

A számításelmélet alapjai I.

2. előadás

előadó: Tichler Krisztián
ktichler@inf.elte.hu

Grammatikák – Ismétlés

Definíció

Egy $G = \langle N, \Sigma, P, S \rangle$ rendezett négyest **grammatikának** nevezünk ha

- ▶ N és Σ diszjunkt véges ábécék (azaz $N \cap \Sigma = \emptyset$). N elemeit **nemterminális**, Σ elemeit pedig **terminális** szimbólumoknak nevezzük.
- ▶ $S \in N$ a grammatika **kezdőszimbóluma**.
- ▶ A P **szabályrendszer** $x \rightarrow y$ alakú szabályok véges halmaza, ahol $x \in (N \cup \Sigma)^* N (N \cup \Sigma)^*$, $y \in (N \cup \Sigma)^*$.

Észrevétel: $(N \cup \Sigma)^* - \Sigma^* = (N \cup \Sigma)^* N (N \cup \Sigma)^*$.

Grammatikák – Ismétlés

Definíció

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

u -ból (több lépésben vagy közvetetten) **levezethető** v , ha $u = v$ vagy van olyan $n \geq 1$ és $w_0, \dots, w_n \in (N \cup \Sigma)^*$, hogy $w_{i-1} \Rightarrow_G w_i$ ($1 \leq i \leq n$) és $w_0 = u$ és $w_n = v$. Jelölés: $u \Rightarrow_G^* v$.

Mondatforma: A kezdőszimbólumból levezethető szó.

\Rightarrow_G illetve \Rightarrow_G^* helyett gyakran röviden \Rightarrow -t illetve \Rightarrow^* -t írunk.

Definíció

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

Grammatikák Chomsky féle osztályozása

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

- ▶ $i = 0$ eset: nincs korlátozás,
- ▶ $i = 1$ eset:
 - (1) P minden szabálya $u_1 A u_2 \rightarrow u_1 v u_2$ alakú, ahol $u_1, u_2, v \in (N \cup \Sigma)^*$, $A \in N$, és $v \neq \varepsilon$,
 - (2) Egyetlen kivétel megengedünk: P tartalmazhatja az $S \rightarrow \varepsilon$ szabályt, de csak abban az esetben, ha S nem fordul elő P egyetlen szabályának jobb oldalán sem.

("Korlátozott ε szabály" vagy röviden "KES")
- ▶ $i = 2$ eset: P minden szabálya $A \rightarrow v$ alakú, ahol $A \in N$ és $v \in (N \cup \Sigma)^*$,
- ▶ $i = 3$ eset: P minden szabálya vagy $A \rightarrow uB$ vagy $A \rightarrow u$, alakú, ahol $A, B \in N$ és $u \in \Sigma^*$.

Jelölje \mathcal{G}_i az i -típusú grammatikák osztályát ($i = 0, 1, 2, 3$).

Grammatikák osztályba sorolása

Példa: Legyen $G = \langle N, \Sigma, P, S \rangle$, ahol $N = \{S, A, B\}$, $\Sigma = \{a, b\}$ és $P = \{S \rightarrow ASB, S \rightarrow \varepsilon, AB \rightarrow BA, BA \rightarrow AB, A \rightarrow a, B \rightarrow b\}$.

Szabályról szabályra nézzük meg, hogy az adott szabály melyik típusba engedi sorolni a grammatikát.

$S \rightarrow ASB$ 0, 1[•], 2

$S \rightarrow \varepsilon$ 0, 1[•], 2, 3

$AB \rightarrow BA$ 0

$BA \rightarrow AB$ 0

$A \rightarrow a$ 0, 1, 2, 3

$B \rightarrow b$ 0, 1, 2, 3

•: Az 1. és 2. szabály **EGYÜTT** megsérti a KES-t. Ha együtt szerepelnek egy grammatikában, akkor az $\notin \mathcal{G}_1$.

A 3. és 4. szabály a 0-son kívül mindent kizár, így $G \in \mathcal{G}_0$ és $G \notin \mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$.

