

A digitális számítás elmélete

3. előadás február 28.

- Grammatikák
- Környezetfüggetlen grammatikák és nyelvek
- Levezetés, levezetési fa, többértelmű CF nyelvek
- Chomsky-féle normálalak
- Környezetfüggetlen nyelvek jellemzése a ‘pumpáló lemmával’
- Chomsky-féle nyelvosztályok
- Programozási nyelvek szintaxisának megadása

Helyes angol mondatok nyelvtana

<SENTENCE> → <NOUN-PH> <VERB-PH>

<NOUN-PH> → <ARTICLE> <NOUN>

<NOUN-PH> → <ARTICLE> <NOUN> <PREP-PH>

<VERB-PH> → <VERB>

<VERB-PH> → <VERB> <NOUN-PH>

<VERB-PH> → <VERB> <NOUN-PH> <PREP-PH>

<PREP-PH> → <PREP> <ARTICLE> <NOUN>

- <ARTICLE> → a | the

- <NOUN> → boy | girl | flower | binocular

- <VERB> → sees | likes | touches

- <PREP> → with

Helyes angol mondatok nyelvtana

the boy sees

<ART> <NOUN> <VERB>

<NOUN-PH> < VERB-PH >

the boy sees a girl

<ART> <NOUN>

<ART> <NOUN> <VERB> <NOUN-PH>

<NOUN-PH> <VERB-PH>

the flower likes a boy

the flower likes the flower

Helyes angol mondatok nyelvtana

the boy	sees	the girl	with the binoculars
<ARTICLE> <NOUN>	<VERB>	<NOUN-PH>	<PREP-PH>
<NOUN-PH>		<VERB-PH>	

the boy	sees	the girl	with the binoculars
<ARTICLE> <NOUN>	<VERB>	<ART.> <NOUN>	<PREP-PH>
<ARTICLE> <NOUN>	<VERB>		<NOUN-PH>
<NOUN-PH>			<VERB-PH>

Formális nyelvek nyelvtana

A $\{0^i 1^i \mid i > 0\}$ nemreguláris nyelv nyelvtana

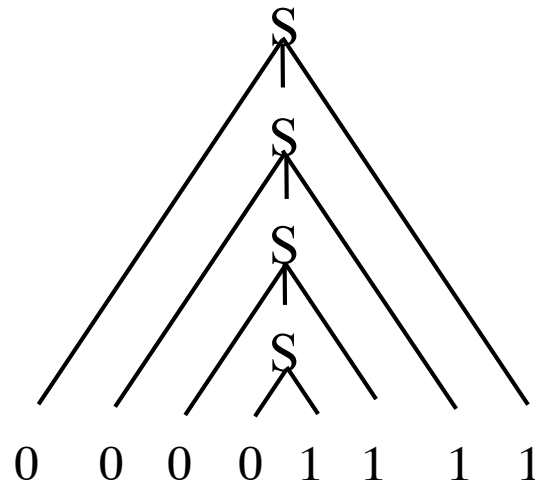
$S \rightarrow 0 S 1$

$S \rightarrow 0 1$

levezetési szabályok

terminálisok (a nyelv abc-je)

grammatikai változók
vagy nemterminális jelek



a **kezdő szimbólum**: egy
kitüntetett nemterminális
jel

Generatív grammatikák

Definíció: Egy $G = (V, \Sigma, R, S)$ négyes generatív grammatika, ahol:

- V egy véges halmaz, a **grammatikai változók vagy nemterminálisok** halmaza,
- Σ egy véges halmaz, diszjunkt V -vel, a **terminálisok halmaza**,
- R egy véges halmaz, a **levezetési szabályok halmaza**, ahol egy levezetési szabály alakja

$$\alpha \rightarrow \beta$$

ahol α és β terminálisok és nemterminálisok tetszőleges sorozata, de α -ban kell legyen egy nemterminális.

- $S \in V$ a **kezdő- vagy mondatszimbólum**.

