



ΗΥ - 280 Θεωρία Υπολογισμού

Διάλεξη 7

Κωνσταντίνος Βάρσος

Τμήμα Επιστήμης Υπολογιστών – Πανεπιστήμιο Κρήτης

Χειμερινό εξάμηνο 2025 - 26

Γλώσσες και Αυτόματα

Τι είδαμε στην Διάλεξη 6

- Κανονικές εκφράσεις.
- Λήμμα Άντλησης για κανονικές γλώσσες.
- Θεώρημα Myhill-Nerode.
- Ελαχιστοποίηση DFA.

Γραμματικές

- Γλωσσολογία: Γλώσσες παράγονται από κανόνες και σύμβολα, **γραμματική**.
- Τα σύμβολα μπορεί να είναι S (πρόταση), N (ουσιαστικό), V (ρήμα), A (επίρρημα).
- Οι κανόνες καθορίζουν πώς μια πρόταση μπορεί να αποσυντεθεί σε μέρη του λόγου, π.χ. $S \rightarrow N V$ ή $S \rightarrow A N V$.
- Σύστημα κανόνων που χρησιμοποιείται για να περιγράψει τη συντακτική δομή μιας γλώσσας, όπου η αντικατάσταση ενός συμβόλου σε ένα κανόνα εξαρτάται μόνο από αυτό το σύμβολο και όχι από τα υπόλοιπα σύμβολα.
- Τέτοιες συλλογές κανόνων ονομάζονται **ασυμφραστικές γραμματικές** (context-free grammars).
- Συστηματικός τρόπος μετασχηματισμού συμβολοσειρών μέσω κανόνων παραγωγής.

Γραμματικές

- **Αλφάβητο**, τερματικά και μη-τερματικά σύμβολα και ένα αρχικό σύμβολο (μη-τερματικό).
- Πεπερασμένο σύνολο κανονων της μορφής $a \rightarrow b$, ορίζουν δυνατότητα αντικατάστασης της συμβολοσειράς a με την συμβολοσειρά b .
- Κάθε (τυπική) γραμματική παράγει μια (τυπική) γλώσσα (formal language), το σύνολο των συμβολοσειρών (με τερματικά σύμβολα μόνο) που παράγονται από το αρχικό σύμβολο.
- Χρησιμοποιούνται στις γλώσσες προγραμματισμού, όταν θέλουμε να περιγράψουμε το συντακτικό μιας γλώσσας.
- Λέγονται και συστήματα μεταγραφής (rewriting systems) αλλά και γραμματικές δομής φράσεων (phrase structure grammars).

Γραμματικές

Σύμβαση για τη χρήση γραμμάτων:

- $a, b, c, \dots \in \Sigma$, πεζά λατινικά, τα αρχικά του αλφαβήτου, συμβολίζουν τερματικά.
- $A, B, C, \dots \in V$, κεφαλαία λατινικά, συμβολίζουν μη-τερματικά. [αναφέρονται και ως παραγωγικά σύμβολα]
- $u, v, w, x, y, z, \dots \in \Sigma^*$, πεζά λατινικά, τα τελευταία του αλφαβήτου, συμβολίζουν συμβολοσειρές τερματικών.
- $\alpha, \beta, \gamma, \dots \in V^*$, ελληνικά συμβολίζουν οποιεσδήποτε συμβολοσειρές (τερματικών και μη).
- ε συμβολίζουμε την κενή συμβολοσειρά.

Γραμματικές

π.χ., Δίνονται οι παρακάτω κανόνες παραγωγής

$$S \rightarrow XY, \quad X \rightarrow 0X1, \quad Y \rightarrow 2Y, \quad X \rightarrow \varepsilon, \quad Y \rightarrow 3.$$

S, X, Y μεταβλητές,

S αρχική μεταβλητή,

$0, 1, 2$ τερματικά σύμβολα.

■ Για να παράξουμε μια συμβολοσειρά από μια γραμματική, ξεκινάμε από το αρχικό σύμβολο και χρησιμοποιούμε **οποιοουσδήποτε** κανόνες και με **οποιαδήποτε σειρά**. Καταλήγοντας σε μια συμβολοσειρά με τερματικά σύμβολα **αποκλειστικά**.

■ Η συμβολοσειρά 001123 παράγεται ως εξής,

$$S \rightarrow XY \rightarrow 0X1Y \rightarrow 00X11Y \rightarrow 00112Y \rightarrow 001123.$$

Γραμματικές

π.χ., Γραμματική για την γλώσσα των περιττών αριθμών

$$S \rightarrow A1$$

S : το αρχικό σύμβολο,

$$A \rightarrow A0$$

A : μη-τερματικό σύμβολο,

$$A \rightarrow A1$$

$0, 1$: τερματικά σύμβολα,

$$A \rightarrow \varepsilon$$

ε : κενή συμβολοσειρά,

■ Τα S και A αντικαθίστανται βάσει των κανόνων.

