# 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 $A \to w$, $w \in V^*$
    (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.

$$V = \{0, 1, S\},\quad T = \{0, 1\},\quad S$$

| Παραγωγές $P$ |
|---|
| $S \to 0S1$ |
| $S \to \varepsilon$ |

| Παραγωγές $P$ |
|---|
| $S \to 0S0 \mid 1S1$ |
| $S \to 0 \mid 1 \mid \varepsilon$ |



## Παράδειγμα

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

$$V = \{0, 1, S\},\quad T = \{0, 1\},\quad S$$

| Παραγωγές $P$ |
|---|
| $S \to 0S1$ |
| $S \to \varepsilon$ |

- 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):

$$V = \{(, ), S\},\quad T = \{(, )\},\quad S$$

| Παραγωγές $P$ |
|---|
| $S \to (S) \mid SS$ |
| $S \to \varepsilon$ |

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

# Παράδειγμα

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

| Παραγωγές $P$ |
|---|
| $S \to 0S0 \mid 1S1 \mid \varepsilon$ |

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

| Παραγωγές $P$ |
|---|
| $S \to 0S0 \mid 1S1 \mid A$ |
| $A \to 1B0 \mid 0B1$ |
| $B \to 0B \mid 1B \mid \varepsilon$ |



# Εκφραστικότητα

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

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

| Παραγωγές $P$ |
|---|
| $E \to O \mid E + O$ |
| $O \to \Pi \mid O \times \Pi$ |
| $\Pi \to (E) \mid x \mid y \mid z \mid \ldots$ |

$$\langle expression \rangle ::= \langle term \rangle \mid \langle expression \rangle + \langle term \rangle$$
$$\langle term \rangle ::= \langle factor \rangle \mid \langle term \rangle \times \langle factor \rangle$$
$$\langle factor \rangle ::= (\langle expression \rangle) \mid x \mid y \mid z \mid \ldots$$



# 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 $(V_1 - T_1) \cap (V_2 - T_2) = \varnothing$
  - Ένωση $L_1 \cup L_2$:
    - Νέο αρχικό σύμβολο S, δύο νέες παραγωγές $S \to S_1 \mid S_2$
    - $G(V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, S, P_1 \cup P_2 \cup \{S \to S_1 \mid S_2\})$
  - Παράθεση $L_1 L_2$:
    - Νέο αρχικό σύμβολο S, νέα παραγωγή $S \to S_1S_2$
    - $G(V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, S, P_1 \cup P_2 \cup \{S \to S_1S_2\})$



# 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 $S' \to S' S \mid \varepsilon$
    - $G(V \cup \{S'\}, T, S, P \cup \{S' \to S' S \mid \varepsilon\})$
- Άσκηση: 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).

## Αριστερότερες Παραγωγές

- Έστω u, v, w ∈ V* για γραμματική G(V, T, S, P).
- Αν A → w παραγωγή, τότε uAv **παράγει άμεσα** uwv :

$$uAv \Rightarrow uwv$$

- u **παράγει** v, u ⇒* v: αν υπάρχουν k ≥ 0 και $u_1, \ldots, u_{k-1}$ ώστε $u \Rightarrow u_1 \Rightarrow \cdots \Rightarrow u_{k-1} \Rightarrow v$ (παραγωγή *k* βημάτων)
- **Αριστερότερη** παραγωγή: σε κάθε βήμα το αριστερότερο μη-τερματικό σύμβολο αντικαθίσταται.
- Αν u ⇒* v υπάρχει τουλάχιστον μία αριστερότερη παραγωγή με την οποία το v παράγεται από το u.



## Διφορούμενες Γραμματικές

- Αν $u \Rightarrow^* v$ υπάρχει τουλάχιστον μία αριστερότερη παραγωγή με την οποία το v παράγεται από το u.
- Αν για κάποια u, v υπάρχουν δύο ή περισσότερες αριστερότερες παραγωγές, η γραμματική λέγεται **διφορούμενη** ή **ασαφής**.
- Διφορούμενη γραμματική: $V = \{(,), S\}$, $T = \{(,)\}$

| Παραγωγές $P$ |
|---|
| $S \to (S) \mid SS \mid \varepsilon$ |

  - Δύο αριστερότερες παραγωγές για ( ).

$$S \Rightarrow (S) \Rightarrow ()$$
$$S \Rightarrow SS \Rightarrow S \Rightarrow (S) \Rightarrow ()$$



## Διφορούμενες Γραμματικές

- Διφορούμενες γραμματικές: συντακτική ανάλυση;
  - Μετατρέπουμε τη διφορούμενη γραμματική σε ισοδύναμη μη-διφορούμενη.
  - Π.χ. $V = \{(,), S\}$, $T = \{(,)\}$

| Παραγωγές $P$ |
|---|
| $S \to (S)S \mid \varepsilon$ |

  - **Εγγενώς διφορούμενες** γραμματικές **δεν** μπορούν να μετατραπούν σε μη-διφορούμενες.
  - Γραμματικές γλωσσών προγραμματισμού **δεν** είναι διφορούμενες.