Generatív grammatikák

Definíció: Egy $G = (V, \Sigma, R, S)$ grammatika alapján a $\kappa \in (V \cup \Sigma)^*$ szóból **közvetlenül levezethető** a $\mu \in (V \cup \Sigma)^*$ szó, ha léteznek olyan $\gamma, \delta \in (V \cup \Sigma)^*$ szavak és $\alpha \rightarrow \beta$ szabály R -ben, hogy:

$$\kappa = \gamma \alpha \delta, \quad \mu = \gamma \beta \delta.$$

Jelölés:

$$\kappa \Rightarrow_G \mu \text{ vagy } \kappa \Rightarrow \mu$$

Generatív grammatikák

Definíció: Tetszőleges u és v $(V \cup \Sigma)^*$ -beli szavak esetén, azt mondjuk, hogy **v levezethető az u -ból** a $G = (V, \Sigma, R, S)$ grammatika alapján, ha $u = v$, vagy léteznek $n \geq 0$ és $u = w_1, \dots, w_n = v$ $(V \cup \Sigma)^*$ -beli szavak úgy, hogy

$$w_1 \Rightarrow_G w_2 \Rightarrow_G \dots \Rightarrow_G w_n.$$

Jelölés:

$$u \Rightarrow_G^* v \text{ vagy } u \Rightarrow^* v$$

Generatív grammatikák

Definíció: A $G = (V, \Sigma, R, S)$ grammatika **generálja az** $L \subseteq \Sigma^*$
nyelvet, ha $L = L(G)$, ahol

$$L(G) = \{ w \in \Sigma^* \mid S \Rightarrow^* w \}$$

a G grammatika által generált nyelv.

Definíció: A G és G' grammatikák **ekvivalensek**, ha $L(G) = L(G')$

Rekurzív grammatikák

Definíció: A $G = (V, \Sigma, R, S)$ grammatika **rekurzív**, ha van olyan

$A \in V$, és $\gamma, \delta \in (V \cup \Sigma)^*$ szavak hogy

$$A \Rightarrow^* \gamma A \delta$$

Ekkor az A egy **rekurzív nemterminálisa** a grammatikának.

Megj:

Nemrekurzív grammatikák véges nyelvet generálnak.

Kérdés: Milyen nyelvet generálnak a rekurzív grammatikák?

Környezetfüggetlen (Context Free, CF) grammatikák

Definíció: A $G = (V, \Sigma, R, S)$ grammatika **környezetfüggetlen**, ha a levezetési szabályainak bal oldalán egyetlen nemterminális szerepel, azaz a szabályok

$$A \rightarrow \beta$$

alakúak, ahol $A \in V$, $\beta \in (V \cup \Sigma)^*$

1. Megj: Egy szabályaival megadott CF grammatika nemterminálisait nagybetűkkel, a terminálisokat kisbetűkkel jelöljük.

2. Megj: A kezdő szimbólum az első szabály bal oldalán szerepel (megállapodás).

3. Megj: $A \rightarrow \beta \mid \gamma \mid \delta$ forma gazdaságos szabályösszevonás.

Környezetfüggetlen (Context Free) grammatikák

1. Példa: $G = (\{S\}, \{a, b\}, R, S)$, és

$$R = \{ S \rightarrow aSb \mid SS \mid \varepsilon \}$$

~~baba~~

~~aba~~

~~bab~~

~~abba~~

abab

aaabbb

aababb

()()

((()))

((()()))

Környezetfüggetlen (Context Free) grammatikák

2. Példa: $G = (V, \Sigma, R, E)$, ahol:

$$V = \{ E, T, F \} \quad \Sigma = \{ (,), +, *, \text{id} \}$$

$$R = \{ E \rightarrow E + T \quad \text{expression} \quad R1$$

$$T \rightarrow T * F \quad \text{term} \quad R3$$

$$T \rightarrow F \quad R4$$

$$F \rightarrow (E) \quad \text{factor} \quad R5$$

$$F \rightarrow \text{id} \} \quad R6$$

Környezetfüggetlen (Context Free) grammatikák

$(id * id + id) * (id + id)$ kifejezése levezetése

$$E \Rightarrow T \Rightarrow T * F \Rightarrow T * (E) \Rightarrow T * (E+T) \Rightarrow$$