■ Κάθε περιττός προκύπτει από το S με κάποια σειρά έγκυρων αντικαταστάσεων.

■ Θα χρησιμοποιούμε την συντομογραφία $A \rightarrow A0 \mid A1 \mid \varepsilon$.

■ Κανονική έκφραση: $(0 + 1)^*1$.

Ιεραρχία Γραμματικών Chomsky

- τύπου-0: γενικές γραμματικές $[\alpha \rightarrow \beta, \alpha \neq \varepsilon]$
- τύπου-1: γραμματικές με συμφραζόμενα ή μονοτονικές $[\alpha \rightarrow \beta, |\alpha| \leq |\beta|, S \rightarrow \varepsilon]$
- τύπου-2: γραμματικές χωρίς συμφραζόμενα (context free) $[A \rightarrow \beta, A \in V]$
- τύπου-3: κανονικές γραμματικές
 δεξιογραμμικές: $A \rightarrow w, A \rightarrow wB$
 αριστερογραμμικές: $A \rightarrow w, A \rightarrow Bw$
 $(w \in \Sigma^*, A, B \in V)$



Noam Chomsky (1928 -)

τύπου 3 $\subset$ τύπου 2 $\subset$ τύπου 1 $\subset$ τύπου 0

Γραμματικές

■ Μια ασυμφραστική γραμματική (context-free grammar, CFG) $G = (V, \Sigma, R, S)$ αποτελείται από:

I. Πεπερασμένο αλφάβητο V από **μη-τερματικά** σύμβολα [εναλλακτικά, **μεταβλητές**]

II. Πεπερασμένο αλφάβητο Σ από **τερματικά** σύμβολα [εναλλακτικά, **σταθερές**]

III. Πεπερασμένο σύνολο R από **κανόνες παραγωγής** [εναλλακτικά, **παραγωγές**], δηλαδή διατεταγμένα ζεύγη (α, β) όπου $\alpha, \beta \in (V \cup \Sigma)^*$ και $\alpha \neq \varepsilon$.

IV. Αρχικό σύμβολο ή αξίωμα $S \in V$.

■ Συχνά αναφερόμαστε στο $V \cup \Sigma$ ως το λεξιλόγιο της γραμματικής.

Παραγωγές

■ Η εφαρμογή των κανόνων παραγωγής ονομάζεται **παραγωγή**, π.χ., από κάποια μορφή εφαρμόζουμε κάποιον κανόνα παραγωγής και εξάφουμε την επόμενη μορφή. Το σύμβολο $\Rightarrow$ σημαίνει **παράγει** (derivation).

■ $\alpha \Rightarrow \beta$, σημαίνει ότι το β μπορεί να παραχθεί από το α με εφαρμογή ενός κανόνας παραγωγής.

■ Επίσης, η $\alpha_1\gamma\alpha_2$ παράγει την $\alpha_1\delta\alpha_2$ και το συμβολίζουμε με $\alpha_1\gamma\alpha_2 \Rightarrow \alpha_1\delta\alpha_2$, αν υπάρχει ο κανόνας παραγωγής $\gamma \rightarrow \delta$ στην γραμματική. [ονομάζονται και ενδιάμεσες μορφές]

■ $\alpha \xRightarrow{k} \beta$, σημαίνει ότι το β μπορεί να παραχθεί από το α με εφαρμογή k κανόνων παραγωγής.

π.χ., $\alpha \xRightarrow{k} \beta$, συνεπάγεται

$$\alpha = \delta_0 \Rightarrow \delta_1 \Rightarrow \dots \Rightarrow \delta_k = \beta.$$

Παραγωγές

■ Για γραμματική $G = (V, \Sigma, R, S)$ και $\alpha, \beta \in (V \cup \Sigma)^*$ μια παραγωγή του β από α είναι μια πεπερασμένη ακολουθία συμβολοσειρών $\gamma_1 \Rightarrow \gamma_2 \Rightarrow \dots \Rightarrow \gamma_k$ για κάποιο $k \in \mathbb{N}$ όπου

- $\gamma_1 = \alpha, \gamma_k = \beta,$
- $\gamma_1, \dots, \gamma_k \in (V \cup \Sigma)^*,$
- $\forall i = 1, \dots, k-1, \gamma_{i+1}$ βρίσκεται από το γ_i αντικαθιστώντας έναν κανόνα παραγωγής.

■ Το $\xRightarrow{*}$ είναι το ανακλαστικό, μεταβατικό κλείσιμο του $\Rightarrow$

■ $\alpha \xRightarrow{*} \beta$, σημαίνει ότι το β μπορεί να παραχθεί από το α με εφαρμογή οσοδήποτε πολλών κανόνων παραγωγής.

