

Μεταγλωττιστές #4

Νίκος Παπασπύρου
nickie@softlab.ntua.gr



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

ΣΑ *top-down*

(i)

- Η συντακτική ανάλυση ξεκινά από τη ρίζα
- Κάθε φορά, αναζητά:
 - το μη τερματικό σύμβολο που θα αντικατασταθεί
 - ⇒ συνήθως επιλέγεται το αριστερότερο
 - τον κανόνα παραγωγής που θα εφαρμοστεί
 - ⇒ βάσει των επόμενων k λεκτικών μονάδων στη συμβολοσειρά εισόδου: $LL(k)$
- Επαναλαμβάνεται μέχρι να εξαντληθούν τα μη τερματικά



Γραμμικές $LL(1)$

- Απαραίτητες προϋποθέσεις:
 - Απουσία **αριστερής αναδρομής** (άμεσης ή έμμεσης)
 - Απουσία **κοινού προθέματος** σε εναλλακτικούς κανόνες
- Μερικές φορές είναι δυνατός ο μετασχηματισμός μιας γραμμτικής σε ισοδύναμη $LL(1)$
 - ⇒ απαλοιφή αριστερής αναδρομής
 - ⇒ αριστερή παραγοντοποίηση

Μετασχηματισμός σε $LL(1)$

- Αντικατάσταση

$$\begin{array}{l} A \rightarrow \alpha_1 \mid \dots \mid \alpha_n \\ B \rightarrow \beta_1 A \beta_2 \end{array} \Rightarrow \begin{array}{l} A \rightarrow \alpha_1 \mid \dots \mid \alpha_n \\ B \rightarrow \beta_1 \alpha_1 \beta_2 \mid \dots \mid \beta_1 \alpha_n \beta_2 \end{array}$$

- Αριστερή παραγοντοποίηση

$$\begin{array}{l} A \rightarrow \alpha \beta_1 \mid \dots \mid \alpha \beta_n \\ A \rightarrow \alpha B \end{array} \Rightarrow \begin{array}{l} B \rightarrow \beta_1 \mid \dots \mid \beta_n \end{array}$$

- Απαλοιφή άμεσης αριστερής αναδρομής

$$\begin{array}{l} A \rightarrow A \alpha_1 \mid \dots \mid A \alpha_n \mid \beta_1 \mid \dots \mid \beta_m \\ A \rightarrow \beta_1 B \mid \dots \mid \beta_m B \\ B \rightarrow \alpha_1 B \mid \dots \mid \alpha_n B \mid \epsilon \end{array} \Rightarrow$$

ΣΑ αναδρομικής κατάβασης

$$A \rightarrow \alpha_1 \mid \dots \mid \alpha_n$$

μετατρέπεται σε κώδικα της μορφής:

αν $token \in \text{FIRST}(\alpha_1)$ τότε

κώδικας για την αναγνώριση της α_1

...

αλλιώς αν $token \in \text{FIRST}(\alpha_n)$ τότε

κώδικας για την αναγνώριση της α_n

αλλιώς αν $\epsilon \notin \text{FIRST}(\alpha_1) \cup \dots \cup \text{FIRST}(\alpha_n)$ τότε

