0$ nyilvanvalóan teljesül. Ha $u_1=vv_1$, ahol $v\in T_0^*$ és $v_1\in T_1^*$ akkor, mivel $u_2\in T_1^+$, a $\varphi_G(u_2)<0$. Mivel $\varphi_G(u)=\varphi_G(u_1)+\varphi_G(u_2)=-1$ és a $\varphi_G(u_2)<0$ a $\varphi_G(u_1)\geq 0$. Tehát $u\in L(G)$ /4.2.1./.

4.5.5.5. lemma: Legyenek a $G_1=(\{S_1\},T,P_1,S_1)$ és a $G_2=(\{S_2\},T,P_2,S_2)$ PN-grammatikák, valamint az $L=L(G_1)\cap L(G_2)\neq\phi$. Ha a következő két feltétel közül valamelyik teljesül, akkor az L nem PN-nyelv.

a, Van olyan $0\leq i,j\leq 2$ indexpár, hogy $i+j\leq 3$ és $C_i\neq\phi$, $D_j\neq\phi$.

b, Az $E_0\neq\phi$ és van olyan $0\leq i,j\leq 2$ indexpár, hogy a $C_i\neq\phi$, $D_j\neq\phi$.

Bizonyitás: Vezessük be a következő jelöléseket:

$L_1=L(G_1)$, $L_2=L(G_2)$, $\varphi_1=\varphi(G_1)$, $\varphi_2=\varphi(G_2)$

a, A metszetképzés szimmetriája miatt elegendő csak a $(0,0)$, $(0,1)$, $(0,2)$, $(1,1)$, $(1,2)$ indexpárokra vizsgálni az állitást. Tegyük fel, tehát, hogy a $C_{01}\neq\phi$ és a $D_{02}\neq\phi$. Legyen az $a\in C_{01}$ és a $b\in D_{02}$. Mivel az $L\neq\phi$ az $E_2\neq\phi$, legyen a $c\in E_2$ /4.5.5.1./. Tekintsük az $u=a^{k_1}b^{k_2}c^{k_3}$ mondatot, ahol $k_1=h_b-l_b$, $k_2=l_a-h_a$, $k_3=(h_b-l_b)(h_a-1)+(l_a-h_a)(h_b-1)+1$. Először igazoljuk, hogy u létezik, vagyis $k_1,k_2,k_3\geq 0$. Mivel az $a\in C_{01}$ és a $b\in D_{02}$, a $k_1>0$, a $k_2>0$ és a $k_3\geq 1$, tehát az u létezik. Egyszerü számolással adódik, hogy $\varphi_1(u)=-1$ és $\varphi_2(u)=-1$. Ebből következik, hogy $u\in L$ /4.5.5.4./. Mivel az

$|\mu|_a = k_1 > 0$ és $|\mu|_b = k_2 > 0$ és $a, b \notin E$ a 4.5.5.2. lemma szerint az L nem PN-nyelv.