■ $(\beta \text{ παράγεται από } \alpha) \iff (\exists \text{ παραγωγή του } \beta \text{ από } \alpha) \iff (\alpha \xRightarrow{*} \beta)$

Παραγωγές

■ Για γραμματική $G = (V, \Sigma, R, S)$, κάθε συμβολοσειρά στο $(V \cup \Sigma)^*$ που παράγεται από το S ονομάζεται **συντακτική μορφή** (sentential form, $SF(G)$).

■ Για γραμματική $G = (V, \Sigma, R, S)$, η γλώσσα $L(G)$ που παράγεται από την G οι συντακτικές μορφές που ανήκουν στο Σ^* , δηλαδή $L(G) = SF(G) \cap \Sigma^*$.

■ **Γλώσσα που παράγεται** από την γραμματική G ,

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

■ Λέμε ότι οι γραμματικές G_1, G_2 είναι **ισοδύναμες** αν $L(G_1) = L(G_2)$.

■ Η $X \rightarrow \varepsilon$ ονομάζεται **ε -παραγωγή**. [ή ε -κανόνας]

■ Η μεταβλητή X λέγεται **μηδενική** αν $\exists$ ακολουθία παραγωγών τ.ω. $X \xRightarrow{*} \varepsilon$.

Κανονικές γραμματικές

Έστω η γραμματική $G = (V, \Sigma, R, S)$.

Λέγεται **δεξιογραμμική** (right-linear):

$$A \rightarrow wB, \quad A \rightarrow w,$$

όπου $w \in \Sigma^*$ και $A, B \in V$.

Λέγεται **αριστερογραμμική** (left-linear):

$$A \rightarrow Bw, \quad A \rightarrow w,$$

όπου $w \in \Sigma^*$ και $A, B \in V$.

π.χ., δεξιογραμμική $G_1: S \rightarrow baS \mid a$, αριστερογραμμική $G_2: S \rightarrow Ra, R \rightarrow Rba \mid \varepsilon$.

■ Μια γραμματική λέγεται **κανονική** αν είναι είτε **δεξιογραμμική** είτε **αριστερογραμμική**.

Κανονικές γραμματικές

π.χ., δεξιογραμμική $G_1: S \rightarrow baS \mid a$,

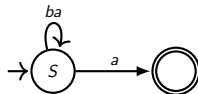
Αν κάνουμε μερικές παραγωγές παίρνουμε,

$$S \Rightarrow baS \Rightarrow babaS \Rightarrow bababaS \Rightarrow bababaa.$$

Παρατηρούμε ότι η G_1 παράγει συμβολοσειρές που τερματίζουν σε a και έχουν οσοδήποτε πολλές επαναλήψεις του ba .

Άρα, η κανονική έκφραση που αντιστοιχεί σε αυτές τις συμβολοσειρές είναι $(ba)^*a$.

Και αντίστοιχο NFA,



Κανονικές γραμματικές

π.χ., αριστερογραμμική G_2 : $S \rightarrow Ra$, $R \rightarrow Rba \mid \varepsilon$,

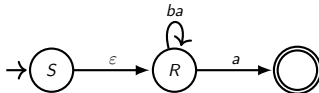
Αν κάνουμε μερικές παραγωγές παίρνουμε,

$$S \Rightarrow Ra \Rightarrow Rbaa \Rightarrow Rbabaa \Rightarrow Rbababaa.$$

Παρατηρούμε ότι η G_2 παράγει συμβολοσειρές που τερματίζουν σε a και έχουν οσοδήποτε πολλές επαναλήψεις του ba .

Άρα, η κανονική έκφραση που αντιστοιχεί σε αυτές τις συμβολοσειρές είναι $(ba)^*a$.

Και αντίστοιχο NFA,



Κανονικές γραμματικές

■ Μια κανονική γραμματική έχει τα ακόλουθα χαρακτηριστικά:

1. Στο αριστερό μέλος της παραγωγής έχουμε μόνο μια μεταβλητή.
2. Οι μεταβλητές μετατρέπονται μόνον σε μια συμβολοσειρά w από τερματικά σύμβολα.
3. Το 1. μπορεί να ακολουθείται από το πολύ μια μεταβλητή.

Κανονικές γραμματικές

■ Μια κανονική γραμματική έχει τα ακόλουθα χαρακτηριστικά:

1. Στο αριστερό μέλος της παραγωγής έχουμε μόνο μια μεταβλητή.
2. Οι μεταβλητές μετατρέπονται μόνον σε μια συμβολοσειρά w από τερματικά σύμβολα.
3. Το 1. μπορεί να ακολουθείται από το πολύ μια μεταβλητή.

π.χ.,

Η γραμματική $S \rightarrow 01A \mid \varepsilon, A \rightarrow B \mid C, B \rightarrow 00A \mid \varepsilon, C \rightarrow 2C \mid B$, είναι κανονική.