συντακτικό σφάλμα

αλλιώς αν $token \notin \text{FOLLOW}(A)$ τότε

συντακτικό σφάλμα

τέλος αν

ΣΑ $LL(1)$

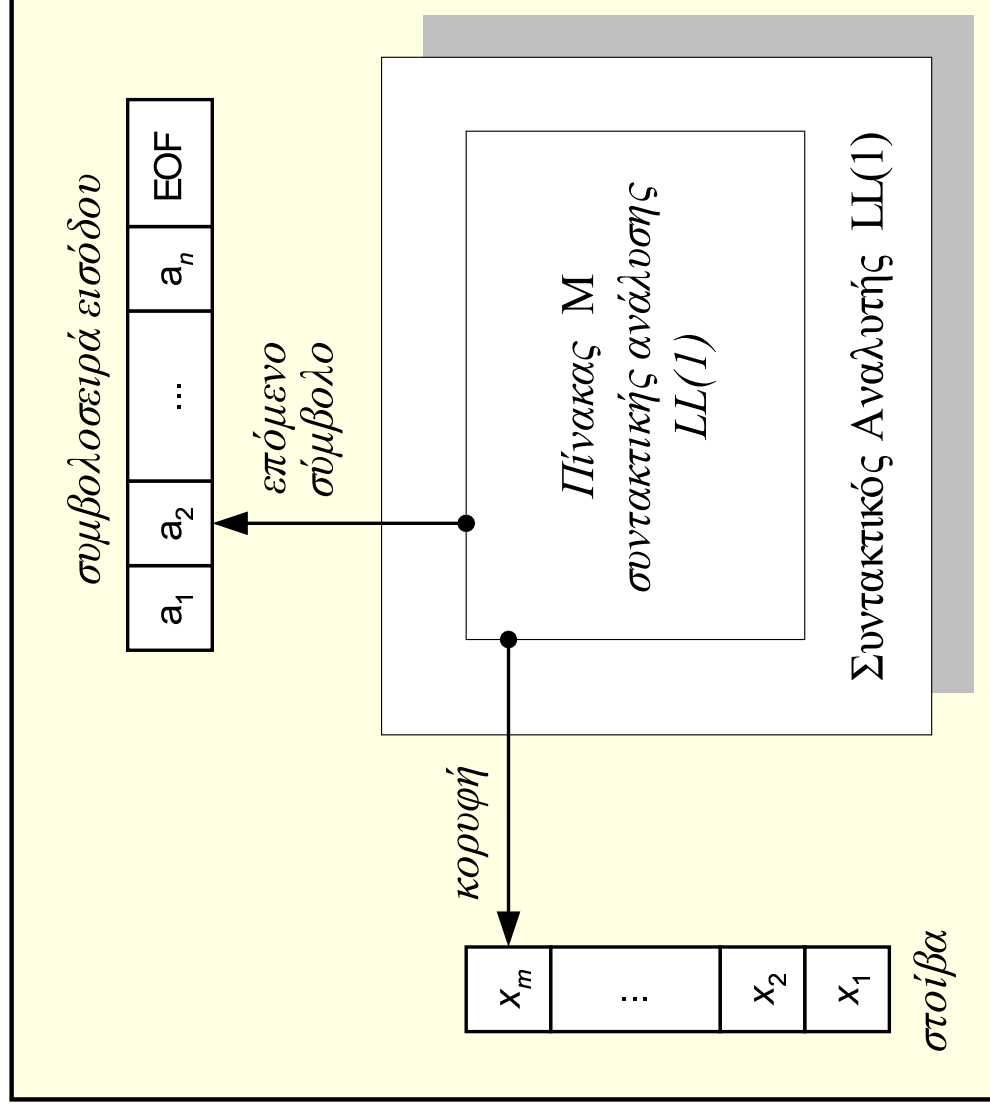
(i)

- Χρησιμοποιούν μια **στοίβα** όπου τοποθετούν σύμβολα της γραμματικής — αρχικά μόνο το S
- Κάθε φορά εξετάζεται η κορυφή της στοίβας:
 - Αν είναι **τερματικό** σύμβολο και είναι το ίδιο με το επόμενο της συμβολοσειράς εισόδου, τότε αφαιρούνται και τα δύο
 - Αν είναι **μη τερματικό σύμβολο**, τότε ανάλογα με το επόμενο της συμβολοσειράς εισόδου εφαρμόζεται κάποιος κανόνας
- **Επιτυχία:** η στοίβα και η συμβολοσειρά εισόδου είναι άδειες

ΣΑ $LL(1)$

Ο αλγόριθμος κατασκευής του πίνακα M ορίζει την οικογένεια των γλωσσών $LL(1)$

(ii)