$$T * (T+T) \Rightarrow T * (F+T) \Rightarrow T * (id+T) \Rightarrow$$

$$T * (id+F) \Rightarrow T * (id+id) \Rightarrow F * (id+id) \Rightarrow$$

$$(E) * (id+id) \Rightarrow (E+T) * (id+id) \Rightarrow (E+F) * (id+id) \Rightarrow$$

$$(E+id) * (id+id) \Rightarrow (T+id) * (id+id) \Rightarrow (T * F+id) * (id+id) \Rightarrow$$

$$(F * F+id) * (id+id) \Rightarrow (F * id+id) * (id+id) \Rightarrow (id * id+id) * (id+id)$$

$E \rightarrow E+T$	R1
$E \rightarrow T$	R2
$T \rightarrow T * F$	R3
$T \rightarrow F$	R4
$F \rightarrow (E)$	R5
$F \rightarrow id$	R6

Egy reguláris nyelv környezetfüggetlen-e?

Tétel: Igen.

Biz: Ha L reguláris, akkor van őt felismerő M dva. Készítünk egy éppen L -et generáló CF grammatikát.

M minden q_i állapotának megfeleltetünk egy R_i nemterminális szimbólumot, a kezdőszimbólum R_0 .

$R_i \rightarrow aR_k$ szabály, ha $\delta(r_i, a) = r_k$

$R_i \rightarrow \varepsilon$ szabály, ha r_i végállapot M -ben.

Az így definiált grammatika éppen az L -beli szavakat generálja valóban.

Megj.: A kapott grammatika egy ún. jobbreguláris grammatika.

Egyértelmű-e mindig egy CF grammatika
szerinti levezetés?

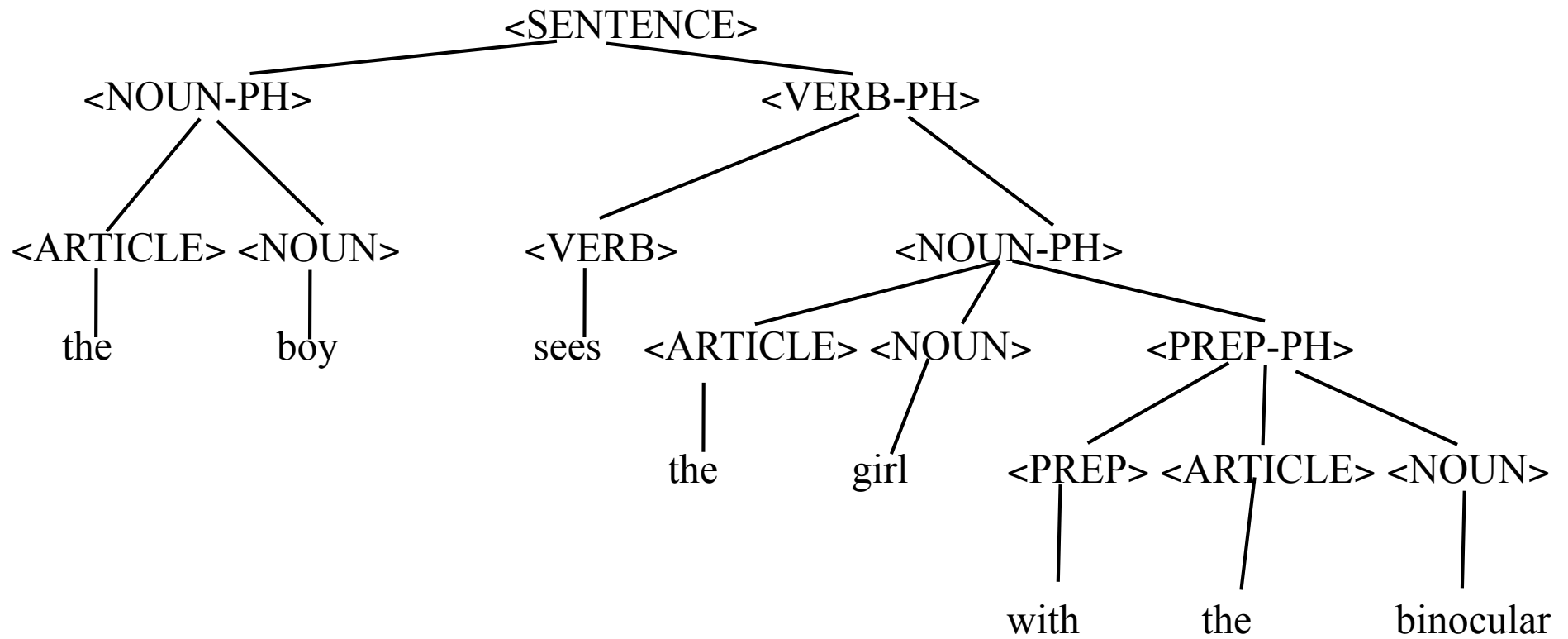
Válasz: Nem mindig, mert a szabályok alkalmazásának sorrendje
tetszőleges lehet.

