

Μεταγλωττιστές #2

Νίκος Παπασπύρου
nickie@softlab.ntua.gr



Εθνικό Μετσόβιο Πολυτεχνείο
Τμήμα Ηλεκτρολόγων Μηχ. και Μηχ. Υπολογιστών
Εργαστήριο Τεχνολογίας Λογισμικού
Πολυτεχνειούπολη, 15780 Ζωγράφου.

Τυπικές γλώσσες

■ Γραμματική $G = (T, N, P, S)$

- T : τερματικά σύμβολα
- N : μη τερματικά σύμβολα
- P : κανόνες παραγωγής
- S : αρχικό σύμβολο

a

A

$\alpha \rightarrow \beta$

■ Παραγωγές: αν $\alpha, \beta, \gamma, \delta \in (T \cup N)^*$

και $(\alpha \rightarrow \beta) \in P$

τότε $\gamma\alpha\delta \Rightarrow \gamma\beta\delta$

■ Γλώσσα: $L(G) = \{ \alpha \in T^* \mid S \Rightarrow^+ \alpha \}$

Ιεραρχία Chomsky

- Τύπου 0: όλες οι γραμματικές, $\alpha \rightarrow \beta$
- Τύπου 1: γραμματικές με συμφραζόμενα (context-sensitive), $\alpha \rightarrow \beta$ με $|\alpha| \leq |\beta|$
- Τύπου 2: γραμματικές χωρίς συμφραζόμενα (context-free) $A \rightarrow \beta$
- Τύπου 3: κανονικές γραμματικές (regular)
 $A \rightarrow aB$ ή $A \rightarrow a$
- Ειδική περίπτωση: γλώσσες που παράγουν την κενή συμβολοσειρά

Αναγνωριστές

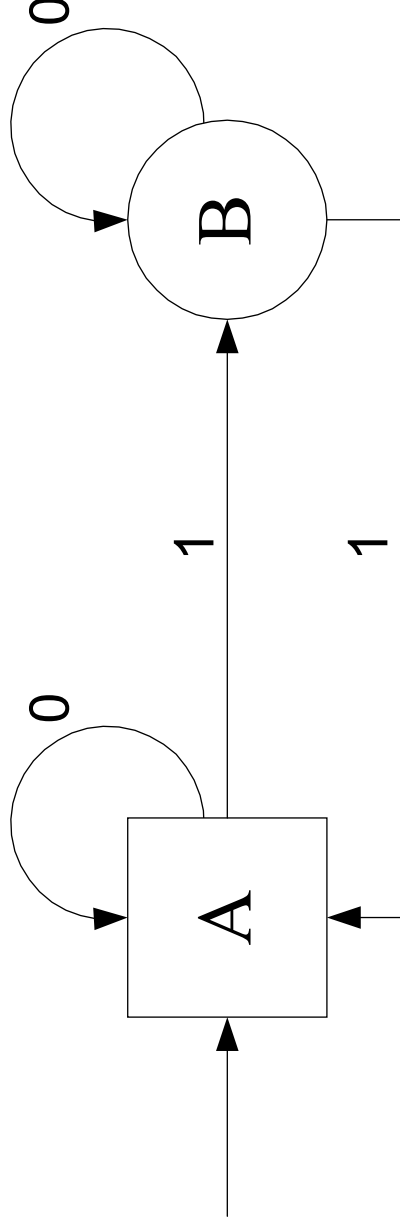
- Τύπου 0: μηχανή Turing
- Τύπου 1: γραμμικά περιορισμένη μηχανή Turing
- Τύπου 2: αυτόματα στοίβας
(push-down automata)
 - Χρήσιμα στη **συντακτική ανάλυση**
- Τύπου 3: πεπερασμένα αυτόματα
(finite automata)
 - Χρήσιμα στη **λεκτική ανάλυση**