b, A feltétel azt jelenti, hogy $E_0 \neq \phi$, $C \neq \phi$ és $D \neq \phi$. Legyen a $d \in E_0$, az $a \in C$ és a $b \in D$. Tekintsük az $\mu = d^{m_0} a^{m_1} b^{m_2} c^{m_3}$ mondatot, ahol az $m_0 = m_1 = k_b - l_b$ az $m_2 = l_a - k_a$ és az $m_3 = (k_b - l_b)(l_d - 1 + l_a - 1) + (l_a - k_a)(l_b - 1) + 1$. Az a és b definiciója miatt $m_0 > 0$, $m_1 > 0$ és $m_2 > 0$. Az m_3 kifejezésben a $k_b - l_b \geq 1$, az $l_d - 1 \geq 1$ és az $l_a - 1 \geq 0$, tehát az összeg első tagja pozitiv. Az összeg második tagja csak akkor negativ, ha $l_b = 0$. Ekkor viszont a $k_b = 1$ és az $m_3 = m_0(l_d - 1) + k_a > 0$ mert az $m_0(l_d - 1) \geq 1$ és a $k_a \geq 0$. Tehát az $m_3 > 0$ minden esetben. Kaptuk, hogy az μ létezik. Figyelembevéve, hogy $l_d = k_d$, egyszerü számolással adódik, hogy $\varphi_1(\mu) = \varphi_2(\mu) = -1$. Az $\mu \in L_1$ igazolására alkalmazható a 4.5.5.4. lemma. Az $\mu \in L_2$ állitás az $a \in C_{01}$ esetre ugyanigy adódik a lemmából. Legyen az $a \in C_2$ és az $\mu = \mu_1 \mu_2$ ahol az $\mu_2 \neq \varepsilon$. Ha az $\mu_1 \in \{d\}^*$, akkor a $\varphi_2(\mu_1) \geq 0$ nyilvánvaló. Legyen az $\mu_1 = d^{m_0} a^k$, ahol $k \leq m_1$. A $\varphi_2(\mu_1) = m_0(k_d - 1) - k = m_1(k_d - 1) - k \geq 0$, mivel a $k_d - 1 \geq 1$. Mivel a $\varphi_2(b) \geq 0$ több esetet nem kell vizsgálni. Tehát az $\mu \in L_1$ és $\mu \in L_2$ igy az $\mu \in L$. Mivel $|\mu|_a = m_1 > 0$, $|\mu|_b = m_2 > 0$, $a, b \in E_1$, az L nem PN-nyelv. /4.5.5.2./

4.5.5.6. lemma : Legyenek a $G_1 = (\{S_1\}, T, P_1, S_1)$ és a $G_2 = (\{S_2\}, T, P_2, S_2)$ PN-grammatikák és az $L = L(G_1) \wedge L(G_2) \neq \phi$. Legyen az $\mu \in L$. Ha az $E_0 = \phi$ és a $C_{01} \cup D_{01} = \phi$, akkor bármely $a \in C_2$ és $b \in D_2$ esetén $|\mu|_a = 0$ és $|\mu|_b = 0$.

Bizonyitás: Tegyük fel, hogy van olyan $\mu \in L$ amelyre nem igaz az állitás. Legyen az $\mu = \mu_1 a \mu_2$, ahol az $a \in C_2 \cup D_2$. Feltehető, hogy az $\mu_1 \in E^*$ vagyis a $C_2 \cup D_2$ halmazbeli elemek legelső előfordulását tekintjük. Tegyük fel, hogy az $a \in C_2$. Mivel az $\mu_1 \in E^*$ és $E_0 = \phi$, a $\varphi_{G_1}(\mu_1) = \varphi_{G_2}(\mu_1) = 0$. Mivel az

$a \in C_2$, $\varphi_{G_2}(a) = -1$ tehát az $u_2 = \varepsilon$ az $u = u_1 a$. Viszont $\varphi_{G_1}(u) \geq 0$, tehát $\varphi_{G_1}(u_1 a) \geq 0$, igy $u = u_1 a \notin L(G_1)$ ellentmondásban a feltétellel. Ha az $a \in D_2$, ugyanilyen módon igazolható az állitás.

4.5.5.7. lemma: Legyenek a $G_1 = (\{S_1\}, T, P_1, S_1)$ és a $G_2 = (\{S_2\}, T, P_2, S_2)$ PN-grammatikák, az $L = L(G_1) \cap L(G_2)$ és az $u \in L$. Ha a $C = \phi$, akkor bármely $b \in D$ esetén $|u|_b = 0$, ha a $D = \phi$ akkor bármely $a \in C$ esetén $|u|_a = 0$.