Levezetési fa

Definíció: Egy CF G grammatika szerinti levezetés olyan fával ábrázolható, melynek **gyökere** a kezdőszimbólum, **levelei** balról jobbra a levezetett szó betűi, további **csomópontjai nemterminálisok**, és az elágazások egy-egy alkalmazott szabálynak felelnek meg.
Ez a **levezetési fa**.

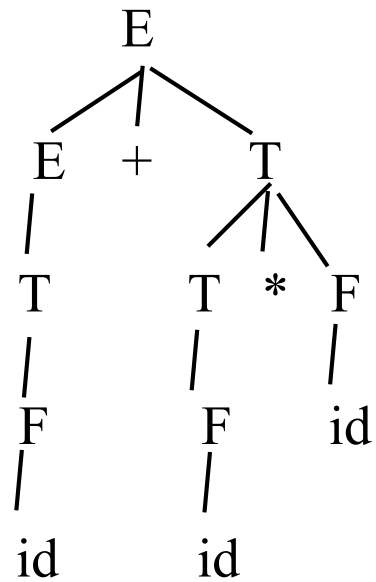
Levezetési fa

1. Pl:



Levezetési fa

2. Pl: $\text{id} + \text{id} * \text{id}$



$E \rightarrow E + T$	R1
$E \rightarrow T$	R2
$T \rightarrow T * F$	R3
$T \rightarrow F$	R4
$F \rightarrow (E)$	R5
$F \rightarrow \text{id}$	R6

Környezetfüggetlen (Context Free) grammatikák

Definíció: Egy **környezetfüggetlen grammatika szerinti levezetés baloldali**, ha a levezetés egyes lépéseiben a legbaloldalibb nemtermimálist helyettesítjük be.

Definíció: Egy nyelvtan **egyértelmű**, ha minden generált szavának egyetlen baloldali levezetése van. Különben **többértelmű**.

Megj: Minden baloldali levezetés meghatároz egy levezetési fát. Tehát egyértelmű nyelvtan esetén minden generált szóhoz egyetlen levezetési fa létezik.

Definíció: Egy CF nyelv **eredendően többértelmű**, ha nem generálható egyértelmű grammatikával.

Egyértelmű-e mindig egy CF grammatika szerinti *baloldali* levezetés?

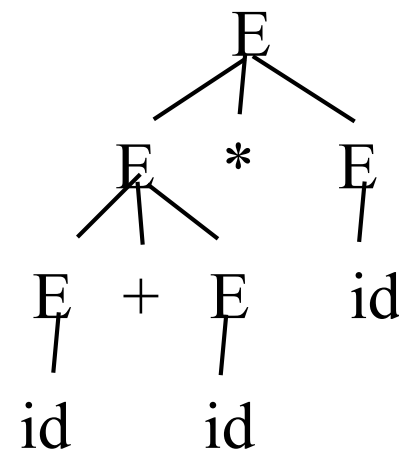
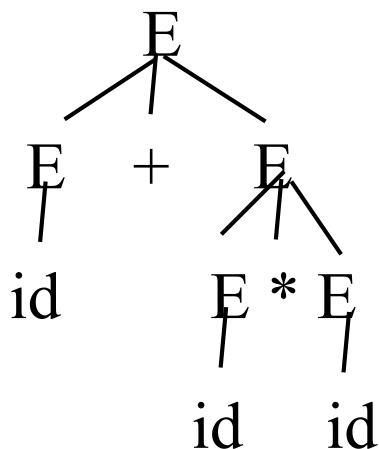
Válasz: Nem mindig.