Η γραμματική $S \rightarrow 01A2 \mid \varepsilon, A \rightarrow B \mid C, C \rightarrow 2C \mid B \mid \varepsilon$, δεν είναι κανονική.

Η γραμματική $S \rightarrow 01A, A \rightarrow 22BC, B \rightarrow 12B \mid \varepsilon$, δεν είναι κανονική.

Η γραμματική $S \rightarrow 01A2 \mid \varepsilon, 1A \rightarrow B \mid C, B \rightarrow 11A \mid \varepsilon, C \rightarrow 0B \mid B$, δεν είναι κανονική.

Κανονικές Γραμματικές και Κανονικές Γλώσσες

Θεώρημα 1. Μια γλώσσα A είναι κανονική ανν παράγεται από κανονική γραμματική.

Απόδειξη (Κατασκευή).

- Έστω DFA $M = (K, \Sigma, \delta, s, F)$ που αναγνωρίζει την A .
- Κανονική γραμματική G τ.ω.
 - $V = K$, [τις μη-τερματικές μεταβλητές της G θα τις λέμε R]
 - $\Sigma = \Sigma$,
 - $S = s$,
 - $R = \{p \rightarrow \sigma q \mid \delta(p, \sigma) = q\} \cup \{f \rightarrow \varepsilon \mid f \in F\}$.

Τέλος, δείχνουμε με μαθηματική επαγωγή ότι $A = L(M)$.

Κανονικές Γραμματικές και Κανονικές Γλώσσες

Θεώρημα 1. Μια γλώσσα A είναι κανονική αν παράγεται από κανονική γραμματική.

Απόδειξη (Κατασκευή, αντίστροφο).

- Έστω κανονική γραμματική $G = (V, \Sigma, S, R)$ [είτε αριστερογραμμική είτε δεξιογραμμική] που παράγει την A με
- Κατασκευάζουμε ένα πεπερασμένο αυτόματο $M = (K, \Sigma, \delta, q_0, F)$ ως εξής:
 - Κάθε μη-τερματικό σύμβολο στη G αντιστοιχεί σε μια κατάσταση $K \in M$.
 - Το αρχικό σύμβολο S της G είναι η αρχική κατάσταση q_0 .
 - Για κάθε παραγωγή $X \rightarrow aY$, προσθέτουμε τη μετάβαση $\delta(X, a) = Y$.
 - Για κάθε παραγωγή $X \rightarrow a$, όπου το a είναι τερματικό σύμβολο, προσθέτουμε τη μετάβαση $\delta(X, a) = f$, όπου το f είναι μια νέα κατάσταση αποδοχής.
 - Αν υπάρχει παραγωγή $X \rightarrow \varepsilon$, τότε χαρακτηρίζουμε την κατάσταση X ως κατάσταση αποδοχής στο M .

Τέλος, δείχνουμε με μαθηματική επαγωγή ότι $A = L(M)$. $\square$

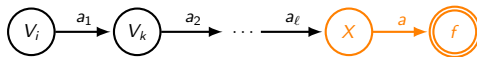
Κανονικές Γραμματικές και Κανονικές Γλώσσες

π.χ., Έστω η δεξιογραμμική γραμματική $G = \langle V, \Sigma, R, S \rangle$ η προηγούμενη κατασκευή σχηματοποιείται ως εξής,

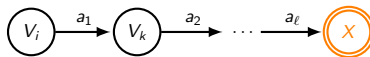
$X \rightarrow aY$,



$X \rightarrow a$,



$X \rightarrow \varepsilon$,



Κανονικές Γραμματικές και Κανονικές Γλώσσες

Θεώρημα. Για το Θεώρημα 1 θ.δ.ο. $A = L(G)$.

Απόδειξη $A \subseteq L(G)$. [$w \in \Sigma^*$, αν $\delta(q_0, w) = q_i \in M$, τότε $R_0 \xRightarrow{*} wR_i \in G$]

Θα αποδείξουμε με επαγωγή στο μήκος του w . Αν $w = \varepsilon$, τότε ισχύει από τον ορισμό του $\xRightarrow{*}$.

Έστω $w = xa$ όπου $a \in \Sigma$. Τότε, $\delta(q_0, w) = \delta(q_0, xa) = \delta(\delta(q_0, x), a) = q_i$.

Έστω $\delta(q_0, x) = q_j$. Με την επαγωγική υπόθεση έχουμε $R_0 \xRightarrow{*} xR_j$. Επίσης, αφού $\delta(q_j, a) = q_i$, έχουμε την παραγωγή $R_j \rightarrow aR_i$ στην G . Επομένως, $R_0 \xRightarrow{*} xaR_i$.

Τώρα, αν $w \in L(M)$, τότε $q_i \in F$, που σημαίνει ότι $R_i \rightarrow \varepsilon$ είναι παραγωγή, και επομένως $R_0 \xRightarrow{*} w$.