Bizonyitás: Tegyük fel, hogy a $C = \phi$ és legyen a $b \in D$, valamint az $u \in L$. Mivel az $b_1 < b_2$, igy a $\varphi_{G_1}(b) < \varphi_{G_2}(b)$. Tegyük fel, hogy $u = u_1 b u_2$. Mivel a $C = \phi$ bármely $v \in T^*$ esetén $\varphi_{G_1}(v) \leq \varphi_{G_2}(v)$. Az $u \in L$ miatt $\varphi_{G_1}(u) = \varphi_{G_2}(u) = -1$.
Viszont $\varphi_{G_1}(u) = \varphi_{G_1}(u_1) + \varphi_{G_1}(b) + \varphi_{G_2}(u_2) < \varphi_{G_2}(u_1) + \varphi_{G_2}(b) + \varphi_{G_2}(u_2)$ ami ellentmondás, tehát $|u|_b = 0$. Hasonlóan bizonyitható a $D = \phi$ eset is.

4.5.5.8. tétel: Bármely G_1 és G_2 PN grammatikák esetén eldönthető, hogy az $L(G_1) \cap L(G_2)$ nyelv
a, üres-e vagy sem.
b, véges-e vagy sem.
c, PN-nyelv-e vagy sem.
Ha az $L(G_1) \cap L(G_2)$ PN-nyelv, akkor megadható egy olyan G PN-grammatika, amelyre az $L(G) = L(G_1) \cap L(G_2)$ teljesül.

Bizonyitás: Legyenek a $G_1 = (\{S_1\}, T, P_1, S_1)$ és a $G_2 = (\{S_2\}, T, P_2, S_2)$ PN grammatikák. Vezessük be a következő rövidiséseket: $L_1 = L(G_1)$, $L_2 = L(G_2)$, $L = L_1 \cap L_2$. A két grammatika ismeretében megadhatók a $C_0, C_1, C_2, D_0, D_1, D_2, E_0, E_1, E_2$ halmazok. Az algoritmus ezeket a halmazokat vizsgálja. Ha az $E_2 = \phi$ akkor az $L = \phi$, ha az $E_2 \neq \phi$ akkor az $L \neq \phi$ /4.5.5.1./. A 4.5.5.3. lemma szerint az L csak akkor véges, ha PN nyelv, ezért először azt döntjük el, hogy az L PN nyelv-e vagy sem. Attól

függően, hogy az $E_0 = \phi$ vagy az $E_0 \neq \phi$ igaz, a megoldó algoritmus két irányban folytatódhat.

1, Legyen az $E_0 \neq \phi$. Ha a $C \neq \phi$ és a $D \neq \phi$ akkor az L nem PN nyelv /4.5.5.5./. Ha a C és D közül legalább az egyik üres, akkor az $L \subseteq E^*$ teljesül /4.5.5.7./. Ez esetben az algoritmus a 3, pontban folytatódik.

2, Legyen az $E_0 = \phi$. Ha a C_{01} és a D_{02} nem üres vagy a C_{02} és a D_{01} nem üres, akkor az L nem PN-nyelv /4.5.5.5./. Ha a fenti két lehetőség egyike sem következik be, akkor két eset lehetséges:

2.1, A $C = \phi$ vagy a $D = \phi$. Ekkor a 4.5.5.7. lemma szerint az $L \subseteq E^*$ és az algoritmus a 3, pontban folytatódik.

2.2, A $C_{01} = \phi$ és a $D_{01} = \phi$ de a $C_2 \neq \phi$ és a $D_2 \neq \phi$. Ekkor a 4.5.5.6. lemma szerint az $L \subseteq E^*$, tehát ugyanaz az eset áll elő mint a 2.1, pontban.

