

Μεταγλωττιστές #3

Νίκος Παπασπύρου
nickie@softlab.ntua.gr



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

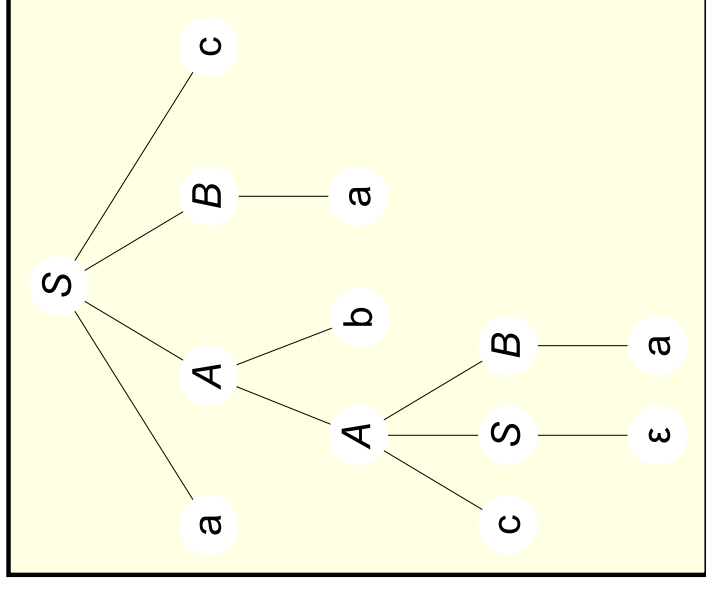
Γλώσσες χωρίς συμφραζόμενα

- Γραμματικές χωρίς συμφραζόμενα: $A \rightarrow \alpha$
- Αριστερότερη / δεξιότερη παραγωγή
(leftmost / rightmost derivation)
- Συντακτικά δέντρα

(parse trees)

$$\begin{aligned} S &\Rightarrow_L aABC \Rightarrow_L aAbBc \\ &\Rightarrow_L acSBbBc \Rightarrow_L acBbBc \\ &\Rightarrow_L acabBc \Rightarrow_L acabac \end{aligned}$$

$$\begin{aligned} S &\Rightarrow_R aABc \Rightarrow_R aAac \\ &\Rightarrow_R aAbac \Rightarrow_R acSBbac \\ &\Rightarrow_R acSabac \Rightarrow_R acabac \end{aligned}$$



Διφορούμενες γραμματικές (i)

- Μια γραμματική είναι **διφορούμενη** (ambiguous) όταν υπάρχουν δύο ή περισσότερα συντακτικά δέντρα για την ίδια παραγόμενη συμβολοσειρά
- Γραμματικές και γλώσσες **εγγενώς διφορούμενες** (inherently ambiguous)
- Παράδειγμα διφορούμενης γραμματικής:

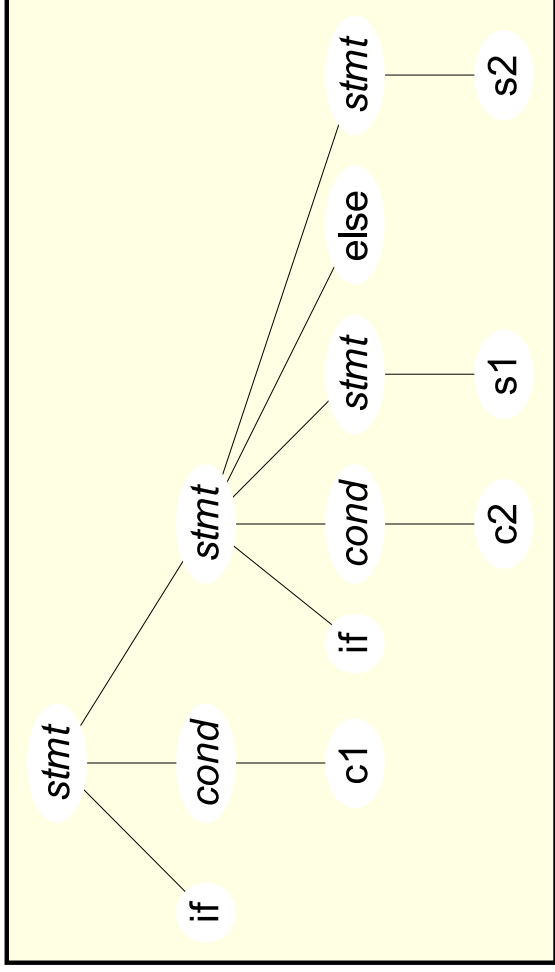
ξεκρέμαστο if (dangling if)