Grammatikák osztályba sorolása

Legyen G egy grammatika. A következő esetek lehetségesek:

- ▶ $G \in \mathcal{G}_3, \mathcal{G}_2, \mathcal{G}_1, \mathcal{G}_0$.
Lokálisan minden szabály 3-as és teljesül a KES.
- ▶ $G \in \mathcal{G}_3, \mathcal{G}_2, \mathcal{G}_0$.
Lokálisan minden szabály 3-as, de NEM teljesül a KES.
- ▶ $G \in \mathcal{G}_2, \mathcal{G}_1, \mathcal{G}_0$.
Lokálisan minden szabály 2-es, van nem 3-as szabály és teljesül a KES.
- ▶ $G \in \mathcal{G}_2, \mathcal{G}_0$.
Lokálisan minden szabály 2-es, van nem 3-as szabály, NEM teljesül a KES.
- ▶ $G \in \mathcal{G}_1, \mathcal{G}_0$.
Lokálisan minden szabály 1-es, van nem 2-as szabály és teljesül a KES.
- ▶ $G \in \mathcal{G}_0$.
Van nem 1-es szabálya vagy minden szabály 1-es, van nem 2-es szabálya és NEM teljesül a KES.

Nyelvek Chomsky féle osztályozása

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



$\mathcal{L}_i := \{L \mid \exists G \in \mathcal{G}_i, \text{ hogy } L = L(G)\}$ jelöli az i -típusú nyelvek nyelvosztályát, elemeit **i -típusú nyelveknek** nevezzük ($i = 0, 1, 2, 3$).

Tehát az i -típusú nyelvek azok, amiket i -típusú grammatikával generálni lehet. ($i = 0, 1, 2, 3$)

A Chomsky féle grammatikaosztályok elnevezése

- ▶ A 0-típusú grammatikákat **mondatszerkezetű** (phrase-structure) grammatikáknak is nevezzük, ami nyelvészeti eredetükre utal.
- ▶ A 1-típusú grammatikák a **környezetfüggő** (context-sensitive) grammatikák, mert szabályai olyanok, hogy, egy A nemterminális valamely előfordulása egy v szóval csak u_1 és u_2 kontextus jelenlétében helyettesíthető.
- ▶ 2-típusú grammatikákat **környezetfüggetlen** (context-free) grammatikáknak mondjuk, mert szabályai olyanok, hogy az A nemterminális v -vel való helyettesítése bármely kontextusban megengedett.
- ▶ A 3-típusú grammatikákat **reguláris** (regular) vagy véges állapotú (finite state) grammatikáknak hívjuk, a véges állapotú automatákkal való kapcsolatuk miatt.

A 0,1,2,3-típusú nyelvek osztályait rendre **rekurzíven felsorolható**, **környezetfüggő**, **környezetfüggetlen**, valamint **reguláris** nyelvosztálynak is mondjuk.

Grammatikák a gyakorlatban

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

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

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

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

A BNF megfelel a környezetfüggetlen grammatikának. Példa:

$$\langle \text{azonosító} \rangle ::= \langle \text{betű} \rangle \mid \langle \text{betű} \rangle \langle \text{azvég} \rangle$$
$$\langle \text{azvég} \rangle ::= \langle \text{betű} \rangle \mid \langle \text{számjegy} \rangle \mid \langle \text{betű} \rangle \langle \text{azvég} \rangle \mid \langle \text{számjegy} \rangle \langle \text{azvég} \rangle$$
$$\langle \text{betű} \rangle ::= a \mid \dots \mid z$$
$$\langle \text{számjegy} \rangle ::= 0 \mid \dots \mid 9$$

A Chomsky-féle hierarchia

A grammatikák alakja alapján $\mathcal{G}_3 \subseteq \mathcal{G}_2 \subseteq \mathcal{G}_0$ és $\mathcal{G}_1 \subseteq \mathcal{G}_0$.

Ebből azonnal következik, hogy $\mathcal{L}_3 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_0$ és $\mathcal{L}_1 \subseteq \mathcal{L}_0$.

Később meg fogjuk mutatni azt is hogy a következő tétel fennáll:

Chomsky nyelvhierarchia tétele

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

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

(2) Nem világos a valódi tartalmazás sem.

(Gyengén) ekvivalens grammatikák és nyelvek

Amíg minden grammatika egyetlen nyelvet generál, ez fordítva nem igaz. Ugyanazt a nyelvet több különböző grammatika is generálhatja.

Definíció

A G_1 és G_2 grammatikák **ekvivalensek** egymással, ha ugyanazt a nyelvet generálják, azaz $L(G_1) = L(G_2)$ teljesül.