3. Ha az $L \subseteq E^*$ teljesül, akkor az L PN-nyelv. Ugyanis ez esetben az $L = L(G_1') \wedge L(G_2')$ ahol a G_1' a G_1 , a G_2' pedig a G_2 grammatikának az E halmazhoz tartozó részgrammatikái /elhagyjuk azokat a szabályokat, amelyekben a terminális jel nem E -beli/. Az E definiciója miatt a G_1' és a G_2' egymással izomorf grammatikák, tehát
$L = L(G_1') = L(G_2')$

Az eddigieket végrehajtva egyértelműen eldöntöttük, hogy az PN-nyelv-e vagy sem. Ha az L nem PN-nyelv, akkor a 4.5.5.3. lemma szerint nem véges, ha PN-nyelv, akkor végessége a 3, pontban leirt G_1' grammatika ismeretében eldönthető /4.5.2.1./.

FÜGGELÉK

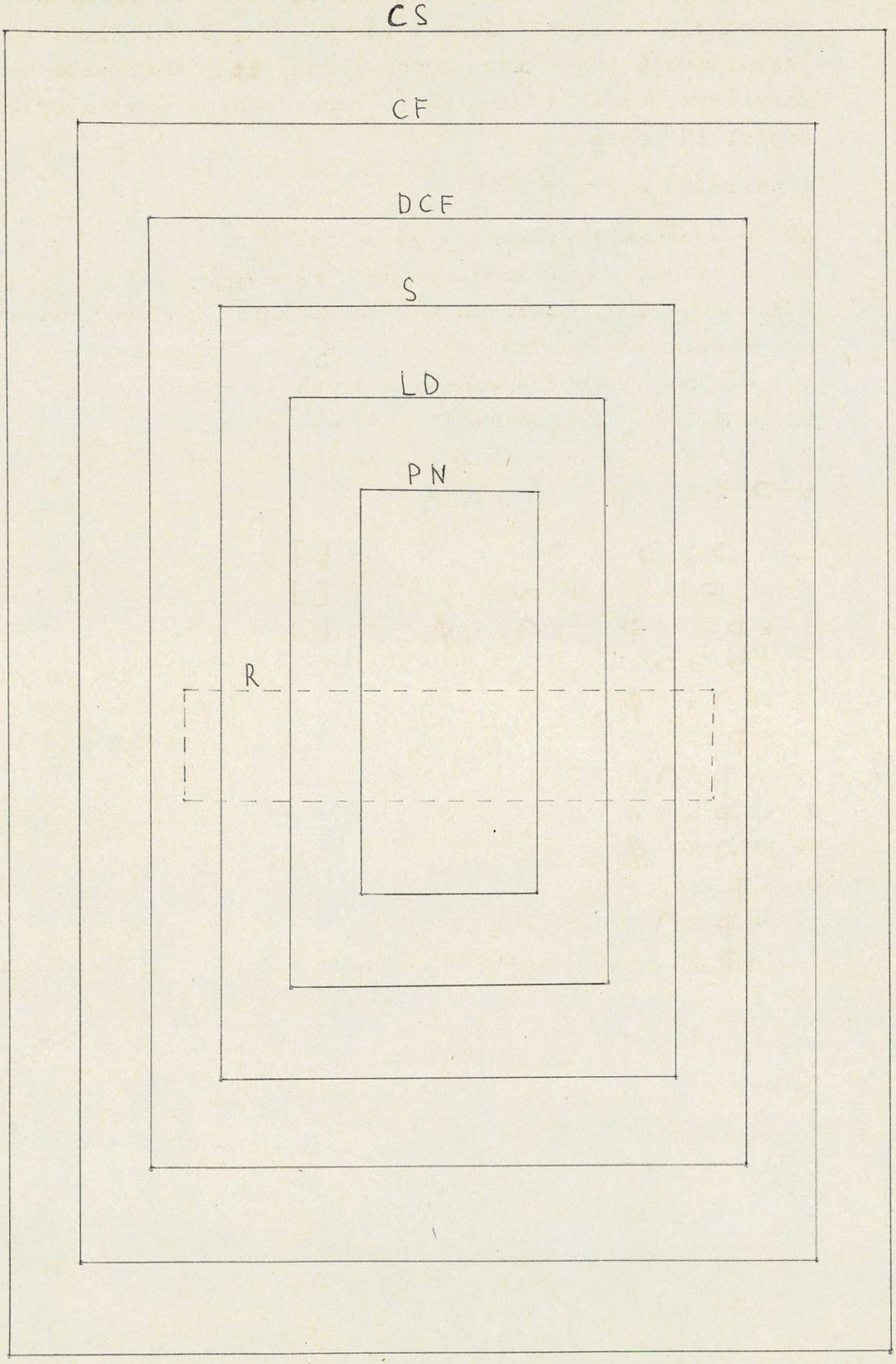
1. A tárgyalt nyelvosztályok közti tartalmazási viszonyok összefoglalása