Άρα, $A \subseteq L(G)$.

Κανονικές Γραμματικές και Κανονικές Γλώσσες

Θεώρημα. Για το Θεώρημα 1 θ.δ.ο. $A = L(G)$.

Απόδειξη $L(G) \subseteq A$. [$w \in \Sigma^*$, αν $R_0 \xRightarrow{*} wR_i \in G$, τότε $\delta(q_0, w) = q_i \in M$]

Θα αποδείξουμε με επαγωγή στο μήκος του w . Αν το μήκος είναι 0, τότε ισχύει.

Έστω $w = xa$ όπου $a \in \Sigma$. Αν $R_0 \xRightarrow{*} xaR_i$ τότε $R_0 \xRightarrow{*} xR_j$ και $R_j \rightarrow aR_i$ είναι παραγωγή.

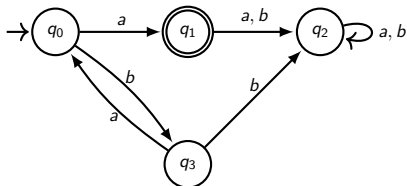
Αυτό συμβαίνει λόγω του τύπου των παραγωγών που έχει η G [έναν τερματικό ακολουθούμενο από μια μεταβλητή στη δεξιά πλευρά].

Με την επαγωγική υπόθεση έχουμε $\delta(q_0, x) = q_j$ και αφού $R_j \rightarrow aR_i$, έχουμε $\delta(q_j, a) = q_i$. Επομένως, $\delta(q_0, xa) = q_i$.

Τώρα, αν $w \in L(G)$, τότε $R_i \rightarrow \varepsilon$ είναι παραγωγή, που σημαίνει ότι $q_i \in F$, και επομένως $\delta(q_0, w) \in F$. Άρα, $L(G) \subseteq A$. □

Κανονικές Γραμματικές και Κανονικές Γλώσσες

π.χ., Έστω το αυτόματο



Κατασκευή δεξιογραμμικής γραμματικής,

$$q_0 \rightarrow a \mid bq_3 \mid aq_1 \quad [S := q_0]$$

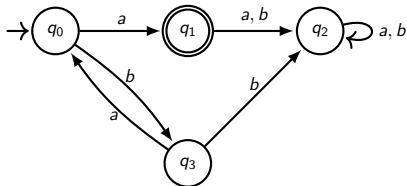
$$q_1 \rightarrow aq_2 \mid bq_2 \mid \varepsilon$$

$$q_2 \rightarrow aq_2 \mid bq_2$$

$$q_3 \rightarrow aq_0 \mid bq_2$$

Κανονικές Γραμματικές και Κανονικές Γλώσσες

π.χ., (συνέχεια) Απαλοιφή συμβόλων που δεν παράγουν τίποτα (q_2) και παραγωγών,



$$q_0 \rightarrow a \mid aq_1 \mid bq_3 \quad [S := q_0]$$

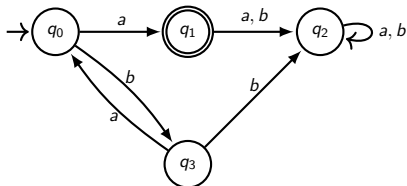
$$q_1 \rightarrow \varepsilon$$

$$q_3 \rightarrow aq_0$$

□

Κανονικές Γραμματικές και Κανονικές Γλώσσες

π.χ., Έστω το αυτόματο



Κατασκευή αριστερογραμμικής γραμματικής,

$$q_1 \rightarrow q_0 a \mid b q_3 \mid q_1 a \quad [S := q_1]$$

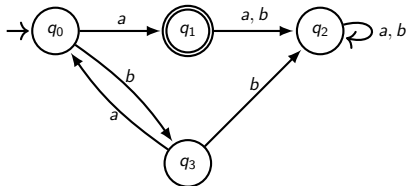
$$q_0 \rightarrow q_3 a \mid \varepsilon$$

$$q_3 \rightarrow q_0 b$$

$$q_2 \rightarrow q_2 b \mid q_b a \mid q_1 b \mid q_1 a \mid q_3 b$$

Κανονικές Γραμματικές και Κανονικές Γλώσσες

π.χ., (συνέχεια) Απαλοιφή συμβόλων που δεν παράγουν τίποτα (q_2) και παραγωγών,



$$q_1 \rightarrow q_0 a \mid \varepsilon \quad [S := q_1]$$

$$q_0 \rightarrow q_3 a$$

$$q_3 \rightarrow q_0 b$$



Γραμματικές χωρίς Συμφραζόμενα

Ορισμός. Μια **γραμματική χωρίς συμφραζόμενα**, context-free grammar (CFG), είναι μια τετράδα $G = (V, \Sigma, R, S)$ όπου:

- V , αλφάβητο (πεπερασμένο σύνολο μεταβλητών).
- Σ , πεπερασμένο σύνολο τερματικών συμβόλων.
- $R \subseteq V \times (V \cup \Sigma)^*$, σύνολο κανόνων. [κανόνες παραγωγής]
- $S \in V$, αρχικό μη-τερματικό σύμβολο.