$stmt \rightarrow \text{if } cond \ stmt \text{ else } stmt \mid \text{if } cond \ stmt \mid s1 \mid s2$

$cond \rightarrow c1 \mid c2$

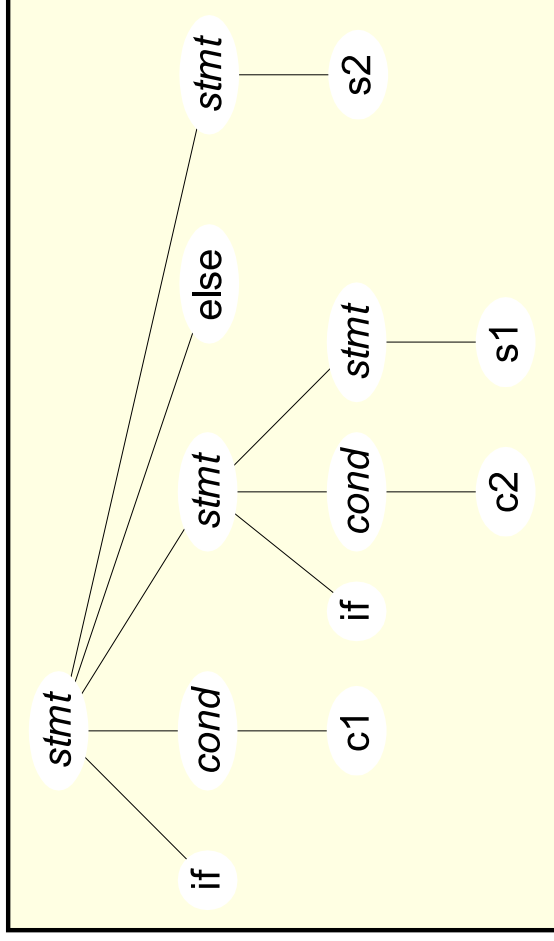
και η συμβολοσειρά: $\text{if } c1 \text{ if } c2 \ s1 \text{ else } s2$

Διφορούμενες γραμματικές (ii)



