

Γλώσσες Χωρίς Συμφραζόμενα

Διδάσκοντες: **Φ. Αφράτη, Δ. Φωτάκης**
Επιμέλεια διαφανειών: **Δ. Φωτάκης**

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Εθνικό Μετσόβιο Πολυτεχνείο



Γλώσσα χωρίς Συμφραζόμενα

- Γραμματική χωρίς Συμφραζόμενα:
 - Παραγωγές $P \subseteq (V - T) \times V^*$ ή $A \rightarrow w, w \in V^*$ (μόνο ένα τερματικό σύμβολο στα αριστερά).
 - Κανόνες εφαρμόζονται ανεξάρτητα από συμφραζόμενα του τερματικού συμβόλου (context-free).

- L είναι γλώσσα χωρίς συμφραζόμενα (ΓχΣ) αν παράγεται από γραμματική χωρίς συμφραζόμενα.

$$\begin{array}{l} V = \{0, 1, S\}, \\ T = \{0, 1\}, \\ S \end{array} \quad \frac{\text{Παραγωγές } P}{\begin{array}{l} S \rightarrow 0S1 \\ S \rightarrow \varepsilon \end{array}}$$

$$\frac{\text{Παραγωγές } P}{\begin{array}{l} S \rightarrow 0S0 \mid 1S1 \\ S \rightarrow 0 \mid 1 \mid \varepsilon \end{array}}$$

Παράδειγμα

- Γραμματική χωρίς συμφραζόμενα $G(V, T, S, P)$:

$$\begin{array}{l} V = \{0, 1, S\}, \quad \underline{\text{Παραγωγές } P} \\ T = \{0, 1\}, \quad \underline{S \rightarrow 0S1} \\ S \quad \quad \quad \underline{S \rightarrow \varepsilon} \end{array}$$

- Γλώσσα $L(G) = \{0^n 1^n : n \geq 0\}$
- Υπάρχουν γλώσσες χωρίς συμφραζόμενα που **δεν** είναι κανονικές.

Παράδειγμα

- Γραμματική χωρίς συμφραζόμενα $G(V, T, S, P)$:

$$\begin{array}{l} V = \{ (,), S \}, \\ T = \{ (,) \}, \\ S \end{array} \quad \begin{array}{l} \hline \text{Παραγωγές } P \\ \hline S \rightarrow (S) \mid SS \\ \hline S \rightarrow \varepsilon \\ \hline \end{array}$$

$$L(G) = \{ w \in \{ (,) \}^* : w \text{ έχει σωστά “ζυγισμένες” παρενθέσεις} \}$$

Παράδειγμα

- Γραμματική χωρίς συμφραζόμενα για γλώσσα
 $L = \{u \in \{0, 1\}^* : u = ww^R\}$ (παλίνδρομα άρτιου μήκους).

$$\begin{array}{c} \text{Παραγωγές } P \\ \hline S \rightarrow 0S0 \mid 1S1 \mid \varepsilon \end{array}$$

- Γραμματική χωρίς συμφραζόμενα για γλώσσα όλων των συμβ/ρών που **δεν** είναι παλινδρομικές.

$$\begin{array}{c} \text{Παραγωγές } P \\ \hline S \rightarrow 0S0 \mid 1S1 \mid A \\ A \rightarrow 1B0 \mid 0B1 \\ B \rightarrow 0B \mid 1B \mid \varepsilon \end{array}$$

Εκφρασιτικότητα

- Γλώσσες χωρίς συμφραζόμενα περιγράφουν:
 - Μικρά υποσύνολα φυσικών γλωσσών.
 - Αριθμητικές εκφράσεις.
 - **Γλώσσες προγραμματισμού** (C, C++, Pascal, ...).
 - Συνήθως σε Backus-Naur form.