1. Pl: the boy sees the girl with the flower

2. Pl: $R' = \{E \rightarrow E+E, E \rightarrow E * E, E \rightarrow (E), E \rightarrow id\}$

$E \rightarrow E+T$
 $E \rightarrow T$
 $T \rightarrow T * F$
 $T \rightarrow F$
 $F \rightarrow (E)$
 $F \rightarrow id$

$id + id * id$ két baloldali levezetése:



Chomsky-féle normál alak (CNF)

Definíció: A $G = (V, \Sigma, R, S)$ CF grammatika **Chomsky-féle normál alakban** adott, ha a szabályok az alábbi alakúak:

$$A \rightarrow BC$$

$$A \rightarrow a$$

$$S \rightarrow \varepsilon$$

ahol $A, B, C \in V$, $a \in \Sigma$, B és C nem kezdőszimbólumok.

Tétel: Tetszőleges CF nyelvhez megadható őt generáló grammatika Chomsky-féle normál alakban.

Chomsky-féle normálalak (CNF)

Bizonyítás:

1. $S_0 \rightarrow S$ új kezdő szimbólum
2. Minden $A \rightarrow \varepsilon$ szabályra:
 - a) töröljük $A \rightarrow \varepsilon$ szabályt, ahol A nem kezdősz.
 - b) $B \rightarrow \alpha A \beta$ szabályokhoz adjunk újabbakat úgy, hogy töröljük A-t, minden lehetséges módon!
pl: $B \rightarrow \alpha A \beta A \gamma$ esetén a köv. 3 szabályt iktatjuk be:
 $B \rightarrow \alpha \beta A \gamma$ $B \rightarrow \alpha A \beta \gamma$ $B \rightarrow \alpha \beta \gamma$
 - c) $B \rightarrow A$ esetén a $B \rightarrow \varepsilon$ szabályt csak akkor iktatjuk be, ha korábban a)-ban nem töröltünk már ilyen szabályt.
3. Minden $A \rightarrow C$ szabályra:
 - a) töröljük $A \rightarrow C$ szabályt
 - b) ha van $C \rightarrow \alpha$ szabály, iktassuk be az $A \rightarrow \alpha$ -t, kivéve ha azt már valamikor töröltük a) folyamán.

Chomsky-féle normálalak (CNF)

4. Végül konvertáljuk a megmaradt és beszűrt szabályokat max 2 hosszú jobboldalas alakra:

a) $A \rightarrow \alpha_1 \alpha_2 \dots \alpha_k$ ahol $k \geq 3$, α_i terminális vagy nemterminális.

Helyettesítjük ezt új nemterminálisok és szabályok beiktatásával:

$$A \rightarrow \alpha_1 A_1, \quad A_1 \rightarrow \alpha_2 A_2, \quad A_2 \rightarrow \alpha_3 A_3, \quad \dots \quad A_{k-2} \rightarrow \alpha_{k-1} \alpha_k$$

b) Minden a)-ban hozzáadott, valamint legalább 2 hosszú jobb oldalú szabályban helyettesítsük α_i terminálist U_i nemterminálissal, és iktassuk be a

$U_i \rightarrow \alpha_i$ szabályt.