A következő tételek és az ezeket szemléletes formában összefoglaló tuloldali ábra a tárgyalt nyelvosztályok közti tartalmazást adják meg. A tételeket itt közvetlenül nem bizonyitjuk, csak hivatkozunk a szakirodalom vagy a dolgozat megfelelő részeire.

A használt röviditések:

CS = kontextusérzékeny nyelvek halmaza
CF = a kontextusmentes nyelvek halmaza
DCF = a determinisztikus kontextusmentes nyelvek halmaza
S = az egyszerü determinisztikus nyelvek halmaza
LD = a baloldali levezetés nyelvek halmaza
PN = a lengyel jelölésmód nyelvek halmaza
R = a reguláris nyelvek halmaza
A $\supset$ jel valódi tartalmazást jelöl

1. $CS \supset CF \supset DCF \supset R$ [1d]
2. $DCF \supset S$, $DCF \supset S \cup R$ [12]
3. $S \not\supset R$, $R \not\supset S$, $S \cap R \neq \phi$ [12]
4. $S \supset LD \supset PN$ /3.2.5. és 4.2.5/
5. $LD \cap R \neq \phi$ /3.2.1. példa/
6. $R \not\supset LD$ /2.4.3. példa és 3.2.1. példa/
7. $LD \not\supset S \cap R$ /3.2.4./
8. $LD \not\supset S - R$ /3.2.4. és 2.4.4. példa/
9. $PN \cap R \neq \phi$ /4.2.3./
10. $R \not\supset PN$ /4.2.2./
11. $PN \not\supset R \cap LD$ /4.2.4./
12. $PN \not\supset LD - R$ /4.2.4. és 2.4.3. példa és 3.21. példa/


CS
CF
DCF
S
LD
PN
R

2. A nyelvosztályok és a müveletek közti összefüggések összefoglaló táblázata

A táblázat használata:

Ha az $\mathcal{N}$ nyelvosztály sorában és az $\mathfrak{m}$ művelet oszlopában az $\mathcal{M}$ nyelvosztály jele áll, ez azt jelenti, hogy az $\mathcal{M}$ a legszükebb olyan nyelvosztály /a felsoroltak közül/, amely az $\mathcal{N}$ bármely /vagy bármely két/ eleme* esetén tartalmazza a művelet eredményét.

Zárójelben adjuk meg a dolgozat megfelelő tételének /vagy tételeinek/ számát. A bekeretezett részben azok az állitások szerepelnek, amelyeknek a részletes bizonyitása /vagy a bizonyitás egy lényeges része/ a dolgozatban szerepel. Ha a "?" is szerepel, ez azt jelenti, hogy csak az bizonyitott, hogy az $\mathcal{M}$ tartalmazza a legszükebb nyelvosztályt.

*A homomorf leképezésnél: "és bármely homomorf leképezés".
A reguláris nyelvvel való metszetképzésnél: "és bármely reguláris nyelv".