$$V = \{E, O, \Pi, (,), \times, +, x, y, z, \dots\}$$

$$T = \{(,), \times, +, x, y, z, \dots\}$$

E

Παραγωγές P
$E \rightarrow O \mid E + O$
$O \rightarrow \Pi \mid O \times \Pi$
$\Pi \rightarrow (E) \mid x \mid y \mid z \mid \dots$

$$\langle expression \rangle ::= \langle term \rangle \mid \langle expression \rangle + \langle term \rangle$$

$$\langle term \rangle ::= \langle factor \rangle \mid \langle term \rangle \times \langle factor \rangle$$

$$\langle factor \rangle ::= (\langle expression \rangle) \mid x \mid y \mid z \mid \dots$$

Κλειστότητα

- Οι γλώσσες χωρίς συμφραζόμενα είναι **κλειστές** ως προς **ένωση και παράθεση** (και Kleene star).
 - Έστω ΓχΣ $G_1(V_1, T_1, S_1, P_1)$ και $L_1 = L(G_1)$.
Έστω ΓχΣ $G_2(V_2, T_2, S_2, P_2)$ και $L_2 = L(G_2)$.
Θεωρούμε ότι $(V_1 - T_1) \cap (V_2 - T_2) = \emptyset$
 - Ένωση $L_1 \cup L_2$:
 - Νέο αρχικό σύμβολο S , δύο νέες παραγωγές $S \rightarrow S_1 \mid S_2$
 - $G(V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, S, P_1 \cup P_2 \cup \{S \rightarrow S_1 \mid S_2\})$
 - Παράθεση $L_1 L_2$:
 - Νέο αρχικό σύμβολο S , νέα παραγωγή $S \rightarrow S_1 S_2$
 - $G(V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, S, P_1 \cup P_2 \cup \{S \rightarrow S_1 S_2\})$

Κλειστότητα

- Οι γλώσσες χωρίς συμφραζόμενα είναι **κλειστές** ως προς **Kleene star** (και ένωση και παράθεση).
 - Έστω ΓχΣ $G(V, T, S, P)$ και $L = L(G)$.
 - Kleene star L^* :
 - Νέο αρχικό σύμβολο S' , νέα παραγωγή $S' \rightarrow S' S \mid \varepsilon$
 - $G(V \cup \{S'\}, T, S, P \cup \{S' \rightarrow S' S \mid \varepsilon\})$
- Παράδειγμα: Ν.δ.ο. $L = \{a^i b^j c^k : j = i + k\}$ είναι ΓχΣ.
 - Έστω $L_1 = \{a^i b^i : i \geq 0\}$ και $L_2 = \{b^k c^k : k \geq 0\}$
 - $L = L_1 L_2$ (παράθεση των δύο γλωσσών).