CF nyelvek jellemzése pumpáló lemmával

Tétel: (Pumpáló lemma, Bar-Hillel lemma):

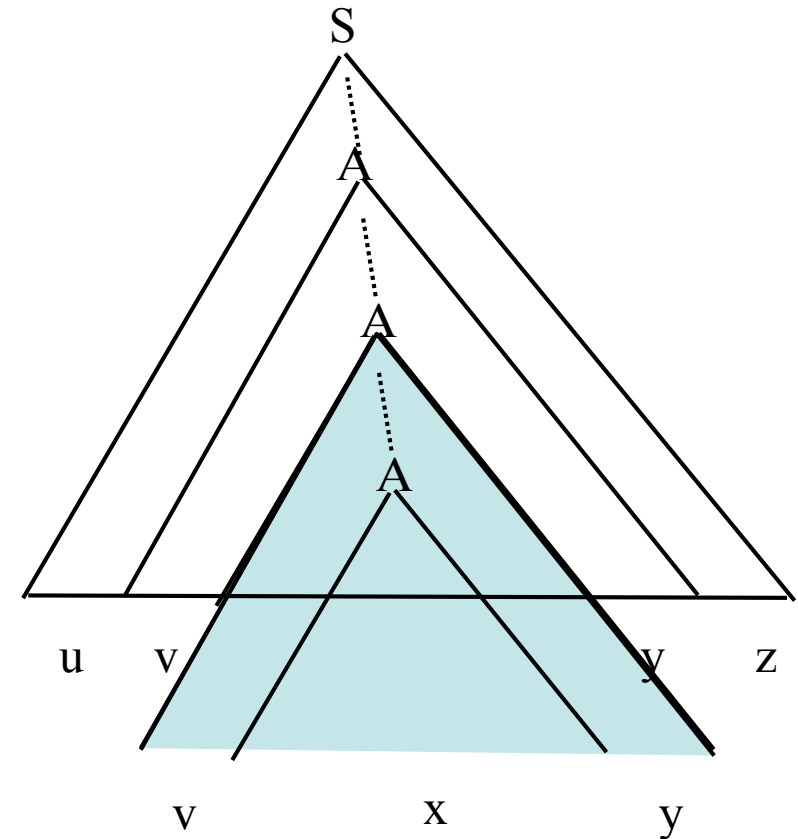
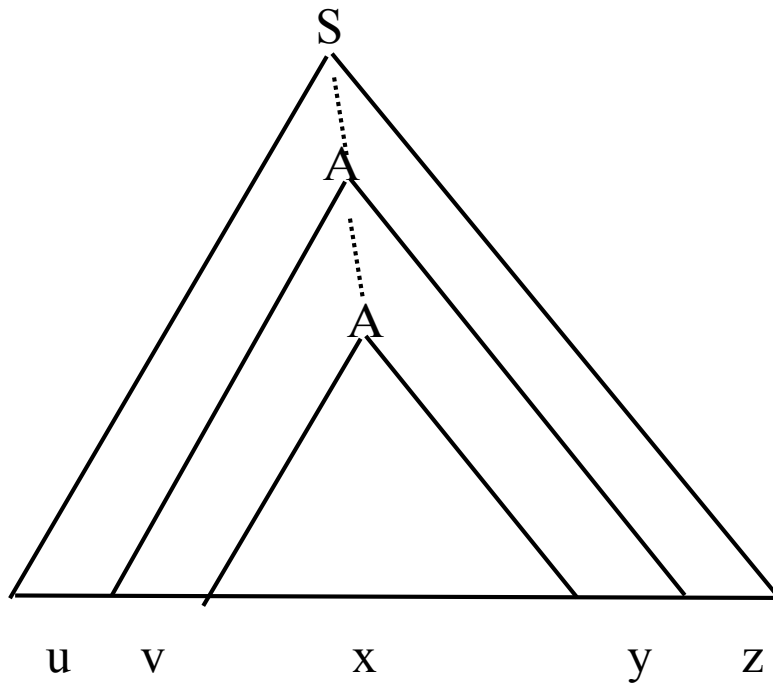
Ha $L \subseteq \Sigma^*$ egy CFL, akkor $\exists p \in \mathbb{N}$ szám, hogy

$\forall s \in L$ esetén, ha s hossza legalább p , akkor s felbontható $s = uvxyz$ alakba úgy, hogy:

1. $uv^nxy^nz \in L$ minden $n \geq 0$ esetén,
2. $|vy| > 0$,
3. $|vxy| \leq p$.