Definíció

Az L_1 és L_2 nyelveket egymással **gyengén ekvivalensnek** nevezzük, ha legfeljebb az ε tartalmazásában térnek el egymástól, azaz ha $(L_1 - L_2) \cup (L_2 - L_1) \subseteq \{\varepsilon\}$.

Definíció

A G_1 és G_2 grammatikákat **gyengén ekvivalensnek** nevezzük, ha egymással gyengén ekvivalens nyelveket generálnak, azaz ha $(L(G_1) - L(G_2)) \cup (L(G_2) - L(G_1)) \subseteq \{\varepsilon\}$.

Ekvivalens grammatikák

Példa

A következő példákban $N \subseteq \{S, S', A, B\}$, $\Sigma = \{a, b\}$ és S a kezdőszimbólum. Csak a szabályrendszert adjuk meg. Ha $\alpha \rightarrow \beta$ és $\alpha \rightarrow \gamma$ két szabály, akkor tömören $\alpha \rightarrow \beta | \gamma$ -t írunk.

a) $S \rightarrow ASB | \varepsilon$, $AB \rightarrow BA$, $BA \rightarrow AB$, $A \rightarrow a$, $B \rightarrow b$

0. típusú

b) $S \rightarrow ASB | \varepsilon$, $AB \rightarrow BA$, $A \rightarrow a$, $B \rightarrow b$

0. típusú

c) $S \rightarrow S' | \varepsilon$, $S' \rightarrow AS'B | AB$, $AB \rightarrow BA$, $A \rightarrow a$, $B \rightarrow b$

0. típusú

d) $S \rightarrow SS | aSb | bSa | \varepsilon$.

2. típusú

$L(G) = L_{\text{EQ}} := \{u \in \{a, b\}^* \mid u\text{-ban ugyanannyi } a \text{ van, mint } b\}$. (4x)

Nem adható 3-as típusú grammatika a nyelvhez (biz. később).

Ekvivalens grammatikák

Példa

A (a) -t már bizonyítottuk, (b) és (c) az alapján könnyen bizonyítható, ezért (d) -t bizonyítjuk:

G szabályai: $S \rightarrow SS \mid aSb \mid bSa \mid \varepsilon$.

$$L(G) \subseteq L_{EQ}$$

A legrövidebb levezetés hosszára vonatkozó teljes indukcióval belátjuk, hogy minden $w \in \{a, b, S\}^*$ mondatformában ugyanannyi a van mint b . Tegyük fel, hogy w legrövidebb S -ből való levezetésének hossza n . Az $n = 0$ eset nyilvánvaló. Tegyük fel, hogy minden n -nél kevesebb lépésben levezethető mondatformára igaz az állítás.

Legyen w' az a mondatforma w egy n hosszú levezetésben, amiből w -t közvetlenül levezettük. Ekkor w' -nek van $n - 1$ hosszúságú levezetése, így az indukciós feltevés szerint w' -re igaz az állítás. Ha w' -ben megegyezik az a -k és a b -k száma, akkor ez G szabályai alapján w -re is fenn fog állni.

Ekvivalens grammatikák

Példa

G szabályai: $S \rightarrow SS \mid aSb \mid bSa \mid \varepsilon$.

$L(G) \supseteq L_{EQ}$

Legyen most $w \in L_{EQ}$, ahol $|w| = n$. n -re vonatkozó indukcióval bizonyítjuk, hogy $w \in L(G)$.

$n = 0$ -ra az állítás nyilvánvaló. Tegyük fel hogy az állítás minden n -nél rövidebb szóra igaz. $n \geq 1$ esetén 3 eset van.

1. eset: $w = aw'b$, valamely $w' \in L_{EQ}$ -ra. Ekkor az indukciós feltevés szerint $S \Rightarrow^* w'$. Így $S \Rightarrow aSb \Rightarrow^* aw'b = w$.

2. eset: $w = bw'a$. Ez az eset az előzőhöz hasonlatos.

3. eset: $w = aw'a$ vagy $w = bw'b$. Ekkor léteznek olyan $u, v \in \{a, b\}^+$ szavak, melyre $w = uv$ és $u, v \in L_{EQ}$. Világos, hiszen w 1 és az $n - 1$ hosszú prefixében különböző előjelű az a -k száma minusz a b -k száma, így kell legyen olyan köztes hossz, amire 0. Ekkor az indukciós feltevés szerint $S \Rightarrow^* u$ és $S \Rightarrow^* v$. Így $S \Rightarrow SS \Rightarrow^* uS \Rightarrow^* uv = w$.