N \ m	Egyesítés	Metszet-képzés	Szorzat-képzés	Komplemen-terképzés	Megforditás	Homomorf leképezés	Metszetkép-zés reguláris nyelvvel
CF	CF /2.5.2/	CS /2.5.1/ /2.5.3/	CF /2.5.2/	CS /2.5.1/ /2.5.3/	CF /2.5.2/	CF /2.5.2/	CF /2.5.2/
DCF	CF /2.5.2/ /2.5.5/	CS /2.5.1/ /3.4.7/	CF /2.5.2/ /2.5.5/	DCF /2.5.4/	CF /2.5.2/ /2.5.5/	CF /2.5.2/ /2.5.5/	DCF /2.5.4/
S	CF /2.5.2/ /3.4.6/	CS /2.5.1/ /3.4.7/	S /2.7.7/	DCF /2.7.8/ /2.5.4/	CF? /2.7.8/	CF /2.5.2/ /3.4.9/	DCF /2.7.8/ /2.5.4/
LD	CF /3.4.6/	CS /3.4.7/	S /3.4.11/	DCF /3.4.10/	CF? /3.4.12/	CF /3.4.9/	DCF /3.4.8/
PN	CF /4.4.5/	CS /4.4.5/	S /4.4.5/	DCF /4.4.5/	DCF /4.4.5/	CF? /2.5.2/ /4.4.3/	DCF /4.4.5/

3. A megoldhatósági kérdések összefoglaló táblázata

A táblázat használata:

A táblázatnak az M nyelvosztályhoz tartozó oszlopában és a k kérdéshez tartozó sorában lévő információ arra vonatkozóan ad tájékoztatást,* hogy van-e olyan algoritmus, amely bármely, M osztálybeli L nyelv /vagy L_1 és L_2 nyelvek/** esetén eldönti a k kérdést, vagyis a megfelelő probléma általánosan megoldható-e.

Az I igenlő, az N a tagadó választ jelenti, a zárójelben a dolgozat megfelelő tételét adjuk meg. Ha csak egy tételre utalunk, I vagy N nélkül, ez azt jelenti, hogy általánosan bizonyitás nem ismeretes, de bizonyos nemtriviális speciális esetekre vonatkozóan a tétel idevágó eredményt ad. Ha semmilyen utalás sem szerepel, ez azt jelenti, hogy nem ismeretes megfelelő tétel. A bekeretezett részben a dolgozatban bizonyitott állitások vannak.

* Kivéve az utolsó sort, amely nem szorul magyarázatra

** A harmadik kérdésnél: "és bármely u mondat"
A negyedik kérdésnél: "és bármely T véges ábécé"
Az ötödik kérdésnél: "és bármely R reguláris nyelv"
A tizenkettedik kérdésnél: "és bármely f homomorf leképezés"
A tizenharmadik kérdésnél: "és bármely R reguláris nyelv".

k \ M	CF	DCF	S	LD	PN
Az L nyelv üres?	I /2.6.2/	I /2.6.2/	I /2.6.2/	I /2.6.2/	H /4.5.2.1/
Az L nyelv véges?	I /2.6.2/	I /2.6.2/	I /2.6.2/	H /2.6.2/	H /4.5.2.1/
Az $\mathcal{U} \in L$ teljesül?	I /2.6.1/	I /2.6.1/	I /2.6.1/	H /2.6.1/	H /2.6.1/
Az $L = T^*$ teljesül?	N /2.6.3/	I /2.6.7/	I /2.7.9/	H /2.7.9/	H /2.7.9/
Az $L = R$ teljesül?	N /2.6.4/	I /2.6.7/	I /2.6.7/	H /2.6.7/	H /2.6.7/
Az L reguláris nyelv?	N /2.6.5/	I /2.6.8/	I /2.6.17/ /2.7.6/	H /2.6.17/ /2.7.6/	H /4.5.2.1/
Az $L_1 \wedge L_2$ nyelv üres?, véges?	N /2.6.9/	N /2.6.9/	N /2.7.11/	N /3.5.1.3/	H /4.5.5.8/
Az $L_1 = L_2$ teljesül?	N /2.6.11/		I /2.7.10/	H /2.7.10/	H /4.5.2.4/
Az $L_1 \subseteq L_2$ teljesül?	N /2.6.10/	N /2.6.10/		/3.5.3.12/	H /4.5.2.4/
Az $\overline{L}$ nyelv is ugyanebbe az osztályba tartozik?	N /2.6.6/	I /2.5.5/	I /2.7.8/	I /2.7.8/	H /2.7.8/