Κανονικές γλώσσες

- Κανονικές εκφράσεις (regular expressions)
 - Κενή συμβολοσειρά: ϵ
 - Κάθε σύμβολο: $a \in \Sigma$
 - Παράθεση δύο κανονικών εκφράσεων: (rs)
 - Διάζευξη δύο κανονικών εκφράσεων: $(r|s)$
 - Κλείσιμο (ή άστρο) Kleene: (r^*)
- Συντομογραφίες:
 - απαλοιφή περιττών παρενθέσεων
 - $r^+ [a_1, a_2, \dots, a_n] r?$

Πεπερασμένα αυτόματα

- Καταστάσεις και μεταβάσεις
- Ντετερμινιστικά (ΝΠΑ), μη ντετερμινιστικά (ΜΠΑ) και ΜΠΑ με κενές μεταβάσεις (ΜΠΑ-ε)



- Η γλώσσα των συμβολοσειρών που αποτελούνται από 0 και 1 και περιέχουν άρτιο αριθμό 1

Αναγωγές και ισοδυναμίες

- κανονική γραμματική \Rightarrow ΜΠΑ- ϵ
- ΜΠΑ- ϵ \Rightarrow κανονική γραμματική
- κανονική έκφραση \Rightarrow ΜΠΑ- ϵ
- ΜΠΑ- ϵ \Rightarrow κανονική έκφραση
- ΜΠΑ- ϵ \Rightarrow ΝΠΑ
- Ελαχιστοποίηση ΝΠΑ

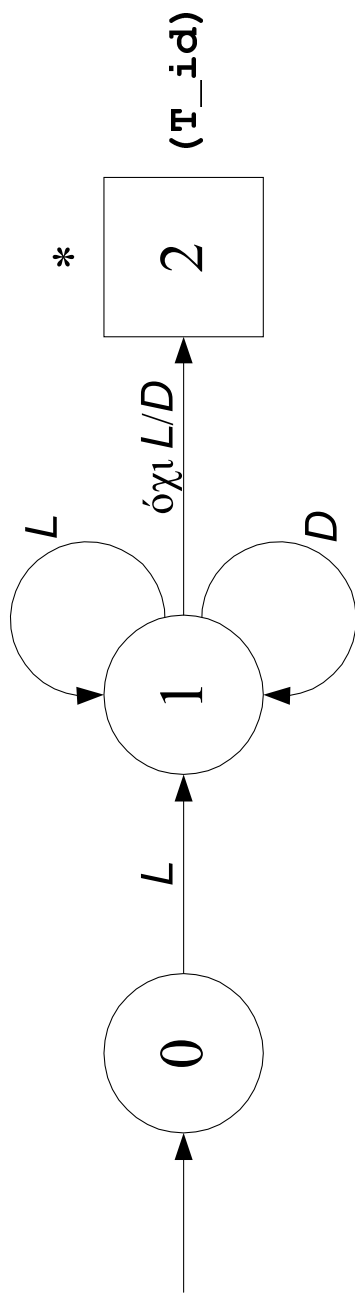
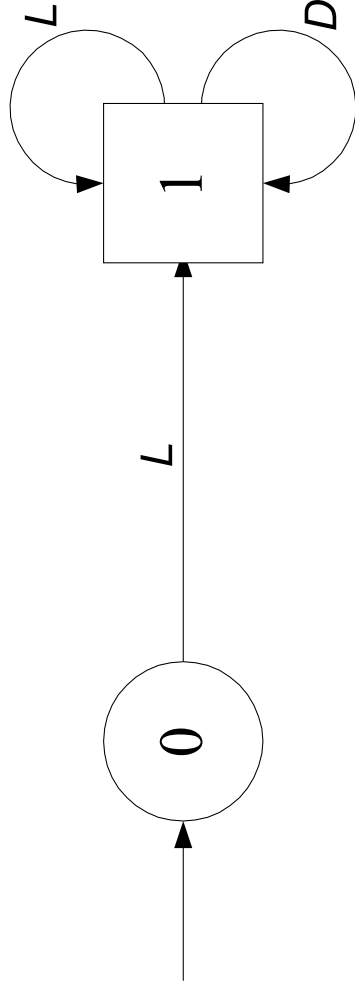
Λεκτική ανάλυση