π.χ.,

Κανόνες R ,

Συνοπτικά, οι κανόνες R γράφονται και

■ $V = \{S, A\},$

$$S \rightarrow 0A1$$

$$S \rightarrow 0A1$$

■ $\Sigma = \{0, 1\},$

$$A \rightarrow 0A11$$

$$A \rightarrow 0A11 \mid 00A1 \mid \varepsilon$$

■ $S = S$

$$A \rightarrow 00A1$$

$$A \rightarrow \varepsilon$$

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

■ Αν G CFG τότε η γλώσσα που παράγεται είναι **γλώσσα χωρίς συμφραζόμενα** (CFL).

■ Όλες οι κανονικές γλώσσες είναι CFL.

π.χ., Έστω η γραμματική $G = (V, \Sigma, R, S)$ με

$$V = \{S\}, \quad \Sigma = \{a, b\}, \quad R = \{S \rightarrow \varepsilon \mid aSb\}$$

Μια δυνατή ακολουθία παραγωγής,

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb. \quad [\text{άρα } S \xRightarrow{*} aaabbb]$$

Η γλώσσα που παράγεται,

$$L(G) := \{a^n b^n \mid n \in \mathbb{N}\}.$$

□

Γλώσσες χωρίς Συμφραζόμενα

■ Μια πρακτική προσέγγιση που σε πολλά παραδείγματα λειτουργεί [αλλά όχι πάντα] είναι να προσπαθήσουμε να βρούμε τη δομή 'φωλιάσματος' των συμβολοσειρών στη γλώσσα. Οι 'φωλιασμένες εξαρτήσεις' πρέπει να δημιουργούνται ταυτόχρονα σε διαφορετικά μέρη της συμβολοσειράς.

■ Οι πράξεις $\cup$, $\circ$, $*$ μπορούν να χρησιμοποιηθούν μέσα σε μια γραμματική χωρίς συμφραζόμενα:

- **Ένωση**, $S \rightarrow S_1 \mid S_2$. [χωρίστε σε διακριτά μέρη]
- **Παράθεση**, $S \rightarrow S_1 S_2$ αν μπορείτε να χωρίσετε τη γλώσσα σε δύο διαδοχικά μέρη, χρησιμοποιήστε αυτή την παραγωγή. [το S είναι το νέο αξίωμα, οι μεταβλητές S_1, S_2 δημιουργούν τα μέρη]
- **Kleene star**, $S \rightarrow S_1 S \mid \varepsilon$. [το S είναι το νέο αξίωμα, η γλώσσα που δημιουργείται από το S_1 επαναλαμβάνεται]

Γλώσσες χωρίς Συμφραζόμενα

π.χ., Θα βρούμε την γραμματική για την γλώσσα $L = \{a^i b^j c^k \mid i = j \vee j = k\}$.

Αρχικά παίρνουμε το υποσύνολο της L τ.ω. $i = j = n$, [δηλαδή, $a^n b^n c^k$], την γλώσσα που προκύπτει την ονομάζουμε L_1 . Η γραμματική για την L_1 είναι,

$$S_1 \rightarrow AB, \quad A \rightarrow aAb \mid ab, \quad B \rightarrow Bc \mid c.$$

Έπειτα παίρνουμε το υποσύνολο της L τ.ω. $j = k = m$, [δηλαδή, $a^n b^m c^m$], την γλώσσα που προκύπτει την ονομάζουμε L_2 . Η γραμματική για την L_2 είναι,

$$S_2 \rightarrow DE, \quad D \rightarrow aD \mid a, \quad E \rightarrow bEc \mid bc.$$

Από ένωση παίρνουμε την γραμματική της L

$$S \rightarrow S_1 \mid S_2.$$

□

Γλώσσες χωρίς συμφραζόμενα

Θεώρημα. Οι γλώσσες CFL είναι κλειστές ως προς την ένωση.

Απόδειξη (Κατασκευή)

- Έστω CFG γραμματικές $G_1 = (V_1, \Sigma_1, R_1, S_1)$ και $G_2 = (V_2, \Sigma_2, R_2, S_2)$.
- Έστω οι γλώσσες που παράγουν $A_1 = L(G_1)$ και $A_2 = L(G_2)$.
- Θεωρούμε $V_1 \cap V_2 = \emptyset$.
- Νέο αρχικό σύμβολο, $S \rightarrow S_1 \mid S_2$.
- Ορίζουμε την CFG, G , που παράγει το $A_1 \cup A_2$ ως εξής

$$G = (V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, R_1 \cup R_2 \cup \{S \rightarrow S_1 \mid S_2\}, S)$$

□

- Κάθε λέξη που παράγει η G είναι είτε μια λέξη στο A_1 είτε μια λέξη στο A_2 .