Περιοδικότητα ΓχΣ

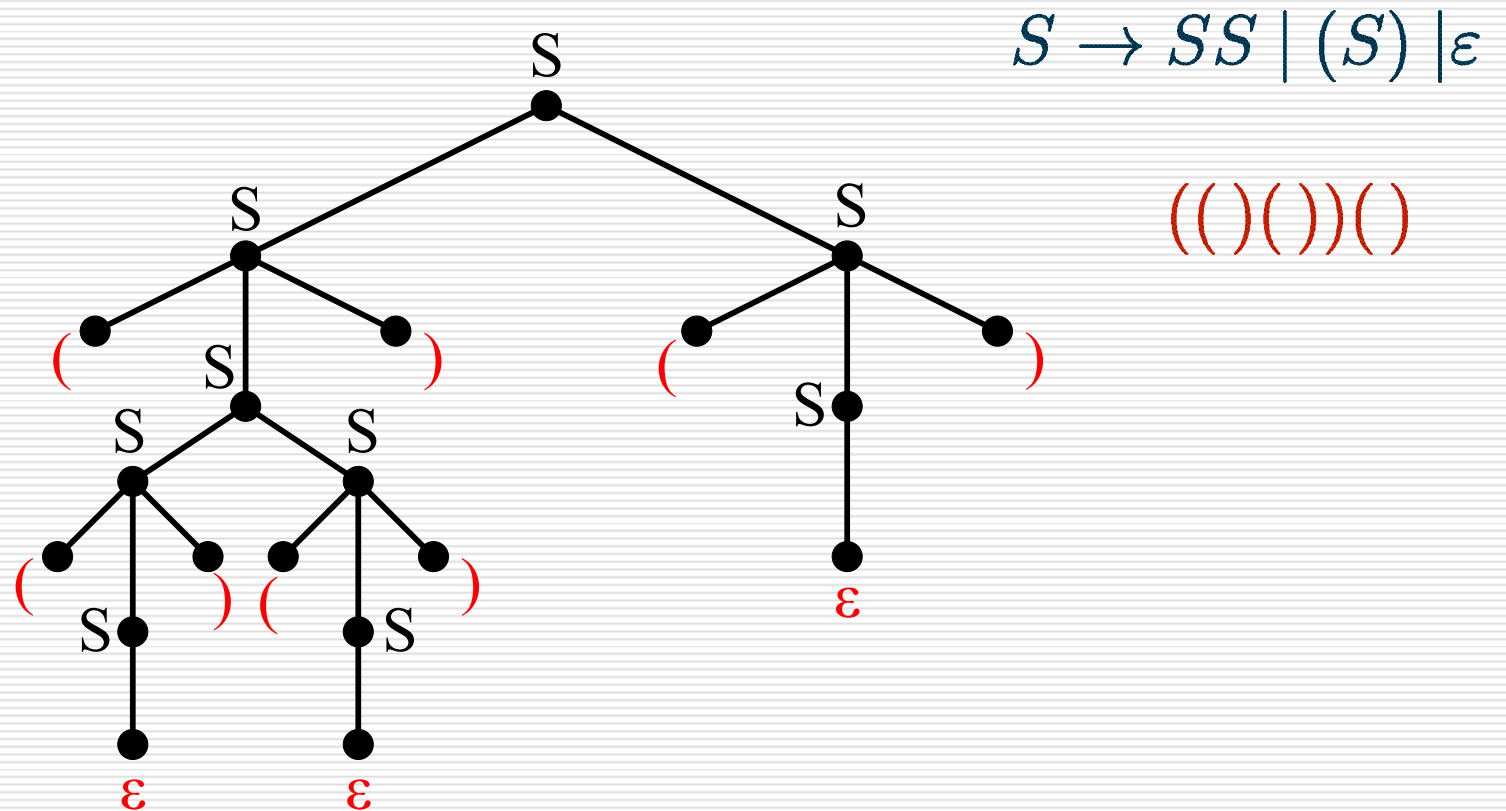
- (Άπειρη) κανονική γλώσσα: μεγάλη συμβ/ρά οδηγεί DFA σε ίδια κατάσταση («κύκλος»).
- Επανάληψη τμήματος συμβ/ράς που αντιστοιχεί σε κύκλο οδηγεί σε τελική κατάσταση (συμβ/ρά της γλώσσας).
- (Άπειρη) ΓχΣ L παράγεται από γραμματική $G(V, T, S, P)$.
- Έστω συμβ/ρά $w = uvxyz$: κατά την παραγωγή της εφαρμόζονται δύο διαφορετικές παραγωγές για μη-τερμ. A :
 $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$ με $A \Rightarrow^* vAy$ και $A \Rightarrow^* x$
- Γλώσσα χωρίς συμφραζόμενα: για κάθε $n \geq 0$, εφαρμογή 1ης n φορές και μετά 2ης δίνει $uv^nxy^n z \in L$
 $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uv^2Ay^2z \Rightarrow^* \dots \Rightarrow^* uv^nxy^n z$

$$S \rightarrow SS \mid (S) \mid \varepsilon$$
$$((()())())$$

ΣΥΝΤΑΚΤΙΚΑ Δέντρα (parse trees)

- ... απεικονίζει την παραγωγή συμβ/ράς από γραμματική $G(V, T, S, P)$:
 - Ρίζα επιγράφεται με αρχικό σύμβολο S .
 - Κόμβοι επιγράφονται με σύμβολα του V .
 - Ενδιάμεσοι κόμβοι επιγράφονται με **μη-τερματικά**.
 - Φύλλα επιγράφονται με **τερματικά** ή ε .
 - (Ενδιάμεσος) κόμβος με επιγραφή A και παράθεση επιγραφών παιδιών του $u \in V^*$: εφαρμογή κανόνα $A \rightarrow u$.
 - Παράθεση επιγραφών φύλλων από αριστερά προς δεξιά δίνει τη συμβ/ρά που παράγεται από συντακτικό δέντρο.