- Λεκτικές μονάδες (tokens)
- Αναγνωρίζονται με **πεπερασμένα αυτόματα** που:
 - διαβάζουν ενδεχομένως περισσότερους χαρακτήρες
 - οπισθοδρομούν αν χρειαστεί
 - διαθέτουν έξοδο που χρησιμοποιείται στη συντακτική ανάλυση
- Ειδικός συμβολισμός: **διαγράμματα μετάβασης**



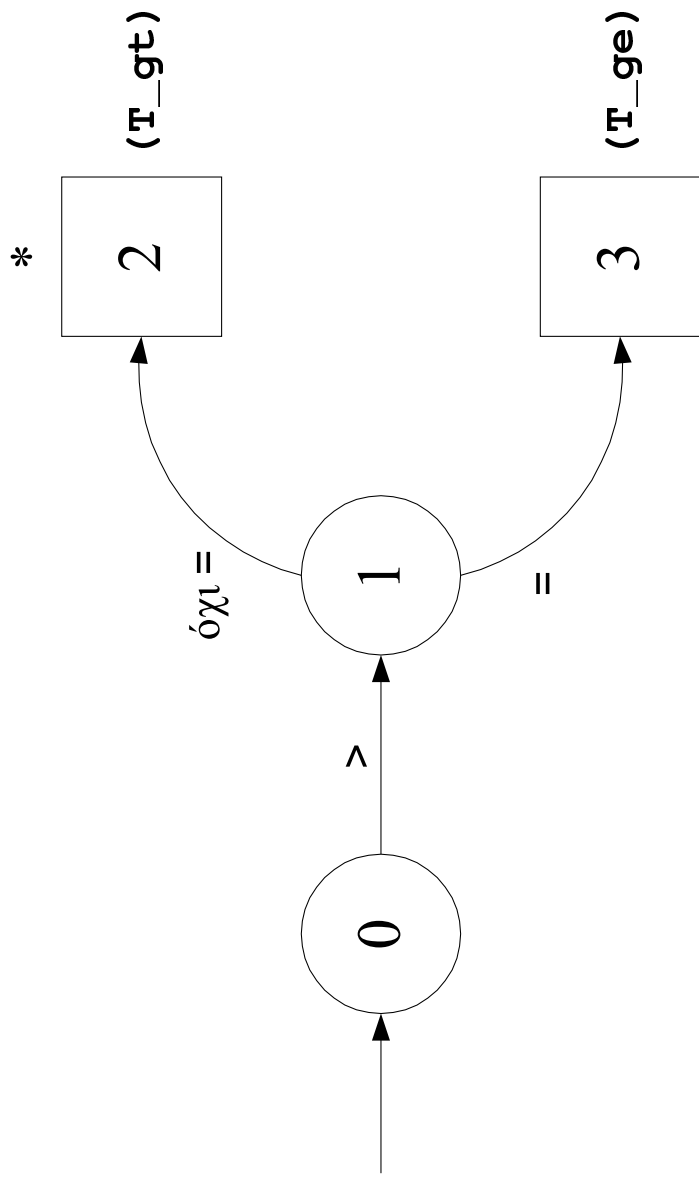
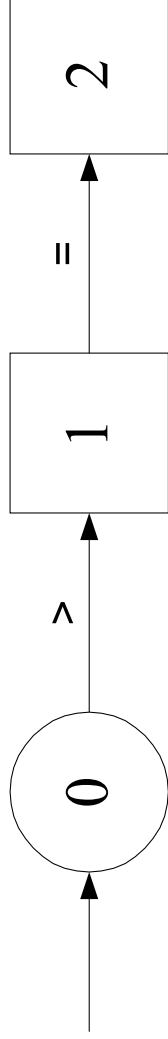
Διαγράμματα μετάβασης (i)

- Αναγνωριστικά της Pascal



Διαγράμματα μετάβασης (ii)