Terminálisok eliminációja a szabályok baloldalán

Tétel:

Minden $G = \langle N, \Sigma, P, S \rangle$ grammatikához van vele ekvivalens és azonos típusú $G' = \langle N', \Sigma, P', S \rangle$ grammatika, melyre $x \in (N')^+$ minden $x \rightarrow y \in P'$ -re.

Bizonyítás: $i = 2, 3$ -ra a grammatika eleve ilyen alakú.

$i = 0, 1$ -re legyen $\bar{\Sigma} = \{\bar{a} \mid a \in \Sigma\}$, Feltehető, hogy $N \cap \bar{\Sigma} = \emptyset$.

$N' := N \cup \bar{\Sigma}$. Helyettesítsünk P -ben minden $a \in \Sigma$ -t \bar{a} -val (a jobboldalakon is!), legyen az így módosított szabályrendszer P_1 . Legyen $\bar{P} = \{\bar{a} \rightarrow a \mid a \in \Sigma\}$ és $P' := P_1 \cup \bar{P}$. Látható, hogy nem fordul elő terminális szabály baloldalán.

Ez a grammatika az eredetivel ekvivalens, azaz $L(G) = L(G')$.

Terminálisok eliminációja a szabályok baloldalán

$L(G) \subseteq L(G')$: Ha $u = t_1 \cdots t_n \in L(G)$, $t_i \in \Sigma$ ($1 \leq i \leq n$), akkor nyilván $S \Rightarrow_{G'}^* \bar{t}_1 \cdots \bar{t}_n$ és ebből \bar{P} szabályainak alkalmazásával $S \Rightarrow_{G'}^* u$, azaz $u \in L(G')$ adódik.

$L(G) \supseteq L(G')$:

Definiáljuk a $h : (N' \cup \Sigma)^* \rightarrow (N \cup \Sigma)^*$ homomorfizmust úgy, hogy $h(x) := x$ minden $x \in (N \cup \Sigma)$ és $h(\bar{a}) := a$ minden $\bar{a} \in \bar{\Sigma}$ esetén.

Ha $u \Rightarrow_{G'} v$, valamely P' -beli szabály alkalmazásával, akkor

- ▶ \bar{P} -beli szabály alkalmazása esetén $h(u) = h(v)$.
- ▶ P_1 -beli szabály alkalmazása esetén $h(u) \Rightarrow_G h(v)$

Tehát $u \Rightarrow_{G'} v$ implikálja $h(u) \Rightarrow_G^* h(v)$. (valójában 0 vagy 1 lépésben)

Terminálisok eliminációja a szabályok baloldalán

Ebből indukcióval könnyen meggondolható, hogy fennáll:

$u \Rightarrow_G^* v$ implikálja $h(u) \Rightarrow_G^* h(v)$.

Azaz, ha $S \Rightarrow_G^* w$, ahol $w \in \Sigma^*$, akkor $S = h(S) \Rightarrow_G^* h(w) = w$.

1. típus esetén a szabályok megfelelő alakúak.

Megjegyzés: A 0. típusú átalakítás a 2. típusra is alkalmazható, így a 0,1,2 típusok esetén feltehető, hogy terminális, csak $A \rightarrow a$ alakban fordul elő ($a \in \Sigma, A \in N$).

A 2-típusú grammatikák normálformájánál használjuk is majd a fenti ötletet.

A bizonyítás során bevezetett új nemterminálisokat, azaz $\bar{\Sigma}$ elemeit ($\bar{\Sigma} = \{\bar{a} \mid a \in \Sigma\}$) **áterminálisoknak** fogjuk nevezni és erre a transzformációs lépésre „**áterminálisok bevezetése**” néven fogunk a későbbiekben hivatkozni.

Zártsági tulajdonságok

Definíció

Az unió, konkatenáció és iteratív lezárás műveleteket **reguláris műveleteknek** nevezzük.

Egy \mathcal{L} nyelvcsalád **zárt** a $\varphi : (L_1, \dots, L_n) \mapsto \varphi(L_1, \dots, L_n)$ n -változós nyelvműveletre nézve, ha minden $L_1, \dots, L_n \in \mathcal{L}$ esetén $\varphi(L_1, \dots, L_n) \in \mathcal{L}$.

Tétel

\mathcal{L}_i zárt a reguláris műveletekre nézve ($i = 0, 1, 2, 3$).