`if c1 (if c2 s1 else s2)`

`if c1 (if c2 s1) else s2`



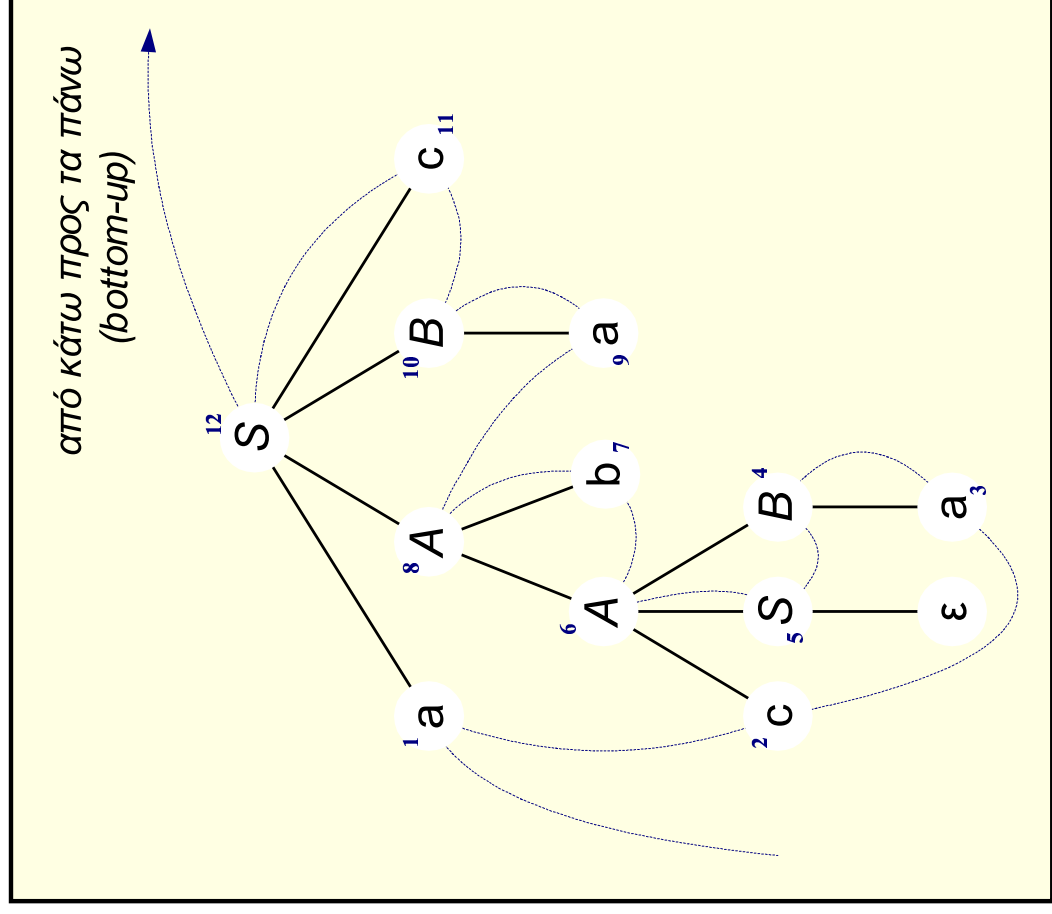
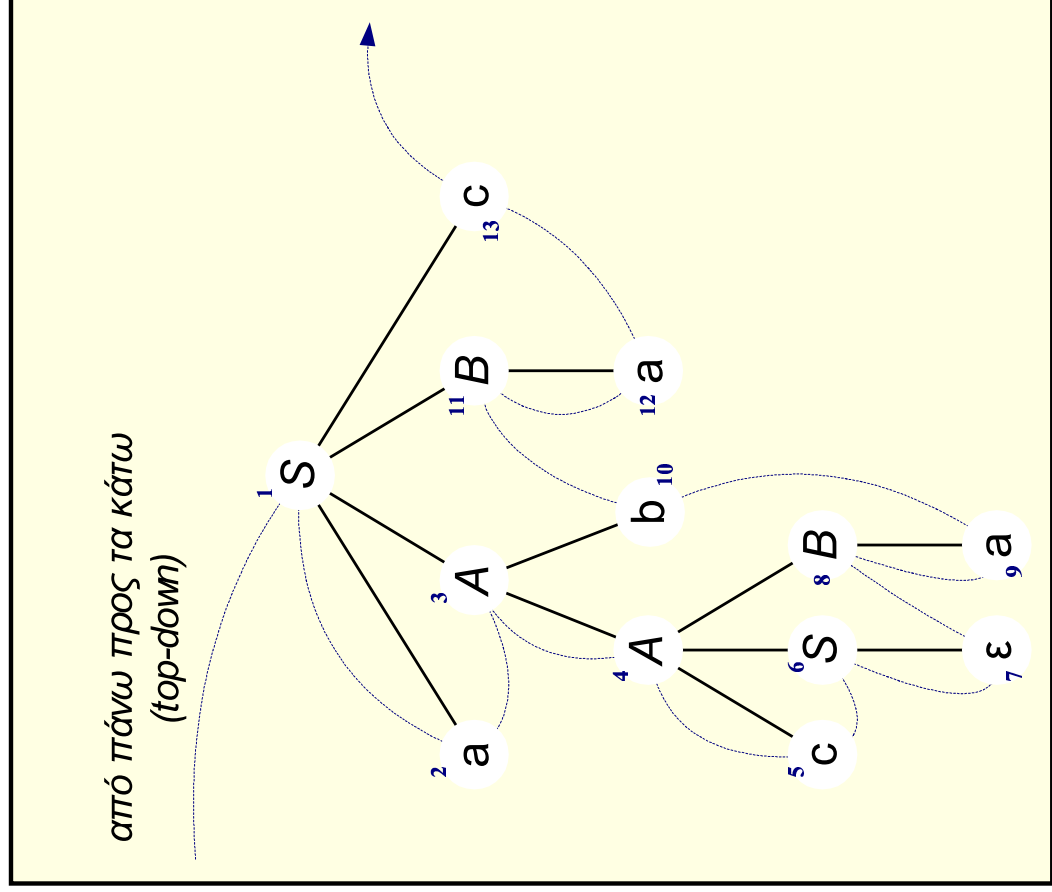
Συμβολισμός *BNF/EBNF*