## 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 $uv^nxy^nz \in L$

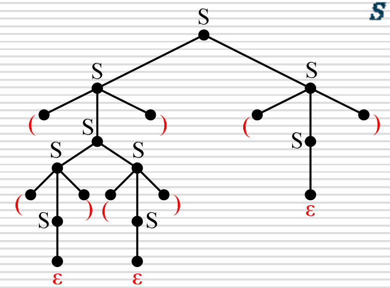
$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uv^2Ay^2z \Rightarrow^* \cdots \Rightarrow^* uv^nxy^nz$$

## Συντακτικά 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 $A \to u$.
  - Παράθεση επιγραφών φύλλων από αριστερά προς δεξιά δίνει τη συμβ/ρά που παράγεται από συντακτικό δέντρο.



## Παράδειγμα




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

(()())()



## Μεγάλες Συμβολοσειρές

- Γραμματική χωρίς συμφραζόμενα G(V, Σ, 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.

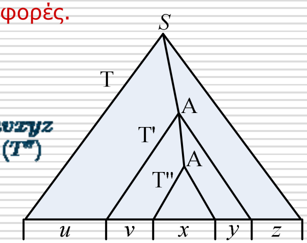


## Μεγάλες Συμβολοσειρές

- Γραμματική χωρίς συμφραζόμενα G(V, Σ, 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:

$$A \Rightarrow^* vAy \ (T')$$
$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$$
$$S \Rightarrow^* uAz \qquad A \Rightarrow^* x \ (T'')$$

  - Για κάθε #εφαρμογών 1ης παραγωγής παίρνουμε συμβ/ρά στη γλώσσα L(G).

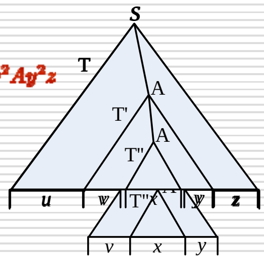


# «Φούσκωμα» Συμβολοσειράς

- Κάθε συμβ/ρά w, $|w| > \varphi(G)^{|N|}$, γράφεται $w = uvxyz$ ώστε $uv^n xy^n z \in L(G)$ 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 $w = uvxyz$ :
  1. για κάθε $n \geq 0$, $uv^n xy^n z \in L$.
  2. $|vy| > 0$ (τουλ. ένα από τα v, y μη-κενό).
  3. $|uxy| \leq k$
- 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 χωρίς συμφραζόμενα.
  - Έστω $k$ ο φυσικός αριθμός του λήμματος άντλησης.
  - Θεωρούμε $w = a^k b^k c^k = uvxyz$ .
  - Αν vy περιέχει τρία διαφορετικά σύμβολα, στη $uv^n xy^n z$ εμφανίζονται σύμβολα εκτός σειράς.
  - Αν vy περιέχει λιγότερα από τρία διαφορετικά σύμβολα, στη $uv^n xy^n z$ υπάρχει διαφορετικός αριθμός a, b, c.
  - Διαισθητικά, χρειάζονται δύο στοίβες για την αναγνώριση αυτής της γλώσσας.



# Εφαρμογές

- Γλώσσα $L = \{a^n : n \text{ είναι πρώτος}\}$ δεν είναι γλώσσα χωρίς συμφραζόμενα.
  - Θεωρούμε $w = uvxyz$ ώστε $|w|$ πρώτος.
  - Έστω $|vy| = t$ και $|uxz| = r$.
  - Ο αριθμός $r+nt$ **δεν** είναι **πρώτος** για κάθε $n \geq 0$. (π.χ. $n = r + t + 1 \Rightarrow (r+t)(t+1)$).
  - Για αλφάβητα ενός συμβόλου, γλώσσες χωρίς συμφραζόμενα είναι κανονικές.

# Εφαρμογές

- Ισχύει ότι η τομή μιας γλώσσας χωρίς συμφραζόμενα με μία κανονική γλώσσα είναι γλώσσα χωρίς συμφραζόμενα.
- Γλώσσα $L = \{w \in \{a, b, c\} : w$ έχει ίσο αριθμό $a, b$ 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.
  - $L_1 = \{a^n b^n c^m : n \geq 0, m \geq 0\}$ χωρίς συμφραζόμενα.
  - $L_2 = \{a^m b^n c^n : m \geq 0, n \geq 0\}$ χωρίς συμφραζόμενα.
  - Τομή $L_1 \cap L_2 = \{a^n b^n c^n : n \geq 0\}$ **δεν** είναι χωρίς συμφραζόμενα.
  - Μη-κλειστότητα ως προς συμπλήρωμα: $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$