CF nyelvek jellemzése pumpáló lemmával

Biz: Ötlet: s jó hosszú L-beli sorozat, ekkor levezetésében 2-szer szerepel egy A



CF nyelvek jellemzése pumpáló lemmával

Bizonyítás:

Legyen G egy, az adott L -t generáló CFG, és b legyen a szabályok jobb oldalán szereplő szimbólumok max. száma, feltételezzük hogy $b \geq 2$. Tetszőleges m mélységű levezetési fában max. b^m levél szerepel.

Legyen $p = b^{|V|+2}$, ahol $|V|$ a G -beli változók száma. Mivel $b \geq 2$, így $p > b^{|V|+1}$, ezért ha s L -beli legalább p hosszú sorozat, akkor minden levezetési fája legalább $|V|+2$ mély kell hogy legyen.

Legyen s . L -beli legalább p hosszú sorozat, legyen τ s egy legkevesebb csomópontú levezetési fája. τ legalább $|V|+2$ mély, ezért a leghosszabb út a gyökértől egy levélig is legalább $|V|+2$. Ez az út legalább $|V|+1$ változót tartalmaz, ekkor a galambdúc elv miatt van olyan változó, mely legalább 2-szer szerepel az út végső, $|V|+1$ hosszú szakaszán. Legyen ez A .

A pumpáló lemma használata

Pl: $L = \{ a^n b^n c^n \mid 0 \leq n \}$ **nem** CFL.

Biz: A pumpáló lemmát használjuk, indirekten bizonyítunk.

Tfel indirekten, hogy L CFL, ekkor p jelölje a pumpáló hosszt.
Legyen $s = a^p b^p c^p$. Ekkor $s = uvxyz$.

1. Ha v is és y is egy-egy fajta terminálist tartalmaznak, akkor uv^2xy^2z nem tartalmaz mindhárom terminálisból ugyanannyit, így uv^2xy^2z nem L -beli.
2. Ha pl. v tartalmaz több terminálist, akkor v^2 -ben azok sorrendje lesz nem megfelelő, így uv^2xy^2z nem L -beli.

A fenti esetek valamelyike fenn kell hogy álljon, de azok mind ellentmondásra vezettek, így az indirekt feltevés hamis.

Chomsky-féle nyelvosztályok

A nyelvet generáló grammatika szabályaira tett megkötések alapján:

3-as nyelvosztály:

reguláris (R)

$A \rightarrow a$, $A \rightarrow aB$ alakú szabályok

2-es nyelvosztály:

környezetfüggetlen (CF)

$A \rightarrow \alpha$ alakú szabályok

1-es nyelvosztály (k:

környezetfüggő (CS)

$\beta A \gamma \rightarrow \beta \alpha \gamma$ alakú szabályok

$\alpha \rightarrow \beta$ alakú szabályok ahol $|\alpha| \leq |\beta|$ nem csökkenő

megj: ezek ekvivalensek!

0-ás nyelvosztály:

tetszőleges alakú szabályok

1.Megj: egyre bővebb nyelvosztályok

2. Megj: nyelv jellemzése az öt generáló grammatika alapján

Backus-Naur Form (BNF)

programozási nyelvek definiálására

- meta-nyelv, eredetileg 59-60-ból, az Algol és Fortran definiálására
- Turing-díj Backusnak
- változatok ma is programozási nyelvek megadására
- <http://cs.wpi.edu/~kal/PLT/PLT2.1.2.html>
- <http://cui.unige.ch/db-research/Enseignement/analyseinfo/BNFweb.html>
- ha egy szintaxis BNF formában adott, vannak eszközök, melyek automatikusan generálnak egy elemzőt, mely a szintaktikus helyességet ellenőrzi, és
levezetési fát produkál, sőt le is fordítja a programot
<http://www.thefreecountry.com/developercity/constructcompilers.shtml>

Miről volt szó ?

- Grammatikák
- Környezetfüggetlen grammatikák és nyelvek
- Levezetés, levezetési fa, többértelmű CF nyelvek
- Chomsky-féle normálalak
- Környezetfüggetlen nyelvek jellemzése a ‘pumpáló lemmával’
- Chomsky-féle nyelvosztályok
- BNF, Programozási nyelvek szintaxisának megadása