- **Backus-Naur Form**
 - Σύμβολο \rightarrow στους κανόνες
 - Μη τερματικά σύμβολα σε γωνιακές παρενθέσεις, π.χ. $\langle \text{expr} \rangle$
 - Σύμβολο $|$ για διάζευξη
- **Extended Backus-Naur Form**
 - Τερματικά σύμβολα σε **εισαγωγικά**
 - **Παρενθέσεις** για ομαδοποίηση
 - **Αγκύλες** για προαιρετικά τμήματα
 - Σύμβολα $*$ και $+$ για επανάληψη

Συντακτική ανάλυση

- Συντακτικό δέντρο (parse tree)
- Κατασκευάζεται με δύο τρόπους:
 - Από πάνω προς τα κάτω (**top-down**)
δηλαδή ξεκινώντας **από τη ρίζα** και
προχωρώντας **προς τα φύλλα**
 - Από κάτω προς τα πάνω (**bottom-up**)
δηλαδή ξεκινώντας **από τα φύλλα** και
προχωρώντας **προς τη ρίζα**

Top-down και bottom-up



Βοηθητικές έννοιες

(i)

- Σύνολα FIRST

- Έστω συμβολοσειρά $\alpha \in (T \cup N)^*$
- Το σύνολο $\text{FIRST}(\alpha) \subseteq (T \cup \{\epsilon\})$ περιέχει τα τερματικά σύμβολα από τα οποία **αρχίζουν** οι συμβολοσειρές που παράγονται από την α
- Αν $\alpha \Rightarrow \epsilon$, τότε $\epsilon \in \text{FIRST}(\alpha)$

Βοηθητικές έννοιες

(ii)

■ Σύνολα FOLLOW

- Έστω μη τερματικό σύμβολο A
- Το σύνολο $\text{FOLLOW}(A) \subseteq (T \cup \{\text{EOF}\})$ περιέχει τα τερματικά σύμβολα που μπορούν να ακολουθούν το A στη διάρκεια μιας παραγωγής
- Αν το A μπορεί να είναι το τελευταίο σύμβολο σε μια παραγωγή, τότε $\text{EOF} \in \text{FOLLOW}(A)$



Υπολογισμός FIRST

(i)

- $\text{FIRST}(\epsilon) = \{\epsilon\}$
- $\text{FIRST}(a\beta) = \{a\}$
- αν $\epsilon \notin \text{FIRST}(A)$
τότε $\text{FIRST}(A\beta) = \text{FIRST}(A)$
- αν $\epsilon \in \text{FIRST}(A)$
τότε $\text{FIRST}(A\beta) = (\text{FIRST}(A) - \{\epsilon\}) \cup \text{FIRST}(\beta)$
- για κάθε κανόνα $A \rightarrow \alpha$, πρέπει
 $\text{FIRST}(\alpha) \subseteq \text{FIRST}(A)$

Υπολογισμός FIRST

- Παράδειγμα

$$\text{FIRST}(E) = \{\text{id}, (\}$$

$$\text{FIRST}(T) = \{\text{id}, (\}$$

$$\text{FIRST}(F) = \{\text{id}, (\}$$

$$\text{FIRST}(E') = \{+, \epsilon\}$$

$$\text{FIRST}(T') = \{*, \epsilon\}$$

(ii)