Bizonyítás: i -típusú nyelv azt jelenti, hogy i -típusú grammatika generálja. A feladat az, hogy ezekből a grammatikákból konstruáljunk egy olyan grammatikát, mely

- (1) a nyelvek unióját, konkatenáltját illetve lezártját generálja
- (2) az új grammatika tartsa meg az eredeti grammatika(k) típusát.

Tehát a bizonyítás $3 \times 4 = 12$ konstrukcióból áll (de lesz köztük közös).

Zártsági tulajdonságok

Az előző tétel alapján feltehető, hogy a szabályok baloldalán nem fordul elő terminális. (Valójában erre csak a 0. típus konkatenáció valamint a 0,1 típusú lezárás műveleteknél lesz szükségünk.)

Feltehető továbbá, hogy az eredeti grammatikák nemterminális ábécéi diszjunktak, és hogy ezek diszjunktak a terminális ábécék uniójától.

Unió:

Legyen $L_k \in \mathcal{L}_i$ és $G_k = \langle N_k, \Sigma_k, P_k, S_k \rangle \in \mathcal{G}_i$ olyan, hogy $L(G_k) = L_k$ ($k = 1, 2$).

$i = 0, 2, 3$ eset:

Legyen $S_0 \notin N_1 \cup N_2$.

$G_{\cup} := \langle N_1 \cup N_2 \cup \{S_0\}, \Sigma_1 \cup \Sigma_2, P_1 \cup P_2 \cup \{S_0 \rightarrow S_1 \mid S_2\}, S_0 \rangle$.

Ekkor G_{\cup} i -típusú és $L(G_{\cup}) = L_1 \cup L_2$.

Zártsági tulajdonságok

Unió:

$i = 1$ eset:

Az előző konstrukcióval az a baj, hogy ha valamelyik grammatikának volt korlátozott ε -szabálya, az most már nem a kezdőszimbólumból van.

Legyen most $G_k = \langle N_k, \Sigma_k, P_k, S_k \rangle \in \mathcal{G}_1$ olyan, hogy $L(G_k) = L_k - \{\varepsilon\}$ ($k = 1, 2$).

Ilyen grammatikák vannak, hiszen L_1, L_2 -t generáló 1-típusú grammatikák a tétel feltételéből következően vannak. Ha van $S \rightarrow \varepsilon$ szabályuk, akkor azt törölve ilyen grammatikát kapunk.

Készítsük el az előbbi G_\cup grammatikát. Amennyiben $\varepsilon \in L_1$ vagy $\varepsilon \in L_2$ akkor adjuk hozzá a szabályrendszerhez az $S \rightarrow S_0 \mid \varepsilon$ szabályokat és S legyen az új kezdőszimbólum.

Zártsági tulajdonságok

Konkatenáció:

Legyen $L_k \in \mathcal{L}_i$ és $G_k = \langle N_k, \Sigma_k, P_k, S_k \rangle \in \mathcal{G}_i$ olyan, hogy $L(G_k) = L_k$ ($k = 1, 2$).

$i = 0, 2$ eset:

Legyen $S_0 \notin N_1 \cup N_2$.

$G_c := \langle N_1 \cup N_2 \cup \{S_0\}, \Sigma_1 \cup \Sigma_2, P_1 \cup P_2 \cup \{S_0 \rightarrow S_1 S_2\}, S_0 \rangle$.

Nyilván $L(G_1)L(G_2) \subseteq L(G_c)$.

$L(G_1)L(G_2) \supseteq L(G_c)$ pedig azért teljesül, mert $N_1 \cap N_2 = \emptyset$ és a szabály-baloldalak terminálistmentessége miatt nem történhet a két szó határán átnyúló szabályalkalmazás.

Valóban, a levezetés hosszára vonatkozó teljes indukcióval bizonyítható, hogy G_c mondatformái $w_1 w_2$ alakúak, ahol $S_k \Rightarrow_{G_k}^* w_k$ ($k = 1, 2$). A levezetés első lépése után ez nyilván teljesül. Tegyük fel, hogy n levezetési lépés után a G_c -beli mondatformák ilyenek. Mivel $P_1 \cup P_2$ minden szabályának baloldala $N_1^+ \cup N_2^+$ -beli ezért ez az $n + 1$. lépés után is állni fog.