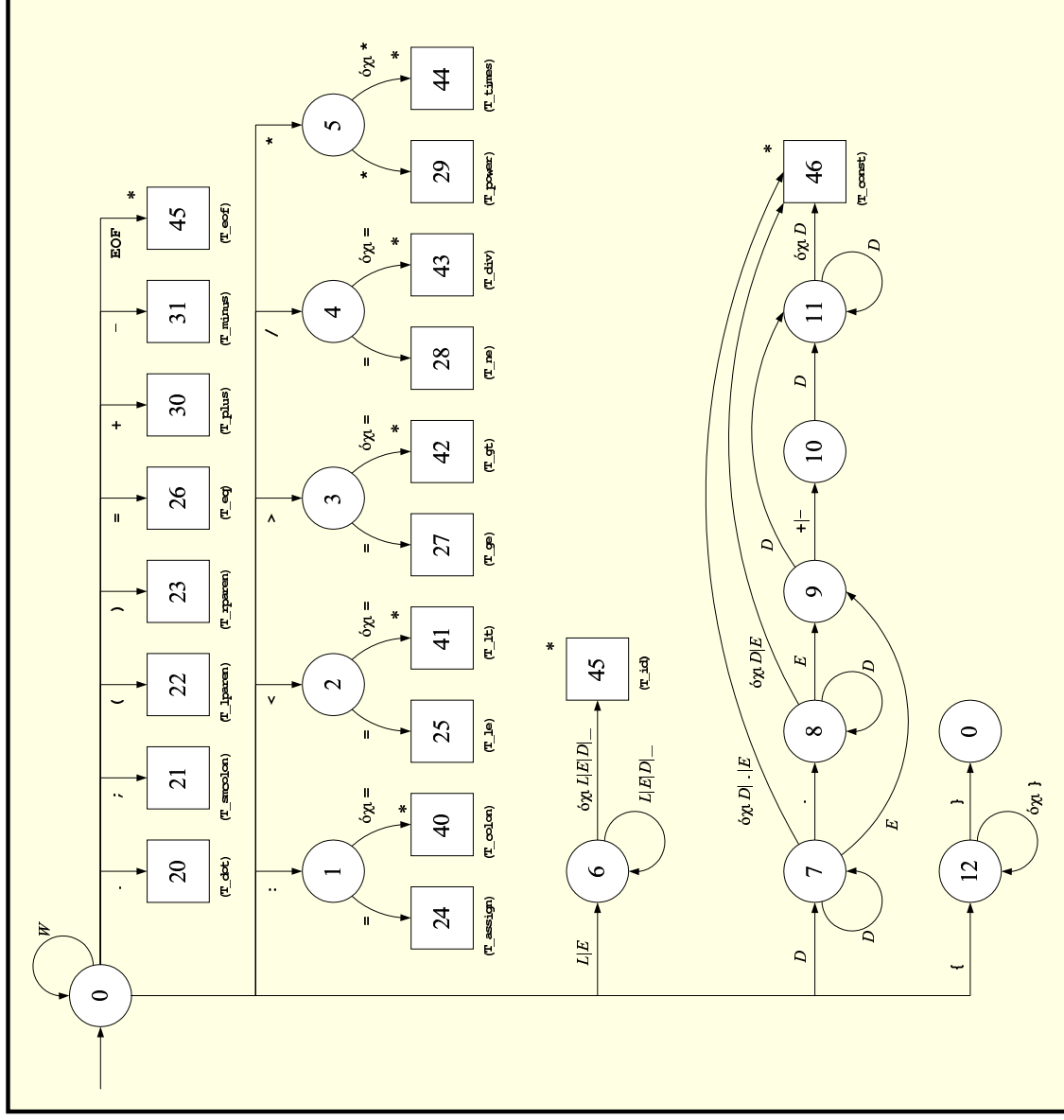
- Τελεστές $>$ και $>=$



Υλοποίηση ΛΑ

(i)

Σχεδίαση
συνολικού
διαγράμματος
μετάβασης



Υλοποίηση ΛΑ

(ii)

- Εναλλακτικοί τρόποι υλοποίησης:
 - Χειρωνακτικά
 - Με πίνακα μεταβάσεων
 - Με το μεταεργαλείο flex