Γλώσσες χωρίς συμφραζόμενα

Θεώρημα. Οι γλώσσες CFL είναι κλειστές ως προς την παράθεση.

Απόδειξη (Κατασκευή)

- Έστω CFG γραμματικές $G_1 = (V_1, \Sigma_1, R_1, S_1)$ και $G_2 = (V_2, \Sigma_2, R_2, S_2)$.
- Έστω οι γλώσσες που παράγουν $A_1 = L(G_1)$ και $A_2 = L(G_2)$.
- Θεωρούμε $V_1 \cap V_2 = \emptyset$.
- Νέο αρχικό σύμβολο, $S \rightarrow S_1 S_2$.
- Ορίζουμε την CFG, G , που παράγει το $A_1 A_2$ ως εξής:

$$G = (V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, R_1 \cup R_2 \cup \{S \rightarrow S_1 S_2\}, S).$$

□

- Κάθε λέξη που παράγει η G είναι μια λέξη στο A_1 ακολουθούμενη από μια λέξη στο A_2 .

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

π.χ., Έστω $L = \{a^m b^n \mid m \neq n\}$. Εναλλακτικά, $L = \{a^m b^n \mid m < n\} \cup \{a^m b^n \mid m > n\}$. Θ.δ.ο L CFL.

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

π.χ., Έστω $L = \{a^m b^n \mid m \neq n\}$. Εναλλακτικά, $L = \{a^m b^n \mid m < n\} \cup \{a^m b^n \mid m > n\}$. Θ.δ.ο L CFL.

Αρχικά, $L_1 = \{a^m b^n \mid m < n\}$. Αν πρέπει να υπάρχουν περισσότερα b , μπορούμε να δημιουργήσουμε ίσους αριθμούς από a και b και να προσθέσουμε τουλάχιστον ένα επιπλέον b : $L_1 = \{a^m b^m b^i \mid m \geq 0, i > 0\} = \{a^m b^m \mid m \geq 0\} \circ \{b^i \mid i > 0\}$. [παράθεση]

Η L έχει και την περίπτωση που έχουμε περισσότερα a , την L_2 . Έπειτα με ένωση παίρνουμε την γραμματική της L :

$S \rightarrow S_1 \mid S_2,$	[ένωση για την γραμματική της L]
$S_1 \rightarrow XB,$	[παράθεση για την γραμματική της L_1]
$X \rightarrow aXb \mid \varepsilon,$	[τόσο για την γραμματική της L_1 όσο και για την γραμματική της L_2]
$B \rightarrow Bb \mid b,$	[για την γραμματική της L_1]
$S_2 \rightarrow AX,$	[παράθεση για την γραμματική της L_2]
$A \rightarrow Aa \mid a,$	[για την γραμματική της L_2]

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

π.χ., Έστω $G, S \rightarrow aS \mid \varepsilon$. Θ.δ.ο. η γλώσσα που παράγει αυτή η γραμματική είναι η $L(G) = \{a^n \mid n \geq 0\}$. [Αποδεικνύουμε $L(G) \subseteq L$ και $L \subseteq L(G)$.]

■ $L \subseteq L(G)$: $S \rightarrow aS \rightarrow a^2S \rightarrow \dots a^nS \rightarrow a^n$.

■ $L(G) \subseteq L$. Πρέπει να αποδείξουμε ότι $S \xRightarrow{*} u$, τότε είτε $u = a^nS$ είτε $u = a^n$, $n \geq 0$. [με επαγωγή]

■ $n = 0$: $S \xRightarrow{0} u \Rightarrow u = S = a^0S$.

■ $n = k$: Έστω ότι ισχύει, $S \xRightarrow{k} u \Rightarrow u = a^mS$ ή $u = a^m$, $m \geq 0$.

■ $n = k + 1$: $S \xRightarrow{k+1} v \Rightarrow S \xRightarrow{k} u \rightarrow v$, ((1) $u = a^mS$, ή (2) $u = a^m$) και $u \rightarrow v$, $m \geq 0$.

Το (2) απορρίπτεται, αν u τερματική, τότε δεν μπορεί να παράξει την v . Άρα, $u = a^mS$ και $u \rightarrow v$, δηλαδή $v = a^{m+1}S$ ή $u = a^{m+1}$. □

Από DFA σε Γραμματικές χωρίς Συμφραζόμενα

Έστω ότι η L είναι μια κανονική γλώσσα που γίνεται αποδεκτή από έναν DFA $M = (K, \Sigma, \delta, q_0, F)$. Θα κατασκευάσουμε μια CFG $G = (V, \Sigma, R, S)$ για τη L .