Zártsági tulajdonságok

Konkatenáció:

$i = 3$ eset: Az előző konstrukcióval az a baj, hogy $S_0 \rightarrow S_1 S_2$ nem 3-as típusú szabály.

$$P_1^+ := \{A \rightarrow uS_2 \mid A \rightarrow u \in P_1, A \in N_1, u \in \Sigma_1^*\} \cup \{A \rightarrow uB \mid A \rightarrow uB \in P_1, A, B \in N, u \in \Sigma_1^*\},$$

$$G_c := \langle N_1 \cup N_2, \Sigma_1 \cup \Sigma_2, P_1^+ \cup P_2, S_1 \rangle.$$

Azaz vegyük a P_1 szabályrendszer azon szabályait, amelyekkel befejezhető egy G_1 -beli levezetés, ezen szabályok végére írjuk oda a G_2 grammatika kezdőszimbólumát, így egy $L(G_1)$ -beli szó levezetése után jobbról konkatenálva elkezdhető egy $L(G_2)$ -beli szó levezetése.

Másrészt legyen vS_2 az $u \in L(G_c)$ terminális szó levezetése során először S_2 -t tartalmazó mondatforma, ekkor $u = vw$ és

$$N_1 \cap N_2 = \emptyset \text{ miatt } S_1 \Rightarrow_{G_1}^* v \text{ valamint } S_2 \Rightarrow_{G_2}^* w.$$

$$\text{Tehát } L(G_c) = L(G_1)L(G_2).$$

Zártsági tulajdonságok

Konkatenáció:

$i = 1$ eset:

A 0-ás és 2-es eset konstrukciójával az a baj, hogy ha valamelyik grammmatikának volt korlátozott ε -szabálya, az most már nem a kezdőszimbólumból van.

Legyen $L'_k = L_k - \{\varepsilon\}$ ($k = 1, 2$). és készítsük al a 0,2 eset G_c konstrukcióját L'_1, L'_2 -hez. Tehát $L'_1 L'_2 \in \mathcal{L}_1$.

Ekkor

		$\varepsilon \notin L_2$	$\varepsilon \in L_2$
$L_1 L_2 =$	$\varepsilon \notin L_1$	$L'_1 L'_2$	$L'_1 L'_2 \cup L'_1$
	$\varepsilon \in L_1$	$L'_1 L'_2 \cup L'_2$	$L'_1 L'_2 \cup L'_1 \cup L'_2 \cup \{\varepsilon\}$

$L_1 L_2 \in \mathcal{L}_1$, mivel $L'_1 L'_2, L'_1, L'_2, \{\varepsilon\} \in \mathcal{L}_1$ és előbb láttuk, hogy \mathcal{L}_1 zárt az unióra.

Zártsági tulajdonságok

Lezárás:

Legyen $L \in \mathcal{L}_i$ és $G = \langle N, \Sigma, P, S \rangle \in \mathcal{G}_i$ olyan, hogy $L(G) = L$.

$i = 2$ eset:

Legyen $S_0 \notin N$ új nemterminális.

$G_* := \langle N \cup \{S_0\}, \Sigma, P \cup \{S_0 \rightarrow SS_0 \mid \varepsilon\}, S_0 \rangle$ L^* -ot generálja.

$i = 3$ eset:

Az előző konstrukcióval az a baj, hogy $S_0 \rightarrow SS_0$ nem 3-as típusú szabály.

$P_* := \{A \rightarrow uS \mid A \rightarrow u \in P, u \in \Sigma^*\},$

azaz vegyük a P szabályrendszer összes befejező szabályát, és ezen szabályok végére írjuk oda a G grammatika kezdőszimbólumát.

Legyen $S_0 \notin N$ új nemterminális.

$G_* := \langle N \cup \{S_0\}, \Sigma, P \cup P_* \cup \{S_0 \rightarrow S \mid \varepsilon\}, S_0 \rangle.$

P_* szabályaival új L -beli szó generálásába kezdhetünk, de meg kell hagyni, az eredeti befejező szabályokat is.

Zártsági tulajdonságok

Lezárás:

$i = 0, 1$ eset:

A 2-es típus konstrukciójával az a gond, hogy mivel lehetnek a grammatikának olyan szabályai, ahol a baloldal nem egyetlen nemterminális az L -beli szavak iteráltján kívül esetleg más szavak is keletkezhetnek. Az 1-es típusnál ezen felül alaki probléma is van, hiszen a KES sérül.

$\varepsilon \notin L$ **eset:** Legyen $S_0, S_1 \notin N$.