E	\rightarrow	$T E'$
E'	\rightarrow	ϵ
E'	\rightarrow	$+ T$
T	\rightarrow	$F T'$
T'	\rightarrow	ϵ
T'	\rightarrow	$* F$
F	\rightarrow	(E)
F	\rightarrow	id

Υπολογισμός FOLLOW

(i)

- $\text{EOF} \in \text{FOLLOW}(S)$
- για κάθε κανόνα $A \rightarrow \alpha B \beta$
 - $(\text{FIRST}(\beta) - \{\epsilon\}) \subseteq \text{FOLLOW}(B)$
 - αν $\epsilon \in \text{FIRST}(\beta)$
τότε $\text{FOLLOW } A \subseteq \text{FOLLOW}(B)$



Υπολογισμός FOLLOW

(ii)

■ Παράδειγμα

$$\text{FOLLOW}(E) = \{ \}, \text{EOF} \}$$

$$\text{FOLLOW}(T) = \{ +, \}, \text{EOF} \}$$

$$\text{FOLLOW}(F) = \{ *, +, \}, \text{EOF} \}$$

$$\text{FOLLOW}(E') = \{ \}, \text{EOF} \}$$

$$\text{FOLLOW}(T') = \{ +, \}, \text{EOF} \}$$

$$\text{FIRST}(E') = \{ +, \epsilon \}$$

$$\text{FIRST}(T') = \{ *, \epsilon \}$$

E	\rightarrow	$T E'$
E'	\rightarrow	ϵ
E'	\rightarrow	$+ T$
T	\rightarrow	$F T'$
T'	\rightarrow	ϵ
T'	\rightarrow	$* F$
F	\rightarrow	(E)
F	\rightarrow	id

ΣΑ bottom-up

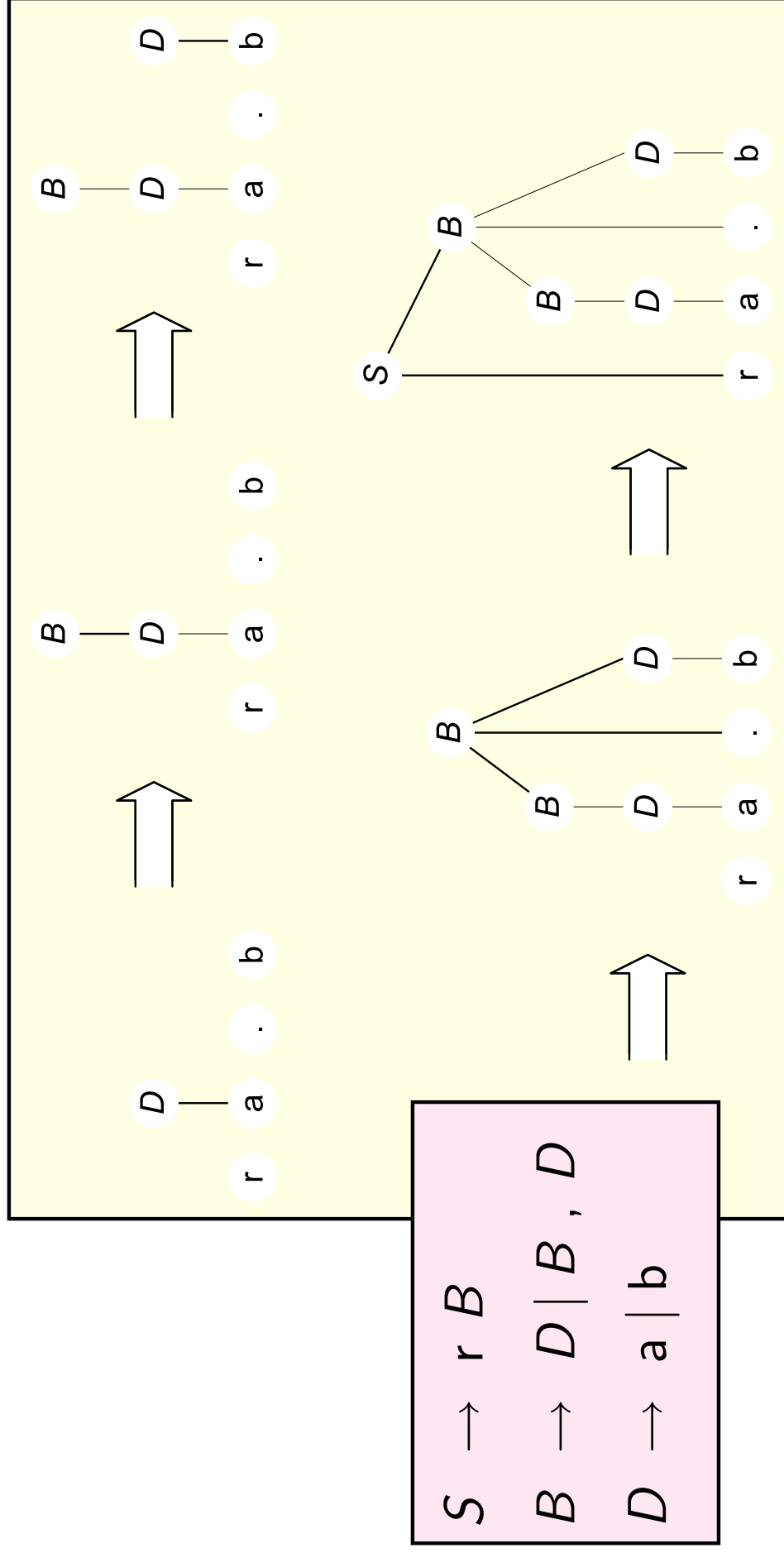
(i)