Αλγόριθμος

- $V = \{R_i \mid q_i \in K, \forall i\}$. Με άλλα λόγια, υπάρχει μια μεταβλητή στη CFG που αντιστοιχεί σε κάθε κατάσταση στο DFA.
- Το σύνολο των τερματικών είναι το ίδιο με το αλφάβητο Σ .
- S είναι η αρχική μεταβλητή, δηλαδή η μεταβλητή που αντιστοιχεί στην αρχική κατάσταση.
- Αν $\delta(q_i, a) = q_j$, τότε ο κανόνας αντικατάστασης $R_i \rightarrow aR_j$ είναι στο R . Επίσης, αν $q_i \in F$, τότε ο κανόνας $R_i \rightarrow \epsilon$ είναι κανόνας αντικατάστασης στο R .

■ $L(M) = L(G)$.

Από DFA σε Γραμματικές χωρίς Συμφραζόμενα

π.χ., Έστω DFA με

$$K = \{q_0, q_1\}, \quad \Sigma = \{a, b\}, \quad s = q_0, \quad F = \{q_1\}$$

και τις μεταβάσεις:

$$\delta(q_0, a) = q_0, \quad \delta(q_0, b) = q_1, \quad \delta(q_1, a) = q_0, \quad \delta(q_1, b) = q_1.$$

Η CFG έχει κανόνες:

$$A_{q_0} \rightarrow aA_{q_0} \mid bA_{q_1}, \quad A_{q_1} \rightarrow aA_{q_0} \mid bA_{q_1} \mid \varepsilon. \quad [S := A_{q_0}]$$



Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

■ X λέγεται **προσβάσιμο** αν $\exists$ παραγωγή $S \xRightarrow{*} \alpha X \beta$, με $\alpha, \beta \in (V \cup \Sigma)^*$.

■ X λέγεται **παραγόμενο** αν $\exists$ παραγωγή $X \xRightarrow{*} w$, με $w \in \Sigma^*$.

Θεώρημα. Κάθε μη-κενή CFL παράγεται από CFG χωρίς μη-προσβάσιμα και μη-παραγόμενα σύμβολα.

Απόδειξη. Απαλείφουμε τα μη-προσβάσιμα και μη-παραγόμενα σύμβολα. $\square$

Θεώρημα. Έστω CFG G , μπορούμε να κατασκευάσουμε CFG $\tilde{G}$ χωρίς μη-προσβάσιμα, μη-παραγόμενα σύμβολα και ε -παραγωγές, τ.ω. $L(\tilde{G}) = L(G) \setminus \{\varepsilon\}$.

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

Θεώρημα. Έστω CFG G , μπορούμε να κατασκευάσουμε CFG $\tilde{G}$ χωρίς μη-προσβάσιμα, μη-παραγόμενα σύμβολα και ε -παραγωγές, τ.ω. $L(\tilde{G}) = L(G) \setminus \{\varepsilon\}$.

π.χ., Έστω η γραμματική

$$S \rightarrow 0Q1Q2, \quad Q \rightarrow 00P0 \mid \varepsilon, \quad P \rightarrow 000.$$

Παίρνουμε,

$$S \rightarrow 0Q1Q2 \mid 01Q2 \mid 0Q12 \mid 012,$$

$$Q \rightarrow 00P0,$$

$$P \rightarrow 000$$



Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

■ Μια παραγωγή της μορφής $A \rightarrow B$ ονομάζεται **μοναδιαία παραγωγή**.

Θεώρημα. Κάθε CFL που δεν περιέχει το ε -παραγωγή παράγεται από μια γραμματική χωρίς μη-προσβάσιμα, μη-παραγόμενα σύμβολα, μοναδιαίες και ε -παραγωγές.

π.χ., Έστω η γραμματική

$$S \rightarrow Q \mid \&\%\%, \quad Q \rightarrow \#\&\%.$$

Παίρνουμε,

$$S \rightarrow \#\&\% \mid \&\%\%.$$



Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

■ Μια παραγωγή της μορφής $A \rightarrow B$ ονομάζεται **μοναδιαία παραγωγή**.

Θεώρημα. Κάθε CFL που δεν περιέχει το ε -παραγωγή παράγεται από μια γραμματική χωρίς μη-προσβάσιμα, μη-παραγόμενα σύμβολα, μοναδιαίες και ε -παραγωγές.

π.χ., Έστω η γραμματική

$$S \rightarrow Q \mid \&\%\%, \quad Q \rightarrow \#\&\%.$$

Παίρνουμε,

$$S \rightarrow \#\&\% \mid \&\%\%.$$

□

■ Βήματα απλοποίησης μιας CFG, απαλοίζουμε: 1) ε -παραγωγές, 2) μη-παραγόμενα σύμβολα, 3) μη-προσβάσιμα, και 4) μοναδιαίες παραγωγές.