M / L	CF	DCF	S	LD	PN
Az L^R nyelv is ugyanebbe az osztályba tartozik?	I /2.5.2/				I /4.5.3.5/
Az $f(L)$ nyelv is ugyanebbe az osztályba tartozik?	I /2.5.2/	N /3.5.1.5/	N /3.5.1.6/	N /3.5.1.7/	I /4.5.4.5/
Az $L \cap R$ nyelv is ugyanebbe az osztályba tartozik?	I /2.5.2/	I /2.5.4/			I /4.5.3.4/
Az $L_1 \cup L_2$ nyelv is ugyanebbe az osztályba tartozik?	I /2.5.2/	N /2.6.12/			I /4.5.3.2/
Az $L_1 \cap L_2$ nyelv is ugyanebbe az osztályba tartozik?	N /2.6.14/	N /2.6.15/			I /4.5.5.8/
Az $L_1 L_2$ nyelv is ugyanebbe az osztályba tartozik?	I /2.5.2/	N /2.6.13/	I /2.7.7/	I /3.5.2.2/	I /4.5.3.3/
Van-e olyan algoritmus, amely bármely CF grammatika esetén eldönti, hogy az általa generált nyelv ebbe az osztályba tartozik-e vagy sem?		N /2.6.16/	N /2.7.12/	N /3.5.1.2/	N /4.5.1.1/

IRODALOMJEGYZÉK

1. CHOMSKY, N. On certain formal properties of grammars. Information and Control 2,2 /1959/ 137-167.

2. CHOMSKY, N. A note on phrase structure grammars. Information and Control 2,4 /1959/, 393-395.

3. CHOMSKY, N. és MILLER, G.A. Finite state languages. Information and Control 1,2 /1958/, 91-112.

4. GINSBURG, S. The mathematical theory of context-free languages. McGraw-Hill, New York, 1966.

5. GINSBURG, S. és GREIBACH, S.A. Deterministic context-free languages. Information and Control 9,6 /1966/ 620-648.

6. GINSBURG, S. és HARRISON, M.A. Bracketed context-free languages. Journal of Computer and System Sciences 1,1 /1967/ 1-23.

7. GREIBACH, S.A. A new normal-form theorem for context--free phrase structure grammars. Journal of the ACM 12,1 /1965/, 42-52.

8. GROSS, M. és LENTIN, A. Mathematische Linguistik. Springer-Verlag, Berlin, 1971.

9. HOPCROFT, J.E. On the equivalence and containment prog-lem for context-free languages. Mathematical System Theory 3,2 /1969/, 119-124.

10. HOPCROFT, J.E. és ULLMAN, J.D. Formal languages and their relation to automata Addison and Wesley, Reading, Mass., 1969.

11. KAIN, R.Y. Automata theory: machines and languages. McGraw-Hill, New York, 1972.

12. KORENJAK, A.J. és HOPCROFT, J.E. Simple deterministic languages. IEEE Annual Switching and Automata Theory Proceedings 7 /1966/, 36-46.

13. LANDWEBER, P.S. Decision problems of phrase structure grammars. IEEE Transactions on Electronic Computers 13 /1964/, 354-362.

14. MINSKY, M.L. Computation: Finite and infinite machines. Prentice-Hall, Engleword Cliffs, N.J. 1967.

15. WOOD, D. The normal form theorem-another proof. Computer Journal 12,2 /1969/, 139-147.