- Η συντακτική ανάλυση ξεκινά από τα φύλλα
- Κάθε φορά, αναζητά:
 - τον **αριστερότερο** κόμβο του δέντρου
 - που δεν έχει ακόμα κατασκευαστεί
 - ενώ όλα τα παιδιά του έχουν κατασκευαστεί
- Επαναλαμβάνει μέχρι να κατασκευαστεί **η ρίζα**
- **Ελάττωση** (reducing): η επιλογή των κόμβων που θα αποτελέσουν τα παιδιά ενός νέου κόμβου



ΣA bottom-up

(ii)



ΣΑ *bottom-up*

(iii)

- ΣΑ ολίσθησης-ελάττωσης (shift-reduce)
 - Χρησιμοποιούν μια (αρχικά κενή) **στοίβα** όπου τοποθετούν σύμβολα της γραμματικής
 - **Ολίσθηση** (shift): μεταφορά ενός συμβόλου από την είσοδο στην κορυφή της στοίβας
 - **Ελάττωση** (reduce): αφαίρεση από την κορυφή της στοίβας του δεξιού μέλους ενός κανόνα και πρόσθεση του αριστερού μέλους
 - **Επιτυχία**: η στοίβα περιέχει μόνο το S και τα σύμβολα της εισόδου έχουν εξαντληθεί



ΣΑ bottom-up

(iv)

βήμα στοίβα είσοδος περιγραφή κίνησης

0	ε	r a , b	ολίσθηση
1	r	a , b	ολίσθηση
2	r a	, b	ελάττωση με $D \rightarrow a$
3	r D	, b	ελάττωση με $B \rightarrow D$
4	r B	, b	ολίσθηση
5	r B ,	b	(όχι ελάττωση με $S \rightarrow r B$)
6	r B , b	ε	ολίσθηση
7	r B , D	ε	ελάττωση με $D \rightarrow b$
			ελάττωση με $B \rightarrow B , D$
			(όχι ελάττωση με $B \rightarrow D$)
8	r B	ε	ελάττωση με $S \rightarrow r B$
9	S	ε	αναγνώριση

shift/reduce
conflict

reduce/reduce
conflict

ΣΑ bottom-up

(v)

- $LR(k)$
- $LR(0)$
- $SLR(1)$
- $LALR(1)$
- $LR(1)$