Κατασκευή ΣΑ LL(1)

$$\begin{aligned} E &\rightarrow T E' \\ E' &\rightarrow + T \mid \epsilon \\ T &\rightarrow F T' \\ T' &\rightarrow * F \mid \epsilon \\ F &\rightarrow (E) \mid \text{id} \end{aligned} \quad \begin{aligned} \text{FIRST}(E) &= \text{FIRST}(T) = \text{FIRST}(F) = \{\text{id}, (\} \\ \text{FIRST}(E') &= \{+, \epsilon\} \\ \text{FIRST}(T') &= \{*, \epsilon\} \\ \text{FOLLOW}(E) &= \text{FOLLOW}(E') = \{ \}, \text{EOF} \\ \text{FOLLOW}(T) &= \text{FOLLOW}(T') = \{+, \epsilon\} \\ \text{FOLLOW}(F) &= \{*, +, \epsilon\}, \text{EOF} \end{aligned}$$

	id	+	*	()	EOF
E	$E \rightarrow TE'$			$E \rightarrow TE'$		
E'		$E' \rightarrow +TE'$			$E' \rightarrow \epsilon$	$E' \rightarrow \epsilon$
T	$T \rightarrow FT'$			$T \rightarrow FT'$		
T'		$T' \rightarrow \epsilon$	$T' \rightarrow *FT'$		$T' \rightarrow \epsilon$	$T' \rightarrow \epsilon$
F	$F \rightarrow \text{id}$			$F \rightarrow (E)$		

Λειτουργία ΣΑ LL(1)

0	E		id	+	id	*	id	E0F	$E \rightarrow T E'$
1	E'	T	id	+	id	*	id	E0F	$T \rightarrow F T'$
2	E'	$T' F$	id	+	id	*	id	E0F	$F \rightarrow \text{id}$
3	E'	$T' \text{id}$	id	+	id	*	id	E0F	
4	E'	T'		+	id	*	id	E0F	$T' \rightarrow \epsilon$
5	E'			+	id	*	id	E0F	$E' \rightarrow + T E'$
6	E'	$T +$		+	id	*	id	E0F	
7	E'	T			id	*	id	E0F	$T \rightarrow F T'$
8	E'	$T' F$			id	*	id	E0F	$F \rightarrow \text{id}$
9	E'	$T' \text{id}$			id	*	id	E0F	
10	E'	T'				*	id	E0F	$T' \rightarrow * F T'$
11	E'	$T' F *$				*	id	E0F	
12	E'	$T' F$			id		id	E0F	$F \rightarrow \text{id}$
13	E'	$T' \text{id}$			id		id	E0F	
14	E'	T'						E0F	$T' \rightarrow \epsilon$
15	E'							E0F	$E' \rightarrow \epsilon$
16	ϵ							E0F	αναγνώριση

Κατηγορίες γραμματικές (i)

- Κατηγορική γραμματική (attribute grammar): γραμματική χωρίς συμφραζόμενα όπου κάθε σύμβολο φέρει ένα σύνολο **κατηγορημάτων**
- Οι τιμές των κατηγορημάτων υπολογίζονται βάσει του συντακτικού δέντρου:
 - **Συνθετικά** κατηγορήματα: οι τιμές τους εξαρτώνται μόνο από κατηγορήματα των παιδιών κάθε κόμβου
 - **Κληρονομούμενα** κατηγορήματα: οι τιμές τους εξαρτώνται μόνο από κατηγορήματα του ‘πατέρα’ και των ‘αδελφών’ κάθε κόμβου

Κατηγορίες γραμμάτικες (ii)

$$E \rightarrow E + T \quad \{ E^1.val := E^2.val + T.val \}$$

$$E \rightarrow T \quad \{ E.val := T.val \}$$

$$T \rightarrow T * F \quad \{ T^1.val := T^2.val * F.val \}$$

$$T \rightarrow F \quad \{ T.val := F.val \}$$

$$F \rightarrow (E) \quad \{ F.val := E.val \}$$

$$F \rightarrow \text{num} \quad \{ F.val := \text{num}.val \}$$

Σημασιολογικοί κανόνες