Παράδειγμα



Μεγάλες Συμβολοσειρές

- Γραμματική χωρίς συμφραζόμενα $G(V, \Sigma, S, P)$:
 - **Εύρος** $\varphi(G)$: μέγιστος #συμβόλων σε δεξιό μέλος κανόνα.
 - Κάθε κόμβος συντακτικού δέντρου έχει $\leq \varphi(G)$ παιδιά.
 - Παραγόμενη συμβ/ρά από συντακτικό δέντρο ύψους h έχει μήκος $\leq \varphi(G)^h$
 - Κάθε συμβ/ρά με μήκος $> \varphi(G)^{|N|}$ παράγεται από συντακτικό δέντρο ύψους $\geq |N| + 1$.
 - $N = V - \Sigma$ σύνολο μη τερματικών, $|N| = \#$ μη τερματικών.
 - Υπάρχει κλάδος με $\geq |N| + 2$ σύμβολα, 1 τερματικό.
 - Υπάρχει κλάδος όπου κάποιο μη-τερματικό σύμβολο εμφανίζεται τουλάχιστον δύο φορές.

Μεγάλες Συμβολοσειρές

- Γραμματική χωρίς συμφραζόμενα $G(V, \Sigma, S, P)$.
Κάθε συμβ/ρά w , $|w| > \varphi(G)^{|N|}$ παράγεται από συντακτικό δέντρο με κλάδο όπου κάποιο μη-τερματικό εμφανίζεται τουλάχιστον δύο φορές.

■ Έστω $w = uvxyz$

■ Παραγωγή αναλύεται:

$$A \Rightarrow^* vAy \ (T')$$

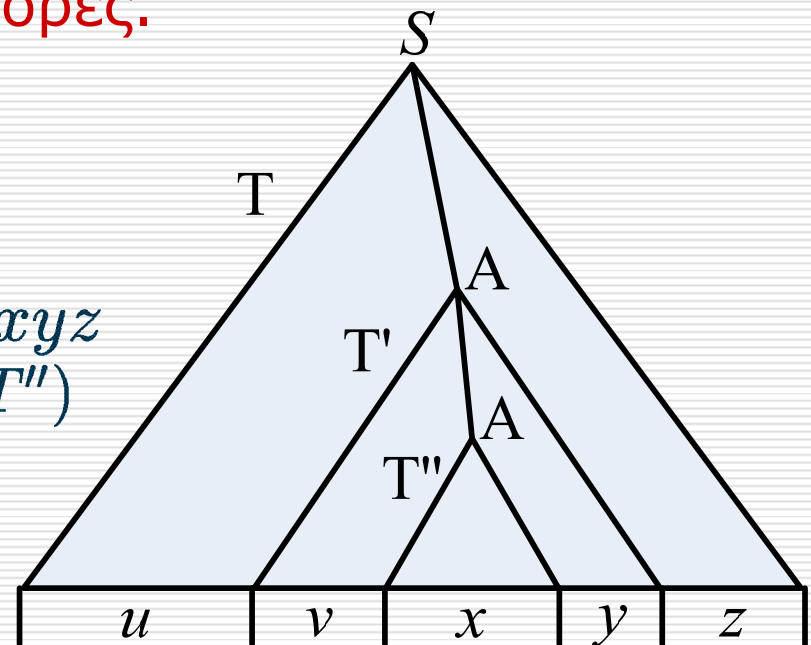
$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$$

$$S \Rightarrow^* uAz \quad A \Rightarrow^* x \ (T'')$$

■ Για κάθε #εφαρμογών

1ης παραγωγής παίρνουμε

συμβ/ρά στη γλώσσα $L(G)$.