$G_* := \langle N \cup \{S_0, S_1\}, \Sigma, P \cup P' \cup P'' \cup P''', S_0 \rangle$, ahol

$P' = \{S_0 \rightarrow \varepsilon \mid S \mid S_1 S\},$ // 0, 1 vagy több iteráció?

$P'' = \{S_1 a \rightarrow S_1 Sa \mid a \in \Sigma\},$ // közbülső iterációk

$P''' = \{S_1 a \rightarrow Sa \mid a \in \Sigma\}.$ // utolsó iteráció

Zártsági tulajdonságok

Állítás: G_* mondatformái éppen az

(a) $S_1 u_1 \cdots u_n$ ($n \geq 1, S \Rightarrow_G^* u_k, 1 \leq k \leq n$) és az

(b) $u_1 \cdots u_n$ ($n \geq 0, S \Rightarrow_G^* u_k, 1 \leq k \leq n$),

alakú szavak, ahol u_k első betűje Σ -beli ($2 \leq k \leq n$).

Az állításból rögtön következik, hogy $L(G_*) = L(G)^*$.

(1) n -re vonatkozó teljes indukcióval bizonyítjuk, hogy az (a) és (b) típusú szavak levezethetők. (a) $n = 1$ -re és (b) $n = 0, 1$ -re ez könnyen ellenőrizhető.

(a) Az indukciós feltevés szerint $S_0 \Rightarrow^* S_1 u_2 \cdots u_n$. Ha u_2 terminálissal kezdődik $S_0 \Rightarrow^* S_1 S u_2 \cdots u_n$. Mivel $S \Rightarrow_G^* u_1$, ezért $S_0 \Rightarrow^* S_1 u_1 u_2 \cdots u_n$.

(b) Ha $n \geq 2$, akkor az indukciós feltevés szerint $S_0 \Rightarrow^* S_1 u_2 \cdots u_n$. Ha u_2 terminálissal kezdődik $S_0 \Rightarrow^* S u_2 \cdots u_n$. Mivel $S \Rightarrow_G^* u_1$, ezért $S_0 \Rightarrow^* u_1 u_2 \cdots u_n$.

Zártsági tulajdonságok

(2) Az (a) vagy (b) típusú szavakból $P' \cup P'' \cup P'''$ -beli szabályokkal (a) vagy (b) típusú lesz.

P -beli szabály esetén a szabály baloldala valamelyik u_k -nak részszoava ($1 \leq k \leq n$) kell legyen, mivel minden szabály baloldala N^+ -beli viszont az iterált szavak határán átnyúló részszoavak tartalmaznának terminálist (u_k terminálissal kezdődik $k \geq 2$ -re).

Így az eredmény (a) vagy (b) típusú.

Ha G 1-típusú, akkor G_* is 1-típusú.

Ezzel az $\varepsilon \notin L$ esettel kész vagyunk.

$\varepsilon \in L$ eset: Vegyük észre, hogy minden L -re $(L - \{\varepsilon\})^* = L^*$, így elég egy tetszőleges 0-típusú vagy 1-típusú G grammatikához egy $L(G) - \{\varepsilon\}$ -t generáló grammatikát megadni (1-típusú grammatika esetén figyelve a típusra).

Zártsági tulajdonságok

$i = 1$ esetén hagyjuk el az $S \rightarrow \varepsilon$ szabályt (ha van).

$i = 0$ esetén legyen $P_\varepsilon = \{p \in P \mid p = \alpha \rightarrow \varepsilon, \alpha \in (N \cup \Sigma)^*\}$ a grammatika ε jobboldalú szabályai.

$P_1 := (P - P_\varepsilon) \cup P'$, ahol

$P' = \{uX \rightarrow X, Xu \rightarrow X \mid X \in (N \cup \Sigma), u \rightarrow \varepsilon \in P\}$.

Ekkor $L(P_1) = L(G) - \{\varepsilon\}$.

Világos, hiszen a nemüres szavak G -beli levezetéseiben a P_ε -beli szabályok alkalmazását mindig helyettesíthetjük egy neki megfelelő, környezetes P' -beli szabály alkalmazásával (a mondatforma állhat mindösszesen az alkalmazandó P_ε -beli szabály baloldalából, mert akkor ε -t generálná ez a levezetés).

Másrészt nyilván G_* minden szabályához van G -ben egy ugyanilyen hatást elérő P -beli szabály: a neki megfelelő, környezet nélküli szabály.