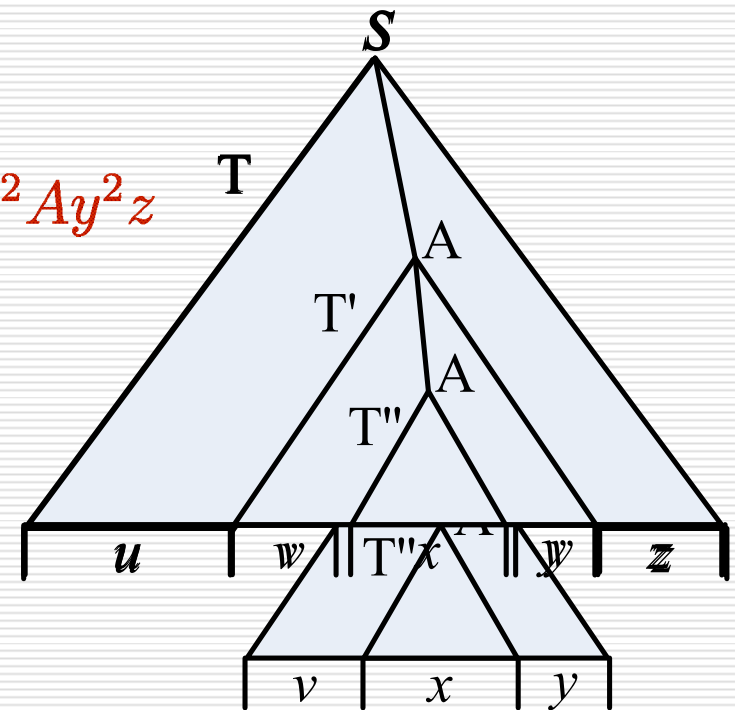


«Φούσκωμα» Συμβολοσειράς

- Κάθε συμβ/ρά w , $|w| > \varphi(G)^{|N|}$, γράφεται $w = uvxyz$ ώστε $uv^nxy^n z \in L(G)$ για κάθε $n \geq 0$.

$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz = w$$

- $n = 0$: $S \Rightarrow^* uAz \Rightarrow^* uxz$
- $n = 1$: εξ' ορισμού.
- $n = 2$: $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uv^2Ay^2z$
- ... ΚΟΚ.



Θεώρημα Άντλησης (Pumping Lemma)

- Για κάθε γλώσσα L χωρίς συμφραζόμενα, υπάρχει $k \geq 1$ ώστε **κάθε** $w \in L$, $|w| > k$, γράφεται $w = u v x y z$:
 1. για κάθε $n \geq 0$, $u v^n x y^n z \in L$.
 2. $|v y| > 0$ (τουλ. ένα από τα v, y μη-κενό).
 3. $|v x y| \leq k$
- Απόδειξη:
 - Αποδείξαμε το (1).
 - Για (2), συντακτικό δέντρο με **ελάχιστο αριθμό κόμβων**.
 - Κάθε κανόνας συνεισφέρει στην τελική συμβ/ρά.
 - Για (3), θεωρούμε δύο **κατώτερες εμφανίσεις A** στον αντίστοιχο κλάδο.

Εφαρμογές

- Γλώσσα $L = \{a^n b^n c^n : n \geq 0\}$ δεν είναι ΓχΣ.
 - Έστω αυθαίρετο $k \geq 0$ (pumping length).
 - Θεωρούμε $w = a^k b^k c^k$ και u, v, x, y, z ώστε
(i) $|v y| > 0$, (ii) $|v x y| \leq k$, και (iii) $w = u v x y z$.
 - Λόγω (i) και (ii), συμβ/ρά $v y$ περιέχει τουλάχιστον 1 και το πολύ 2 σύμβολα από a, b, c .
 - Άρα $u v^2 x y^2 z \notin L$ γιατί περιέχει διαφορετικό αριθμό a, b, c .

Εφαρμογές

- Γλώσσα $L = \{a^i b^j c^m : 0 \leq i \leq j \leq m\}$ δεν είναι ΓχΣ.
 - Έστω αυθαίρετο $k \geq 0$ (pumping length).
 - Θεωρούμε $w = a^k b^k c^k$ και u, v, x, y, z ώστε
(i) $|v y| > 0$, (ii) $|v x y| \leq k$, και (iii) $w = u v x y z$.
 - Λόγω (i) και (ii), συμβ/ρά $v y$ περιέχει τουλάχιστον 1 και το πολύ 2 σύμβολα από a, b, c .
 - Αν $v y$ περιέχει c (οπότε δεν περιέχει a), τότε $u v^0 x y^0 z \notin L$ γιατί περιέχει περισσότερα a από b ή c .
 - Αν $v y$ δεν περιέχει c , τότε $u v^2 x y^2 z \notin L$ γιατί περιέχει λιγότερα c από a ή b .

Εφαρμογές

- Γλώσσα $L = \{ww : w \in \{0, 1\}^*\}$ δεν είναι ΓχΣ.
 - Έστω αυθαίρετο $k \geq 0$ (pumping length).
 - Θεωρούμε $w = 0^k 1^k 0^k 1^k$ και u, v, x, y, z ώστε
(i) $|v y| > 0$, (ii) $|v x y| \leq k$, και (iii) $w = u v x y z$.
 - Αν $v x y$ περιορίζεται στο 1^ο μισό, δηλ. $|u v x y| \leq 2k$:
 - $u v^2 x y^2 z \notin L$ γιατί πρώτο σύμβολο του 2^{ου} μισού είναι 1.
 - Παρόμοια αν $v x y$ περιορίζεται στο 2^ο μισό.
 - Αν $v x y$ κατανέμεται και στα 2 μισά:
 - $u v^0 x y^0 z = 0^k 1^i 0^j 1^k \notin L$ γιατί $i < k, j \leq k$ ή $i \leq k, j < k$.

Εφαρμογές

- Γλώσσα $L = \{1^n : n \text{ είναι πρώτος}\}$ δεν είναι ΓχΣ.
 - Έστω αυθαίρετο $k \geq 0$ (pumping length).
 - Θεωρούμε $w = 1^p$, p ελάχιστος πρώτος $> k$, και u, v, x, y, z :
(i) $|v y| > 0$, (ii) $|v x y| \leq k$, και (iii) $w = u v x y z$.
 - Έστω $|v y| = t > 0$.
 - Τότε $|u v^n x y^n z| = p + (n-1)t$.
 - Για $n = p+1$, ο αριθμός $p + pt = p(t+1)$ **δεν** είναι **πρώτος**.
 - Άρα $u v^{p+1} x y^{p+1} z \notin L$
 - Για αλφάβητα ενός συμβόλου, ΓχΣ είναι κανονικές.

Εφαρμογές

- Ισχύει ότι η **τομή** μιας γλώσσας **χωρίς συμφραζόμενα** με μία **κανονική** γλώσσα είναι γλώσσα **χωρίς συμφραζόμενα**.
- Γλώσσα $L = \{w \in \{a, b, c\} : w \text{ έχει ίσο αριθμό } a, b \text{ και } c\}$ δεν είναι χωρίς συμφραζόμενα.
 - Η γλώσσα $L_1 = \{a^n b^n c^n : n \geq 0\}$ είναι τομή της L με **κανονική γλώσσα** $a^* b^* c^*$.
 - Αν L ήταν ΓχΣ, θα ήταν και η L_1 . **Άτοπο.**

Μη-Κλειστότητα

- Η κλάση των γλωσσών χωρίς συμφραζόμενα **δεν** είναι κλειστή ως προς συμπλήρωμα και τομή.
 - $L_1 = \{a^n b^n c^m : n \geq 0, m \geq 0\}$ γλώσσα χωρίς συμφραζόμενα.
 - $L_2 = \{a^m b^n c^n : m \geq 0, n \geq 0\}$ γλώσσα χωρίς συμφραζόμενα.
 - Τομή $L_1 \cap L_2 = \{a^n b^n c^n : n \geq 0\}$ **δεν** είναι γλώσσα χωρίς συμφραζόμενα.
 - Μη-κλειστότητα ως προς συμπλήρωμα: $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$